Signal perception and belief updating

Ilke Aydogan[a], Aurélien Baillon*[a], Emmanuel Kemel[b], and Chen Li[a]

[a]*Erasmus School of Economics - Erasmus University Rotterdam*[†]

[b]*GREGHEC - CNRS and HEC Paris*[‡]

January 26, 2017

Abstract

This paper introduces a theory of signal perception to study how people update their beliefs. By allowing perceived signals to deviate from actual signals, we identify the probability that people miss or misread signals, giving indices of conservatism and confirmatory bias. In an experiment, we elicited perceived signals from choices and obtained a structural estimation of the indices. The subjects were conservative and acted as if they missed 65% of the signals they received. Also they exhibited confirmatory bias by misreading 17% of the signals contradicting their prior beliefs.

Keywords: non-Bayesian updating; conservatism; confirmatory bias; perceived signals; belief elicitation.

1 Introduction

In standard economic models, from game theory to macroeconomics, decision makers incorporate new information using the rational gold standard of belief updating: the Bayes' rule. Yet, studies from the psychology literature highlighted regular deviations from Bayesian updating. Famous examples are the

*Corresponding author: baillon@ese.eur.nl

†P.O. Box 1738, Rotterdam, 3000 DR, The Netherlands.

‡Campus HEC W08, 1 Rue de la Libération, 78350 Jouy-en-Josas, France.

confirmatory bias (see Oswald and Grosjean, 2004, for a review), in which people tend to neglect or even misinterpret signals contradicting their prior beliefs, and the conservatism bias (Phillips and Edwards, 1966; Edwards, 1968), in which people fail to sufficiently incorporate new information, resulting in posteriors that are *too* close to their priors.

Since the end of the 1990s, economists have proposed models to incorporate deviations from Bayesian updating. For instance, Rabin and Schrag (1999) modeled confirmatory bias as decision makers misreading signals that contradict their priors, which may give rise to behavioral biases such as overconfidence. Epstein (2006) provided an axiomatic foundation for non-Bayesian updating through a retroactively changing prior. Wilson (2014) modeled a decision maker with bounded memory, which can lead to the emergence of confirmatory bias and conservatism in belief formation.

Alternatively, the literature on *motivated beliefs* (see Bénabou et al., 2016, for a review) models deviations from Bayesian updating through the decision maker's tradeoff between the accuracy and desirability of their beliefs. Strategies to cope with this tradeoff includes reality denial and wishful thinking. The motivated-belief approach is appealing in situations when people are motivated to attach values to their beliefs, such as when they think of their own abilities or of important aspects of their life (Bénabou and Tirole, 2002, 2006). It is not obvious though whether it would predict deviations from Bayesian updating when beliefs concern external, 'neutral' factors.

In this paper, we propose a theory of signal perception to model belief updating. We introduce two indices that are derived from the difference between people's *perceived signals*, revealed from their choices, and the actual signals received. Rabin and Schrag (1999) models the confirmatory bias as a probability of *misreading* a contradicting signal as confirming. Our first index, q, captures the same confirmatory tendency in belief updating. In particular, when $0 < q < 1$, it has the same interpretation as in Rabin and Schrag's model. Empirical evidence shows that people can also exhibit the opposite pattern (Eil and Rao, 2011). Our index captures disconfirmatory bias as well. Our second index p captures people's tendency of *missing* a signal, regardless of its agreement with the prior. Our model, combining q and p, is a portable extension of Bayesian updating in the sense of Rabin (2013). It can be incorporated in any model from macroeconomics or game theory by re-coding actual signals into perceived signals using transformations based on q and p.

After presenting our model, we show how perceived signals can be revealed from choices. In an experiment, we elicited subjects' beliefs and obtained a structural estimation of the indices, demonstrating the tractability of the model. We found clear evidence for both conservatism and confirmatory bias, showing that deviations of Bayesian updating may occur even in the absence of clear motivation. On average, subjects missed 65% of the signals and misread 17% of the signals contradicting their prior beliefs. We further explored factors influencing the indices. Consistent with previous findings (Griffin and Tversky, 1992), moderately informative signals led to more conservatism.

2 Perceived signal theory

2.1 Setup and perceived signals

We model a simple signal setup, in which a decision maker faces a mechanism producing independent and identically distributed binary signals. It produces *successes* with an unknown probability s (and *failures* with probability $1-s$). The decision maker is interested in learning about the success rate s. We consider an initial state of ignorance, represented by a uniform probability measure $Prob(s)$ defined over $\mathcal{S} \subset (0,1)$. We assume that the support $\mathcal{S}$ is symmetric around 0.5, i.e., $p \in \mathcal{S} \Rightarrow (1-p) \in \mathcal{S}$.

Before receiving a specific set of signals, the decision maker has a prior sample with α_0 successes and β_0 failures in his memory. Hence, his *prior beliefs* are $\Lambda(s;\alpha_0,\beta_0) = Prob(s|\alpha_0,\beta_0)$, abbreviated as $\Lambda(\alpha_0,\beta_0)$. When $\alpha_0 = \beta_0$, the mean of $\Lambda(\alpha_0,\beta_0)$ is equal to 0.5. The initial state of ignorance is an hypothetical construct that allows us to interpret the decision maker's beliefs in terms of signals. Departures from uniformity in prior beliefs are modeled by (possibly hypothetical) signals in the decision maker's mind.

After receiving a sequence of signals, his *posterior beliefs* becomes $\Lambda(\alpha_1,\beta_1)$. Define $\alpha = \alpha_1 - \alpha_0$, $\beta = \beta_1 - \beta_0$, and $\eta = \alpha + \beta$. These parameters measure how much the decision maker has updated his beliefs and therefore, how many signals (successes, failures) he has perceived. We call η the *perceived number of signals*, α the *perceived number of successes*, and β the *perceived number of failures.*

Consider a Bayesian updater with a uniform prior over $(0,1)$, which is equivalent to $Beta(1,1)$. If he observes a success, his posterior will also be a beta distribution, given by $Beta(2,1)$. It would be $Beta(1,2)$, had he observed a failure. After each success (failure), the first (second) parameter of the beta distribution is incremented by one. If the prior belief is $Beta(\alpha_0,\beta_0)$, with α_0 and β_0 possibly different from 1, then the expected probability of success is given by $\frac{\alpha_0}{\eta_0}$ with $\eta_0 = \alpha_0 + \beta_0$. Hence, the decision maker will expect success and failure to be equally likely iff $\alpha_0 = \beta_0$. In our application, we will use such a setting with beta distributions but the theory below does not rely on it.[1]

For a Bayesian updater, all signals are perceived without distortion: receiving n signals consisting of a successes and b failures implies $\alpha = a$, $\beta = b$, and $\eta = n$. However, this is not true for non-Bayesian updaters. Deviations from Bayesian updating can therefore be captured by differences between people's perceived signals (α,β, and η) and the actual signals they observe (a, b, and n). We study two sources of deviations: confirmatory bias and conservatism bias. Confirmatory bias captures people's tendency to "misread evidence as additional support for initial hypotheses" (Rabin and Schrag, 1999), whereas conservatism captures people's tendency to miss evidence and to not update enough their beliefs, without discriminating different types of signals.

