Quantitative Economics 16 (2025), 329–365 1759-7331/20250329

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

Ilke Aydogan
IESEG School of Management, Univ. Lille, CNRS, UMR 9221 - LEM - Lille Economie Management

Aurélien Baillon
emlyon business school, CNRS, Université Lumière Lyon 2, Université Jean Monnet Saint-Etienne, GATE

Emmanuel Kemel
GREGHEC, CNRS and HEC Paris

Chen Li
Erasmus School of Economics, Erasmus University Rotterdam

Belief-updating biases hinder the correction of inaccurate beliefs and lead to suboptimal decisions. We complement Rabin and Schrag's (1999) portable extension of the Bayesian model by including conservatism in addition to confirmatory bias. Additionally, we show how to identify these two forms of biases from choices. In an experiment, we found that the subjects exhibited confirmatory bias by misreading 19% of the signals that contradicted their priors. They were also conservative and acted as if they missed 28% of the signals.

Keywords. Non-Bayesian updating, conservatism, confirmatory bias, perceived signals, belief elicitation.

JEL classification. C91, D83.

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 to biased beliefs that underlie societal issues such as global

Ilke Aydogan: i.aydogan@ieseg.fr
Aurélien Baillon: baillon@em-lyon.com
Emmanuel Kemel: emmanuel.kemel@greg-hec.com
Chen Li: c.li@ese.eur.nl
We thank Peter Wakker and Han Bleichrodt for helpful comments and discussions. Baillon acknowledges the financial support of NWO Vidi grant 452-13-013. Aydogan acknowledges the support of the Region Haut-de-France (2021.00865 CLAM) and the European Union's Horizon Europe research and innovation programme under grant agreement (101056891 CAPABLE). Li acknowledges the financial support of NWO Veni grant VI.Veni.191E.024.

[1] See, for instance, Phillips and Edwards (1966), Edwards (1968), Tversky and Kahneman (1974), El-Gamal and Grether (1995), Oswald and Grosjean (2004), Möbius, Niederle, Niehaus, and Rosenblat (2022), Bénabou and Tirole (2016), Ambuehl and Li (2018).

Available at http://qeconomics.org. https://doi.org/10.3982/QE2094

warming (Deryugina (2013), Howe and Leiserowitz (2013)) and gender inequality (Sarsons (2017), Bohren, Imas, and Rosenberg (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, Rabin, and Raymond (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 is either qualitative classifications (El-Gamal and Grether (1995)) or reduced-form (Charness and Dave (2017), Coutts (2019), Möbius et al. (2022)).

We propose a portable model based on Rabin and Schrag (1999), and an empirical approach providing a structural estimation of biases from choices. The model has two parameters, which can be easily interpreted, plugged into 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 over or underweight new information relative to their priors. Common biases of this type include conservatism (Phillips and Edwards (1966), Edwards (1968)), which leads to underweighting of new information, and overinference, which corresponds to the overweighting of new information due to, for instance, the representativeness heuristics (Tversky and Kahneman (1974), Bordalo, Coffman, Gennaioli, and Shleifer (2016)).

Both types of biases are present in most decisions, yet existing studies mostly focus on specific biases relevant to decision situations under consideration.[3] We extend the model of Rabin and Schrag (1999) to separate and quantify both asymmetric and symmetric biases. Our separation strategy taps into the following conceptual difference

[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, Rabin, and Raymond (2016) is not portable because the ordering and timing of signals are needed, unlike in traditional updating. The 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 occur 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.

between the two types of biases. While asymmetric biases, such as 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. The method we employ also allows for individual heterogeneity without a precommitment regarding the direction of biases, hence achieving 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 over prior beliefs.

Our model remains parsimonious with two parameters—one for each type of bias. We refer to the first index as the *confirmatory bias index* (q) and the second 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 contradicting signals as confirming, whereas the conservatism index as the probability of missing new signals. These interpretations, though convenient, are not necessary for our approach to remain valid. The indexes can also be interpreted as over 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, Abramson, Hyde, and Hankin (2004)).

Our experiment concerns an ego-neutral learning environment unrelated to individuals' self-image. We found that, on average, subjects exhibited confirmatory bias by misreading 19% (95% credible interval [0.12, 0.26]) of the signals that contradicted their prior beliefs. Moreover, they were conservative and acted as if they missed approximately 28% (95% credible interval [0.15, 0.40]) of the signals received. Our findings suggest that confirmatory bias and conservatism can occur even in the absence of clear motivation, such as enhancing one's self-perception.[5] Our method further allowed us to obtain complete prior and posterior distributions, not only point estimates. The technique also enabled us to uncover individual heterogeneity, where 27% exhibited overinference and 15% the disconfirmatory bias.

1.1 *Contribution*

This paper quantifies symmetric and asymmetric belief-updating biases, both theoretically and empirically. We borrow Rabin and Schrag's (1999) way of modeling the confirmatory bias and apply it to model not only the confirmatory bias, but also other types of asymmetric and symmetric biases. Hence, our method enriches Rabin and Schrag's model in the following three aspects.

[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.

[5]This observation is consistent with the results of Charness and Dave (2017), who found that confirmatory bias is amplified by motivated reasoning but still exists in the absence of motives.

First, our model encompasses a richer set of belief-updating biases. To our best knowledge, this study is the first to make a formal distinction between asymmetric and symmetric biases, and unify 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. We can thereby provide the first structural estimates of the biases. By providing a quantitative assessment of the information lost and/or distorted in the updating process, our estimations are unique despite the abundance of reduced-form evidence on how biases affect behaviors. This analysis can help understand and predict when people will underreact to information and may delay important actions regarding, for instance, climate change, gender inequality, or discrimination in an organization. The tendency to ignore signals independently of prior suggests that learning will be slow, even when it reinforces prior beliefs. In a series of experiment, Campos-Mercade and Mengel (2024) showed that due to conservatism in updating, non-Bayesian statistical discrimination was responsible for 40% of the hiring gap between a disadvantaged and an advantaged group.

Third, we expand the empirical relevance of the original model. Rabin and Schrag (1999) modeled 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 an over or 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.

2. Modeling 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 $\text{Beta}(\alpha_0, \beta_0)$, 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, Bleichrodt, Kemel, and l'Haridon (2021)) to model beliefs over a success rate. Beta distributions are also flexible and can take a wide array of shapes with different locations and dispersions for various parameters.[6]

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 nonzero 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 set of signals, his *posterior belief* becomes $\text{Beta}(\alpha_1, \beta_1)$. 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 situation does not hold 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).

In Bayesianism, updating after having observed an entire set of signals and after each signal of the set both lead, ultimately, to the same posterior. When deviating from Bayesianism, however, this tenet no longer needs to be true. We assume that updating occurs after receiving the entire set of signals. This modeling assumption ensures the portability of the model, that is, we do not require more than the number of successes and failures between two decision times to apply the model, as would be necessary with the Bayesian model. This modeling assumption is further discussed in Section 7.1.

[6]The beta family was also used in Moreno and Rosokha (2016), Esponda, Vespa, and Yuksel (2024), Bland and Rosokha (2021) for modeling symmetric deviations from Bayesian updating without considering asymmetric ones such as confirmatory bias.

We study two sources of deviations: asymmetric and symmetric belief-updating biases. Asymmetric biases distort the relative proportions of successes and failures, that is, distort the actual sample means $\frac{a}{a+b}$ (and $\frac{b}{a+b}$) into perceived sample means $\frac{\alpha}{\alpha+\beta}$ (and $\frac{\beta}{\alpha+\beta}$). In comparison, symmetric biases affect all signals without distinguishing between successes and failures, that is, they distort n into η. Conceptually, asymmetric biases change the balance of evidence by distorting the sample mean, whereas symmetric biases affect the weight assigned to the sum of the evidence.

In Section 2.2, we first introduce an index for the most well-known asymmetric bias—confirmatory bias, and its negative counterpart capturing 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 Section 2.3, we introduce an index for symmetric biases, such as conservatism. Section 2.4 combines 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 contradicting signals 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 modeling, 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, which we call *disconfirmatory bias*. This bias can be modeled as the probability q_d to misread confirming signals 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 ($q_d \frac{b}{a+b}$).

From observing perceived signals, either $q_c > 0$ or $q_d > 0$ can be determined whenever $\alpha_0 \neq \beta_0$, by comparing perceived signals to actual signals. Consider the case where successes are believed to be more likely than failures, that is, $\alpha_0 > \beta_0$. 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 outcome suggests 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 no longer a probability 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 signals in a confirmatory way and values in $[-1, 0]$ as minus probabilities to misread signals 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 from disconfirmatory biases (as is done for other attitude measures, such as risk aversion). Figure 8 reports the estimated distribution for our experimental subjects.

Our modeling approach can also be adapted to quantify other types of asymmetric biases. For instance, the self-serving bias can be modeled as the probability to misread signals that damage self-image or confidence as self-enhancing ones, independently 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

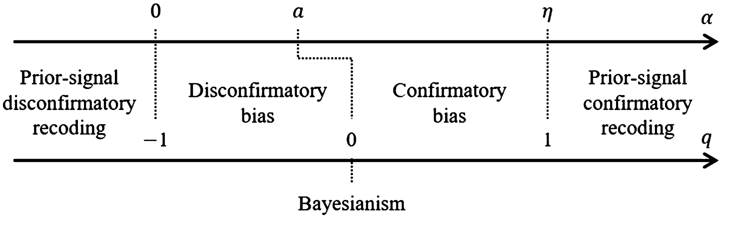

FIGURE 1. Interpretation of q and relationship with α when $\alpha_0 > \beta_0$. The case $q = 0$ corresponds to $\alpha = a$. The edge on the graph indicates that a need not be at equal distance to 0 and η.

