How much do we learn?
Measuring symmetric and asymmetric deviations from Bayesian updating through choices

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

[a] *IESEG School of Management Lille* [†]

[b] *Erasmus School of Economics - Erasmus University Rotterdam*[‡]

[c] *GREGHEC - CNRS and HEC Paris*[§]

February 14, 2022

Abstract

Belief updating biases hinder the correction of inaccurate beliefs and lead to sub-optimal decisions. We complement Rabin and Schrag's (1999) portable extension of the Bayesian model by including conservatism in addition to confirmatory bias, and we show how to identify these two forms of biases from choices. In an experiment, we found that people exhibited confirmatory bias by misreading 18% of the signals contradicting their priors. They were also conservative and acted as if they missed 61% of the signals.

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

1 Introduction

Beliefs are the basis of our actions, and the way we process information shapes our beliefs. However, belief updating biases are known to prevent us from optimally learning from information,[1] leading

*Corresponding author: baillon@ese.eur.nl

†Department of Economics and Quantitative Methods, 65 Rue Philippe-Laurent Roland, 59000 Lille, France

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

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

[1] See for instance Phillips and Edwards (1966); Edwards (1968); Tversky and Kahneman (1974); El-Gamal and Grether (1995); Oswald and Grosjean (2004); Mobius et al. (2014); Bénabou and Tirole (2016); Ambuehl and Li (2018).

to biased beliefs that underlie societal issues such as global warming (Deryugina, 2013; Howe and Leiserowitz, 2013) and gender inequality (Sarsons, 2017; Bohren et al., 2019). Despite the accumulating evidence on belief-updating biases, gauging the extent to which people are biased and the resulting economic consequences remains challenging due to the gap between theoretical models of empirical evidence. Most existing non-Bayesian belief models (Epstein, 2006; Wilson, 2014), although capturing the empirical phenomenon under consideration with rigor and elegance, are too complex to be empirically estimated or contain theoretical constructs that are not easily observable. For instance the model of Benjamin et al. (2016) can explain how beliefs can be in contradiction with the law of large numbers, but it requires additional information about the timing of signals (Rabin, 2013; Benjamin, 2019) and is therefore not a portable extension of the Bayesian model.[2] As a result of these limitations, empirical evidence are either qualitative classifications (El-Gamal and Grether, 1995) or reduced-form (Mobius et al., 2014; Coutts, 2019; Charness and Dave, 2017).

We propose a portable model, based on Rabin and Schrag (1999), and an empirical approach providing structural estimation of biases from choices. The model has two parameters, which can be easily interpreted, plugged in any theoretical work using Bayesian updating, and estimated from choice data.

The two parameters capture two forms of deviations from Bayesian updating. The first type captures asymmetry in how decision makers incorporate confirming and contradicting information conditional on their priors. The most prominent example of this kind is the confirmatory bias (Rabin and Schrag, 1999), in which people tend to neglect or even misinterpret signals contradicting their prior beliefs. Although less common, the disconfirmatory bias also exists, especially when disconfirming is beneficial and enhances one's self perception, leading to inaccurate but motivated beliefs (Bénabou and Tirole, 2002, 2006; Eil and Rao, 2011; Bénabou and Tirole, 2016). The second type captures the extent to which decision makers overweigh or underweigh new information relative to their priors. Common biases of this type include conservatism (Phillips and Edwards, 1966; Edwards, 1968) which leads to underweighing of new information, and over-inference which corresponds to overweighing of new information due to, for instance, the representative heuristics (Tversky and Kahneman, 1974; Bordalo et al., 2016).

Both types of biases are present in most decisions, yet existing studies mostly focus on specific biases relevant for decision situations under consideration.[3] We extend Rabin and Schrag (1999)'s

[2]See Rabin (2013) for the definition of a portable extension of an existing model. Benjamin (2019) argues that, unlike the model of Rabin and Schrag (1999), that of Benjamin et al. (2016) is not portable because the ordering and timing of signals is needed, unlike in traditional updating. Timing of updating may matter in the model of Rabin and Schrag (1999) if the hypothesis favored by the prior changes, but the updating is assumed to take place when decisions are made, which preserves the portability of the model.

[3]A notable exception is Charness and Dave (2017). See also Benjamin (2019) for a review of virtually all belief biases.

model to separate and quantify both asymmetric and symmetric biases. Our separation strategy taps into the following conceptual difference between the two types of biases. While asymmetric biases, such as the confirmatory bias, make people overweight one type of evidence over the other, symmetric biases affect the weighting of the sum of evidence with no further distinction. This distinction is crucial in our identification strategy. Our method also allows for individual heterogeneity without pre-committing to the direction of biases, and hence, achieves increased descriptive validity. We implement our approach without making assumptions about people's prior information, showing that it can be applied in situations where researchers have no control of prior beliefs.

Our method remains parsimonious. The model has two parameters, one for each type of biases. We refer to the first index as the *confirmatory bias index* (q) and the second index as the *conservatism index* (p), following the index naming convention by using the more common pattern within each type.[4] In the special case where both indexes are between 0 and 1, the confirmatory bias index can be interpreted as the probability of misreading a contradicting signal as confirming, whereas the conservatism index as the probability of missing a new signal. These interpretations, though convenient, are not necessary for our approach to remain valid. The indexes can also be interpreted as overweighting and underweighting indexes, distorting either the balance of evidence (asymmetric biases) or the weight assigned to the sum of evidence (symmetric biases) in the updating process. Furthermore, we show how the method can be adapted to other asymmetric biases, such as the self-serving bias (Miller and Ross, 1975; Mezulis et al., 2004).

Our experiment concerns an ego-neutral learning environment that does not relate to individuals' self-image. We found that people exhibited confirmatory bias by misreading 18% of the signals contradicting their prior beliefs. Also they were conservative and acted as if they missed approximately 61% of the signals they received. Our findings suggest that deviations from Bayesian updating occurs even in the absence of clear motivation such as enhancing one's self-perception. Our method further allowed us to obtain the complete prior and posterior distributions, and not only point estimates, showing that even though initial priors were mostly symmetric, they were not uniform. It also enabled us to uncover individual heterogeneity, where 22% exhibited over-inference and 3% the disconfirmatory bias.

[4] For instance, it is also conventional to refer to the index capturing risk attitudes as the risk aversion index, even though it also captures risk seeking and risk neutrality.

2 Modelling biases

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 assume that the decision maker's prior belief Λ about s, defined over $(0,1)$, follows a Beta distribution $Beta(\alpha_0, \beta_0)$, where a uniform prior corresponds to $\alpha_0 = \beta_0 = 1$. Our approach can be applied to any distribution with a conjugate prior whose parameters can be expressed as a function of the received signals. The Poisson distribution and its Gamma conjugate prior or the multinominial distribution with its Dirichlet conjugate prior are examples with a discrete support. The beta family is both natural (Moreno and Rosokha, 2016) and tractable (Abdellaoui et al., 2014) to model beliefs over a success rate. Beta distributions are also flexible and can take a wide array of shapes with different locations and dispersion for different parameters.

The parameters of the Beta distribution can be directly interpreted in terms of signal samples, allowing us to summarize the decision maker's beliefs as samples of signals. 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. The uniform case ($\alpha_0 = \beta_0 = 1$) means that the decision maker knows that both successes and failures may happen, and assigns equal probability mass to all non-zero probabilities of successes or failures. Departures from uniformity in prior beliefs are modeled by (possibly hypothetical) signals in the decision maker's mind. 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 $\alpha_0 = \beta_0$.

After receiving a sequence of signals, his *posterior belief* becomes $Beta(\alpha_1, \beta_1)$. Under Bayesian updating, every single observation of success (failure) increments the first (second) parameter of the beta distribution by one, no matter what the initial parameters were. 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. Following Rabin and Schrag (1999), we call η the *perceived number of signals*, α the *perceived number of successes*, and β the *perceived number of failures.*

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 $\eta = n$. This does not hold 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: asymmetric vs. symmetric belief updating biases. Asymmetric biases distort the relative proportions of successes and failures, i.e. distort the actual sample means $\frac{a}{a+b}$ (and $\frac{b}{a+b}$) into perceived sample means $\frac{\alpha}{\alpha+\beta}$ (and $\frac{\beta}{\alpha+\beta}$). Symmetric biases affect all signals without distinguishing between successes and failures, i.e. distort n into η. Conceptually, the asymmetric biases change the balance of evidence by distorting the sample mean whereas the symmetric biases affect the weight assigned to the sum of evidence.