2.2 Confirmatory bias

Following Rabin and Schrag (1999), we model the *confirmatory bias* as the probability q_c to misread a contradicting signal as confirming prior expectations. By symmetry, the opposite bias, that we called *disconfirmatory bias*, can be modelled as the probability q_d to misread a signal as contradicting prior expectations. If, according to the decision maker's prior belief, successes are more

[1]For instance, the support $\mathcal{S}$ can be discrete.

likely than failures (i.e. $\alpha_0 > \beta_0$), the confirmatory bias gives:

$$\begin{cases} \alpha = a + q_c b \\ \beta = (1 - q_c) b \end{cases}, \tag{1}$$

whereas the disconfirmatory bias gives

$$\begin{cases} \alpha = (1 - q_d) a \\ \beta = b + q_d a \end{cases}. \tag{2}$$

If, according to the decision maker's prior belief, successes are less likely than failures (i.e. $\alpha_0 < \beta_0$), the confirmatory bias gives:

$$\begin{cases} \alpha = (1 - q_c) a \\ \beta = b + q_c a \end{cases}. \tag{3}$$

whereas the disconfirmatory bias gives

$$\begin{cases} \alpha = a + q_d b \\ \beta = (1 - q_d) b \end{cases}, \tag{4}$$

If successes and failures are equally likely according to the decision maker's prior belief, the perceived number of success and failure is not affected by confirmatory bias.

From observing perceived signals, either q_c or q_d can be determined whenever $\alpha_0 \neq \beta_0$. Consider the case $\alpha_0 > \beta_0$. If $a \leq \alpha \leq \eta$, there is evidence for confirmatory bias and q_c can be computed. In practice, we may even observe $q_c > 1$ when $\eta < \alpha$ (and therefore $\beta < 0$). In such a case, q_c is not a probability anymore but can still be used as an index of confirmatory bias. The case $q_c > 1$ indicates that the decision maker exhibits an extreme form of confirmatory bias, in which he even recodes the signals from his prior. We call such a case *prior-signal confirmatory recoding*. Figure 1 depicts all possible cases. The interpretation of the decision maker's perceived signals depend on his prior beliefs (α_0 and β_0) and his perceived number of successes α. Moreover, we can combine q_c and q_d into a unique index of confirmatory bias q defined as:

$$q = \begin{cases} q_c & \text{if } (\alpha_0 > \beta_0 \text{ and } a \leq \alpha) \text{ or } (\alpha_0 < \beta_0 \text{ and } b \leq \beta) \\ -q_d & \text{if } (\alpha_0 > \beta_0 \text{ and } a \geq \alpha) \text{ or } (\alpha_0 < \beta_0 \text{ and } b \geq \beta) \end{cases}. \tag{5}$$

Values of q in $[0, 1]$ can be directly interpreted as probabilities to misread signal in a confirmatory way and values in $[-1, 0]$ as minus probabilities to misread signal in a disconfirmatory way. The global index q is useful for empirical purposes. For instance, its distribution for the population can be estimated at once, without separating confirmatory biases from disconfirmatory biases (as is done for other attitude measures such as risk aversion).

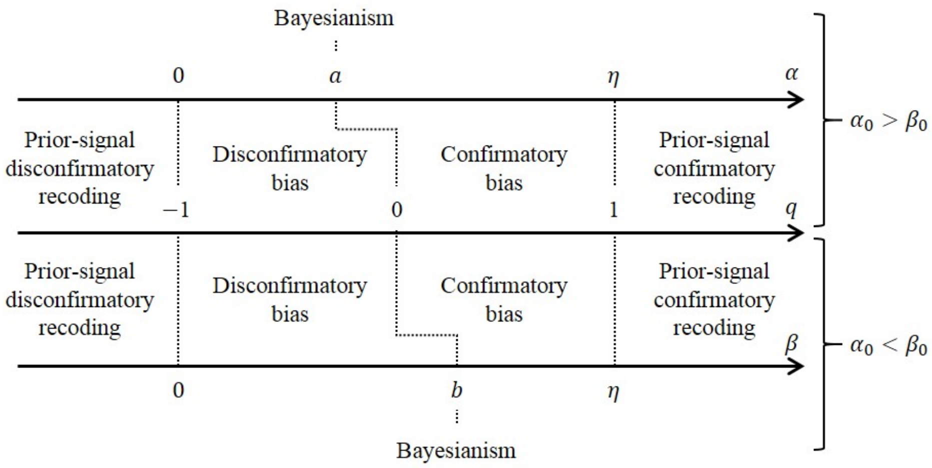

Figure 1: Interpretation of q and relationship with α

2.3 Conservatism bias

We expand the confirmatory bias model by also considering a conservative decision maker's tendency to ignore signals. In this subsection, we introduce a measure of conservatism bias, and in the next subsection, we present a model in which both biases are combined in one model.

A conservative decision maker places too little weight on the sample information while updating and thereby tends to ignore some of the relevant information. We model the *conservatism bias* as a probability p to miss a signal. Hence, $\eta = (1 - p)n$. The conservatism bias can affect both types of signals indistinguishably, leading to $\alpha = (1-p)a$ and $\beta = (1-p)b$. Bayesian updating implies $p = 0$. If $p = 1$, there is no updating at all.

Interestingly, p can also be interpreted if it lies outside the unit interval, but obviously not as a probability. The case $p > 1$ captures situations where the perceived number of signal is negative, suggesting that the decision maker received information that undermined his prior. For instance, a decision maker whose prior was too extreme, expecting successes almost exclusively, might be less confident in his beliefs after observing a few failures. In our perceived signal theory, such behavior corresponds to *prior signal destruction*.

By contrast, $p < 0$ means that the decision maker perceived too many signals. It can be further illustrated in the case of a Beta distribution. The posterior mean $\frac{\alpha_0+\alpha}{\eta_0+\eta}$ can be decomposed in terms of prior mean and sample mean:

$$\begin{aligned}\frac{\alpha_0+\alpha}{\eta_0+\eta} &= \frac{\alpha_0+(1-p)a}{\eta_0+(1-p)n} \\ &= \frac{\eta_0}{\eta_0+(1-p)n}\cdot\frac{\alpha_0}{\eta_0} + \frac{(1-p)n}{\eta_0+(1-p)n}\cdot\frac{a}{n} \\ &= \frac{\eta_0}{\eta_0+(1-p)n}\cdot \text{ prior mean } + \frac{(1-p)n}{\eta_0+(1-p)n}\cdot \text{ sample mean}\end{aligned} \quad . \qquad (6)$$

Bayes rule requires $p = 0$, i.e. the actual and the perceived number of signals match. A positive $p (< 1)$ decreases the impact of the sample mean, implying conservatism. The decision maker underweights the sample information and overweights the prior information. Negative p corresponds to *base rate neglect*, the decision maker assigning too much weight to the sample and neglecting his prior beliefs. Such behavior can be explained by the *representativeness heuristic* (Tversky and Kahneman, 1974), when decision makers assume that a sample must resemble the process it originates from and therefore tend to equate the process mean too much with the sample mean.

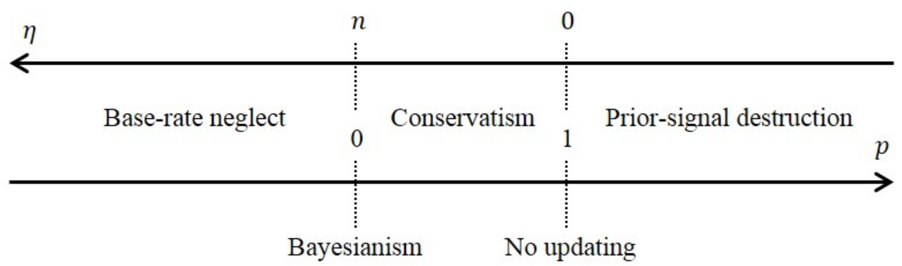

Figure 2: Interpretation of p and relationship with η

Figure 2 depicts the relationship between the perceived number of signal η and the conservatism index p. It shows that p is a simple rescaling of η such that p is independent of the actual sample size n.

