Learning gaps? Described versus Experienced Signals

Job Market Paper

David Gonzalez-Jimenez[a]

[a] *Erasmus School of Economics, Erasmus University Rotterdam* *

December 12, 2024

Abstract

We study the effect of learning the relationship between signals and payoff-relevant events by description or by experience. Employing a novel method for eliciting ambiguity attitudes, where beliefs and attitudes are disentangled, two gaps are identified. The first gap concerns attitudes: aversion is lower in experience and decreases with higher signal informativeness, while insensitivity is higher for experience but reduced by informativeness of the signal in both conditions. The second gap pertains to belief updating: subjects are more responsive to new information in experience than in description. The findings show that direct experience brings people closer to Bayesian rationality than description does.

Keywords: Decision from Experience; Description-Experience gap; Learning; Updating; Beliefs; Risk Attitude; Ambiguity Attitude.

*P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. Email: gonzalezjimenez@ese.eur.nl. I thank Peter Wakker, Aurelien Baillon and Chen Li for their support and guidance in this project, and Mohammed Abdellaoui for his hospitality during my visit to the department of Economics and Decision Sciences at HEC where an early version of this project benefited from his insightful remarks. This project was preregistered in as predicted platform https://doi.org/10.17605/OSF.IO/MUZFH

1 Introduction

Standard decision theory does not distinguish whether risk information is obtained from description or from experience. The underlying premise is that the context and process through which information is learned are not important; what matters is the payoffs and the information about them. However, a recent strand of literature has challenged this assumption by comparing the decisions individuals make when facing traditional lotteries, where they are presented by probability-payoff pairs (Decision from Description, DfD), with decisions made after discovering the distribution through sampling (Decision from Experience, DfE) (Hertwig, 2015). The most common finding is that individuals weight rare events differently depending on how they acquire risk information; this is known as the Description-Experience gap.

Although DfD and DfE both focus on how individuals interpret and respond to payoff distributions, there are cases where decision makers must not only know which contextual cues are relevant before making a decision, but also assess their degree of relevance. The importance of these situations becomes apparent when examining the literature on information economics, where events that determine payoffs of interest, referred to as Payoff-Relevant Events (PR-events), are not observable to an agent. Instead, decision-makers rely on observable events, known as signals, which do not (directly) impact the payoff but provide information through their relationship with PR-events (Milgrom, 1981). This relationship is expressed as a conditional probability, which individuals then use to calculate the probability that the event of interest, the PR-event, occurs through Bayes' theorem. An example can be diagnosing viral infections given the symptoms, or degrees, that should signal the productivity of individuals, which is unobservable, to employers (Crawford and Sobel, 1982). As was previously the case with the risk literature, the literature on updating to information has neglected the context and process in which beliefs are formed. Both aspects, context and process, are central in the studies of the DfD-DfE gap.

Building on the DfD-DfE gap identified in the literature, this work introduces the role of learning from signals into this framework. This addition provides a new dimension to understanding decision-making, as it examines how individuals develop associations between signals and PR-events, how the strength of these associations impacts their decisions, and how these processes differ between DfE and DfD. We aim first to understand the differences in belief distortion from uncertainty attitudes and belief updating between DfD and DfE. Second, we aim to understand how levels of signal informativeness affect uncertainty attitudes within both DfD and DfE.

As mentioned before, the DfD-DfE literature finds systematic differences in attitudes when considering PR-events, but in this work two sources of differences may arise. The first concern belief distortions due to ambiguity attitudes and the second one concerns biases in belief updating. Subjects in the experience condition do not know the true distribution and never really know it, thus distortions on their beliefs appear due to their ambiguity attitudes (Abdellaoui et al., 2011). As Subjects in DfD have all available information, they differ in their weighting of rare events which produces the gap (Hertwig and Erev, 2009), thus can be referred to as an attitudes gap. On the other hand, subjects are found to often deviate from Bayes theorem by under-reacting to information (conservatism; Benjamin 2019). However, experience influences belief updating, where professionals familiar with conditional probabilities align more with Bayesian principles than the general population (Araujo et al., 2024). These differences may stem from the sequential and engaging nature of experience (Bohren et al., 2024). To accurately assess experience's effects on belief updating, distinguishing between beliefs and attitudes is essential; otherwise, conclusions regarding attitudes towards uncertainty or underreaction may be biased.