In subsection 2.2 we first introduce an index for the most well-known asymmetric bias - the confirmatory bias, and its negative counterpart capturing the disconfirmatory bias. We further show how this index can be adapted to capture other types of asymmetric biases such as the self-serving bias. In subsection 2.3 we introduce an index for symmetric biases such as conservatism. In subsection 2.4, we combine both asymmetric and symmetric biases into one model.

2.2 Asymmetric biases

We start with *confirmatory bias*. Following Rabin and Schrag (1999), we model it as the probability q_c to misread a contradicting signal as confirming prior expectations. For a decision maker who believes that successes are more likely than failures (i.e. $\alpha_0 > \beta_0$), confirmatory bias implies:

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

Although, we adopt Rabin and Schrag's way of modelling, we are not committed to their interpretation of misreading disconfirming signals. We can also interpret q_c as a parameter that captures how much the decision maker distorts the weightings of confirming and disconfirming signals in his sample. In this case, the decision maker discounts the observations of failure signals by q_c (as they are disconfirming his prior expectations) and assigns excess weight to success signals, distorting their relative frequency by $\frac{\alpha}{\alpha+\beta} - \frac{a}{a+b} = q_c \frac{b}{a+b}$. If the decision maker expects failures to be more likely (i.e. $\alpha_0 < \beta_0$), then

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

and the relative frequency of failures is distorted by $\frac{\beta}{\alpha+\beta} - \frac{b}{a+b} = q_c \frac{a}{a+b}$. Again, the decision maker assigns disproportionately more weight to confirming signals (failures) where the extra-weight increases with q_c.

Next we extend Rabin and Schrag's model to include the opposite bias, that we call *disconfirmatory bias*. It can be modelled as the probability q_d to misread a confirming signal as contradicting prior expectations. This means

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

if $\alpha_0 > \beta_0$, yielding underweighting of success signals $(-q_d \frac{a}{a+b})$ and

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

when $\alpha_0 < \beta_0$, yielding overweighting success signals $(\frac{q_d \times b}{a+b})$.

From observing perceived signals, either q_c or q_d can be determined whenever $\alpha_0 \neq \beta_0$. Consider the case where successes are believed to be more likely than failures, i.e. $\alpha_0 > \beta_0$. If, after observing the signals, the perceived number of successes is revealed to be greater than the actual number of successes ($a \leq \alpha$), this suggest 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.* Moreover, we can combine q_c and q_d into a unique index of confirmatory bias q defined as:

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

Figure 1 depicts all possible cases when $\alpha_0 > \beta_0$. The corresponding figure for $\beta_0 > \alpha_0$ can be obtained by replacing α by β and a by b in Figure 1. Values of q in $[0, 1]$ can be directly interpreted as probabilities to misread a 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 8 reports the estimated distribution for our experimental subjects.

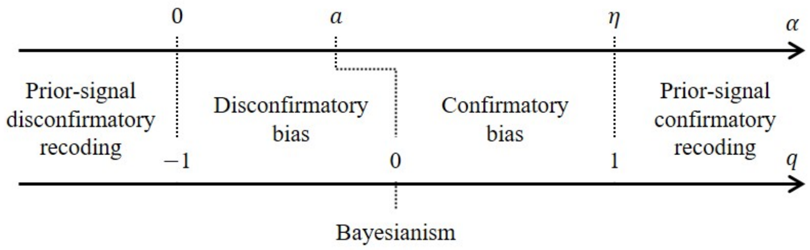

Figure 1: Interpretation of q and relationship with α when $\alpha_0 > \beta_0$. The case $q = 0$ corresponds to $\alpha = a$. The edge on the graph indicates that a need not be at equal distance to 0 and η.

Whereas quantifying (dis)confirmatory bias is the main focus of this paper, our way to model (dis)confirmatory bias can also be adapted to quantify other types of asymmetric biases. For instance, the self-serving bias can be modelled as the probability (q_s) to misread signals that damage self-image or confidence as self-enhancing ones. Let a, b and n denote the number of self-enhancing, self-harming, and total actual signals, and α, β, and η denote the perceived ones. Self-serving bias results in:

$$\begin{cases} \alpha = a + q_s b \\ \beta = (1 - q_s) b \end{cases}, \tag{6}$$

independent of the decision maker's prior beliefs. In ego-relevant decision situations, both confirmatory and self-serving biases may play a role. In our experiment, we focus on an ego-neutral decision situation, where self-serving biases or motivated beliefs are unlikely to arise, to derive a benchmark level of confirmatory bias. We leave the quantification of other asymmetric biases for future research.

2.3 Symmetric biases

We next consider symmetric biases, where a decision maker's tendency to overweight or underweight signals, regardless of whether they are successes or failures. Such an approach, in line with Rabin and Schrag (1999), was also used by Moreno and Rosokha (2016), who compared perceived signals with actual signals to study conservatism.

We start with a conservative decision maker who 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, the decision maker perceives on average only $1 - p$ of all actually received signals, i.e. $\eta = (1 - p)n$. The conservatism bias affects 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, when $p > 1$, even though it cannot be interpreted as a probability, it still has meaningful empirical interpretation. 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. We call this case *prior signal destruction.* Although it is commonly assumed that more information makes people having more precise beliefs, in real life, experience may be surprising, causing people to be less sure of their (unjustifiably) narrow priors. In our model, this phenomenon is translated into negative signals, which indicate doubts about prior beliefs.

By contrast, $p < 0$ means that, contrary to being conservative and underweighting signals, the decision maker placed too much weights on the signals. This can be illustrated by making use of the properties of the mean of Beta distributions. The posterior mean $\frac{\alpha_0+\alpha}{\eta_0+\eta}$ can be decomposed in terms of prior mean and sample mean:

$$\begin{aligned}\frac{\alpha_0+\alpha}{\eta_0+\eta} &= \frac{\alpha_0+(1-p)a}{\eta_0+(1-p)n} \\ &= \frac{\eta_0}{\eta_0+(1-p)n}\cdot\frac{\alpha_0}{\eta_0}+\frac{(1-p)n}{\eta_0+(1-p)n}\cdot\frac{a}{n} \\ &= \frac{\eta_0}{\eta_0+(1-p)n}\cdot \text{ prior mean } + \frac{(1-p)n}{\eta_0+(1-p)n}\cdot \text{ sample mean}\end{aligned} \quad . \tag{7}$$

A negative p corresponds to *over-inference*, where 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.

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, where the decision maker underweights the sample information and overweights the prior information. Note that conservatism is modelled as a deviation from Bayesianism that is linear in terms of not only in signals but also of means. The latter is also an implication of the model of Kovach (2021), in which conservatism is represented as a linear combination of the prior and the Bayesian posterior.

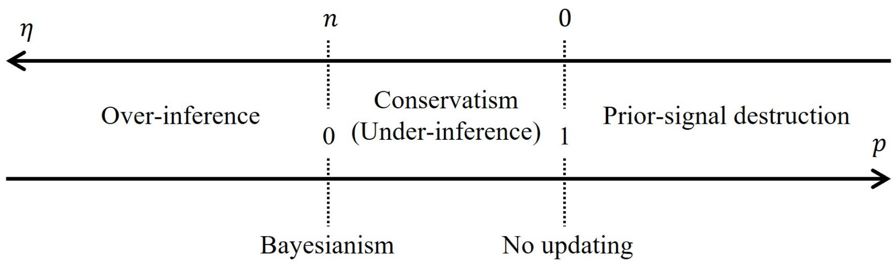

Figure 2: Interpretation of p and relationship with η

Figure 2 depicts the relationship between the perceived number of signals 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 also misread those he did not miss (confirmatory bias). For $\alpha_0 > \beta_0$, replacing a and b in equations of (1) respectively by $(1-p)a$ and $(1-p)b$, the confirmatory bias in presence of conservatism bias gives (replacing q_c by q):

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

Similarly, replacing a and b in equations of (2) respectively by $(1-p)a$ and $(1-p)b$, disconfirmatory bias with conservatism gives (replacing q_d by $-q$)

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

The case $\alpha_0 < \beta_0$ is symmetric. If successes and failures were equally likely according to the decision maker's prior belief, confirmatory bias plays no role:

$$\begin{cases} \alpha = (1-p)a \\ \beta = (1-p)b \end{cases} . \tag{10}$$

In terms of observability, if $\alpha_0 \neq \beta_0$ and $\alpha + \beta \neq 0$, then q can be obtained by comparing $\frac{\alpha}{\alpha+\beta}$ and $\frac{a}{a+b}$. Further, p can always be obtained by comparing η with n. Note that, p does not affect the estimation of q because symmetric biases do not distort the perceived relative frequencies but only the relative weightings of prior and sampling information.