2.4 Combining biases

In the combined model, the decision maker may miss signals (conservatism bias) and then misread those he did not miss (confirmatory bias). If, according to the decision maker's prior belief, successes are more likely than failures ($\alpha_0 > \beta_0$), the confirmatory bias in presence of conservatism bias gives (replacing q_c by q):

$$\begin{cases} \alpha = (1-p)a + q(1-p)b \\ \beta = (1-q)(1-p)b \end{cases} , \tag{7}$$

whereas the disconfirmatory bias gives (replacing q_d by $-q$)

$$\begin{cases} \alpha = (1+q)(1-p)a \\ \beta = (1-p)b - q(1-p)a \end{cases} . \tag{8}$$

The case $\alpha_0 < \beta_0$ is symmetric. If successes and failures were equally likely according to the decision maker's prior belief:

$$\begin{cases} \alpha = (1-p)a \\ \beta = (1-p)b \end{cases} . \tag{9}$$

In terms of observability, p can always been obtained by comparing η with n. Further, if $\alpha_0 \neq \beta_0$ and $p \neq 1$, then q can be obtained by first correcting

a and b for conservatism (multiplying them by $1-p$) and then applying the adequate equation of (dis)confirmatory bias. If after a first round of signals, the decision maker receives a second round of signals, p and q can be determined again using the posterior of the first round as the prior of the second round.

2.5 Measuring informativeness

Literature shows that deviations from Bayesian updating depend on various situational factors such as the strength of evidence. Griffin and Tversky (1992) find that moderate signals lead to insufficient updating while extreme signals lead to overreaction. The same set of signals may be deemed extremely informative by a decision maker but less so by another. The informativeness of signals thus depends both on the signals themselves and the prior of the decision maker. Hence a measure of informativeness should depend on α_0, β_0, a, and b.

We define our measure of informativeness as the *information gain (IG)*, also known as relative entropy or Kullback-Leibler divergence, between the prior $\Lambda(\alpha_0, \beta_0)$ and the posterior the decision maker would have if he were Bayesian $\Lambda(\alpha_0 + a, \beta_0 + b)$. Let $g(s)$ be the density function of prior, and $h(s)$ be the density of the Bayesian posterior. The IG is calculated as:

$$IG(\alpha_0, \beta_0, a, b) = \int_{[0,1]} h(s) \log \frac{h(s)}{g(s)} ds. \tag{10}$$

The IG measure captures how much the signals should influence the decision maker's beliefs. It allows us to examine the impact of signal informativeness on belief updating biases.

3 Revealing perception through choices

To reveal people's perception of signals, it is necessary to make their beliefs observable. Belief elicitation methods in the literature, such as proper scoring rules (see Schotter and Trevino, 2014, for a survey in economics), often rely on the descriptive validity of expected value or expected utility to reveal people's true beliefs. In this paper, we consider two methods that do not rely on expected utility.

We are interested in the decision maker's belief about the unknown success rate s. Let $\mathcal{P}$ denote the σ-algebra on $(0,1)$, which is the domain of s. *Events*, $E \in \mathcal{P}$, of interest to the decision maker are subsets of $(0,1)$. The decision maker faces *(binary) acts*, denoted by $\gamma_E\delta$, which pays a positive money amount γ if event E happens and δ otherwise. The decision maker also faces *(binary) lotteries* $\gamma_\lambda\delta$, yielding γ with probability λ and δ otherwise.

Assume that the decision maker whose behavior towards lotteries can be represented by a function V satisfying first order stochastic dominance. The function V need not be expected utility and it therefore allows for deviations from expected utility such as in the paradoxes suggested by Allais (1953). The decision maker is *probabilistically sophisticated* (Machina and Schmeidler 1992) if his behavior towards acts can be entirely explained by V and a probability measure Λ over $\mathcal{P}$. In other words, the assumption of a probabilistically sophisticated decision maker guarantees that choices are consistent with a probability measure and therefore is a sufficient condition to observe beliefs from choices.

We present two methods to elicit Λ irrespective of V. The first method to observe belief involves measuring *matching probabilities*, i.e. λ such that $\gamma_E\delta \sim \gamma_\lambda\delta$. Under probabilistic sophistication, this indifference implies $V(\gamma_{\Lambda(E)}\delta) = V(\gamma_\lambda\delta)$ and thus, $\Lambda(E) = \lambda$, thereby revealing beliefs. Many studies used matching probabilities to elicit people's beliefs (Raiffa, 1968; Spetzler and Stael von Holstein, 1975; Holt, 2007; Karni, 2009). The second method we consider involves elicitation of *exchangeable events*, events E and F, such that $\gamma_E\delta \sim \gamma_F\delta$. If probabilistic sophistication holds, the elicited indifference implies, $V(\gamma_{\Lambda(E)}\delta) = V(\gamma_{\Lambda(F)}\delta)$, and thus, $\Lambda(E) = \Lambda(F)$, providing constraints on the belief function. For instance, if they are complementary, then $\Lambda(E) = \Lambda(F) = \frac{1}{2}$. This method is based on the original idea of Ramsey (1931) (called ethically neutral events) and of De Finetti (1937) and has been long-known in decision analysis (Raiffa, 1968; Spetzler and Stael von Holstein, 1975). Recent experimental implementations can be found in Baillon (2008) and Abdellaoui et al. (2011).

Both methods have advantages and are therefore implemented in our experiment. Matching probabilities directly reveals the probability of an event whereas exchangeable events only reveal that two events are equally likely. Yet, matching probabilities require that the function V is the same for lotteries and for acts. If the decision maker tend to prefer lotteries to acts, exhibiting ambiguity aversion (Ellsberg, 1961), matching probabilities may be biased. Eliciting exchangeable events, which do not require the use of lotteries, is robust to this problem.[2] Implementing both methods will allow us to assess the possible impact of ambiguity attitude.

For empirical tractability, we assume that decision makers' beliefs follow a beta distribution. The beta family is both natural to model beliefs over a success rate and very tractable. Beta distributions are flexible and can take a wide array of shapes with different locations and dispersion for different parameters. Before and after they receive a set of signals, we elicit their priors and posteriors using the methods described above. We then estimate their perceived signals and are able to construct measures of their conservatism and confirmatory biases.

4 Experimental design

4.1 Subjects

Seven experimental sessions were conducted at the Erasmus School of Economics Rotterdam. The number of participants in each session varied between 20 and 27, summing up to 157 in total. Subjects were bachelor and master students at Erasmus University Rotterdam, with an average age of 21.3. Each session lasted one hour and fifteen minutes including instructions and payment.

4.2 Stimuli

During the experiment, subjects faced choice situations about acts whose payoffs depended on the actual color composition of a spinning wheel. The spinning wheel was covered by two (and only two) colors: yellow and brown. The color composition was randomly drawn from an opaque bag at the beginning of the experiment in front of all subjects by an *implementer* – one randomly selected

[2] Ambiguity is sometimes assumed to be equivalent to the absence of probabilistic beliefs. As demonstrated theoretically by Chew and Sagi (2008) and empirically by Abdellaoui et al. (2011), one can preserve the existence of a belief function expressed in probabilistic terms and allow for the Ellsberg paradox. The decision maker is *within-source probabilistically sophisticated* if there exists a probability measure Λ defined over $\mathcal{P}$ and a function W satisfying first order stochastic dominance such that $\gamma_E\delta$ is evaluated $W(\gamma_{\Lambda(E)}\delta)$. Under this model, exchangeable events E and F still satisfy $\Lambda(E) = \Lambda(F)$.

subject.

The experiment consisted of alternating periods of choice and sampling (see Figure 3 for the flow). It started with a choice period in which subjects made choices without any knowledge about the color composition of the wheel. Then, the implementer spinned the wheel three times and reported the resulting colors. Having acquired this new information, subjects made choices in the same choice situations (but potentially in different orders) again. The same procedure was repeated two more times.