To illustrate, consider a doctor diagnosing a patient. If its a fairly well-known disease, she relies on well-established clinical guidelines for diagnostic pathways, that is, which is the disease, its possible severity, and actions to take. One example is diabetes, the patient's blood sugar level is an observable feature, and through decades of research, there is a clear and known association between blood sugar levels and diabetes, giving her specific probabilities to diagnose the disease and its path in the presence of certain blood sugar levels. However, there are diseases with vaguer symptoms such as mild fever and fatigue, which could be related to the flu or many other conditions. In this case, the doctor may know the marginal distributions of flu cases during a particular season and the marginal distribution of symptoms in patients, but the likelihood that this particular combination of symptoms points to flu may not be as clear. Lacking the same kind of robust data as in the case of diabetes, she relies more on personal experience and subjective judgment to assess the diagnosis and the path of the disease.

Traditional decision-making theories would suggest that if both scenarios provide the same information on the likelihood of the disease, her beliefs update should be equivalent, derived from clinical guidelines or personal experience. However, it is not far fetched to assume that when using her own experience with a given population, she may give the appropriate weight to the evidence she observes. Moreover, when based on clinical guidelines, her diagnosis can underreact to observed evidence, resulting in less decisive action being taken.

The opposite effect might come from how they perceive uncertainty; she may feel more confident when they use well-established medical guidelines because the joint probability of symptoms and disease is explicitly known and quantified. This offers a level of precision that reduces uncertainty. In contrast, relying on personal experience without such concrete data, she might be more hesitant in a diagnosis, and its assessment of paths may feel vaguer, even if the likelihoods were similar. This might result in less decisive action looking as if she underreacted to evidence thought she did not. This scenario highlights how attitudes towards uncertainty can influence decisions beyond mere belief updates, emphasizing the need to distinguish between attitudes and beliefs in our analyses.

To separate attitudes from beliefs, we use the belief hedge method of Baillon et al. (2021). The DfD-DfE framework is modified in such a way that, using this method, it is possible to show the effect of changing the underlying informativeness of the signal on belief updating and ambiguity attitude. Regarding attitudes, we study two components: aversion and insensitivity. Aversion captures the motivational component of uncertainty attitudes, reflecting how much individuals approach or avoid uncertainty. In probability weighting representations, it is associated with the elevation parameter (Wakker, 2010). Insensitivity reflects a lack of discriminatory power between different levels of belief. It is related to the curvature of the probability weighting function (Dimmock et al., 2016). Once these two components are measured, beliefs can be recovered under certain assumptions (Li et al., 2019). The belief hedge method, therefore, provides independent measures of attitudes and beliefs. In other words, it corrects belief measurements for confounding ambiguity attitudes.

In the experiment, subjects were faced with three urns with balls that were colored and marked. The balls could be of three possible colors and were marked with letters A or B. The three urns had an equal 1/3 proportion of balls of each color and an equal 1/2 proportion of letters. The color of the ball was the PR-event, whereas the letter served merely as signals of the color. The correlations between colors and letters varied between different urns and were only partially revealed by sampling with replacement in DfE. Thus, with the marginal distributions of colors and of letters known and fixed, we focus on the joint distribution of signals and PR-events. Hence, three informativeness levels for the signal were introduced: no information signal, weak signal, and strong signal. In the Description arm, individuals were provided with full information about the urn composition, including how weak (i.e. noisy) the signal was. In the Experience arm, subjects sampled with replacement and hence had to infer the informativeness of the signal. After a ball was randomly selected from the urn and the letter on the ball was announced, matching probabilities were elicited, from which attitudes and beliefs could be derived. This procedure was applied for all three urns.