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's 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 beliefs 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 pros and cons and are therefore implemented in our experiment. Both methods give subjective probabilities without the influence of risk attitude, with V dropping out from the equations, as seen above. Hence, eliciting and correcting for risk attitudes are not necessary. 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 and makes use of an external device (the lottery), which may be confusing. Eliciting exchangeable events, which do not require the use of lotteries, is robust to this problem.

Before and after decision makers receive a set of signals, we elicit their priors and posteriors using the methods described above. We fit Λ with a beta distribution whose parameters are expressed as functions of our conservatism and confirmatory bias indexes using a system of structural equations.

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 21 and 28, summing up to 164 in total. In each session, one subject is randomly selected as the *implementer* of the session, who assisted the transparent and fair implementation of uncertainty resolution during the experiment. More details on the implementer's role are in appendix A. Excluding the implementers, we collected choice data from 157 subjects 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, the subjects faced choice situations that involved 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 the implementer.

The experiment consisted of alternating rounds of choice and periods of sampling (see Figure 3 for the flow). It started with round 0 in which subjects made choices without any knowledge about the color composition of the wheel. Then, the implementer spun 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 (round 1). The same procedure was repeated two more times, ending with choice round 3.

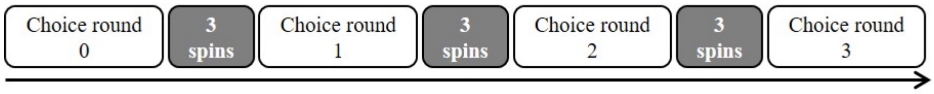

Figure 3: Experimental flow

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

4.2.1 Matching probability elicitation tasks

Figure 4 presents a choice list to elicit a matching probability. In each choice question, subjects had to choose between an option W(heel) and an option C(ard). The payoff of option W depended on the actual proportion of yellow (or brown) on the wheel being within a certain interval. For example, in line 2, it depended on the yellow proportion being between 0% and 4%.[5] The payoff of option C(ard) 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.

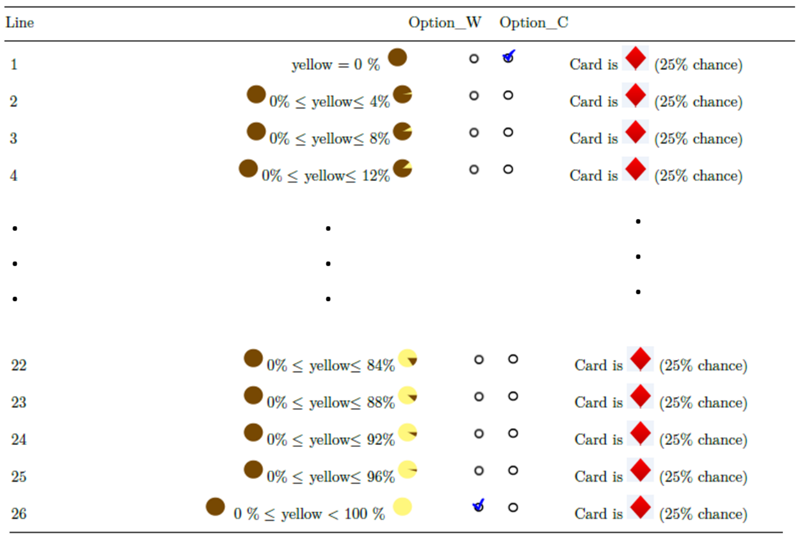

Figure 4: Choice list to elicit matching probabilities

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 yellow cannot be 0% (otherwise there is only

[5]To control for possible suspicion effects, each subject chose their own color to bet on in this option in the beginning of the experiment. For further details, see appendix A.

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 $y_{0.75}$, where the events $(0, y_{0.5}]$ and $(0, y_{0.75}]$ have 0.5 and 0.75 matching probabilities respectively. The choice lists for these elicitations 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 tasks

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 yellow 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.

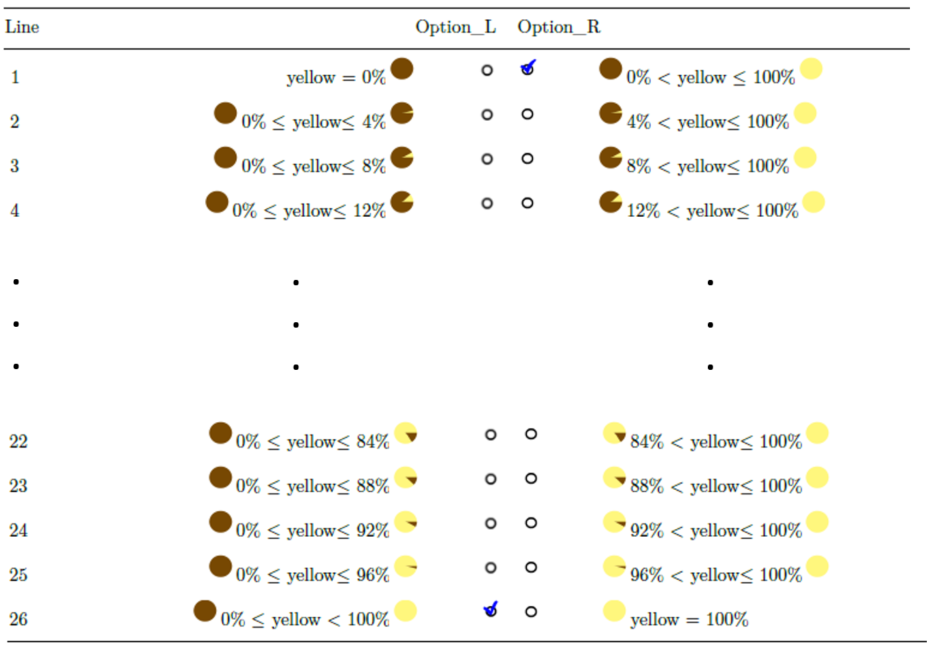

Figure 5: Choice list to elicit exchangeable events

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 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 $20_{(0,y^-_{\text{median}}]}0 \prec 20_{(y^-_{\text{median}},1)}0$ and $20_{(0,y^+_{\text{median}}]}0 \succ 20_{(y^+_{\text{median}},1)}0$. Therefore, for some $y_{\text{median}} \in [y^-_{\text{median}}, y^+_{\text{median}}]$, we have $20_{(0,y_{\text{median}}]}0 \sim 20_{(y_{\text{median}},1)}0$. 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). A summary of experimental elicitation is in Table 1.

Method	**Value of Elicitation**	**Indifference**	**Beliefs**
Matching Probability	$y_{0.25}$	$20_{(0,y_{0.25}]}0 \sim 20_{0.25}0$	$\Lambda((0, y_{0.25}]) = 0.25$
Matching Probability	$y_{0.5}$	$20_{(0,y_{0.5}]}0 \sim 20_{0.5}0$	$\Lambda((0, y_{0.5}]) = 0.5$
Matching Probability	$y_{0.75}$	$20_{(0,y_{0.75}]}0 \sim 20_{0.75}0$	$\Lambda((0, y_{0.75}]) = 0.75$
Exchangeable Events	y_{median}	$20_{(0,y_{\text{median}}]}0 \sim 20_{(y_{\text{median}},1)}0$	$\Lambda((0, y_{\text{median}}]) = \Lambda((y_{\text{median}}, 1))$
Exchangeable Events	y_{low}	$20_{(0,y_{low}]}0 \sim 20_{(y_{low},0.5]}0$	$\Lambda((0, y_{\text{low}}]) = \Lambda((y_{\text{low}}, 0.5])$
Exchangeable Events	y_{high}	$20_{[0.5,y_{high}]}0 \sim 20_{(y_{high},1)}0$	$\Lambda([0.5, y_{\text{high}}]) = \Lambda((y_{\text{high}}, 1))$

Table 1: Summary of Experimental Elicitation

4.2.3 Task orders

For each choice round, subjects received a separate questionnaire, each containing the same set of choice tasks summarized in Table 1. In each questionnaire, first the order of the two types of tasks (matching probabilities and exchangeable events), and then the order of the choice lists within each type were randomized.