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 over 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 signals. Hence, the decision-maker perceives on average only $1-p$ of all actually received signals, that is, $\eta=(1-p)n$. 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.

The case $p>1$ cannot be interpreted as a probability but still has a meaningful empirical interpretation. It captures situations where the perceived number of signals is negative, suggesting that the decision-maker received (surprising) information that undermined his prior.

We call this case *prior-signal destruction*. By contrast, $p<0$ corresponds to *over-inference*, where the decision-maker assigns too much weight to the sample and too little to 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 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). Both biases are applied to the entire set of signals received, and not sequentially after each signal. For instance, in the

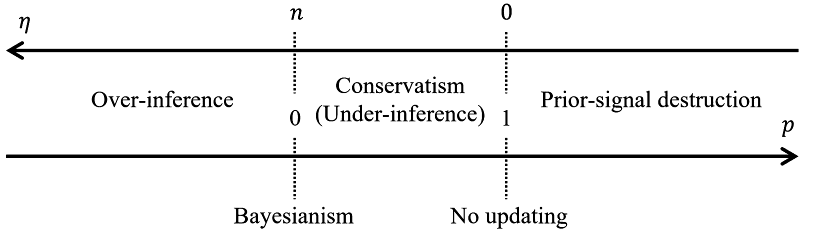

FIGURE 2. Interpretation of p and relationship with η.

case $\alpha_0 > \beta_0$, replacing a and b in equation (1) respectively by $(1-p)a$ and $(1-p)b$, the confirmatory bias in the presence of conservatism gives (replacing q_c by q):

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

Conservatism p can always be identified by comparing η with n. When $\alpha_0 \neq \beta_0$ and $\alpha + \beta \neq 0$, then index q is observed by comparing $\frac{\alpha}{\alpha+\beta}$ and $\frac{a}{a+b}$. The indexes hence capture different aspects of the perceived signals: symmetric biases influencing the total number of perceived signals for p versus asymmetric biases influencing the relative proportion of perceived successes and failures for q.

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 toward 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 toward 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*, namely 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, Baillon, Placido, and Wakker (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 acts and uses 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.

We elicit their priors and posteriors using the methods described above before and after decision-makers receive a set of signals. 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. The direction of the confirmatory bias can change for each set of signals, depending on the beliefs prior to that set of signals.

4. Experimental design

4.1 *Subjects*

Seven experimental sessions were conducted at the Erasmus School of Economics Rotterdam. The number of subjects 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, we used a spinning wheel that was covered by two (and only two) colors: yellow and brown. The subjects faced choice situations that involved acts whose payoffs depended on the actual color composition, namely the proportion of the wheel covered by each color. Stickers displaying a variety of color compositions were placed in an opaque bag. One of the stickers was drawn by the implementer at the beginning of the experiment, hence determining the color composition, which remained unknown to the subjects until the end of the experiment. Pictures of the wheel and detailed procedure can be seen in Appendix A.

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

FIGURE 3. Experimental flow.

spun the wheel three times and reported the color falling under a fixed arrow. Hence, the probability that a color was reported was equal to the proportion of that color on the wheel. 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.

The color composition of the wheel stayed the same and was 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.

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

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 option W(heel) and 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, choosing option W would only give a payoff of €20 if the yellow portion of the wheel was anywhere from 0% to 4%. On the other hand, receiving or not the €20 payoff for choosing option C was determined by randomly selecting one of four cards, each rep-

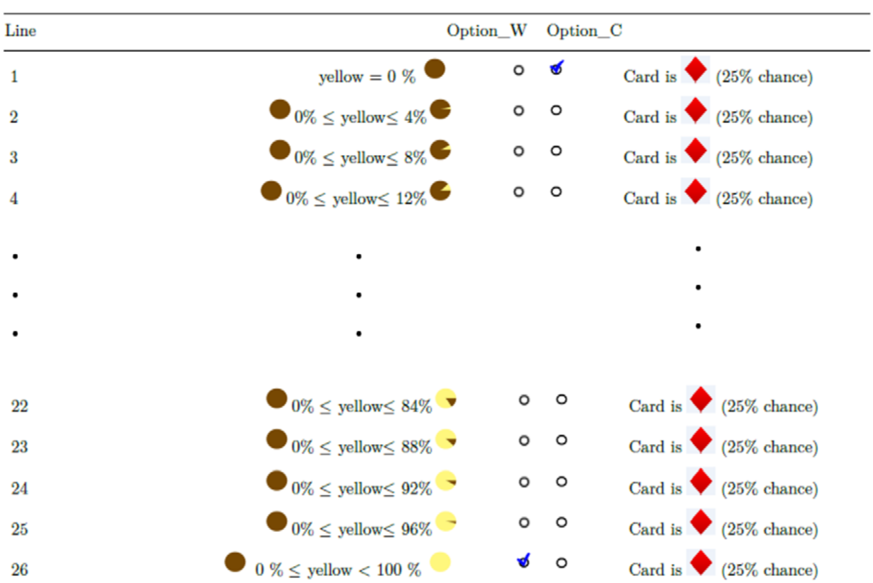

FIGURE 4. Choice list to elicit matching probabilities.

resenting a different suit (hearts, diamonds, clubs, and spades) from a deck. Each card had an equal chance of being drawn, setting the probability to get the reward at 25%.

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 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 the choice list 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. For instance, in line 4 of the list, if the wheel's yellow portion was 12% or less, choosing the Left Option (L) would yield €20. If the yellow portion was more than 12%, then picking the Right Option (R) would yield €20. People had to decide which option to pick by guessing how much yellow would be on the wheel.

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

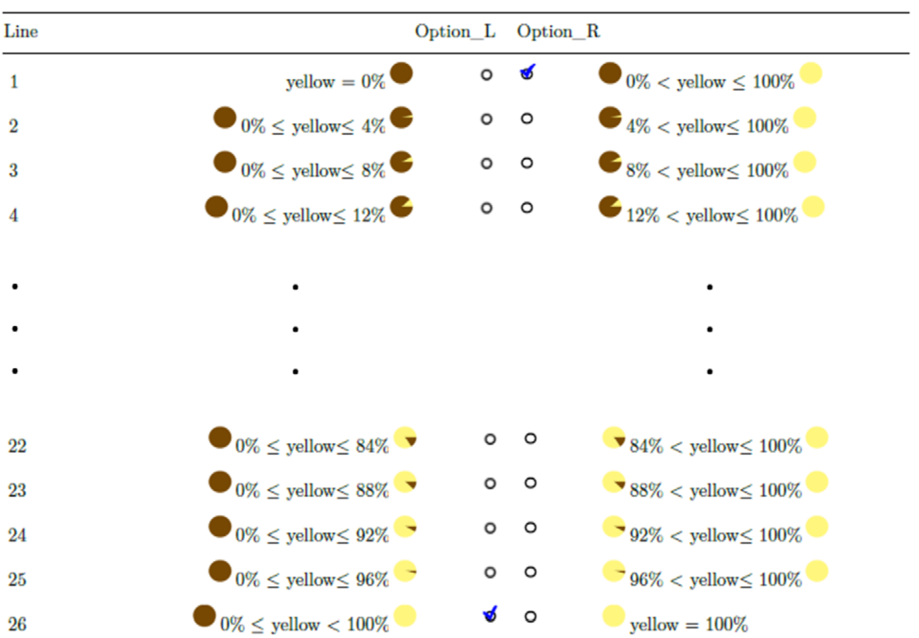

FIGURE 5. Choice list to elicit exchangeable events.

Table 1. Summary of experimental elicitation.

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}	$20_{(0,y_{\text{low}}]}0 \sim 20_{(y_{\text{low}},0.5]}0$	$\Lambda((0, y_{\text{low}}]) = \Lambda((y_{\text{low}}, 0.5])$
Exchangeable Events	y_{high}	$20_{[0.5,y_{\text{high}}]}0 \sim 20_{(y_{\text{high}},1)}0$	$\Lambda([0.5, y_{\text{high}}]) = \Lambda((y_{\text{high}}, 1))$

L became better, and option R became worse. At some point, they might want to switch from preferring option R to option L.

Where the subject switched in the choice list depicted 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 $20_{(0,y_{\text{low}}]}0 \sim 20_{(y_{\text{low}},0.5]}0$, and $20_{[0.5,y_{\text{high}}]}0 \sim 20_{(y_{\text{high}},1)}0$, following the method of Abdellaoui et al. (2021). 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.

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, Baillon, Bleichrodt, Li, Van Dolder, and Wakker (2021)) 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.

Table 2. Description of sessions.

Session	# Subjects	Received Signals Between Rounds: 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

5. Raw data

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 $y_{\text{median}} = \frac{y^-_{\text{median}}+y^+_{\text{median}}}{2}$.