We find that the informativeness of the signal reduces aversion in DfE but not in DfD. For both Description and Experience, the informativeness decrease insensitivity. Aversion was lower in the experience condition for the three levels of signal informativeness, while insensitivity was higher in the experience arm for the no information and weak signals, but not for the strong signal. In terms of beliefs, individuals exhibited an underreaction on average in the description arm. Results on the Experience arm depend on the priors assumed; assuming highly informative priors leads to no difference between arms while assuming noninformative priors leads reactions closer to the Bayesian benchmark. These results reveal an attitude gap and hint at learning gaps, as they show how distinctively attitudes react to informativeness and how the more intuitive learning environment of DfE impacts more beliefs than DfD.

The contribution of this work to the DfD-DfE literature is as follows. We adapt the DfD-DfD framework to study learning from signals, whereas before this literature has focused on studying belief formation for PR-events. When only studying PR-events, the description-experience gap is quite robust and persists even when there is no under-sampling (Wulff et al., 2018). In economics, the literature on cognitive constraints has shown interest in this gap and its determinants (Bohren et al., 2024; Oprea and Vieider, 2024). In this work, by using signals, we study a broader type of learning where individuals, though uncertain whether an PR-event will happen, can associate contextual cues with the likelihood of the PR-event happening and differentiate between the strengths of these associations. This aligns with principles from the literature on associative learning, which emphasizes that learning is shaped by how well signals predict payoffs (Le Pelley et al., 2016). By separating beliefs and attitudes, a necessity that only became possible after Baillon et al.'s, not only an attitudes gap can be evaluated, but also the existence of a learning (from signals) gap, i.e. difference in the updating on both arms. In addition, it can be evaluated how ambiguity attitudes are influenced by the strength of association between signals and PR-events, thereby extending the DfD-DfE literature's focus beyond PR-events.

This work closely relates to recent literature that has emphasized the role of uncertain signals and how they affect individuals' choices. Epstein and Halevy (2024) find that when signals are uncertain, more individuals have posterior beliefs that do not average to the prior, indicating signal ambiguity. These authors also analyzed patterns of updating and their deviation from Bayesian updating. However, they noted that the patterns found might be distorted because beliefs and ambiguity attitudes were not separated. Similarly, Liang (2019) attempted to account for the confound of attitudes and beliefs by deriving theoretical predictions of how certainty equivalents behave in updating problems where the prior, the signal, or both were uncertain, and then compared these patterns with patterns found in experimental data. However, Liang's approach did not allow it to estimate how much reaction or underreaction to information there is, only to elucidate to which model they resembled more. The current work resolves the problems in the two aforementioned studies, because we can separate beliefs from attitudes we can thus test how individuals' beliefs deviate from Bayesian updating.

The main finding in the literature on deviations from Bayesian updating is that individuals underreact to information (conservatism) (Benjamin, 2019). Moreover, the characteristics of both the signals and the payoffs may affect how much individuals deviate from the Bayesian benchmark. For example, if information is good or bad news (Eil and Rao, 2011), and how informative the signal is affects how much individuals underreact to information (Ambuehl and Li, 2018). More recently, the literature on cognitive noise has turned its attention to biases in Bayesian updating, documenting how complexity, signal structure, and state-space structure may explain the biases found. The sequential nature of our DfE treatment has some resemblance to a treatment in Esponda et al. (2024) where they give feedback to individuals in a conditional problem without revealing the primitives. Relevant for this work are Ba et al. (2024), which documented the effect of complexity of the state space on updating, and Augenblick et al. (2024), which found overreaction to weak signals and underreaction to strong signals. However, the elicitation methods of these studies are not designed to capture attitudes toward risk and control for them. Again, this paper presents a method that obtains ambiguity attitudes and the degree to which individuals update their beliefs. Also note that DfE is typically seen as a task of lower complexity (Lejarraga and Hertwig, 2021), so comparing DfD with DfE is relevant in this literature.