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 round (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 list, 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 round would matter, the implementer randomly drew a number from one to four. Further details about the implementation are reported in the appendix.

5 Raw data

		Received signals between rounds:		
Session	# Subjects	0&1	1&2	2&3
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 2: Description of sessions

Out of the 157 subjects, ten subjects exhibited multiple switching patterns systematically in more than one choice list. We removed these subjects from the analysis as they exhibited consistent violations of monotonicity. Another five subjects exhibited a multiple switching pattern in only one out of six choice lists. These subjects were included in the analysis but only the inconsistent observations were discarded for them.

Table 2 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 from the choice lists 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 (i.e. y_{median} in the exchangeability method and y_{50} in the matching method) about the yellow proportion and the Bayesian benchmark.

A positive (negative) difference corresponds to an overestimation (underestimation) of the yellow proportion. In sessions with balanced signals (e.g. in sessions 2, 4 and 5), subjects' median beliefs did not deviate much from the Bayesian benchmark, however, in sessions where the subjects received extreme signals (e.g. in sessions 1 and 6), deviations were high. For instance, in session 1, subjects only received Brown signals. The median deviations in this section were positive, suggesting an overestimation of the yellow proportion on the wheel. This can be explained by conservatism as it suggests that the subjects did not incorporate the signals sufficiently. A similar pattern was observed in session 6 where the subjects only received yellow signals and underestimated the yellow proportion on the wheel.

Similarly, Figure 7 shows how the dispersion in subjects' beliefs, measured as $y_{\text{high}} - y_{\text{low}}$ with the exchangeability method and $y_{75} - y_{25}$ with 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 indexes while taking individual differences into account.