Out of the 157 subjects, ten subjects exhibited multiple switching patterns systematically in more than one choice list. Another eight subjects exhibited a multiple switching pattern in only one out of six choice lists. To better discern the patterns in the raw data, here we calculate the descriptive statistics based on consistent observations while excluding the former ten subjects and the inconsistent observations from the latter eight. Our econometric analysis in the next section is run without any exclusion, and a robustness analysis is performed by excluding inconsistent observations (Appendix C).[7]

We consider 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

[7]As the data are coded as indifference intervals, in case of multiple switching, one can code the smallest interval that contains all the observed switching points as the indifference interval. As described in the next Section 6.2, our econometric analysis can account for multiple switching patterns by adopting an interval regression specification.

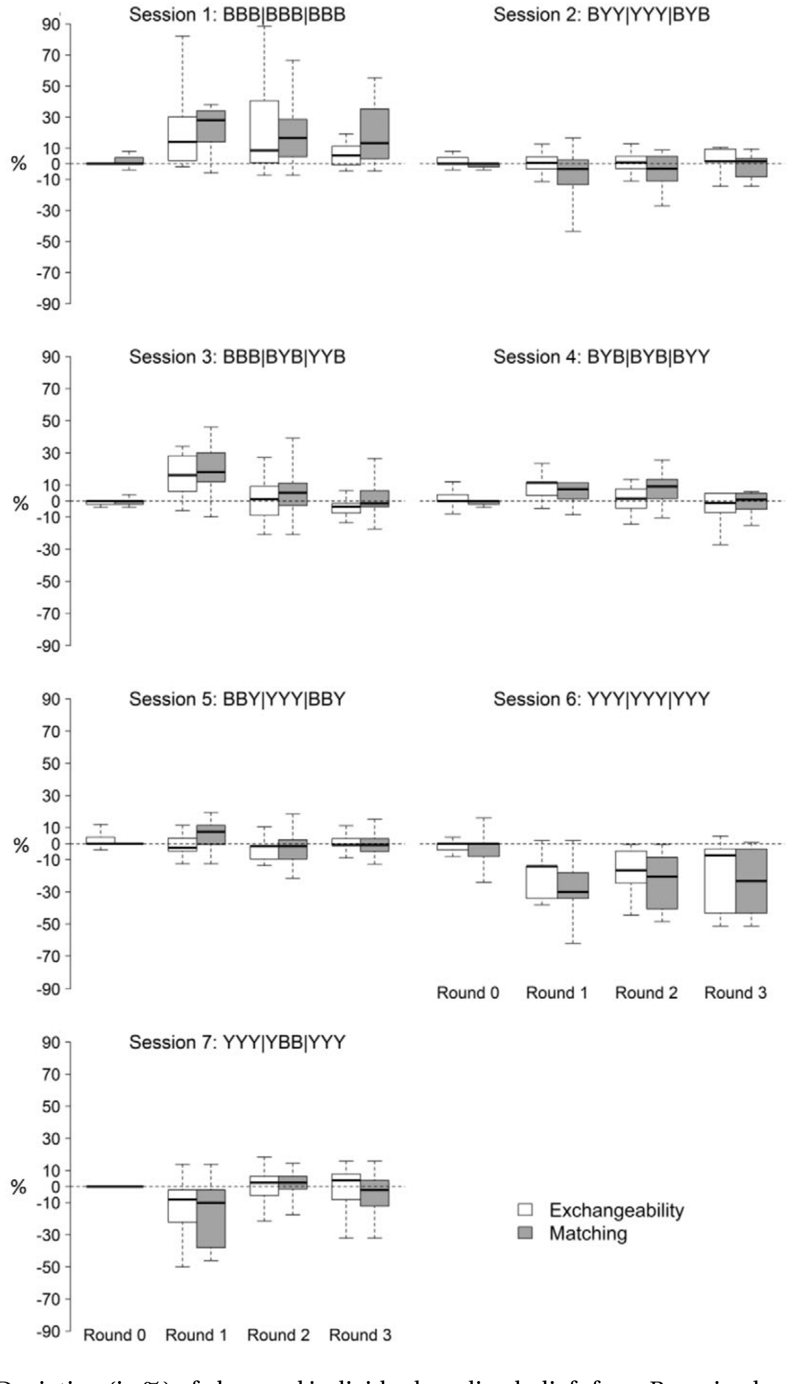

FIGURE 6. Deviation (in %) of observed individual median beliefs from Bayesian benchmark, for each session, round and measurement method.

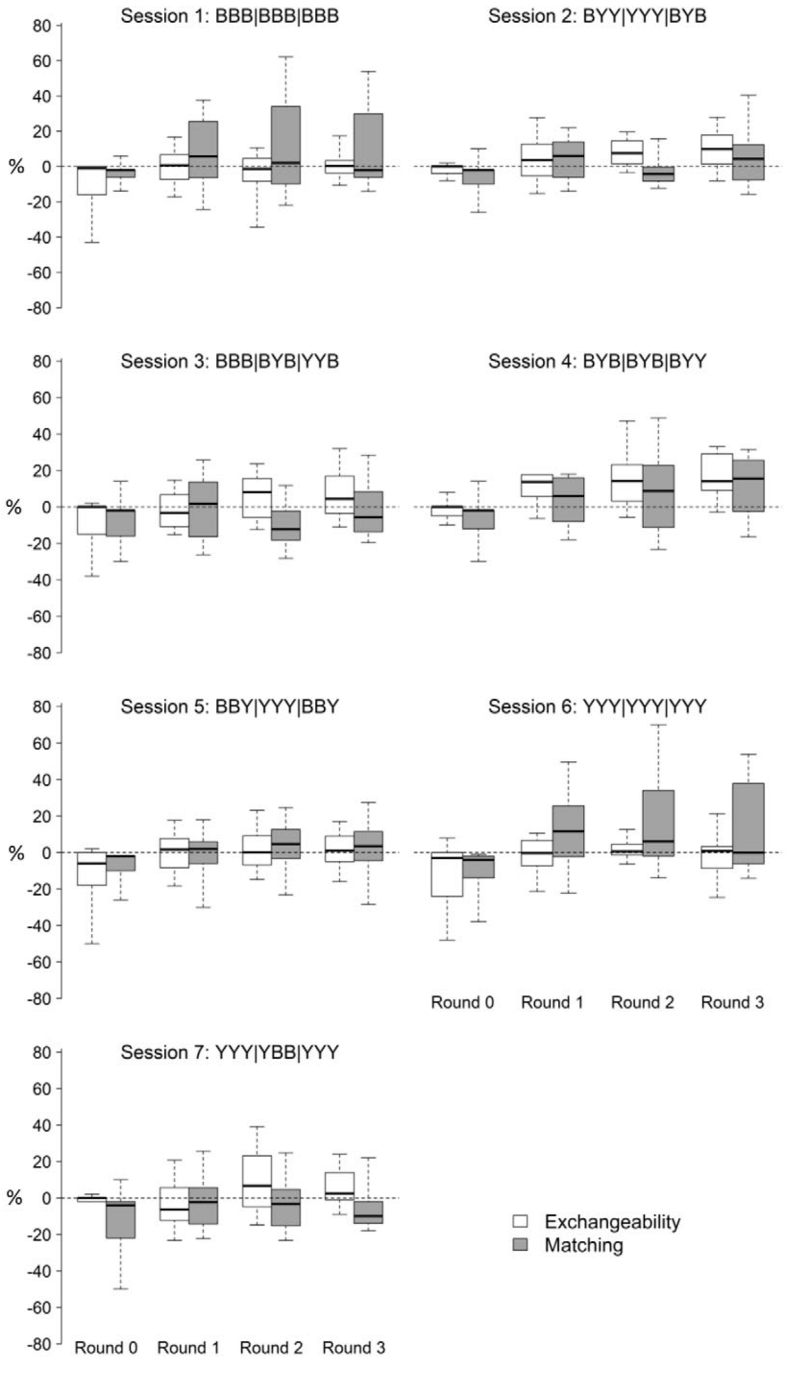

FIGURE 7. Deviation (in %) of observed individual dispersion in beliefs from Bayesian benchmark, for each session, round and measurement method.

from the Bayesian benchmark. A positive (negative) difference shows that subjects are underprecise (overprecise) 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 accounting for individual differences.

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 $\Lambda(\cdot) = \text{Beta}(\cdot|\alpha_{i,j}, \beta_{i,j})$ where $\alpha_{i,j}$ ($\beta_{i,j}$) is the number of successes (failures) perceived by subject i at round j. Each subject chose the color to bet on, at the beginning of the experiment. For each subject, successes (failures) are signals corresponding to the chosen (other) color.

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 that subjects use 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 equation (6). 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 $f(a_{i,j}, b_{i,j}, \alpha_{i,j-1}, \beta_{i,j-1}, p_{i,j}, q_{i,j}) = b_{i,j}$.

Before observing any signals, subjects had no reason to believe that 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. The robustness subsection below further provides evidence for this assumption. The direction of the confirmatory bias at rounds 2 and 3 is determined by the beliefs prior to these rounds (round 1 and round 2 beliefs, respectively) and, therefore, the direction of the confirmatory bias at rounds 2 and 3 need not be the same. 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 from 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.