Figure 3: Experimental flow

The color composition of the wheel stayed the same and unknown throughout the experiment, which means that in later choice periods, subjects made choices based on accumulated knowledge about the same wheel. For example, the last questionnaire was filled relying on the information of nine spins in total.

4.2.1 Matching probabilities

Figure 4 presents a choice list to elicit a matching probability. In each choice question, subjects had to choose between option W(heel) whose payoff depended on the actual color composition of the same spinning wheel, and option C(ard) whose payoff depended on a random draw from a deck of four cards of different suits: aces with heart, diamond, club, and spade, each with 25% probability.

Line		Option_W	Option_C	
1	brown = 0 %	○	✓	Card is ♦ (25% chance)
2	0% ≤ brown≤ 4%	○	○	Card is ♦ (25% chance)
3	0% ≤ brown≤ 8%	○	○	Card is ♦ (25% chance)
4	0% ≤ brown≤ 12%	○	○	Card is ♦ (25% chance)
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
22	0% ≤ brown≤ 84%	○	○	Card is ♦ (25% chance)
23	0% ≤ brown≤ 88%	○	○	Card is ♦ (25% chance)
24	0% ≤ brown≤ 92%	○	○	Card is ♦ (25% chance)
25	0% ≤ brown≤ 96%	○	○	Card is ♦ (25% chance)
26	0 % ≤ brown < 100 %	✓	○	Card is ♦ (25% chance)

Figure 4: Choice list to elicit matching probabilities

The choice in the first line was pre-ticked for the subjects by the experimenters, as in this case, option C dominates option W since the proportion of brown cannot be 0% (otherwise there is only one color on the wheel). Similarly, the last line was also pre-ticked. Subjects were informed that as they move down the list, option W became better while option C stayed the same. Therefore, at one point, they may switch from preferring option C to option W.

The subjects' switching pattern in Figure 4 gave an interval $[y^-_{0.25}, y^+_{0.25}]$ for $y_{0.25}$ such that $20_{[0,y^-_{0.25}]}0 \prec 20_{0.25}0$ and $20_{[0,y^+_{0.25}]}0 \succ 20_{0.25}0$, implying that 0.25 was the matching probability of event $[0, y_{0.25}]$. We also elicited the corresponding intervals for $y_{0.5}$ and $y_{0.75}$. The choice lists were similar, except that the card options had more winning suits – two winning suits for 50% and three for 75%.

4.2.2 Exchangeable events

Figure 5 presents a choice list used to elicit exchangeable events. In each choice question, subjects had to choose between two lotteries. Payoffs of both lotteries depended on the actual color composition of the spinning wheel. Take line 4 of the list as an example, Option L(eft) pays €20 if the actual brown proportion is no more than 12%, whereas Option R(ight) pays €20 if it is more than

12%. Subjects had to choose between the two lotteries in each line of the list, depending on their subjective judgment of the actual color composition of the wheel.

Line		Option_L	Option_R	
1	brown = 0%	○	✓	0% < brown ≤ 100%
2	0% ≤ brown≤ 4%	○	○	4% < brown≤ 100%
3	0% ≤ brown≤ 8%	○	○	8% < brown≤ 100%
4	0% ≤ brown≤ 12%	○	○	12% < brown≤ 100%
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
22	0% ≤ brown≤ 84%	○	○	84% < brown≤ 100%
23	0% ≤ brown≤ 88%	○	○	88% < brown≤ 100%
24	0% ≤ brown≤ 92%	○	○	92% < brown≤ 100%
25	0% ≤ brown≤ 96%	○	○	96% < brown≤ 100%
26	0% ≤ brown < 100%	✓	○	brown = 100%

Figure 5: Choice list to elicit exchangeable events

The first and the last lines were pre-ticked by similar dominance arguments as for matching probabilities, and subjects were told that as they move down from the list, option L became better and option R became worse. At some point, they may switch from preferring option R to option L.

Where subject switched in Figure 5 provided an interval $[y^-_{\text{median}}, y^+_{\text{median}}]$ for y_{median} such that $20_{[0,y^-_{\text{median}}]}0 \prec 20_{(y^-_{\text{median}},100]}0$ and $20_{[0,y^+_{\text{median}}]}0 \succ 20_{(y^+_{\text{median}},100]}0$. Therefore, for some $y_{\text{median}} \in [y^-_{\text{median}}, y^+_{\text{median}}]$, we have $20_{[0,y_{\text{median}}]}0 \sim 20_{(y_{\text{median}},100]}0$. The events $[0, y_{\text{median}}]$ and $(y_{\text{median}}, 1]$ were both exchangeable and complementary, meaning that the subjects assigned them probability $\frac{1}{2}$. Similarly, we elicited intervals for y_{low} and y_{high} such that $20_{[0,y_{low}]}0 \sim 20_{(y_{low},0.5]}0$, and $20_{[0.5,y_{high}]}0 \sim 20_{(y_{high},1]}0$, following the method of Abdellaoui et al. (2014). Choice lists to elicit y_{low} and y_{high} were similar, but with different start and end points of proportion intervals (from 0% to 50% for the former, and 50% to 100% for the latter).

4.3 Incentives

Each subject received a €5 show-up fee and a variable amount of €20 depending on one of his choices in one choice period (the implementer received a flat payment of €15). A prior incentive system (Johnson et al., 2015) was implemented to determine for each subject which choice would matter for his final payment. Before the experiment started, each subject randomly drew a sealed envelope from a pile of 156 sealed envelopes each containing one choice question (subjects faced in total 6 choice situations, each with 26 choice questions). Subjects were informed that the question that would matter for their payment was in their envelope, and were told not to open their envelopes until the end of the experiment. To determine which choice period would matter, the implementer randomly drew a number from one to four. Further implementation detailed are reported in the appendix.[3]

5 Raw data

Session	# Subjects	Received signals between rounds: 1&2	2&3	3&4
1	24	BBB	BBB	BBB
2	27	BYY	YYY	BYB
3	20	BBB	BYB	YYB
4	20	BYB	BYB	BYY
5	23	BBY	YYY	BBY
6	20	YYY	YYY	YYY
7	23	YYY	YBB	YYY

Table 1: Description of sessions

Table 1 summarizes the number of subjects and the color of spins in sampling periods in each session. For results reported in this section, we take the mid point of the elicited intervals as the indifference values. For instance, we take $y_{\text{median}} = \frac{y^-_{\text{median}} + y^+_{\text{median}}}{2}$.

Take the belief of a Bayesian updater with a uniform prior as the Bayesian benchmark. Figure 6 plots the difference between subjects' median belief (y_{median} in the exchangeability method and y_{50} in the matching method) of the yellow

[3]In particular, we also controlled for possible suspicion effects by letting the subjects choose on which color they would be betting.

proportion and the Bayesian benchmark. A positive (negative) difference corresponds to an overestimation (underestimation) of the yellow proportion. In sessions with balanced signals, subjects' median beliefs did not deviate much from the Bayesian benchmark, however, in sessions (e.g. session 1 and 7) where they received extreme signals, deviations were high. For instance, in session 1, subjects only received Brown signals. Their median deviations were positive, suggesting an overestimation of the yellow proportion on the wheel. The overestimation can be caused by conservatism: subjects did not incorporate the signals sufficiently. A similar pattern was observed in session 6 where subjects only received yellow signals and underestimated the yellow proportion on the wheel.