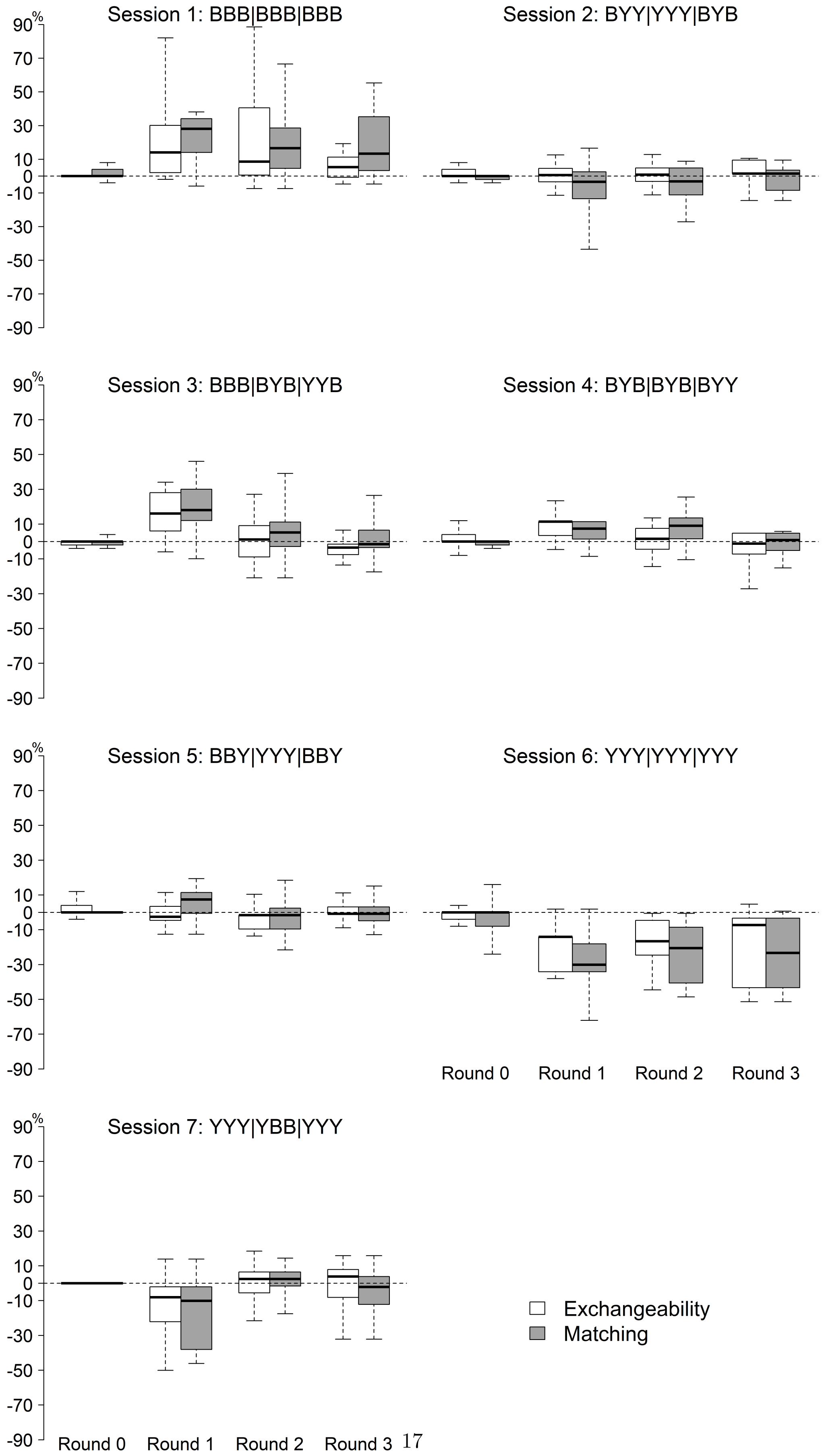

Figure 6: Deviation of Median Beliefs From Bayesian Benchmark

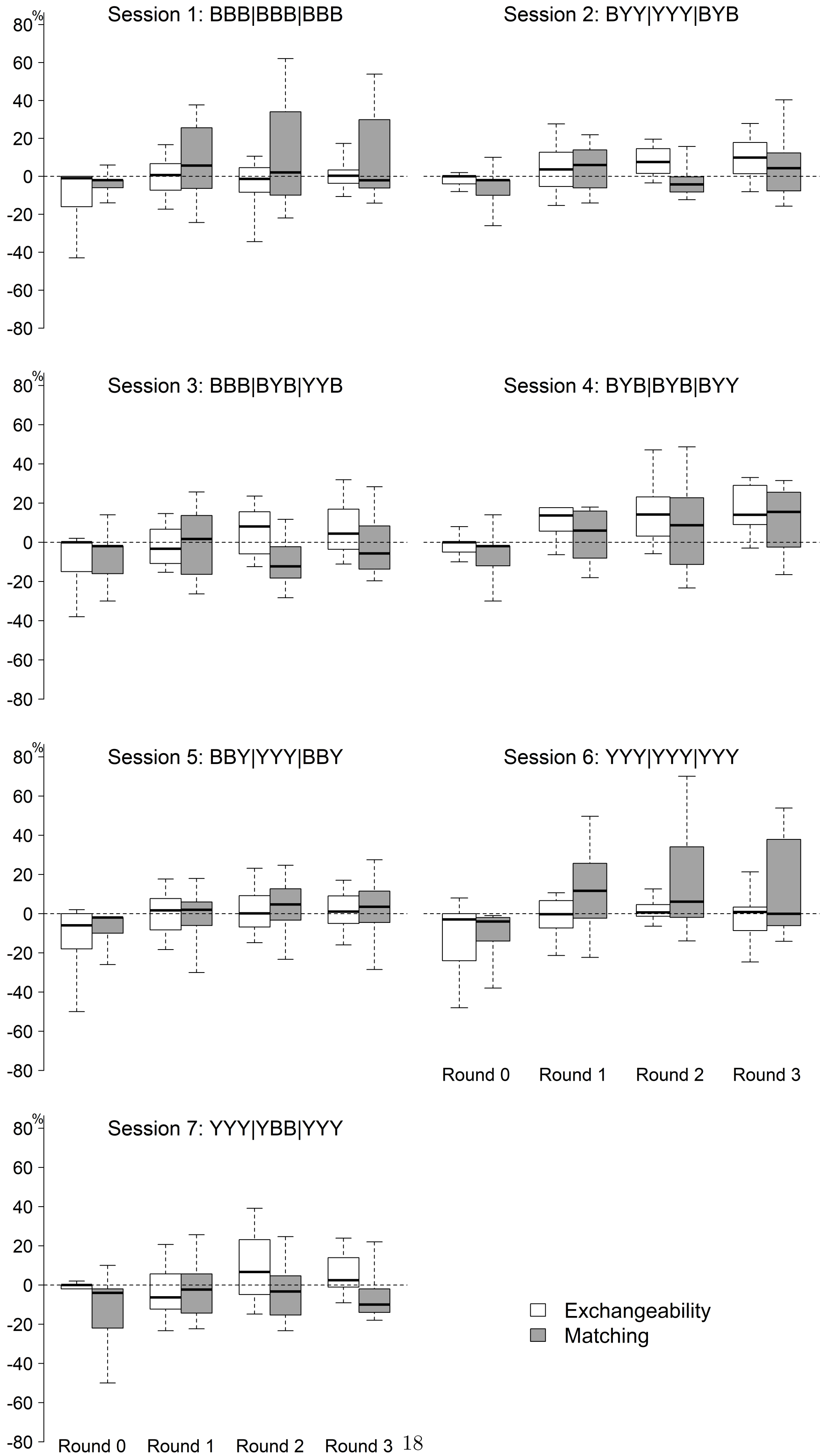

Figure 7: Deviation of Dispersion in Beliefs From Bayesian Benchmark

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 $\Lambda(.) = Beta(.|\alpha_{i,j}, \beta_{i,j})$. The prior of subject i at round 0, determined by $\alpha_{i,0}$ and $\beta_{i,0}$, is assumed to be exogenous and will be estimated. We assume that in round j, subjects uses their posteriors in round $j-1$ as priors. Then for rounds $j > 0$, we have:

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

where s and f are the functions that determine respectively the perceived successes and failures, as modeled by equations 8 to 10. These functions depend on the prior beliefs parameters $\alpha_{i,j-1}$ and $\beta_{i,j-1}$ in round j, the actual received signals $a_{i,j}$ and $b_{i,j}$ and the indexes of deviations from Bayesian updating, $p_{i,j}$ and $q_{i,j}$. For a Bayesian decision maker, $s(a_{i,j}, b_{i,j}, \alpha_{i,j-1}, \beta_{i,j-1}, p_{i,j}, q_{i,j}) = a_{i,j}$ and $f(a_{i,j}, b_{i,j}, \alpha_{i,j-1}, \beta_{i,j-1}, p_{i,j}, q_{i,j}) = b_{i,j}$. Prior to observing signals, subjects had no reason to believe one color is more likely than another. Since confirmatory bias plays no role when there is no asymmetry in the prior, we only estimate confirmatory bias in rounds 2 and 3. In order to account for heterogeneity in prior beliefs (at round 0) we allow parameters $\alpha_{i,0}$, $\beta_{i,0}$ to vary across subjects. We also allow for heterogeneity in deviations form Bayesian updating, by allowing indexes $p_{i,j}$ and $q_{i,j}$ to vary across subjects. Eventually, deviations may also vary from one round to another, due to learning or fatigue. We thus allow for $p_{i,j}$ and $q_{i,j}$ to vary between rounds.

In order to account for these three sources of variations, we assume the following structural equations

$$\begin{aligned} \alpha_{i,0} &\sim LN(\bar{\alpha}_0, \sigma_{\alpha_0}) \\ \beta_{i,0} &\sim LN(\bar{\beta}_0, \sigma_{\beta_0}) \end{aligned} \tag{11}$$

$$\begin{aligned} p_{i,j} &= p_i + \Delta_{p,2}\gamma_{j=2} + \Delta_{p,3}\gamma_{j=3} \\ \text{with } p_i &\sim N(\bar{p}, \sigma_p) \end{aligned} \tag{12}$$

$$q_{i,j} = q_i + \Delta_{q,3}\gamma_{j=3}$$
$$\text{with } q_i \sim N(\bar{q}, \sigma_q) \tag{13}$$

Parameters $\alpha_{i,0}$ and $\beta_{i,0}$ are non-negative and are assumed to be log-normally distributed with mean $\bar{\alpha}_0$ and $\bar{\beta}_0$ and standard deviation σ_{α_0} and σ_{β_0}. Individual parameters p_i and q_i are normally distributed with mean $\bar{p}$ and $\bar{q}$ and standard deviation σ_p and σ_q. Variables $\gamma_{j=k}$ are dummy variables that denote the round and coefficients Δ capture variations of indexes across rounds.

6.2 Estimating the model

Our set of structural equations define a non-linear random-parameter model. The model is estimated using simulated maximum likelihood (Train 2009). In what follows, we present the likelihood function that is used for the estimation.

Under our specification, the beliefs of a subject i at round j take the form of a probability distribution $\Lambda(.|\theta, X_{i,j})$ where θ is a vector of coefficients and $X_{i,j}$ contains the rounds, the received signals and the perceived signals at round $j-1$. Lighter notation $\Lambda(.)$ is used in the rest of this section. 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 λ_k and measures their corresponding values $y_k^\star$ such that $\Lambda\left((0, y_k^\star]\right) = \lambda_k$. More precisely, these choice lists determine two values y_k^- and y_k^+ such that $20_{(0,y_k^-]}0 \prec 20_{\lambda_k}0$ and $20_{(0,y_k^+]}0 \succ 20_{\lambda_k}0$ 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 $\Lambda((0, n_k]) - \Lambda((0, y_k^\star]) = \Lambda((0, y_k^\star]) - \Lambda((0, m_k])$, i.e. $\Lambda((0, y_k^\star]) = \frac{\Lambda((0,m_k])+\Lambda((0,n_k)]}{2}$. 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 $20_{[m_k,y_k^+]}0 \succ 20_{[y_k^+,n_k]}0$ 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 $y_{i,j,k}^{th}(\theta, X_{i,j,k})$ where θ is the vector of coefficients of our decision model, and $X_{i,j,k}$ is the set of variables containing choice lists characteristics and the round in which it was completed. In order to account for subject and/or specification errors, we assume that $y_{i,j,k}^\star = y_{i,j,k}^{th} + \epsilon_{i,j,k}$ with $\epsilon_{i,j,k} \sim N(0, \sigma_i^2)$. Using this error specification, the likelihood of the observations provided by a given choice list is

$$
\begin{aligned}
p(y^{\star}_{i,j,k} \in [y^{-}_{i,j,k}; y^{+}_{i,j,k}]) =& p(\epsilon_{i,j,k} \in [y^{-}_{i,j,k} - y^{th}_{i,j,k}(\theta, X_{i,j,k});\ y^{+}_{i,j,k} - y^{th}_{i,j,k}(\theta, X_{i,j,k})] \\
=& \Phi(\frac{y^{+}_{i,j,k} - y^{th}_{i,j,k}(\theta, X_{i,j,k})}{\sigma}) - \Phi(\frac{y^{-}_{i,j,k} - y^{th}_{i,j,k}(\theta, X_{i,j,k})}{\sigma_i}) \\
=& l(\theta | y^{+}_{i,j,k}, y^{-}_{i,j,k}, X_{i,j,k})
\end{aligned}
$$

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

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

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

In order to account for heterogeneity in behavior we assume that θ varies across individuals according to a multivariate distribution[6] of mean $\bar{\theta}$ and diagonal variance covariance matrix Ω_θ. In words, Ω_θ contains the variance of individual parameters and captures the heterogeneity in these parameters. The models are estimated by maximizing the sum of the log of simulated individual likelihoods, using 1000 Halton draws. Table 6.3 presents the model estimates. Model 1 is a representative-agent model that does not account for individual heterogeneity in model parameters. In model 2, heterogeneity in priors is also accounted for. Model 3 introduces heterogeneity in bias indices p and q. Model 4 augment model 3 with fixed effects capturing the evolution of the means of these indices across rounds: Δp_2 (Δp_3) measures the difference of mean in indice p between round 2 (3) and round 1 (2). Similarly, Δq_3 measures the difference of mean in indice q between round 3 and round 2.[7] For all the estimations, maximization is ran with the BFGS algorithm, and standard errors are computed from the cross-product of individual scores, thereby accounting for individual-level clustering.