We assume the following structural equations to account for these three sources of variations:

$$\begin{aligned} &\alpha_{i,0} \sim LN(\bar{\alpha}_0, \sigma_{\alpha_0}), \\ &\beta_{i,0} \sim LN(\bar{\beta}_0, \sigma_{\beta_0}), \end{aligned} \tag{7}$$

$$\begin{aligned} &p_{i,j} = p_i + \Delta_{p,2}\gamma_{j=2} + \Delta_{p,3}\gamma_{j=3}, \\ \text{with } &p_i \sim N(\bar{p}, \sigma_p), \end{aligned} \tag{8}$$

$$q_{i,j} = q_i + \Delta_{q,3}\gamma_{j=3}, \\ \text{with } q_i \sim N(\bar{q}, \sigma_q). \tag{9}$$

Parameters $\alpha_{i,0}$ and $\beta_{i,0}$ are nonnegative 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 defines a nonlinear, 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 $j-1$. The 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 $\Lambda((0, y_k^\star]) = \lambda_k$. 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 $20_{(0,y_k^+]}0 \succ 20_{\lambda_k}0$, that is, $y_k^\star \in [y_k^-; y_k^+]$.

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

Overall, our two types of choices lists produce intervals $[y_k^-; y_k^+]$ that contain an indifference value $y_k^\star$. The size of these intervals informs the precision with which the indifference value is measured by the choice lists. The interval size can also accommodate that indifference values should be considered less precise in the case of multiple switching in a choice list. In the case of multiple switching, we considered the smallest interval containing all the switching points. The choice lists with multiple switching thus provide intervals with larger sizes, thereby indicating weaker precision in the measurement of the indifference value.

For each individual i, round j and choice list k, the structural equation model provides a theoretical value $y_{i,j,k}^{\text{th}}(\theta, X_{i,j,k})$ 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. To account for subject and/or specification errors, we assume that $y_{i,j,k}^\star = y_{i,j,k}^{\text{th}} + \epsilon_{i,j,k}$ 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

$$\begin{aligned} p\big(y^\star_{i,j,k} \in \big[y^-_{i,j,k}; y^+_{i,j,k}\big]\big) &= p\big(\epsilon_{i,j,k} \in \big[y^-_{i,j,k} - y^{\text{th}}_{i,j,k}(\theta, X_{i,j,k}); y^+_{i,j,k} - y^{\text{th}}_{i,j,k}(\theta, X_{i,j,k})\big]\big) \\ &= \Phi\left(\frac{y^+_{i,j,k} - y^{\text{th}}_{i,j,k}(\theta, X_{i,j,k})}{\sigma}\right) - \Phi\left(\frac{y^-_{i,j,k} - y^{\text{th}}_{i,j,k}(\theta, X_{i,j,k})}{\sigma_i}\right) \\ &= l\big(\theta | y^+_{i,j,k}, y^-_{i,j,k}, X_{i,j,k}\big). \end{aligned} \tag{10}$$

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 round (indexed by j), writes

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

To account for heterogeneity in behavior, we assume that θ varies across individuals according to a multivariate distribution[8] of the mean $\bar{\theta}$ and the variance-covariance matrix Ω_θ. In particular, the diagonal elements of Ω_θ are the variance of individual parameters and capture the heterogeneity in these parameters. Table 3 presents the model estimates. Model 0 is a representative-agent model that accounts for neither individual heterogeneity nor deviation from Bayesian updating (i.e., assuming $p = q = 0$). Model 1 is also a representative-agent model but it allows for deviation from Bayesian updating ($p = q = 0$). Model 2 introduces individual heterogeneity in priors as well as in bias indices p and q. Model 3 augments Model 2 with fixed effects capturing the evolution of the means of these indices across rounds: Δp_2 (Δp_3) measures the difference of mean in index p between round 2 (3) and round 1 (2). Similarly, Δq_3 measures the difference of mean in index q between round 3 and round 2.[9]

The models are estimated using the Hierarchical Bayes (HB) procedure presented by Train (2009) and implemented in the RSGHB R package.[10] The procedure draws from the posterior distributions using Gibbs sampling. We use an MCMC of 60,000 draws. The first 10,000 are burnt, then over the 50,000 following draws, we keep one of every five to avoid serial correlation. The means (standard deviations) of posterior distributions of each parameter are taken as the estimates (standard errors). The likelihood is taken as the product of individual likelihoods estimated at the mean parameter value. A useful feature of HB procedures is that the posteriors of parameters $\bar{\theta}$ and Ω_θ are simulated jointly with the posteriors of individual parameters θ_i. It thus provides not only estimates of the characteristics of parameter distribution in the sample, but also estimates

[8]Distributions are assumed to be log-normal for nonnegative parameters and normal for other parameters.

[9]If priors are symmetrical, there is no confirmation bias at round 2. This explains why a nonzero q is introduced for rounds 2 and 3 only.

[10]It assumes that elements of $\bar{\theta}$ have normal priors with large variance (i.e., flat, noninformative priors), and that Ω_θ has an inverse Wishart prior where the degrees for freedom and scale are the number of parameters and the identity matrix, respectively.

of the individual parameters (taken as the means of the posteriors of the individual distributions).

6.3 *Results*

This section presents the estimated indexes of biases, and their variations in our sample, without excluding any subjects. The results of the estimations are presented in Table 3. Parameter estimates (standard errors) are the means (standard deviation) of their posterior distribution and the precision of estimations is measured by the 95% credible interval.

We start with the representative-agent models where no between-individual heterogeneity is assumed neither in priors nor in model parameters, to provide a first picture of the mean patterns. In Model 0, indexes p and q are fixed at 0, meaning that the representative agent must update her beliefs according to Bayes' rule without bias. The α and β parameters of the prior distribution take values 1.6 ([1.49, 1.65]) and 1.5 ([1.42, 1.57]). The similarity of these two values suggests that the prior distribution of our representative subject was roughly symmetric. Consistently with the instructions we provided, the subjects did not expect one color to be more likely than the other, before receiving any signal. We note however that priors were not perfectly uniform. The parameters were larger than 1, 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.

Model 1 is a representative-agent model where the deviation indexes p and q are introduced and assumed to be the same for all the subjects. Regarding the measures of conservatism bias (p) and confirmatory bias (q), the estimations indicated strong conservatism: the representative subject behaved as if they neglected half of the actual signals (51.3%, 95%CI = [0.47, 0.55]). We also observed evidence for confirmatory bias: the representative subject behaved as if they misinterpreted 14% (95%CI = [0.11, 0.19]) of the signals that contradicted their prior beliefs.

Model 0 is nested in model 1, imposing $p = q = 0$. Its likelihood is lower than that of model 1 (−9956.6 vs. −9777.0). Using a test from classical statistics, a likelihood ratio test would reject Model 0 in favor of Model 1 ($p < 0.001$).

Model 2 extends Model 1 by allowing for heterogeneity in priors and indexes. In other words, Model 2 accounts for the fact that not all subjects may have the same priors, nor exhibit the same degree of deviations from Bayesian updating. The estimations suggested a large degree of between-subject variations. The average value of p in the sample was estimated as 0.28 (95%CI = [0.15, 0.40]), and the average value of q has an estimate of 0.19 (95%CI = [0.12, 0.26]). The estimated standard deviations of the distributions of the indexes are large when compared to the estimated means ($\sigma_p = 0.47$ and $\sigma_p = 0.09$, respectively). This comparison suggests that these parameters are highly dispersed over our subject sample, meaning that the magnitude of updating biases is highly heterogeneous, particularly for parameter p. We further discuss this heterogeneity at the end of this section.

Model 3 addresses the variations of the mean indexes across rounds. The model tests whether subjects become more (or less) conservative or exhibit stronger (or weaker)

TABLE 3. Model 0 is a representative-agent model without updating bias. Model 1 is a representative-agent model with updating biases. Model 2 allows for heterogeneity in prior parameters and bias parameters. Model 3 also includes round fixed effects for p and q.

	Model 0			Model 1			Model 2			Model 3		
	Estimate	Std Error	95%CI	Estimate	Std Error	95%CI	Estimate	Std Error	95%CI	Estimate	Std Error	95%CI
$\bar{\alpha}_0$	1.565	0.040	[1.490; 1.646]	1.302	0.025	[1.255; 1.353]	1.648	0.073	[1.516; 1.800]	1.643	0.071	[1.514; 1.788]
$\bar{\beta}_0$	1.494	0.037	[1.424; 1.569]	1.254	0.024	[1.210; 1.301]	1.506	0.060	[1.396; 1.630]	1.500	0.058	[1.394; 1.621]
$\bar{p}$				0.511	0.019	[0.472; 0.548]	0.277	0.064	[0.148; 0.399]	0.284	0.064	[0.153; 0.403]
$\bar{q}$				0.145	0.020	[0.106; 0.186]	0.194	0.035	[0.124; 0.262]	0.228	0.038	[0.155; 0.306]
σ_{α_0}							0.477	0.114	[0.296; 0.742]	0.463	0.106	[0.292; 0.706]
σ_{β_0}							0.311	0.070	[0.199; 0.471]	0.301	0.066	[0.196; 0.450]
σ_p							0.472	0.086	[0.327; 0.667]	0.474	0.085	[0.334; 0.665]
σ_q							0.092	0.036	[0.049; 0.175]	0.068	0.012	[0.048; 0.097]
Δp_2										−0.024	0.012	[−0.048; −0.003]
Δp_3										−0.002	0.008	[−0.018; 0.013]
Δq_3										−0.046	0.043	[−0.135; 0.036]
LL	−9956.627			−9777.009			−7888.125			−7880.273		