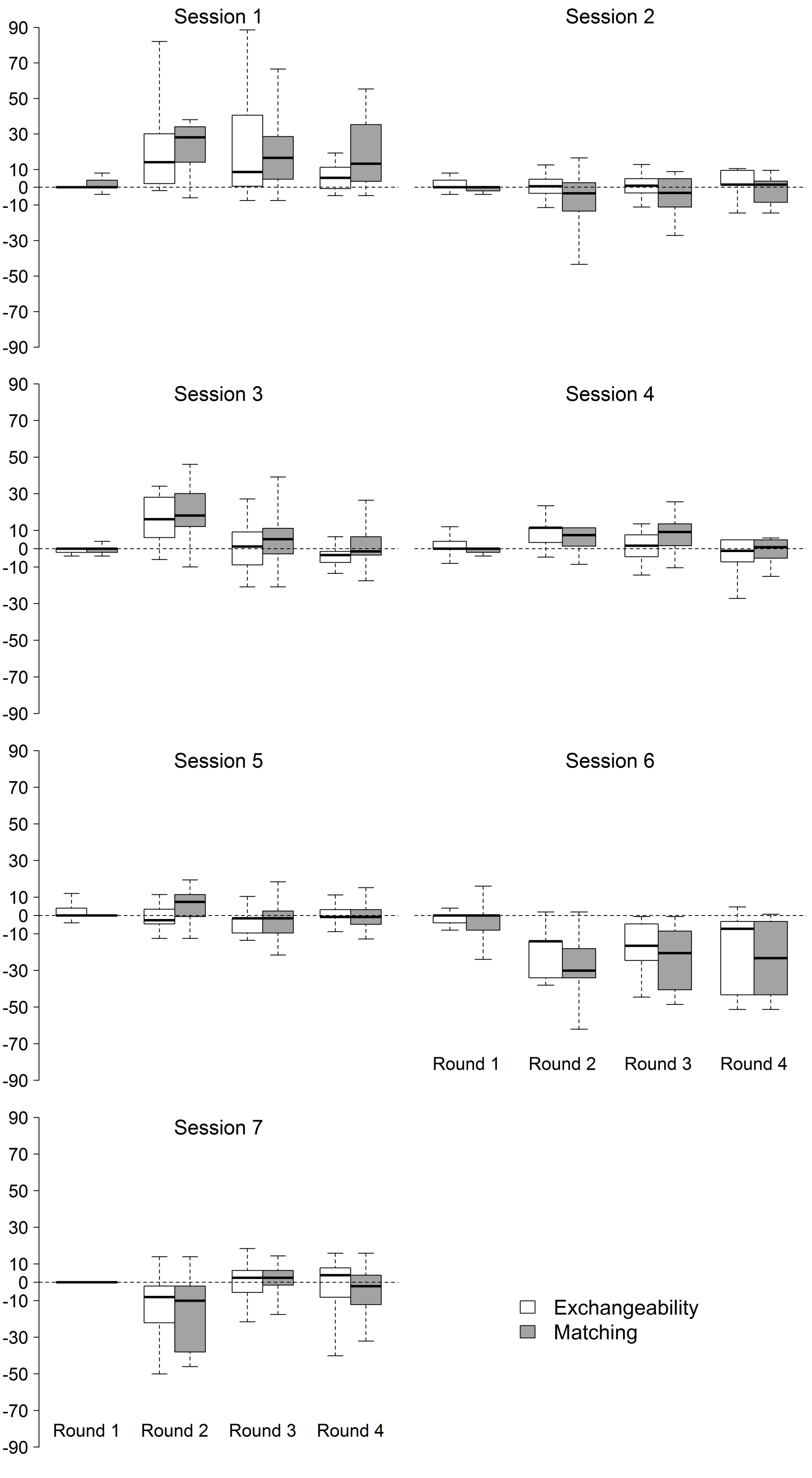

Figure 6: Median Deviation From the Bayesian Benchmark

Similarly, Figure 7 shows how the dispersion in subjects' beliefs ($s_{\text{high}} - s_{\text{low}}$ for the exchangeability method and $s_{75} - s_{50}$ for the matching method) differs from the Bayesian benchmark. A positive (negative) difference shows that subjects are under-precise (over-precise) as compared to the Bayesian benchmark. For both median and dispersion deviations, we observed persistent individual heterogeneity. In our structural model, we estimate the confirmation and conservatism indices while taking individual differences into account.

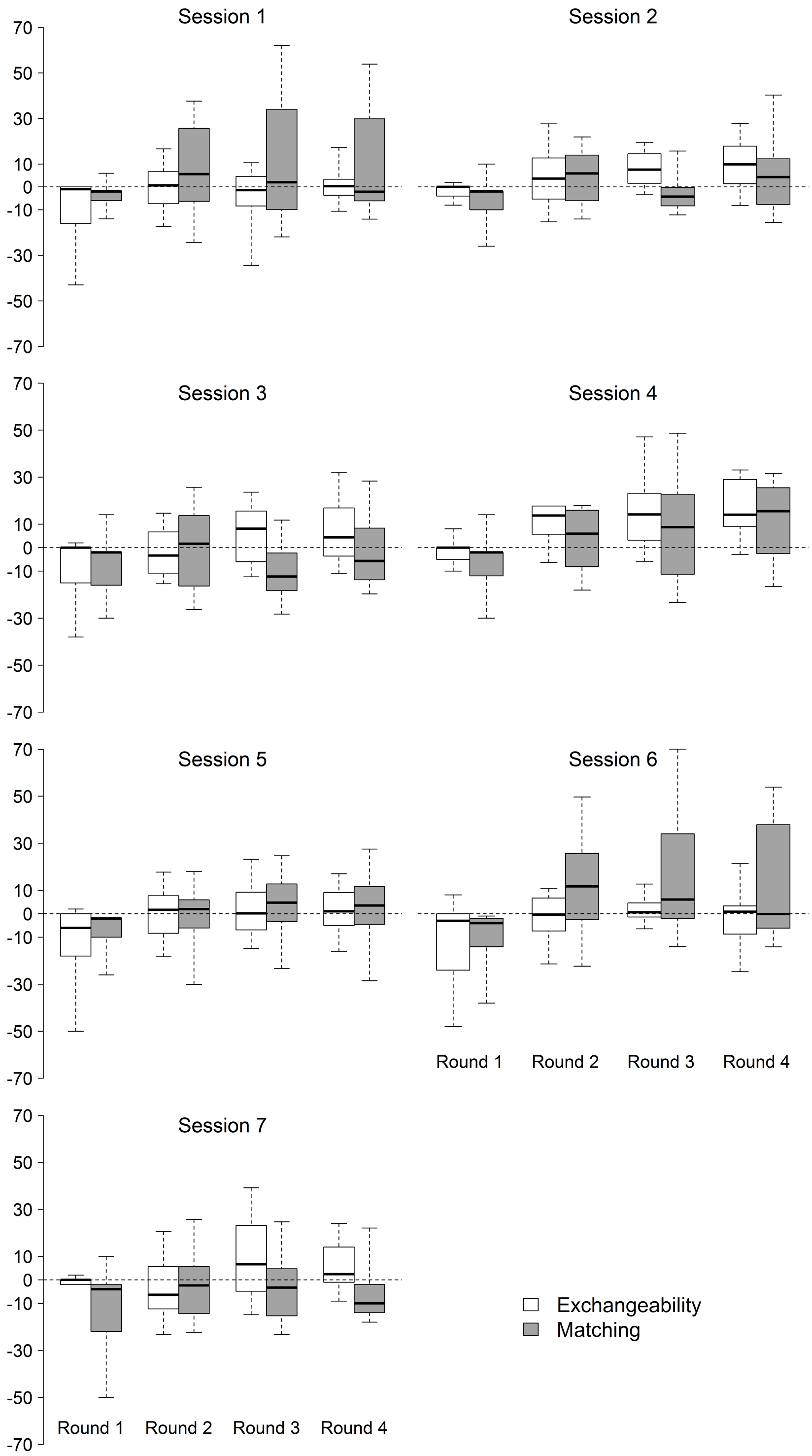

Figure 7: Dispersion Deviation From Bayesian updating

6 Econometric Analysis

6.1 Econometric model

6.1.1 Measuring beliefs and deviations from Bayesian updating

The beliefs of a subject i at round j are assumed to follow a Beta distribution $Beta(.|\alpha_{i,j}, \beta_{i,j})$. The prior of subject i at round 1, determined by $\alpha_{i,1}$ and $\beta_{i,1}$, is assumed to be exogenous and will be estimated. Then, for rounds $j > 1$:

$$\alpha_{i,j} = \alpha_{i,j-1} + s(a_{i,j}, b_{i,j}, \alpha_{i,j-1}, \beta_{i,j-1}, p_{i,j}, q_{i,j})$$
$$\beta_{i,j} = \beta_{i,j-1} + f(a_{i,j}, b_{i,j}, \alpha_{i,j-1}, \beta_{i,j-1}, p_{i,j}, q_{i,j})$$

where s and f are the functions that determine respectively the perceived successes and failures, as modeled by equations 7 to 9. These functions depend on the current beliefs parameters $\alpha_{i,j-1}$ and $\beta_{i,j-1}$, the received signals $a_{i,j}$ and $b_{i,j}$ and the indices of deviations from Bayesian updating, $p_{i,j}$ and $q_{i,j}$. For a Bayesian, $s(a_{i,j}, b_{i,j}, \alpha_{i,j-1}, \beta_{i,j-1}, p_{i,j}, q_{i,j}) = a_{i,j}$ and $f(a_{i,j}, b_{i,j}, \alpha_{i,j-1}, \beta_{i,j-1}, p_{i,j}, q_{i,j}) = b_{i,j}$.

According to Figure 6, there is little or no heterogeneity in prior beliefs as measured in round 1. We therefore assume that α_1 and β_1 are constant across subjects. Much more heterogeneity in beliefs is observed for later rounds, both between and within sessions. Heterogeneity between sessions can be due to session-specific received signals that can be more or less surprising. Heterogeneity within sessions can be due to subjects characteristics. Eventually, biases may also vary from one round to another, due to learning or fatigue. We attempt to account for these three possible sources of heterogeneity in our econometric analysis. To do so, we built a structural model that includes several explanatory variables of the deviation indices. Specifically, we assume that

$$p_{i,j} = p + \Gamma_p X_{i,j}$$
$$q_{i,j} = q + \Gamma_q X_{i,j}$$

where p and q are the intercepts of the deviation indices. When $\Gamma = 0$, these intercepts measure the aggregated indices over sessions, individuals and rounds. They should be equal to 0 if subjects perceived signals according to Bayes rule. Γ_p (resp. Γ_q) is the vector of coefficients that measure the impact of variables $X_{i,j}$ on index p (resp. q). The following explanatory variables are considered:

- a categorical variable denoting the round of the experiment,[4]
- the information gain of the received signals based on the individual specific prior,
- the squared information gain,
- subjects characteristics: including gender, and field of studies (economics or econometrics versus other).

We denote θ, the vector of coefficients to be estimated, with $\theta = (\alpha^1, \beta^1, p, q, \Gamma_p, \Gamma_q)$.

6.2 Estimating the model

Under our specification, the beliefs of a subject i at round j take the form of a probability distribution $\Lambda(.|\theta, a, b)$ where θ is a vector of coefficients and a and b are the received signals. This probability distribution is revealed by a series of choices, grouped within choice lists. Two types of choices lists are used. The first type, eliciting matching probabilities, considers a series of quantiles q_k and measures their corresponding values $y_k^\star$ such that $\Lambda(y_k^\star) = q_k$. More precisely, these choice lists determine two values y_k^- and y_k^+ such that $20_{[0,y_k^-]}0 \prec 20_{q_k}0$ and $20_{[0,y_k^+]}0 \succ 20_{q_k}0$ i.e. $y_k^\star \in [y_k^-; y_k^+]$.

The other type of choice lists, eliciting exchangeable events, considers intervals $[m_k, n_k]$ and measures the corresponding values $y_k^\star$ such that $\Lambda(m_k) - \Lambda(y_k^\star) = \Lambda(y_k^\star) - \Lambda(n_k)$ i.e. $\Lambda(y_k^\star) = \frac{\Lambda(m_k)-\Lambda(n_k)}{2}$. Here again, the choice lists determine two values y_k^- and y_k^+ such that $20_{[m_k,y_k^-]}0 \prec 20_{[y_k^-,n_k]}0$ and $20_{[m_k,y^+]}0 \succ 20_{[y^+,n_k]}0$ i.e. $y_k^\star \in [y_k^-; y_k^+]$.

For each individual i, round j and choice list k, the structural equation model provides a theoretical value $y_k^{th}(\theta, X_k)$ where θ is the vector of coefficients

[4] for p we consider round 2 as the reference and introduce dummy variables for round 3 and round 4. For q, preliminary analysis revealed that $alpha_1 = beta_1$, meaning that q is not defined for round 2. We considered round 3 as the reference and introduced a dummy variable for round 4.

of our decision model, and X_k is the set of variables containing choice lists characteristics and other explanatory variables. In order to account for subject and/or specification errors, we assume that $y_k^\star = y_k^{th} + \epsilon_k$ with $\epsilon_k \sim N(0, \sigma^2)$. Using this error specification, the likelihood of the observations provided by a given choice list is

$$\begin{aligned} p(y_k^\star \in [y_k^-; y_k^+]) =& p(\epsilon_k \in [y_k^- - y_k^{th}(\theta, X_k);\ y_k^+ - y_k^{th}(\theta, X_k)] \\ =& \Phi(\frac{y_k^+ - y_k^{th}(\theta, X_k)}{\sigma}) - \Phi(\frac{y_k^- - y_k^{th}(\theta, X_k)}{\sigma}) \\ =& l(\theta | y_k^+, y_k^-, X_k) \end{aligned}$$

This equation defines the likelihood of the vector of coefficients to be estimated, given the observations provided by choice lists and exogenous variables.

For a given individual i, the likelihood of a series of responses to choice lists (indexed by k), for each rounds (indexed by j), writes

$$l_i(\theta) = \prod_j \prod_k l(\theta | y_{i,j,k}^+, y_{i,j,k}^-, X_{i,j,k})$$

We estimate the vector of coefficients θ by maximizing the sum of log-likelihoods over individuals: $LL(\theta) = \sum_i l_i(\theta)$. This log-likelihood function is maximized by the BFGS algorithm.[5] In order to account for heterogeneity in individual error terms across rounds, specific error variances are estimated for each round. Inference is based on the (subjects) clustered standard-errors, computed from the variance-covariance matrix of individual scores.

6.3 Results

This section presents the estimated indices of deviation from Bayesian updating, and their explanatory variables. The results of the estimations are presented in Table 2.