6.3 Results

This section presents the estimated indexes of biases, and their variations among our sample of subjects. The results of the estimations are presented in Table 3.

Model 1 is a representative-agent model, where no between-individual heterogeneity is assumed

[6] Distributions are assumed to be log normal for non-negative parameters and normal for other parameters.

[7] If priors are symmetrical, there is no confirmation bias at round 2. This explains why a non-zero q is introduced for round 2 and 3 only.

		Model 1			Model 2			Model 3			Model 4		
		Estimate	Std error	p-value	Estimate	Std error	p-value	Estimate	Std error	p-value	Estimate	Std error	p-value
Means	$\bar{\alpha}_0$	1.298	0.019	0.000	1.496	0.041	0.000	1.246	0.015	0.000	1.274	0.016	0.000
	$\bar{\beta}_0$	1.252	0.019	0.000	1.401	0.033	0.000	1.149	0.011	0.000	1.200	0.013	0.000
	$\bar{p}$	0.493	0.010	0.000	0.513	0.009	0.000	0.607	0.015	0.000	0.636	0.016	0.000
	$\bar{q}$	0.142	0.011	0.000	0.207	0.011	0.000	0.183	0.018	0.000	0.209	0.018	0.000
Std	σ_{α_0}				0.413	0.018	0.000	0.116	0.007	0.000	0.157	0.009	0.000
	σ_{β_0}				0.319	0.019	0.000	0.093	0.006	0.000	0.157	0.009	0.000
	σ_p							0.745	0.028	0.000	0.757	0.025	0.000
	σ_q							0.155	0.021	0.000	0.201	0.018	0.000
Fixed effects	Δp_2										-0.102	0.015	0.000
	Δp_3										-0.027	0.013	0.035
	Δq_3										-0.051	0.031	0.102
	LL	-9480.911			-9359.655			-8567.744			-8558.403		

Table 3: Maximum likelihood estimations. Model 1 is a representative-agent model. Model 2 allows for heterogeneity in prior parameters α and β_0. Model 3 further accounts for heterogeneity in the bias parameters p and q. Model 4 includes round fixed effects for p and q.

in model parameters. It provides a first picture of the mean patterns.

The mean of the parameters α_0 and β_0 characterizing priors at round 0 and before receiving any signal, had very similar estimates: 1.298 and 1.252. The similarity of these two values suggests that the prior distribution of our representative subject was roughly symmetric. Consistent with the instructions we provided, subjects did not expect one color to be more likely than the other, before receiving signals. We note however that priors were not perfectly uniform. The parameters were larger than 1 ($p < 0.001$ for each of them), revealing that the representative subject exhibited a bell-shaped prior distribution and gave more probability weight to central than to extreme values of the [0,1] interval.

Regarding the measures of conservatism bias (p) and confirmatory bias (q), both indexes differed from 0, and a likelihood ratio test rejects the joint hypothesis that the indices are both equal to 0 ($p < 0.001$). The estimations indicated strong conservatism: the representative subject behaved as if he neglected half of the actual signals (49.3%). We also observed evidence for confirmatory bias: the representative subject behaved as if he misinterpreted 14% of the signals that contradicted their prior beliefs.

Model 2 refines the analysis by allowing for heterogeneity in prior parameters α_0 and β_0. The estimated standard deviations of the distributions of these parameters indeed showed significant heterogeneity although this did not change the estimations of the bias indexes qualitatively: $\bar{p}$ was estimated as 0.51, and the estimate of $\bar{q}$ increased to 0.21.

Because not all subject may exhibit the same degree of deviations from Bayesian updating, model 3 also allows for heterogeneity in indexes p and q. Although the estimated means of the indexes were still qualitatively similar to those estimated under model 1 and model 2, the estimations of model 3 suggested a large degree of between-subject variation. On average, $\bar{p}$ was estimated as 0.61, and the estimate of $\bar{q}$ increased to 0.18, the estimated standard deviations of the distributions of the indexes were close in magnitude to the estimated means ($\sigma_p = 0.75$ and $\sigma_p = 0.16$, respectively). We also note that accounting for heterogeneity in the indexes produced a large improvement in the goodness of fit.

Model 4 further allows for variations in the means of p and q across rounds and tests whether subjects become more (or less) conservative or exhibit stronger (or weaker) confirmatory bias in later rounds. The fixed effects showed that the conservatism index p decreased across rounds with the largest decrease observed between rounds 2 and 3. The confirmatory index q did not vary across rounds. Having accounted for the round effects, Model 4 provides the best goodness of fit to our data (Likelihood ratio tests, all $p < 0.001$).

In order to illustrate the heterogeneity of deviation indexes across subjects, Figure 8 shows the histogram of estimated individual indexes p and q under Model 3. We report model 3 estimates as they are average values across all rounds.

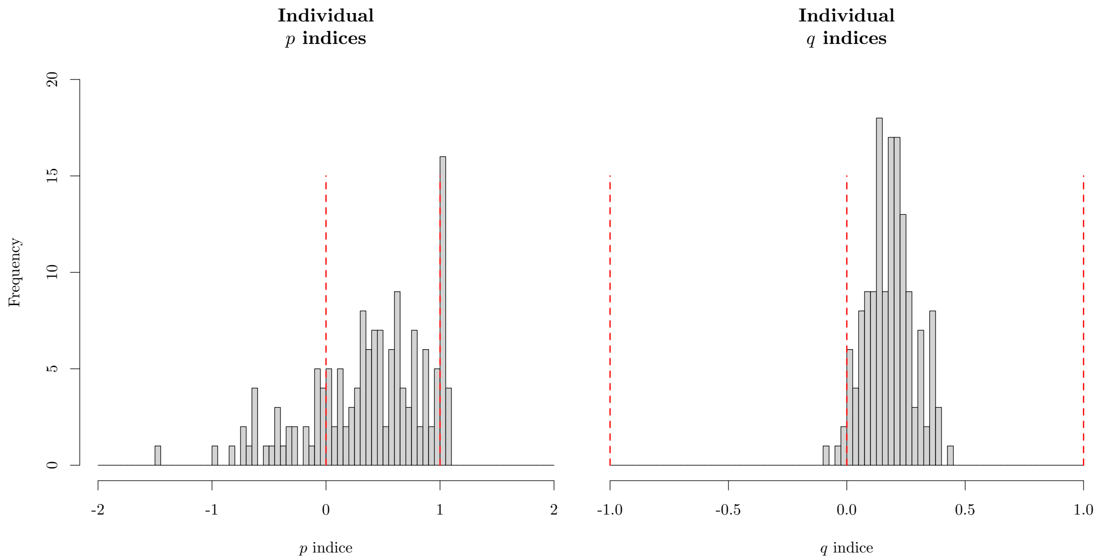

Figure 8: Individual distributions of indexes

We observed that the majority (65%) of our subjects exhibited conservatism ($0 \leq p \leq 1$), and the second most common pattern (22%) was over-inference ($p < 0$). We also found some evidence for prior-signal destruction ($p > 1$), the extreme case of conservatism, which held for 13% of our subjects. The mode (approximately) at 1 indicates that a substantial proportion of the subjects did not update their beliefs. For the confirmatory index q, the large majority (97%) of our subjects exhibited confirmatory bias ($0 \leq q \leq 1$), and the rest exhibited disconfirmatory bias ($q < 0$). No extreme cases of this index was observed in our data.