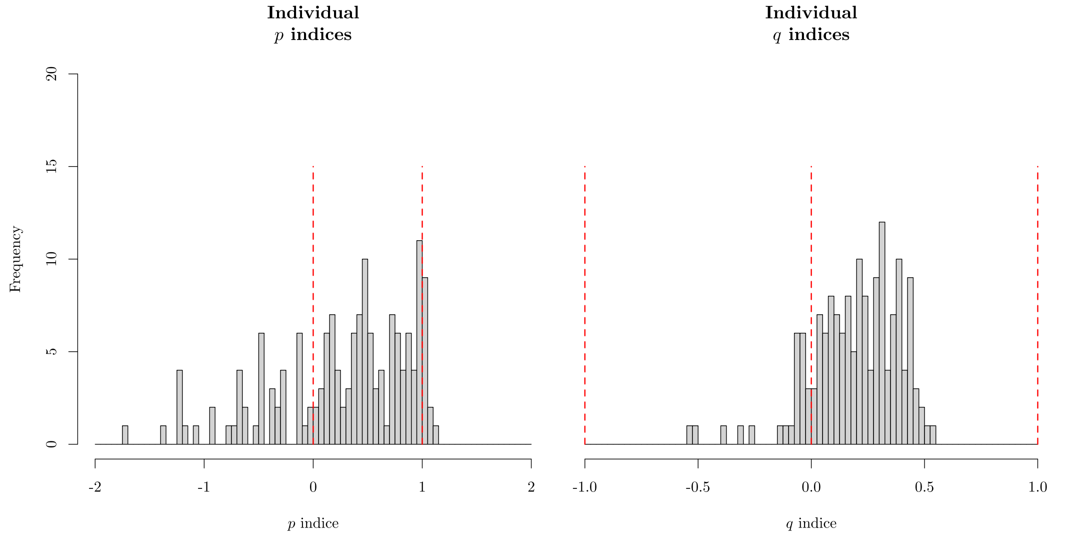

FIGURE 8. Histograms of estimates of indexes p (on the left) and q (on the right).

confirmatory bias in later rounds. The fixed effects showed that the conservatism index p decreased between rounds 1 and 2 ($\Delta p_2 = -0.02$, 95%CI $= [-0.048, -0.003]$), even though the difference is weakly significant in terms of behavior, given its low magnitude. The confirmatory index q did not vary significantly across rounds. A likelihood ratio test comparing Model 2 to Model 3 would show that the dummy variables capturing fixed effects across rounds are significant ($p < 0.001$).

Overall, Models 0 to 3 are nested one into another. Likelihood ratio tests comparing models two by two always favor the most complex model. Compared to Model 3, Model 2 has the advantage of providing population means of p and q over the overall experiment, which we use to make general statements on the two biases. Similarly, we use the Model 2 estimates to illustrate the heterogeneity of deviation indexes across subjects in Figure 8.

We observed that the majority (65%) of our subjects exhibited conservatism ($0 \leq p \leq 1$), and the second most common pattern (27%) was overinference ($p < 0$). We also found some evidence for prior-signal destruction ($p > 1$)—the extreme case of conservatism—which held for 8% of our subjects. The presence of a local 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 (85%) 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 were observed in our data.

In sum, our subjects on average exhibited conservatism and confirmation bias, but a large heterogeneity was captured at the individual level. Aside from the modal patterns, individual patterns revealed that overinference 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 five aspects: (1) the measurement method, (2) the exclusion of subjects exhibiting multiple switching patterns in the choice lists, (3) the presence of extreme signal sets in some sessions, (4) the estimation procedure, and (5) the presence of confirmatory bias at round 1. We focus on the representative-agent Model 1 and the most general Model 3.

Our experiment used two 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 reestimated the representative-agent model (Model 1) and the most flexible model (Model 3) with fixed effects capturing the differences in mean estimates across methods. The details of these estimations are reported in Appendix B.[11] We observed that the average patterns of the deviations from Bayesian updating were qualitatively stable across the two measurement methods. More precisely, the fixed effects capturing the impact of the method on p and q are not significant, suggesting that the estimation of the biases is robust to the measurement method. This is not the case with prior beliefs. The prior Beta distribution parameters 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 nonneutral ambiguity attitudes, which could potentially affect the results. Empirically, we found that the matching probability method did introduce differences in estimations. Nevertheless, the method still captured the main qualitative patterns of deviations from Bayesianism (p and q).

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. Among the fifteen subjects who exhibited multiple switching, five did so for only one choice list. To evaluate the possible impact of the inclusion of subjects without multiple switching, and for comparison with the results reported in Section 5, we ran additional estimations of Models 1 and 3 where we excluded all the observations of the 15 subjects who exhibited multiple switching in at least one choice list. The estimations are reported in Appendix C. The results are virtually identical to those reported in the main analysis.

The third aspect deals with the possible impact of extreme signal observations in our experimental sessions. For example, subjects in sessions 1 and 6 always observed the same color across all the rounds, and subjects in 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 reestimated Models 1 and 3 (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.

[11] We also ran separate estimations on the data produced by the two belief-measurement methods. They gave results consistent with the pooled data but with higher standard errors. See Appendix B.

The fourth aspect regards the estimation procedure of our models. Models 2 and 3 allow for heterogeneity in prior beliefs and in indices p and q. Our modeling approach assumes that the related parameters are distributed in the sample, and that the moments (mean and variance) of their distribution can be estimated. This approach is called the random parameter approach (as parameters are assumed to be randomly distributed in the sample). Two usual ways for estimating such a model are the Hierarchical Bayes estimations and the Maximum Simulated Likelihood estimations (MSL, Train (2009)). To assess the stability of our results across estimation methods,[12] we report the estimations of Models 0 to 3 using MSL in Appendix E.[13] The results are virtually identical for Models 0 and 1 and qualitatively identical for Models 2 and 3, in terms of means and variance of indexes p and q. The two parameters are significant and largely dispersed, and p has a greater mean and variance than q.

Our main analysis assumes that confirmatory bias, captured by index q is at play for rounds 2 and 3 but not for round 1. The normative explanation for this is that at round 1, not having observed any signals yet, the subjects do not have any objective reason to believe that one color is more likely than the other, and thus should not show confirmatory bias toward a given color. As a robustness check, we take a behavioral perspective and allow for confirmation bias at round 1. Subjects may indeed engage in wishful thinking and believe that the color they have chosen as the winning color is more likely. We reestimated Models 1 and 3, including the effect of q at round 1 (see Appendix F). In Model 1, the estimate of q is lower while the estimate of p remains unchanged. This suggests that the effect of q at round 1 is weaker than at the other rounds. Model 3 refines the analysis by capturing round-specific means of indexes p and q, taking their values at round 1 as a reference. Focusing on round 1, the mean of q loses statistical significance, which accords with no confirmatory bias at round 1. Similar patterns are also observed in the MSL estimations.[14]

7. Discussion

7.1 *Discussion of the theory*

Grether (1980) models 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

[12] Previous versions of this paper only reported the MSL estimates. Convergence issues in additional analysis during the revision of the paper made us switch to HB.

[13] The likelihoods are simulated using 1000 Halton draws and the standard errors are computed using the sandwich estimator of the variance-covariance matrix.

[14] However, this part of the analysis led to convergence issues with the MSL procedure. Estimations tended to be starting-value specific, suggesting local optima. Surprisingly, and in contradiction with all other models and robustness checks, p decreases drastically and becomes nonsignificant with HB (not with MSL). We suspect that it created a form of multicollinearity or identification issues. One reason for this is that our experimental design introduces exogenous variation in beliefs at rounds 1 to 3 by providing different signals across rounds and sessions. This was not the case at round 0 where all the subjects received the same description of the wheel and had very symmetric priors.

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, with the recent exception of Howe, Perfors, Walker, Kashima, and Fay (2022).

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 the total number and distribution of perceived signals. In a nutshell, two dimensions (sum and relative proportion) can only give two parameters (without further assumptions).

Identification issues may also arise if we open up the possibility that decision-makers update after each signal or after a few signals. In our theoretical framework and in the empirical implementation, (biased) updating was applied to a set of signals. In other words, we assumed that beliefs are formed when we observe them, at the decision time. At a given round, it would not be possible to identify both the parameters and whether updating was applied sequentially or on the whole set. We chose the assumption that maximizes portability. Assuming that beliefs are updated continuously, after each signal, would make the model less portable. In a Bayesian setup, only summary statistics (number of successes and failures) are enough to derive beliefs, and neither knowing when signals were received nor figuring out how they were dealt with (as a whole or sequentially) is necessary to implement the Bayesian model. Our extension shares this feature.

Although we provide intuitive interpretations of our parameters, 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, Rabin, and Raymond (2016), Falk and Zimmermann (2017)).

7.2 *Discussion of the empirics*

To the best of our knowledge, all previous studies on belief updating elicited only the mean of the belief distribution (see, for instance, Goeree, Palfrey, Rogers, and McKelvey (2007), Moreno and Rosokha (2016), Esponda, Vespa, and Yuksel (2024), Bland and Rosokha (2021), Möbius et al. (2022)), and the vast majority considered only two possible states (e.g., two competing hypotheses for a probability of success and failure) rather than the full distribution over all possible states Esponda, Vespa, and Yuksel (2024), Möbius et al. (2022); for a review, see also Section 4 in Benjamin (2019)).