The remainder of the paper is organized as follows. Section 2 outlines the experimental design and framework, Section 3 introduces the measures and indices, the hypothesis, and the statistical analysis. Section 4 presents the main results of the experiment, while in Section 5 we discuss possible explanations and implications.

2 Experimental design

2.1 Procedure

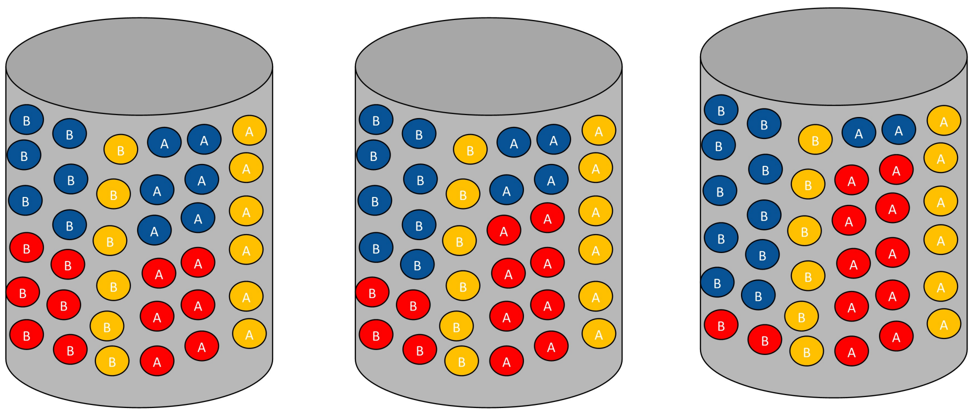
The experiment consists of a between-subjects design with two arms: the Experience arm (DfE) and the Description arm (DfD). Subjects in both treatments faced three urns that differed in the level of informativeness associated with the underlying signal: no (for no information), weak, and strong signal urns. The urns contained balls that were colored balls, blue, yellow, and red in equal proportions of 1/3, and marked with a letter, A, or B in equal proportions of 1/2. The correlation between colors and letters differed across the urns, hence varying the degree of informativeness.

Subjects in both conditions were informed at the beginning about the marginal distribution of colors and letters within the urns. They were also told that each color would have at least one ball marked with A and one ball marked with B. The varying signal probabilities conditional on the colors of each urn are shown in Table 1:
In informative cases (the "weak" and "strong" urn), letter A was always a signal favoring blue and red disfavoring, and letter B was the opposite. The letters never signaled anything about yellow.

Subjects were paid a show-up fee of $5.00 USD and had the opportunity to win a bonus of $20 USD. One of their choices was selected at random to determine the bonus payment. The question implemented for real at the end was selected before any choices were made (but kept hidden) to ensure that uncertainty attitudes did not affect the incentive compatibility of the payment.

	Urn Informativeness		
	No	*Weak*	*Strong*
$P(A\|Blue)$	1/2	1/3	1/6
$P(B\|Blue)$	1/2	2/3	5/6
$P(A\|Yellow)$	1/2	1/2	1/2
$P(B\|Yellow)$	1/2	1/2	1/2
$P(A\|Red)$	1/2	2/3	5/6
$P(B\|Red)$	1/2	1/3	1/6

Table 1: Signal likelihood of signal condition on color by urn



(a) No (b) Weak (c) Strong

Figure 1: Urns color and letter distribution

The procedure was as follows: First, subjects received common information about the three urns they would encounter. This information included the proportions of balls of colors and letters in the urns. Subjects were also informed that each color had at least one ball marked with each letter.

The subjects then received information on the joint distribution of colors and letters in the first urn according to their assigned arm. The computer then performed a random draw from the urn. The color of the drawn ball was kept hidden, but the letter on it was announced. subjects then made a series of choices between two types of prospects: one prospect paid \$20 if the color of the retrieved ball matched a specified color ($20_{Col}0$), and the other prospect was a lottery with a specified probability p of winning \$20 ($20p0$). To explain the mechanics of the lottery, let us look at a particular decision in which the subject chose between the event and a lottery with a chance of winning of 50 out of 100. Before the decision is made, a random number between 1 and 100 was picked but it was not revealed. If the subject chose the lottery and the selected number was smaller or equal to 50 (that is, the chances of winning), then the subject would receive the bonus of \$20. It is well known that probabilities are easier to understand if expressed as relative frequencies (Gigerenzer, 1996).