[5]In order to avoid local maxima, for each estimations, suitable starting values were computed using grid search over 1000 possible vectors. After convergence, 10 additional estimations were run around estimated coefficients

Coefficients	No Explanatory	Rounds	Rounds and Information	Round, information and subject's characteristics
σ_1	8.221 (0.086)***	8.133 (0.085)***	8.192 (0.095)***	8.184 (0.094)***
σ_2	12.143 (0.204)***	12.121 (0.204)***	12.237 (0.213)***	12.081 (0.211)***
σ_3	13.782 (0.27)***	13.696 (0.266)***	13.331 (0.263)***	13.496 (0.283)***
σ_4	13.389 (0.255)***	13.386 (0.265)***	13.229 (0.259)***	13.285 (0.28)***
α_1	1.449 (0.03)***	1.451 (0.031)***	1.485 (0.035)***	1.46 (0.034)***
β_1	1.368 (0.029)***	1.369 (0.03)***	1.378 (0.033)***	1.363 (0.034)***
$p_{Intercept}$	0.648 (0.009)***	0.673 (0.012)***	0.43 (0.053)***	0.374 (0.075)***
$q_{Intercept}$	0.166 (0.019)***	0.18 (0.028)***	0.414 (0.107)***	0.272 (0.17)ns
$p_{round=3}$		-0.09 (0.024)***	-0.256 (0.03)***	-0.119 (0.024)***
$p_{round=4}$		-0.01 (0.043)ns	-0.103 (0.069)ns	0.05 (0.063)ns
$q_{round=4}$		0.057 (0.075)ns	-0.339 (0.212)ns	-0.502 (0.108)***
p_{ig}			2.895 (0.263)***	1.435 (0.362)***
p_{ig^2}			-4.294 (0.358)***	-1.846 (0.433)***
q_{ig}			0.177 (0.568)ns	0.735 (3.198)ns
q_{ig^2}			-0.867 (0.711)ns	-2.984 (9.529)ns
$p_{gender=female}$				0.069 (0.014)***
$p_{major=Econ}$				0.085 (0.021)***
$q_{gender=female}$				0.047 (0.06)ns
$q_{major=Econ}$				0.156 (0.061)*

Clustered standard errors are reported between brackets.
Stars report significance levels: ns for $p \geq 0.05$, $*$ for $p < 0.05$, $**$ for $p < 0.01$, $***$ for $p < 0.001$.

Table 2: Results of Econometric Estimations

Whatever the set of explanatory variables, the parameters α and β characterizing priors at round 1 and before receiving any signal, had very similar estimates: 1.5 and 1.4. The similarity of these two values suggests that the belief distribution of our representative subject was symmetrical. Consistently with the provided instructions, subjects did not expect one color to be more likely than the other, before receiving signals. It is nevertheless worth to note that priors were not perfectly uniform either, they exhibited a smaller variance and give sightly more probability weight to central than to extreme values of the [0,1] interval.

The first model introduces overall measures of conservatism bias (p) and confirmatory bias (q) for our representative subject. Both indices differed from 0. According to the estimated values, subjects exhibited a pronounced tendency to conservatism: they behaved as if they neglected 65% of the sample size of actual signals. Evidence for confirmatory bias was also observed: subjects behaved as if they misinterpreted 17% of signals contradicting their beliefs.

The other models (columns 2 to 4) enrich the analysis by introducing explanatory variables for the bias indices. When allowing biases to vary across rounds, we observed that conservatism bias was smaller for round 3 than for round 2, but the dummy variable for round 4 was not significant. For confirmatory bias, no significant differences were observed between rounds 3 and 4.[6]

The model of column 3 accounts for heterogeneity across rounds and across sessions by including information gain as an explanatory variables. Exploratory analysis suggested that the impact of information gain might be non-linear, and therefore a polynomial effect was considered. Information gain was not found to impact confirmatory bias, but impacted conservatism bias significantly. The coefficients associated to the two degrees of the polynomial were significant and suggested that the relationship was not monotonic, but inverse-U shaped. The shape of the estimated effect is represented in Figure 8. Moderately informative signals increased the biases, whereas very poorly or highly informative ones reduced them. It is noteworthy that very surprising signals were able to reverse the sign of the conservatism index, possibly leading to prior signals destruction.

[6] The q index was not estimated for round 2 because people hold (approximately) symmetric beliefs in round 1.

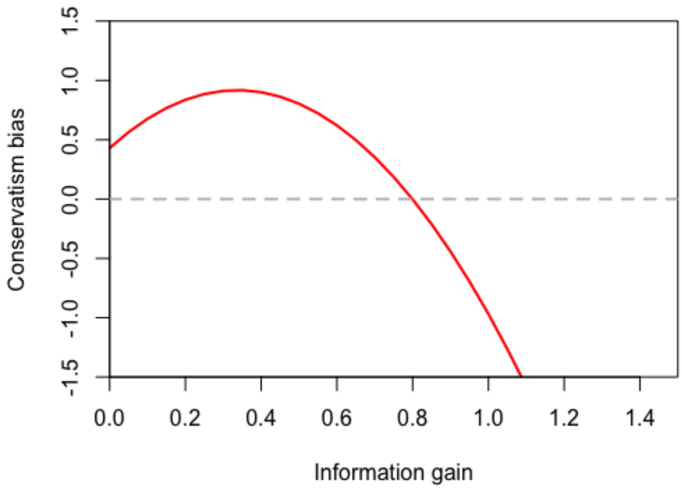

Figure 8: Polynomial effect of information gain on p

.

6.4 Stability of results across measurement methods

Two different methods were used to measure beliefs. Estimations presented in the previous section pool observations from the two methods, assuming that they give similar patterns. We also tested this assumption.

The first and last models from Table 2 were re-estimated with all explanatory variables interacting with a dummy variable coding for the method used for estimations. For the simple model containing only intercept values of p and q, the dummy variable was found to have a significant impact on p. When measured with the exchangeability method (that is robust to ambiguity attitudes), the index of conservatism bias was lower by 0.133 (clustered standard error: 0.026). Regarding the confirmatory bias, the difference between measurement methods was estimated as 0.07 (clustered standard error: 0.10) but was not statistically significant.

The aforementioned result suggest that ambiguity can amplify conservatism. Their posterior beliefs, measured with matching probabilities, remain too diffuse. Controlling for ambiguity by the use of exchangeable events leads to less deviations from Bayesian updating.

The dummy variable denoting the measurement method was also included in the model with explanatory variables for biases (column 4), and possible interactions were allowed. The estimated model contains 32 coefficients. A likelihood test was run to check whether adding interactions between dummy variable for method and other explanatory variables increased the likelihood significantly. The p-value of the test is 0.07. This suggests that allowing for coefficients to interact with the method dummy does not increase the goodness of fit significantly. Therefore, the coefficients of the explanatory variables do not vary significantly with the measurement method.

7 Discussion

This paper models belief updating when a combination of conservatism and confirmatory bias may distort people's perception of signals received, thus incorporating new information insufficiently or asymmetrically. Our model provides an intuitive interpretation of the biases and makes them observable from revealed preferences. It extends Rabin and Schrag's 1999 model by accounting for more patterns of deviations from Bayesian updating.

The experiment illustrated how the indices could be estimated in a tractable manner. Thus, it provided the first structural estimation of the two well-known biases. The results showed evidence for both confirmatory bias and conservatism at aggregate level. On average, the confirmatory bias index was estimated as 0.17 suggesting that an opposite signal may be misread with 17% chance. The conservatism index was 0.65 suggesting a strong stickiness to priors. Furthermore, the conservatism bias also depended on the informativeness of signals as measured by how surprising the signals were given the prior beliefs. In particular, subjects were more conservative when the new signals were neither extremely surprising nor extremely non-surprising. This pattern is consistent with the previous findings indicating the tradeoff between the strength and the credibility of evidence (Griffin and Tversky, 1992; Massey and Wu, 2005).

Our experiment contributes to the empirical investigation of confirmatory bias. Despite the abundance of theoretical models on confirmatory bias in economics literature, the main empirical findings for confirmatory bias mainly come from the psychology literature (for reviews, see Klayman (1995); Nickerson (1998); Oswald and Grosjean (2004)). However, the subjective nature of the psychological experiments do not allow a formal investigation of confirmatory bias due to the lack a normative benchmark for comparison of revised beliefs. Although there are a few field studies documenting evidence on confirmatory bias (Andrews et al., 2015; Sinkey, 2015; Christandl et al., 2011), there is still lack of evidence in standard Bayesian updating experiments. Several recent studies document evidence on asymmetric processing of information in Bayesian updating as in confirmatory bias, when the information has a valence or it is self-relevant (Coutts et al., 2016; Eil and Rao, 2011; Ertac, 2011). Different from our ego-neutral setting, these studies employ ego-related settings where subjects make inferences about their scores on intelligence tests or their physical attractiveness rated by other subjects in the same experimental session. Eil and Rao 2011 argue that confirmation of prior beliefs happens only when the confirming evidence supports a positive ego image. Specifically, people are more responsive to positive feedback compared to negative feedback about themselves regardless of their prior beliefs. Our results show that confirmatory bias can also arise in an ego-neutral setting. The direction of the bias, however, then depends on the informativeness of signals.