In order to structurally estimate our model and empirically disentangle the distinct impact of symmetric and asymmetric biases on choices, we need estimates of decision makers' full belief distributions regarding the underlying uncertain data generating process. These were, however, not available in existing data sets. Therefore, we designed out tailor-made experiment that allowed for structurally estimate our model. First, instead of only two possible states, we considered a refined state space (from the success rate $s = 0\%$ to $s = 100\%$ with increment of 1%). Second, we elicited several quantiles of the distribution, rather than just the mean or the median. This enabled us to fit the shape of the full belief distribution, which is crucial for parameter identification as we have shown in Section 2. Third, our experimental approach is robust to confounds that may bias belief elicitations, including nonneutral risk and ambiguity attitudes. Lastly, all of our belief measures were real-incentivized, which enhanced the internal validity of our data.

Antoniou, Harrison, Lau, and Read (2015) and Moreno and Rosokha (2016) adopt a similar empirical approach to ours but only investigate conservatism (a symmetric bias). It is possible that parts of the conservatism they document is due to missing "disconfirming signals," a symptom of the asymmetric confirmatory bias, however, their approach does not allow for empirical disentanglement of these two types of biases. Möbius et al. (2022) and Coutts (2019) consider both types of deviations from Bayesian updating, similarly as we do. However, their model considers only two possible states, and their empirical approach elicited only the mean beliefs. As a result, their model cannot be structurally estimated, and hence it produces evidence only in reduced form. Buser, Gerhards, and Van Der Weele (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.

Another common practice in existing studies is that priors are assumed to be uniform (Goeree et al. (2007), Möbius et al. (2022)), even when the assumption is sometimes unrealistic (Esponda, Vespa, and Yuksel (2024)). Similarly as in Moreno and Rosokha (2016) and Bland and Rosokha (2021), we elicit and estimate subjects' priors in round 0, instead of imposing such assumptions. In this sense, our approach is fully subjective and more generally applicable in many real-life decision situations, where control over prior beliefs is often unavailable.

The caveat to bear in mind is that, in the absence of prior information, the decision situation may be perceived as ambiguous, where nonneutral ambiguity attitudes, such as ambiguity aversion (Ellsberg (1961)), may distort belief elicitation. This could cause a problem for matching probabilities method (also referred to as the cross-over method) used in many studies (Grether (1992), Holt and Smith (2009), Coutts (2019), Möbius et al. (2022)). The exchangeable events method is robust to this problem (Abdellaoui et al. (2011)), while being more sensitive to more extreme beliefs. We included both methods in our experiment to tap into the strengths of these methods. Our results suggest that the exchangeable events method is better at capturing nonuniform priors, however, both methods gave similar estimates of belief updating biases.

In our neutral setting, decision-makers exhibited conservatism. Evidence in the literature shows that depending on various situational factors, people may under or overreact to novel information (Griffin and Tversky (1992), Baeriswyl and Cornand (2014), Luo, Nie, and Young (2015)). Our index of symmetric biases can capture not only underinference (conservatism) but also overinference, making it suitable to identify various factors that affect decision-makers' reactions to information.

Our study also contributes to the empirical literature on confirmatory bias. Despite the abundance of theoretical models on confirmatory bias in the 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, Fetchenhauer, and Hoelzl (2011), Sinkey (2015), Andrews, Logan, and Sinkey (2018)). Nevertheless, the same problem of missing a clear Bayesian benchmark remains for these studies. In addition to the recent study of Buser, Gerhards, and Van Der Weele (2018), several other studies have investigated the 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), Coutts (2019), Coutts, Gerhards, and Murad (2024), Möbius et al. (2022)). Differing 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. For example, Eil and Rao (2011) argue that confirmation of prior beliefs occurs 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. In particular, our findings 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 model and estimate the full belief distribution regarding the data generating mechanism, in a setup similar to real-life situations. We provided natural interpretations of well-known asymmetric and symmetric updating biases and made them observable from choices. Our empirical approach, tailor-made for structural estimation of the model parameters, provided quantitative estimates of both types of biases. Empirical identification of these biases was made possible through their orthogonal impact on observable 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 suggest the empirical validity of our model and method by reproducing common findings on Bayesian updating in a general updating environment where subjective priors were not restricted to be over a limited set of data generating mechanisms and subjects could reveal their full subjective beliefs over all possible states of the world. Our portable model and empirical approach can thus be applied to investigate subjective beliefs and their updating under wide range of situations in different 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 execute the randomization tasks to ensure 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 replication package.[15] Throughout the experiment, subjects could ask questions when anything was unclear. A training wheel was used during the instructions for illustration purposes. The training wheel was covered in 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 in blue and red, and there were no other colors on the wheel. He then spun 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 an 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. In practice, he drew a sticker from a bag containing many stickers and put the sticker on the wheel. Figure A.1 shows the wheel without and with a sticker. 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 every choice round, and the sampling period proceeded. The outcome of every spin was announced by the implementer, and written down on the white board by the experimenter. New questionnaires were handed out after each sampling period.

[15] https://doi.org/10.5281/zenodo.14040733

Figure A.1. Wheel without and with sticker.

At the end of the experiment, the color composition of the wheel, the card suit, and the choice round drawn for the payment stage were revealed to the subjects by the implementer. The subjects were requested to open their envelopes and 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.

Appendix B: Consistency of the results across the different measurement methods

Table B.1 presents the estimations of Model 1 and Model 3 to test differences across different measurement methods. The estimates reported in the first rows are based

Table B.1. Comparison of estimates across measurement methods.

	Model 1			Model 3		
	Estimate	Std Error	95%CI	Estimate	Std Error	95%CI
$\bar{\alpha}_0$	1.182	0.031	[1.125; 1.246]	1.586	0.073	[1.453; 1.737]
$\bar{\beta}_0$	1.142	0.029	[1.090; 1.202]	1.428	0.059	[1.320; 1.551]
$\bar{p}$	0.485	0.030	[0.427; 0.542]	0.255	0.065	[0.123; 0.378]
$\bar{q}$	0.187	0.029	[0.127; 0.244]	0.258	0.042	[0.177; 0.339]
σ_{α_0}				0.488	0.115	[0.305; 0.756]
σ_{β_0}				0.301	0.070	[0.190; 0.461]
σ_p				0.505	0.088	[0.358; 0.699]
σ_q				0.078	0.013	[0.056; 0.107]
Δp_2				−0.025	0.011	[−0.049; −0.005]
Δp_3				−0.008	0.010	[−0.029; 0.012]
Δq_2				−0.011	0.033	[−0.070; 0.055]
ΔM_{α_0}	0.301	0.055	[0.186; 0.406]	0.135	0.023	[0.093; 0.181]
ΔM_{β_0}	0.286	0.053	[0.178; 0.387]	0.154	0.024	[0.107; 0.202]
ΔM_p	0.017	0.035	[−0.050; 0.086]	0.011	0.005	[0.002; 0.020]
ΔM_q	−0.062	0.035	[−0.129; 0.008]	−0.105	0.023	[−0.149; −0.061]
LL	−9749.621			−7846.532		

Table B.2. Model 1 with different measurement methods.

	MP			Exch		
	Estimate	Std Error	95%CI	Estimate	Std Error	95%CI
$\bar{\alpha}_0$	1.446	0.048	[1.353; 1.544]	1.171	0.024	[1.124; 1.219]
$\bar{\beta}_0$	1.390	0.045	[1.303; 1.480]	1.131	0.023	[1.087; 1.177]
$\bar{p}$	0.551	0.028	[0.493; 0.601]	0.453	0.027	[0.396; 0.505]
$\bar{q}$	0.135	0.031	[0.076; 0.199]	0.197	0.025	[0.147; 0.247]
LL	−4457.278			−5011.836		

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.

Table B.2 and Table B.3 present the estimations of Model 1 and Model 3 based on different measurement methods separately.

Table B.3. Model 3 with different measurement methods.

	MP			Exch		
	Estimate	Std Error	95%CI	Estimate	Std Error	95%CI
$\bar{\alpha}_0$	1.903	0.116	[1.698; 2.148]	1.418	0.064	[1.303; 1.553]
$\bar{\beta}_0$	1.782	0.111	[1.585; 2.020]	1.345	0.056	[1.244; 1.465]
$\bar{p}$	0.285	0.065	[0.149; 0.406]	0.177	0.077	[0.017; 0.318]
$\bar{q}$	0.184	0.047	[0.093; 0.276]	0.263	0.040	[0.183; 0.340]
σ_{α_0}	0.885	0.268	[0.491; 1.516]	0.342	0.090	[0.199; 0.551]
σ_{β_0}	0.895	0.258	[0.505; 1.505]	0.262	0.067	[0.155; 0.414]
σ_p	0.394	0.074	[0.273; 0.561]	0.561	0.122	[0.367; 0.842]
σ_q	0.083	0.017	[0.054; 0.119]	0.076	0.017	[0.051; 0.117]
Δp_2	−0.006	0.006	[−0.019; 0.004]	−0.007	0.010	[−0.025; 0.013]
Δp_3	−0.009	0.009	[−0.020; 0.011]	0.005	0.010	[−0.015; 0.025]
Δq_2	−0.083	0.052	[−0.183; 0.014]	−0.025	0.043	[−0.122; 0.056]
LL	−3162.449			−3764.469		

Appendix C: Robustness to multiple switching

Table C.1 reports the results of estimations of Model 1 and Model 3 (presented in Table 3) with an exclusion rule for subjects exhibiting multiple switching: all subjects with multiple switching for at least one choice list (i.e., 18 subjects) are removed from the analysis.