For example, as in Figure 2, a participant might choose between winning the bonus if the ball marked A is red and playing a lottery with a chance of winning 50%. The matching probability is the probability p at which the participant is indifferent between the two options $20_{Col}0 \sim 20p0$. It will depend on the participant's belief that the event would occur and the participant's attitudes towards uncertainty. The matching probabilities between the lottery and the bet on the color were derived using the bisection method, and the matching probabilities are defined as the midpoint between switching points. A further explanation of the bisection method can be found in Appendix A. Six matching probabilities were elicited: three for single-event prospects (e.g., the retrieved ball is red) and three for composite events (e.g., the retrieved ball is red or blue). Monotonicity of matching probabilities of composite events (exceeding the matching probabilities of contained single events) was enforced in the choices. After the six matching probabilities were elicited, subjects continued to the next urn until they completed this process for all three urns. The order of presentation of the urn was randomized.

Figure 2: Example of a matching probability

2.2 Information arms

As previously mentioned, the treatment arms differed only in the type of information provided on the contents of the urn before the matching probabilities were elicited. In the DfD arm, subjects were given exact information about how many balls of each color were marked with A and how many were marked with B before the ball that would determine the state of the world was drawn. Figure 3 gives an example.

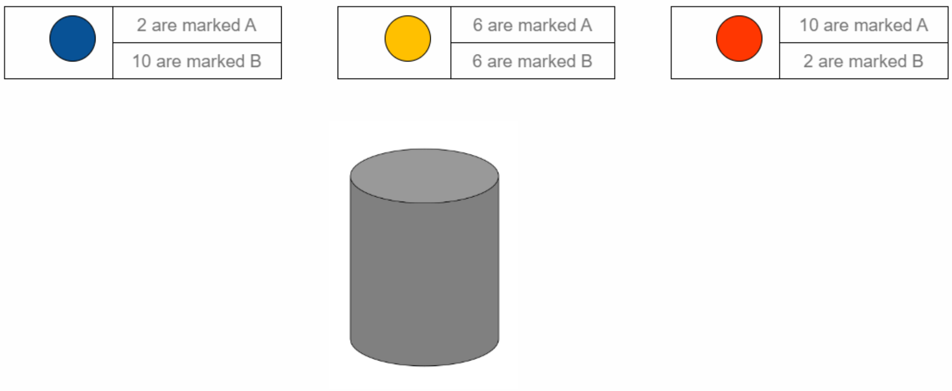


Figure 3: Interface of the information in the urn in the DfD arm

In the DfE arm, subjects were not informed about the exact composition of the urn. Instead, they were given the opportunity to sample with replacement from the urn as much as they wanted. In line with findings from the DfD-DfE literature, which has shown that memory deficiencies can be a source of probability distortion (Cubitt et al., 2022), a memory aid was provided displaying the entire sampling history to avoid memory deficiencies. Figure 4 displays an example of the interface used in the Experience arm.

3 Analysis

Eliciting the six matching probabilities allows us to adopt the belief hedge method (Baillon et al., 2021). In particular, we use the specification found in Baillon et al. (2018). This method allows for the measurement of ambiguity attitudes without knowing the subjects' beliefs and the derivation of such beliefs. This separation will allow us to identify the two types of effects that the signal has, first in the decision maker's attitudes towards uncertainty and second in their beliefs. For example in the case of the strong signal, if it is A individuals should change their beliefs in the direction of the ball being Blue with a higher probability, while decreasing the probability of Red. At the same time,