To sum up, our subjects on average exhibited conservatism and confirmation bias, but a large heterogeneity was captured at individual level. Aside from the modal patterns, individual patterns revealed that over-inference and disconfirmatory bias were also exhibited by some of the subjects. There was also heterogeneity in prior beliefs, but accounting for this heterogeneity did not impact the estimations of the indexes qualitatively.

6.4 Robustness checks

We assess the robustness of our findings considering the following three aspects: the measurement method, the inclusion or exclusion of subjects exhibiting multiple switching patterns in the choice lists, and the sessions with extreme signals.

Our experiment used two different methods to measure beliefs: probability matching and exchangeability. Estimations presented in the previous section pool observations from the two methods, assuming that they do not differ. We also tested this assumption. For this, we re-estimated the representative-agent model (model 1) and our "best model" (model 4) with fixed effects capturing the differences in mean estimates across methods. The details of these estimations are reported in Appendix B. We observed that the average patterns of the deviations from Bayesian updating were qualitatively stable across the two measurement methods. Conservatism and confirmatory bias remained the most prevalent type. However, under the matching probability method, the estimations with model 4 indicated on average higher conservatism (2% more, $p < 0.001$) and lower confirmatory bias (5% less, $p = 0.02$). The prior Beta distribution parameters also differed between the two methods. When priors were measured from exchangeable events, they appeared to be closer to the uniform distribution than when they were measured by matching probabilities. The matching probability method is not robust to non-neutral ambiguity attitudes, which could potentially affect the results. Empirically, we found that the matching-probability method did introduce differences in estimations. Nevertheless, it still captured the main qualitative patterns of deviations from Bayesianism.

The second aspect concerns the elicitation of preferences by using choice lists. Monotonicity implies that subjects should exhibit only one switching point per list. This was always the case for 142 (out of 157) subjects. Five subjects who exhibited multiple switching for only one choice list were included in the main analysis (while discarding only those choice lists with multiple switching patterns), and other subjects who exhibited multiple switching in more than one choice lists were excluded. In order to evaluate a possible impact of this exclusion, we ran two additional estimations of model 1. We considered a stricter exclusion criterion where we excluded all choice lists for the subjects who exhibited multiple switching in at least one choice list, and a weaker exclusion criterion where we only excluded choice lists with multiple switching, and the other with a The estimations are reported in Appendix C. Results are virtually identical under different exclusion criteria.

The third aspect deals with the possible impact of extreme signal observations in our experimental sessions. For example, subjects from session 1 and 6 always observed the same color across all the rounds, and subjects from all the other sessions (except for session 4) experienced at least one sampling round where all the spins resulted in the same color. To test the role of these extreme observations in our results, we re-estimated model 1 (from Table 3) by excluding those rounds where extreme signals were observed. We present the results in Appendix D. This analysis replicates the same patterns as in our main analysis, suggesting that our findings are not driven or distorted by the cases of extreme signals.

7 Discussion

7.1 Relation to Rabin and Schrag (1999)

This paper quantifies symmetric and asymmetric belief updating biases, both theoretically and empirically. We borrow Rabin and Schrag's way of modelling the confirmatory bias and apply it to model not only the confirmatory bias, but also other types of asymmetric and symmetric biases. Our method hence enriches Rabin and Schrag's model in the following three aspects.

First, it encompasses a richer set of belief updating biases. To our best knowledge, our study is the first to make a formal distinction between asymmetric and symmetric biases, and unified them in one model. We theoretically demonstrate how asymmetric biases affect the balance of evidence by overweighting one type of evidence over the other, while symmetric biases affect the weight assigned to the sum of evidence without distinguishing between different types of evidence. Our model sheds light on how these biases affect belief updating in distinct ways.

Second, whereas the original model was introduced to illustrate the impact of the confirmatory bias on beliefs and choices, we focus more on retrieving bias estimates from beliefs revealed by choices. Thereby we are able to provide the first structural estimates of the biases: People on average misread 18% contradictory signals as confirming, while ignoring 60% of the signals overall. Providing a quantitative assessment of the information lost and/or distorted in the updating process, our estimations make our method unique despite the abundance of reduced-form evidence on how biases affect behaviors.

Third, we expand the empirical relevance of the original model. Rabin and Schrag (1999) modelled the confirmatory bias index as a probability of misreading contradictory evidence as confirming. This interpretation is mostly applicable when the signals are ambiguous and open for interpretation. In a rather neutral setting like the one in our experiment, it is less realistic to assume that people would misread the signals. However, the lack of room for misreading signals does not guarantee a lack of confirmatory bias. It can still exhibit itself through overweighting confirming signals relative to the disconfirming ones. We show how the model is compatible with such a overweighting/underweighting interpretation, hence making it more empirically relevant. Also, the initial probabilistic interpretation confines the parameters in the unit interval $[0, 1]$. We allow the parameters to be any real numbers, and provide interpretations for values outside of the unit interval.

Adding conservatism to the model of Rabin and Schrag (1999) preserves its portability. Probabilities of misreading and missing signals can be incorporated in theoretical and empirical studies that use Bayesian updating. As demonstrated in our experiment, these parameters can be recovered from choices, without more data than those necessary to analyze the same choices with Bayes' rule.

7.2 Further modeling considerations

Grether (1980) model conservatism together with base-rate neglect. In principle, our model could be extended to also include base-rate neglect by introducing a probability of neglecting prior information in favor of ignorance or uniform prior. A drawback of this approach however is that it makes parameter identification difficult without further assumptions. In particular, as mentioned in Benjamin (2019), most studies assume uniform prior beliefs.

This becomes even more challenging when adding confirmatory bias. Adding base-rate neglect to our analysis made parameter estimates unstable. In the analysis reported above, a clean identification of the parameters arose from the independent impact of symmetric and asymmetric biases on the total number of signals and their relative distribution respectively. Incorporating the model of Grether (1980) would lead the various biases to interact when influencing total number and distribution of perceived signals. In a nutshell, two dimensions (sum and relative proportion) can only give two parameters (without further assumptions).

While we provide intuitive interpretations of our parameters, it is worth noting that our method adopts an as-if approach. The current study does not claim that the interpretations of the bias indexes reflect necessarily the exact cognitive processes in the decision maker's mind. The underlying reasons why people may exhibit such belief distortions were investigated in other studies (Benjamin et al., 2016; Falk and Zimmermann, 2017).

7.3 Relation to the empirical literature

Antoniou et al. (2015) and Moreno and Rosokha (2016) adopts a similar empirical approach as ours, but only investigates conservatism. Mobius et al. (2014) and Coutts (2019) consider both types of deviations from Bayesian updating, similar as we do. However, their model cannot be structurally estimated, and hence produce evidence only in reduced-form. Buser et al. (2018) provides individual measurements of conservatism and asymmetry in belief updating but their measures are based on interpersonal comparisons in an ego-related setting rather than based on deviations from the Bayesian benchmark. Different from these studies, instead of imposing assumptions on prior beliefs, we also elicit subjects' priors in round 0 as Moreno and Rosokha (2016) also does. In this sense, our approach is fully subjective and more generally applicable as in many real-life decision situations, where there is often no control over prior beliefs. A drawback is that, in the absence of prior information, the situation may be perceived as ambiguous. This may cause a problem for matching probabilities, where ambiguous acts and objective lotteries were presented side by side. If the decision maker tend to prefer lotteries to acts, exhibiting ambiguity aversion (Ellsberg, 1961), matching probabilities may be biased. Eliciting exchangeable events is robust to this problem (Abdellaoui et al., 2011).

In our neutral setting, decision makers exhibited conservatism. Evidence in the literature shows that depending on various situational factors, people may under-react or over-react to new information (Griffin and Tversky, 1992; Baeriswyl and Cornand, 2014; Luo et al., 2015). Our index of symmetric biases can capture not only under-inference (conservatism) but also over-inference, making it suitable to identify different factors that affect decision makers' reaction to information.