Table C.1. Models 1 and 3 without outliers.

	Model 1			Model 3		
	Estimate	Std Error	95%CI	Estimate	Std Error	95%CI
$\bar{\alpha}_0$	1.290	0.025	[1.243; 1.342]	1.639	0.075	[1.503; 1.794]
$\bar{\beta}_0$	1.248	0.024	[1.203; 1.297]	1.519	0.062	[1.405; 1.649]
$\bar{p}$	0.493	0.020	[0.453; 0.531]	0.259	0.067	[0.122; 0.386]
$\bar{q}$	0.137	0.020	[0.098; 0.175]	0.186	0.041	[0.108; 0.267]
σ_{α_0}				0.510	0.122	[0.318; 0.794]
σ_{β_0}				0.335	0.077	[0.213; 0.511]
σ_p				0.471	0.088	[0.327; 0.666]
σ_q				0.079	0.014	[0.054; 0.112]
Δ_{p_2}				−0.014	0.008	[−0.029; 0.002]
Δ_{p_3}				0.006	0.008	[−0.010; 0.022]
Δ_{q_3}				−0.024	0.040	[−0.107; 0.046]
LL	−8668.825			−6905.541		

Appendix D: Robustness to extreme signals

To ensure that our results are not driven by rounds where subjects received extreme signals (i.e., no success or failure), we reestimated Model 1 and Model 3 based on a data set from which data corresponding to such rounds were removed. The results are presented in Table D.1.

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

Table D.1. Models 1 and 3 without extreme signals.

	Model 1			Model 3		
	Estimate	Std Error	95%CI	Estimate	Std Error	95%CI
$\bar{\alpha}_0$	1.121	0.017	[1.091; 1.157]	1.465	0.056	[1.364; 1.580]
$\bar{\beta}_0$	1.095	0.016	[1.066; 1.129]	1.336	0.044	[1.256; 1.430]
$\bar{p}$	0.470	0.034	[0.398; 0.534]	0.227	0.080	[0.062; 0.376]
$\bar{q}$	0.110	0.023	[0.065; 0.155]	0.129	0.041	[0.044; 0.208]
σ_{α_0}				0.236	0.059	[0.145; 0.370]
σ_{β_0}				0.131	0.032	[0.082; 0.206]
σ_p				0.367	0.087	[0.232; 0.574]
σ_q				0.074	0.016	[0.049; 0.110]
Δp_2				−0.038	0.024	[−0.078; 0.010]
Δp_3				−0.005	0.022	[−0.046; 0.043]
Δq_3				−0.003	0.051	[−0.109; 0.095]
LL	−5155.124			−4324.244		

Appendix E: Alternative estimation method

Table E.1 reports the maximum simulated likelihood estimations. Model 0 is a representative-agent model with neither conservatism nor confirmatory bias ($p = q = 0$). Model 1 is a representative-agent model allowing for conservatism and confirmatory bias. Model 2 allows for heterogeneity in prior parameters α and β_0, as well as bias indices p and q. Model 3 includes round fixed effects for p and q.

Table E.1. Maximum simulated likelihood estimations.

	Model 0			Model 1			Model 2			Model 3		
	Estimate	Std Error	p Value	Estimate	Std Error	p Value	Estimate	Std Error	p Value	Estimate	Std Error	p Value
$\bar{\alpha}_0$	1.563	0.150	<0.001	1.302	0.073	<0.001	1.311	<0.001	<0.001	1.312	0.020	<0.001
$\bar{\beta}_0$	1.492	0.134	<0.001	1.253	0.063	<0.001	1.184	0.003	<0.001	1.206	0.024	<0.001
$\bar{p}$				0.513	0.069	<0.001	0.369	0.005	<0.001	0.336	0.080	<0.001
$\bar{q}$				0.144	0.037	<0.001	0.180	0.041	<0.001	0.196	0.019	<0.001
σ_{α_0}							0.172	0.005	<0.001	0.177	0.021	<0.001
σ_{β_0}							0.178	<0.001	<0.001	0.181	0.046	<0.001
σ_p							0.496	<0.001	<0.001	0.546	0.065	<0.001
σ_q							0.156	0.039	<0.001	0.153	0.014	<0.001
Δp_2										−0.063	0.012	<0.001
Δp_3										−0.036	0.019	0.064
Δq_3										−0.049	0.051	0.342
LL	−10,214.888			−10,027.768			−9061.891			−9054.101		

Appendix F: Allowing for confirmatory bias at round 1

Unlike in the main text, where confirmatory bias was only estimated for rounds 2 and 3, Tables F.1 and F.2 include it in round 1 as well.

Table F.1. Estimations of Models 1 and 3 accounting for confirmatory bias (q) at round 1.

	Model 1			Model 3		
	Estimate	Std Error	95%CI	Estimate	Std Error	95%CI
$\bar{\alpha}_0$	1.286	0.025	[1.236, 1.335]	1.552	0.065	[1.431; 1.689]
$\bar{\beta}_0$	1.246	0.023	[1.201, 1.292]	1.417	0.054	[1.316; 1.530]
$\bar{p}$	0.501	0.021	[0.459, 0.542]	0.089	0.083	[−0.080; 0.247]
$\bar{q}$	0.042	0.014	[0.013, 0.069]	−0.028	0.033	[−0.091; 0.038]
σ_{α_0}				0.380	0.088	[0.240; 0.585]
σ_{β_0}				0.248	0.056	[0.159; 0.376]
σ_p				0.669	0.128	[0.457; 0.963]
σ_q				0.064	0.010	[0.048; 0.085]
Δp_2				−0.012	0.011	[−0.033; 0.009]
Δp_3				−0.002	0.011	[−0.024; 0.018]
Δq_2				0.201	0.030	[0.142; 0.260]
Δq_3				0.239	0.029	[0.185; 0.295]
LL	−9801.830			−7954.480		

Table F.2. MSL of Models 1 and 3 accounting for confirmatory bias (q) at round 1.

	Model 1			Model 3		
	Estimate	Std Error	p Value	Estimate	Std Error	p Value
$\bar{\alpha_0}$	1.285	0.048	<0.001	1.193	0.006	<0.001
$\bar{\beta_0}$	1.245	0.042	<0.001	1.128	<0.001	<0.001
$\bar{p}$	0.504	0.074	<0.001	0.196	0.042	<0.001
$\bar{q}$	0.041	0.032	0.205	0.042	0.025	0.088
σ_{α_0}				0.075	0.006	<0.001
σ_{β_0}				0.016	<0.001	<0.001
σ_p				0.625	0.030	<0.001
σ_q				0.174	0.014	<0.001
Δp_2				−0.022	0.018	0.223
Δp_3				−0.023	0.018	0.204
Δq_3				0.162	0.045	<0.001
Δq_2				0.222	0.057	<0.001
LL	−10,050.749			−8997.823		

References

Abdellaoui, Mohammed, Aurélien Baillon, Laetitia Placido, and Peter P. Wakker (2011), "The rich domain of uncertainty: Source functions and their experimental implementation." *The American Economic Review*, 101 (2), 695–723. [0338, 0354]

Abdellaoui, Mohammed, Han Bleichrodt, Emmanuel Kemel, and Olivier l'Haridon (2021), "Measuring beliefs under ambiguity." *Operations Research*, 69 (2), 599–612. [0333, 0341]

Allais, Maurice (1953), "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine." *Econometrica: Journal of the Econometric Society*, 21 (4), 503–546. [0337]

Ambuehl, Sandro and Shengwu Li (2018), "Belief updating and the demand for information." *Games and Economic Behavior*, 109, 21–39. [0329]

Andrews, Rodney J., Trevon D. Logan, and Michael J. Sinkey (2018), "Identifying confirmatory bias in the field: Evidence from a poll of experts." *Journal of Sports Economics*, 19 (1), 50–81. [0355]

Antoniou, Constantinos, Glenn W. Harrison, Morten I. Lau, and Daniel Read (2015), "Subjective Bayesian beliefs." *Journal of Risk and Uncertainty*, 50 (1), 35–54. [0354]

Baeriswyl, Romain and Camille Cornand (2014), "Reducing overreaction to central banks' disclosures: Theory and experiment." *Journal of the European Economic Association*, 12 (4), 1087–1126. [0354]

Baillon, Aurélien (2008), "Eliciting subjective probabilities through exchangeable events: An advantage and a limitation." *Decision Analysis*, 5 (2), 76–87. [0338]

Bénabou, Roland and Jean Tirole (2002), "Self-confidence and personal motivation." *The Quarterly Journal of Economics*, 117 (3), 871–915. [0330]

Bénabou, Roland and Jean Tirole (2006), "Belief in a just world and redistributive politics." *The Quarterly Journal of Economics*, 121 (2), 699–746. [0330]

Bénabou, Roland and Jean Tirole (2016), "Mindful economics: The production, consumption, and value of beliefs." *Journal of Economic Perspectives*, 30 (3), 141–164. [0329, 0330]

Benjamin, Daniel J. (2019), "Errors in probabilistic reasoning and judgment biases." In *Handbook of Behavioral Economics: Applications and Foundations 1*, Chapter 2, 69–186. [0330, 0353]