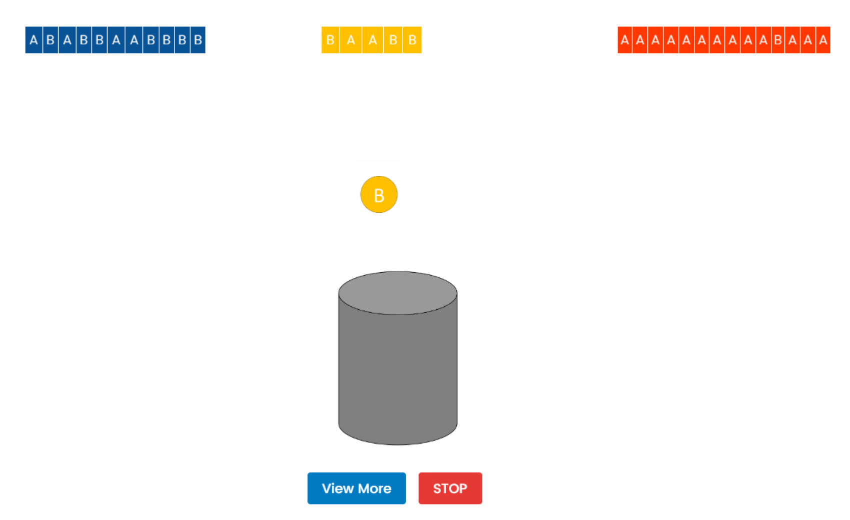


Figure 4: Interface for the Sampling stage of DfE arm

the ease with which subjects can associate the letter with a color in such a high information signal must make individuals more certain. Such an effect should be more noticeable especially in the Experience arm where distributions are not known, thus reducing ambiguity attitudes. However, in a low or no information situation, individuals overreact to signals (Ba et al., 2024) changing their beliefs when they should not and their attitudes are more pronounced. Disentangling the attitude and belief effects allows us to properly estimate them.

3.1 Ambiguity attitudes

Individuals in both the Experience and Description conditions face three urns $u \in \{N, W, S\}$, where N stands for the no information urn, W for the urn with the weak signal and S for the urn with the strong signal. For each urn, a ball is drawn, and then the letter on the ball σ is announced, $\sigma \in \{A, B\}$. The event of the ball being of a particular color is represented by $Col = k$ for simple events and $Col = k \cup l$ for composite events, where $k \neq l$ and $k, l \in \{Red, Yellow, Blue\}$. For simplicity, I will write $m_{u,k}$ instead of $m_u(k|\sigma_u)$, as all matching probabilities elicited are conditioned on the same letter for a given urn.

If individuals were Subjective Expected Utility Maximizers, then the matching probability would just give us their belief about the event $m_u(Col) = P(Col|\sigma_u)$. Such beliefs would be congruent with probabilities, so $m_u(Col) + m_u(Col^c) = 1$. Typically, individuals do not comply with SEU meaning that their elicited beliefs are contaminated by their attitudes. For example, an averse individual would reveal $m_u(Col) + m_u(Col^c) < 1$. It was the case that in order to affirm that the subjects' beliefs were contaminated by attitudes, it was necessary to know the uncontaminated beliefs beforehand; this is not necessary now when using Baillon et al.'s method.

Let $\overline{m_u}_s = (m_{u,Blue} + m_{u,Red} + m_{u,Yellow})/3$ be the average matching probability for all single events and let all composite event matching probabilities average be $\overline{m_u}_c = (m_{u,Blue\cup Yellow} + m_{u,Yellow\cup Red} + m_{u,Blue\cup Red})/3$. The proposed index of ambiguity aversion is as follows

$$b_u = 1 - (\overline{m_u}_s + \overline{m_u}_c)$$

For Subjective Expected Utility maximizers we have $\overline{m_u}_s = 1/3$ and $\overline{m_u}_c = 2/3$ and thus $b = 0$ the neutrality point. Typically, the index is in $[-1, 1]$. It is an anti-index that captures the average elevation of the matching probabilities. Low elevations mean that individuals even prefer lotteries with low probabilities to the prospect with the event, i.e., they dislike the events and are ambiguity averse. The opposite means a systematic preference for the events. In the DfD, low elevation means avoidance of the complex "urn" risk in favor of simpler risks, whereas in the DfE arm its avoidance of uncertainty in favor of known risk.