Our study also contributes to the empirical literature on 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, these psychology experiments do not allow a formal investigation of confirmatory bias due to the lack of a normative benchmark for a comparison with revised beliefs. A few recent field studies document evidence on confirmatory bias (Christandl et al., 2011; Andrews et al., 2018; Sinkey, 2015), yet, the same problem of missing clear Bayesian benchmark remains for these studies. In addition to the recent study of Buser et al. (2018), several other studies investigate asymmetric processing of information in Bayesian updating as in confirmatory bias, when the information has a valence or is self-relevant (Eil and Rao, 2011; Ertac, 2011; Mobius et al., 2014; Coutts, 2019; Coutts et al., 2019). Different from our ego-neutral setting, these studies employ ego-related settings where subjects make inferences about their scores on some performance tasks or about 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. Our findings also demonstrate that, to obtain evidence on pure ego-relevant biases, it is important to control for ego-neutral biases. Our model facilitates such control, by comparing bias estimates between ego-relevant and ego-neutral settings.

8 Conclusion

This paper studied biases in people's belief updating from a descriptive perspective. We showed how to quantify asymmetric and symmetric biases in updating, based on a model allowing perceived signals to differ from the signals people actually receive. It provided a natural interpretation of well-known biases and made them observable from choices. Our approach thus adhered to the revealed-preference approach of economics.

In our experiment, confirmatory bias and conservatism were modal patterns, while a minority of subjects exhibited the opposite biases. 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. Our portable model and empirical approach can be applied to investigate subjective beliefs and their updating in various contexts.

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. The slides and oral explanations are available in the online appendix.[8] 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 blue and red, and there were no other colors on the wheel. He then span 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 0 to 3. 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 round, 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

[8] Link to the online appendix:
https://www.dropbox.com/s/9bv30pszrjljtw2/Experimental%20Instructions.pdf?dl=0

every choice round, 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 round 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.

B. Consistency of the results across the different measurement methods

The estimates reported in the first rows are based on exchangeable-event questions. The fixed effects, ΔM capture the differences in the estimates of the mean parameters between the matching-probability method and the exchangeable-event method.

Parameter		Model 1 with fixed effects for the methods			Model 4 with fixed effects for the methods		
		Estimate	Std error	p-value	Estimate	Std error	p-value
Means	$\bar{\alpha}_0$	1.172	0.025	0.000	1.181	0.019	0.000
	$\bar{\beta}_0$	1.134	0.024	0.000	1.105	0.018	0.000
	$\bar{p}$	0.471	0.015	0.000	0.655	0.016	0.000
	$\bar{q}$	0.189	0.023	0.000	0.222	0.019	0.000
Std	σ_{α_0}				0.137	0.009	0.000
	σ_{β_0}				0.129	0.008	0.000
	σ_p				0.713	0.029	0.000
	σ_q				0.148	0.017	0.000
Fixed effects	Δp_2				-0.090	0.012	0.000
	Δp_3				-0.022	0.011	0.046
	Δq_3				0.019	0.030	0.531
Differences between methods	ΔM_p	0.011	0.028	0.694	0.022	0.004	0.000
	ΔM_q	-0.069	0.040	0.081	-0.049	0.020	0.015
	ΔM_α	0.293	0.060	0.000	0.284	0.033	0.000
	ΔM_β	0.280	0.056	0.000	0.304	0.031	0.000
LL		-9346.763			-8460.828		

Table 4: Comparison of estimates across measurement methods

C. Robustness to multiple switching

Table 5 reports the results of estimations of Model 1 (presented in Table 3) with different exclusion rules for subjects exhibiting multiple switching. In the "stricter exclusion" estimations, all subjects with multiple switching for at least one choice list (i.e 15 subjects) are removed from the analysis. In the "weaker exclusion" estimation all the subjects are considered, while only the choice lists with multiple switching are excluded.

	Stricter exclusion			Weaker exclusion		
	Estimate	Std Error	p-value	Estimate	Std Error	p-value
$\hat{\alpha}_0$	1.29	0.02	0.00	1.34	0.02	0.00
$\hat{\beta}_0$	1.25	0.02	0.00	1.29	0.02	0.00
p_0	0.50	0.01	0.00	0.50	0.01	0.00
q_0	0.14	0.01	0.00	0.14	0.01	0.00
LL	-9084.26			-10629.20		

Table 5: Model with different exclusion rules regarding multiple swithcing

D. Robustness to extreme signals

In order to ensure that our results are not driven by rounds where subjects received extreme signals (i.e. no success or no failure), we re-estimated Model 1 based on a dataset from which data corresponding to such rounds were removed. The results are presented in Table 6

Parameter	Estimate	Std Error	p-value
$\hat{\alpha_0}$	1.12	0.02	0.00
$\hat{\beta_0}$	1.10	0.02	0.00
p	0.46	0.02	0.00
q	0.11	0.02	0.00
LL	-5191.71		

Table 6: Model 1 without extreme signals

The estimates of the indices p and q are still significant in this analysis, and their magnitudes are in line with our estimations of Model 1 based on the unrestricted data set. This result suggests that our main findings are not driven by extreme signals.

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.

Ambuehl, S. and Li, S. (2018). Belief updating and the demand for information. *Games and Economic Behavior*, 109:21–39.

Andrews, R. J., Logan, T. D., and Sinkey, M. J. (2018). Identifying confirmatory bias in the field: Evidence from a poll of experts. *Journal of Sports Economics*, 19(1):50–81.

Antoniou, C., Harrison, G. W., Lau, M. I., and Read, D. (2015). Subjective bayesian beliefs. *Journal of Risk and Uncertainty*, 50(1):35–54.

Baeriswyl, R. and Cornand, C. (2014). Reducing Overreaction to Central Banks' Disclosures: Theory and Experiment. *Journal of the European Economic Association*, 12(4):1087–1126.

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

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1*, 2:69–186.

Benjamin, D. J., Rabin, M., and Raymond, C. (2016). A model of nonbelief in the law of large numbers. *Journal of the European Economic Association*, 14(2):515–544.

Bohren, J. A., Imas, A., and Rosenberg, M. (2019). The dynamics of discrimination: Theory and evidence. *American economic review*, 109(10):3395–3436.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*, 131(4):1753–1794.

Buser, T., Gerhards, L., and Van Der Weele, J. (2018). Responsiveness to feedback as a personal trait. *Journal of Risk and Uncertainty*, 56(2):165–192.

Charness, G. and Dave, C. (2017). Confirmation bias with motivated beliefs. *Games and Economic Behavior*, 104:1–23.

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. (2019). Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, 22(2):369–395.

Coutts, A., Gerhards, L., and Murad, Z. (2019). No one to blame: Biased belief updating without attribution. Working Paper.

De Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. 7(1):1–68.

Deryugina, T. (2013). How do people update? the effects of local weather fluctuations on beliefs about global warming. *Climatic change*, 118(2):397–416.

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.

El-Gamal, M. A. and Grether, D. M. (1995). Are people bayesian? uncovering behavioral strategies. *Journal of the American statistical Association*, 90(432):1137–1145.

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.

Falk, A. and Zimmermann, F. (2017). Information processing and commitment. *The Economic Journal*, 128(613):1983–2002.

Grether, D. M. (1980). Bayes rule as a descriptive model: The representativeness heuristic. *The Quarterly journal of economics*, 95(3):537–557.

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.

Howe, P. D. and Leiserowitz, A. (2013). Who remembers a hot summer or a cold winter? the asymmetric effect of beliefs about global warming on perceptions of local climate conditions in the us. *Global environmental change*, 23(6):1488–1500.

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.

Kovach, M. (2021). Conservative updating. *arXiv preprint arXiv:2102.00152*.

Luo, Y., Nie, J., and Young, E. R. (2015). Slow Information Diffusion and the Inertial Behavior of Durable Consumption. *Journal of the European Economic Association*, 13(5):805–840.

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., and Hankin, B. L. (2004). Is there a universal positivity bias in attributions? a meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological bulletin*, 130(5):711.

Miller, D. T. and Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological bulletin*, 82(2):213.

Mobius, M. M., Niederle, M., Niehaus, P., and Rosenblat, T. S. (2014). Managing self-confidence: Theory and experimental evidence. Technical report, National Bureau of Economic Research.

Moreno, O. M. and Rosokha, Y. (2016). Learning under compound risk vs. learning under ambiguity–an experiment. *Journal of Risk and Uncertainty*, 53(2-3):137–162.

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.

Sarsons, H. (2017). Interpreting signals in the labor market: evidence from medical referrals. *Job Market Paper*.

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.