Benjamin, Daniel J., Matthew Rabin, and Collin Raymond (2016), "A model of nonbelief in the law of large numbers." *Journal of the European Economic Association*, 14 (2), 515–544. [0330, 0353]

Bland, James R. and Yaroslav Rosokha (2021), "Learning under uncertainty with multiple priors: Experimental investigation." *Journal of Risk and Uncertainty*, 62, 157–176. [0333, 0353, 0354]

Bohren, J. Aislinn, Alex Imas, and Michael Rosenberg (2019), "The dynamics of discrimination: Theory and evidence." *American Economic Review*, 109 (10), 3395–3436. [0330]

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer (2016), "Stereotypes." *The Quarterly Journal of Economics*, 131 (4), 1753–1794. [0330]

Buser, Thomas, Leonie Gerhards, and Joël Van Der Weele (2018), "Responsiveness to feedback as a personal trait." *Journal of Risk and Uncertainty*, 56 (2), 165–192. [0354, 0355]

Campos-Mercade, Pol and Friederike Mengel (2024), "Non-Bayesian statistical discrimination." *Management Science*, 70 (4), 2549–2567. [0332]

Charness, Gary and Chetan Dave (2017), "Confirmation bias with motivated beliefs." *Games and Economic Behavior*, 104, 1–23. [0330, 0331]

Christandl, Fabian, Detlef Fetchenhauer, and Erik Hoelzl (2011), "Price perception and confirmation bias in the context of a vat increase." *Journal of Economic Psychology*, 32 (1), 131–141. [0355]

Coutts, Alexander (2019), "Good news and bad news are still news: Experimental evidence on belief updating." *Experimental Economics*, 22 (2), 369–395. [0330, 0354, 0355]

Coutts, Alexander, Leonie Gerhards, and Zahra Murad (2024), "What to blame? Self-serving attribution bias with multi-dimensional uncertainty." *The Economic Journal*, 134 (661), 1835–1874. [0355]

De Finetti, Bruno (1937), "La prévision: Ses lois logiques, ses sources subjectives." *Annales de l'Institut Henri Poincaré*, 7 (1), 1–68. [0338]

Deryugina, Tatyana (2013), "How do people update? The effects of local weather fluctuations on beliefs about global warming." *Climatic change*, 118 (2), 397–416. [0330]

Edwards, Ward (1968), "Conservatism in human information processing." *Formal Representation of Human Judgment*, 17, 51. [0329, 0330]

Eil, David and Justin M. Rao (2011), "The good news-bad news effect: Asymmetric processing of objective information about yourself." *American Economic Journal: Microeconomics*, 3 (2), 114–138. [0330, 0355]

El-Gamal, Mahmoud A. and David M. Grether (1995), "Are people Bayesian? Uncovering behavioral strategies." *Journal of the American statistical Association*, 90 (432), 1137–1145. [0329, 0330]

Ellsberg, Daniel (1961), "Risk, ambiguity, and the savage axioms." *The quarterly journal of economics*, 75 (4), 643–669. [0354]

Epstein, Larry G. (2006), "An axiomatic model of non-Bayesian updating." *The Review of Economic Studies*, 73 (2), 413–436. [0330]

Ertac, Seda (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. [0355]

Esponda, Ignacio, Emanuel Vespa, and Sevgi Yuksel (2024), "Mental models and learning: The case of base-rate neglect." *American Economic Review*, 114 (3), 752–782. [0333, 0353, 0354]

Falk, Armin and Florian Zimmermann (2017), "Information processing and commitment." *The Economic Journal*, 128 (613), 1983–2002. [0353]

Goeree, Jacob K., Thomas R. Palfrey, Brian W. Rogers, and Richard D. McKelvey (2007), "Self-correcting information cascades." *The Review of Economic Studies*, 74 (3), 733–762. [0353, 0354]

Grether, David M. (1980), "Bayes rule as a descriptive model: The representativeness heuristic." *The Quarterly journal of economics*, 95 (3), 537–557. [0352, 0353]

Grether, David M. (1992), "Testing Bayes rule and the representativeness heuristic: Some experimental evidence." *Journal of Economic Behavior & Organization*, 17 (1), 31–57. [0354]

Griffin, Dale and Amos Tversky (1992), "The weighing of evidence and the determinants of confidence." *Cognitive Psychology*, 24 (3), 411–435. [0354]

Holt, Charles A. (2007), *Markets, Games, and Strategic Behavior: Recipes for Interactive Learning*. Pearson Addison Wesley. [0337]

Holt, Charles A. and Angela M. Smith (2009), "An update on Bayesian updating." *Journal of Economic Behavior & Organization*, 69 (2), 125–134. [0354]

Howe, Peter D. and Anthony Leiserowitz (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. [0330]

Howe, Piers Douglas Lionel, Andrew Perfors, Bradley Walker, Yoshihisa Kashima, and Nicolas Fay (2022), "Base rate neglect and conservatism in probabilistic reasoning: Insights from eliciting full distributions." *Judgment and Decision Making*, 17 (5), 962–987. [0353]

Johnson, Cathleen, Aurélien Baillon, Han Bleichrodt, Zhihua Li, Dennie Van Dolder, and Peter P. Wakker (2021), "Prince: An improved method for measuring incentivized preferences." *Journal of Risk and Uncertainty*, 62 (1), 1–28. [0341]

Karni, Edi (2009), "A mechanism for eliciting probabilities." *Econometrica*, 77 (2), 603–606. [0337]

Klayman, Joshua (1995), "Varieties of confirmation bias." *Psychology of learning and motivation*, 32, 385–418. [0355]

Luo, Yulei, Jun Nie, and Eric R. Young (2015), "Slow information diffusion and the inertial behavior of durable consumption." *Journal of the European Economic Association*, 13 (5), 805–840. [0354]

Machina, Mark J. and David Schmeidler (1992), "A more robust definition of subjective probability." *Econometrica: Journal of the Econometric Society*, 60 (4), 745–780. [0337]

Mezulis, Amy H., Lyn Y. Abramson, Janet S. Hyde, and Benjamin L. Hankin (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. [0331]

Miller, Dale T. and Michael Ross (1975), "Self-serving biases in the attribution of causality: Fact or fiction?" *Psychological Bulletin*, 82 (2), 213. [0331]

Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat (2022), "Managing self-confidence: Theory and experimental evidence." *Management Science*, 68 (11), 7793–7817. [0329, 0330, 0353, 0354, 0355]

Moreno, Othon M. and Yaroslav Rosokha (2016), "Learning under compound risk vs. learning under ambiguity–an experiment." *Journal of Risk and Uncertainty*, 53 (2–3), 137–162. [0333, 0336, 0353, 0354]

Nickerson, Raymond S. (1998), "Confirmation bias: A ubiquitous phenomenon in many guises." *Review of General Psychology*, 2 (2), 175. [0355]

Oswald, Margit E. and Stefan Grosjean (2004), "Confirmation bias." In *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*, Chapter 4. Psychology Press. [0329, 0355]

Phillips, Lawrence D. and Ward Edwards (1966), "Conservatism in a simple probability inference task." *Journal of Experimental Psychology*, 72 (3), 346. [0329, 0330]

Rabin, Matthew (2013), "An approach to incorporating psychology into economics." *The American Economic Review*, 103 (3), 617–622. [0330]

Rabin, Matthew and Joel L. Schrag (1999), "First impressions matter: A model of confirmatory bias." *Quarterly journal of Economics*, 114 (1), 37–82. [0329, 0330, 0331, 0332, 0333, 0334, 0336]

Raiffa, Howard (1968), *Decision Analysis: Introductory Lectures on Choices Under Undertainty*. Addison-Wesley. [0337, 0338]

Ramsey, Frank P. (1931), "Truth and probability (1926)." In *The Foundations of Mathematics and Other Logical Essays*, 156–198. [0337]

Sarsons, Heather (2017), "Interpreting signals in the labor market: Evidence from medical referrals." Job Market Paper. [0330]

Schotter, Andrew and Isabel Trevino (2014), "Belief elicitation in the laboratory." *Annual Review of Economics*, 6 (1), 103–128. [0337]

Sinkey, Michael (2015), "How do experts update beliefs? Lessons from a non-market environment." *Journal of Behavioral and Experimental Economics*, 57, 55–63. [0355]

Spetzler, Carl S. and Carl-Axel S. Stael von Holstein (1975), "Exceptional paper-probability encoding in decision analysis." *Management Science*, 22 (3), 340–358. [0337, 0338]

Train, Kenneth E. (2009), *Discrete Choice Methods With Simulation*. Cambridge University Press. [0346, 0352]

Tversky, Amos and Daniel Kahneman (1974), "Judgment under uncertainty: Heuristics and biases." *Science*, 185 (4157), 1124–1131. [0329, 0330, 0336]

Wilson, Andrea (2014), "Bounded memory and biases in information processing." *Econometrica*, 82 (6), 2257–2294. [0330]

Co-editor Garance Genicot handled this manuscript.

Manuscript received 14 February, 2022; final version accepted 11 November, 2024; available online 5 December, 2024.

The replication package for this paper is available at https://doi.org/10.5281/zenodo.14040733. The Journal checked the data and codes included in the package for their ability to reproduce the results in the paper and approved online appendices.