The aversion index has to do with a motivational aspect, that is, pursuance or avoidance of uncertain events. The next index captures the cognitive side of uncertainty attitudes. In particular, it captures to what extent ambiguity confuses individuals, ultimately making them have an all-or-nothing attitude. It has been called the a-insensitivity[1] index (Dimmock et al., 2016).

$$a_u = 3 \times \left(\frac{1}{3} - (\overline{m_u}_s - \overline{m_u}_c)\right)$$

Expected utility maximizers have $a_u = 0$. Most of the indices will lie in $[0, 1]$.

As this work uses measurements of attitudes that the Description-Experience gap has not used before, it is worth relating it to the parametric forms used in this literature. The most popular was proposed by Goldstein and Einhorn (1987). In this form two there are two parameters, an elevation parameter and a curvature parameter. The aversion index used in this work relates to the elevation parameter in probability weighting functions, the a-insensitivity index is closely related to the curvature parameter.

Following the literature from the Description-Experience gap that has found that individual decisions differ under both conditions (Wulff et al., 2018; Hertwig, 2015), we perform a comparison between arms. Thus, if an attitudes gap is expected, the general averages of both indexes should differ in such a between-arms comparison.

$$\overline{a}^{DfD} \neq \overline{a}^{DfE}$$

[1] a-insensitivity stands for Ambiguity generated likelihood insensitivity

$$\overline{b}^{DfD} \neq \overline{b}^{DfE}$$

In addition, we compare each urn with its counterpart in the other arm. We expect the attitudes gap to be also present in this comparison within the urns.

$$\overline{a_u}^{DfD} \neq \overline{a_u}^{DfE}$$

$$\overline{b_u}^{DfD} \neq \overline{b_u}^{DfE}$$

Finally, we perform an within-arm analysis, in which we expect that as the signal informativeness varies between urns, attitudes within the same arm must differ between urns.

$$\overline{b}_N^j \neq \overline{b}_W^j \neq \overline{b}_S^j$$

$$\overline{a}_N^j \neq \overline{a}_W^j \neq \overline{a}_S^j$$

with $j \in \{DfD, DfE\}$.

We expect that the indexes decrease as the signal becomes less noisy because better information is more liked and reduces insensitivity and aversion.

3.2 Beliefs

The method used in this work allows for measurement of the a-neutral probabilities as well. They are additive probabilities that capture beliefs once corrected for ambiguity attitudes[2] (Li et al., 2019). The a-neutral probabilities conditional on the signal are:

$$\pi(Col = i|\sigma_u) = \frac{m_{u,i\cup l} + m_{u,i\cup k} - m_{u,k} - m_{u,l}}{2(1-a)} \tag{1}$$

where $i, k, l \in \{Blue, Yellow, red\}$.

The analysis of the updating of a-neutral probabilities follows methods in the literature that estimate under- or overweighting of probabilities.

[2]Li et al. (2019) characterize these probabilities as the belief that the ambiguity neutral twin of our subjects will have and show the assumptions necessary to obtain them

3.2.1 Updating Analysis

Given the additive beliefs, it is now possible to analyze how they react to information in both DfD and DfE. In order to estimate the parameter quantifying the weight individuals give to information, we assume the widely applayed quasi-Bayesian model (Grether, 1992; Holt and Smith, 2009; Ambuehl and Li, 2018; Coutts, 2019; Möbius et al., 2022; Augenblick et al., 2024) and we use a log-odds specification to derive it. Such specifications are known as the Grether regressions, but in most studies using them the state space is binary, and log-odds completely capture it. Instead, here we calculate the Grether regressions for the log-odds of each color and its complement[3], as seen in Equation 2, where ρ captures the weight that individuals give to new information. When $\rho = 1$ individuals update as a Bayesian would. when $\rho < 1$, individual underreacts to information and when $\rho > 1$ they overreact to it.