8 Conclusion

This paper studies biases in people's belief updating from a descriptive perspective. We modeled deviations from Bayesian updating by allowing perceived signals to differ from the signals people actually receive. It provides a natural interpretation of well-known biases and makes them observable from choices. Our model thus adheres to the revealed-preference approach of economics.

In our experiment, confirmatory bias and conservatism were dominant at the aggregate level, while individual heterogeneity persisted. The opposite of conservatism arose in situations where the signals were extremely surprising. This finding illustrates the relevance of allowing for different deviation patterns. Overall, our results replicated previous findings on Bayesian updating, suggesting that our model and the method are empirically valid.

Appendix

A. Detailed experimental procedure

Every subject received a subject ID upon arrival. In each session the subject whose ID started with M was invited to the front and introduced to all subjects as the implementer of that session. The implementer was then guided to a desk at the rear end of the room isolated by a wooden panel. The implementer would implement the randomization tasks to make sure that they were conducted in a fair and transparent manner.

Each session started with oral instructions by one of the experimenters – the instructor – using slides. Throughout the experiment, subjects could ask questions when anything was unclear. A training wheel was used during the instructions for illustration purpose. The training wheel was covered by blue and red, instead of brown and yellow to avoid potential misunderstandings and biases. The implementer first confirmed that the training wheel hidden behind the panel was covered by brown and brown, and there were no other colors on the wheel. He then spinned the wheel three times and reported the resulting colors. These colors were written down on the white board so that all subjects could see during the instruction. Subjects then received a training questionnaire with all choice situations that they would face during the experiment. The instructor went through them with the subjects, and the subjects filled in the training questionnaires based on the sample information from the practice wheel as a practice.

After all subjects were familiarized with the experimental tasks, the instructor explained to the subjects how their final payment would be determined with an example envelope content. The oral instructions ended with the explanation of the structure of the experiment.

After the instructions and before the start of the actual experiment, each subject drew a sealed envelope and the implementer randomly drew a period number from 1 to 4. Then, the implementer randomly drew a card from the deck of four cards. The selected period number and the card were sealed in two envelopes and only revealed at the end of the experiment. The implementer then drew a color composition for the wheel. He confirmed to all subjects that the wheel was covered by two and only two colors: yellow and brown.

Before handing out the questionnaires for the first choice period, each subject could state his preference between betting on yellow proportion and betting on brown proportion during the experiment. He received questionnaires with that color throughout the experiment. The subjects were requested to write their subject IDs on every questionnaire that they filled in so that their choices could be tracked down over the periods. The questionnaires were collected at the end of every choice period, and the sampling period proceeded. The outcome of every spin were announced by the implementer, and written down on the white board by the experimenter. New questionnaires were handed out after each sampling period.

At the end of the experiment, the color composition of the wheel, the card suit, and the choice period drawn for the payment stage was revealed to the subjects by the implementer. The subjects were requested to open their envelopes, and to proceed to the payment desk, where they got paid according to the outcome of their preferred lottery in the choice question that came out of their envelopes.

References

Abdellaoui, M., Baillon, A., Placido, L., and Wakker, P. P. (2011). The rich domain of uncertainty: Source functions and their experimental implementation. *The American Economic Review*, 101(2):695–723.

Abdellaoui, M., Bleichrodt, H., Kemel, E., and L'Haridon, O. (2014). Beliefs and attitudes for natural sources of uncertainty. Technical report, Working paper.

Andrews, R. J., Logan, T. D., and Sinkey, M. J. (2015). Identifying confirmatory bias in the field evidence from a poll of experts. *Journal of Sports Economics*, page 1527002515617511.

Baillon, A. (2008). Eliciting subjective probabilities through exchangeable events: An advantage and a limitation. *Decision Analysis*, 5(2):76–87.

Bénabou, R. and Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, 117(3):871–915.

Bénabou, R. and Tirole, J. (2006). Belief in a just world and redistributive politics. *The Quarterly Journal of Economics*, 121(2):699–746.

Bénabou, R., Tirole, J., et al. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, 30(3):141–64.

Chew, S. H. and Sagi, J. S. (2008). Small worlds: Modeling attitudes toward sources of uncertainty. *Journal of Economic Theory*, 139(1):1–24.

Christandl, F., Fetchenhauer, D., and Hoelzl, E. (2011). Price perception and confirmation bias in the context of a vat increase. *Journal of Economic Psychology*, 32(1):131–141.

Coutts, A. M., White, E. C., Tapper, N. J., Beringer, J., and Livesley, S. J. (2016). Temperature and human thermal comfort effects of street trees across three contrasting street canyon environments. *Theoretical and Applied Climatology*, 124(1-2):55–68.

De Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. In *Annales de l'institut Henri Poincaré*, volume 7, pages 1–68.

Edwards, W. (1968). Conservatism in human information processing. *Formal representation of human judgment*, 17:51.

Eil, D. and Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2):114–138.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The quarterly journal of economics*, pages 643–669.

Epstein, L. G. (2006). An axiomatic model of non-bayesian updating. *The Review of Economic Studies*, 73(2):413–436.

Ertac, S. (2011). Does self-relevance affect information processing? experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior & Organization*, 80(3):532–545.

Griffin, D. and Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive psychology*, 24(3):411–435.

Holt, C. A. (2007). *Markets, games, and strategic behavior: recipes for interactive learning.* Pearson Addison Wesley.

Johnson, C., Baillon, A., Bleichrodt, H., Li, Z., van Dolder, D., and Wakker, P. P. (2015). Prince: An improved method for measuring incentivized preferences. *Available at SSRN.*

Karni, E. (2009). A mechanism for eliciting probabilities. *Econometrica*, 77(2):603–606.

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of learning and motivation*, 32:385–418.

Massey, C. and Wu, G. (2005). Detecting regime shifts: The causes of under-and overreaction. *Management Science*, 51(6):932–947.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2):175.

Oswald, M. E. and Grosjean, S. (2004). Confirmation bias. In *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory*, chapter 4. Psychology Press.

Phillips, L. D. and Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of experimental psychology*, 72(3):346.

Rabin, M. (2013). An approach to incorporating psychology into economics. *The American Economic Review*, 103(3):617–622.

Rabin, M. and Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. *Quarterly journal of Economics*, pages 37–82.

Raiffa, H. (1968). *Decision Analysis: Introductory Lectures on Choices Under Undertainty.* Addison-Wesley.

Ramsey, F. P. (1931). Truth and probability (1926). *The foundations of mathematics and other logical essays*, pages 156–198.

Schotter, A. and Trevino, I. (2014). Belief elicitation in the laboratory. *Annu. Rev. Econ.*, 6(1):103–128.

Sinkey, M. (2015). How do experts update beliefs? lessons from a non-market environment. *Journal of Behavioral and Experimental Economics*, 57:55–63.

Spetzler, C. S. and Stael von Holstein, C.-A. S. (1975). Exceptional paper-probability encoding in decision analysis. *Management science*, 22(3):340–358.

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

Wilson, A. (2014). Bounded memory and biases in information processing. *Econometrica*, 82(6):2257–2294.