$$ln\left(\frac{\pi(Col|\sigma_u)}{1-\pi(Col|\sigma_u)}\right) = ln\left(\frac{P(Col)}{1-P(Col)}\right) + \rho ln\left(\frac{P(\sigma_u|Col)}{P(\sigma_u|Col^c)}\right) \tag{2}$$

Given that the unconditional probabilities of the colors are always 1/3, then the log odds of the priors is always $ln(1/2)$. Thus, the constant in this specification will just reflect the log-odds of the priors as seen in Eq. 3. Hence, it is important that individuals understand the composition of the urn. Because of this, a comprehension question[4] was added about the contents of the urn and subjects who did not correctly answer it were removed from the analysis.

$$ln\left(\frac{\pi(Col|\sigma_u)}{1-\pi(Col|\sigma_u)}\right) = \rho_0 + \rho ln\left(\frac{P(\sigma_u|Col)}{P(\sigma_u|Col^c)}\right) + \epsilon \tag{3}$$

This equation will be estimated for both experience and description independently,in order to estimate if there exists conservatism in the conditions. To explore any possible differences between arms, that is a learning (from signals) gap, the two arms are pooled together and the following regression is estimated

$$ln\left(\frac{\pi(Col|\sigma_u)}{1-\pi(Col|\sigma_u)}\right) = \rho_0 + \rho_1 DfE + \rho_2 ln\left(\frac{P(\sigma_u|Col)}{P(\sigma_u|Col^c)}\right) + \rho_3 DfE * ln\left(\frac{P(\sigma_u|Col)}{P(\sigma_u|Col^c)}\right) + \epsilon \tag{4}$$

where DfE is a binary variable that indicates if the subject was assigned to the DfE arm. Thus, ρ_3 captures if there is any difference between the log-odds of the prior between the DfD and the DfE arms. On the other hand, ρ_1 captures the average response to information of all subjects, while ρ_2 captures any difference in the reaction to information about being in the Experience arm.

[3]Ba et al. (2024) argue against using the Grether equations for more than three spaces because multiple log-odds ratios are needed to capture all possible combinations of states, thus implicitly assuming that agents distort each pair of prior odds and signal likelihoods in an identical way. We relax this by considering the log-odds of the state and its complement, thereby reducing the number of necessary log-odds combinations to equal the number of states.

[4]The question can be found in appendix B Figure 11

3.2.2 Signal Probabilities

In order to stablish under- or overreaction by means of Eq.3 and Eq. 4 the probabilities that individuals give to the signal given each state space, that is $P(\sigma_u|Col)$,are needed; in the case of the description arm, since individuals know the contents of the urn, it is natural that this probability corresponds to the actual proportion of letter σ_u in color k of the urn. However, in the case of the Experience arm, individuals do not observe this information directly, thus using the actual likelihood of the urn distribution can lead to bias in the estimation of reaction to information.

In order to derive a likelihood given their actual sampling histories under DfE it is important to look at the type of problem individuals are facing. It can be summarized as an inference problem about the joint probabilities while keeping the marginals fixed. As explained in section 2.1, individuals were informed about all marginals. This means that what is unknown is the conjunction, and here is where the inference problem lies.

We modeled this problem as if the subjects were inferring from which of the urns that comply with the description of the contents in Section 2.1 are they sampling. In total, only 91 different urn compositions are possible. By making the urns the object of inference, the subjects keep marginal distribution constant while inferring a different composition of the joint distribution of colors and letters.

Subjects are assumed to have a prior η about the urn U being the actual urn u, $\eta = P(U = u)$, as the joint probabilities of color and letters characterize each urn U, therefore a belief about the urns can be seen as a belief over a set of 6 probabilities. Hence, it is intuitive to assume that the prior η follows a Dirichlet distribution $D(\alpha_m)$, $m = \{1, 2, 3, 4, 5, 6\}$, where α_m are the concentration parameter of the Dirichlet distribution. As there is no reason to think that subjects before sampling might believe that one combination of color and letters is more prevalent than the other, we assume that all parameters α are equal. This assumption guarantees that the prior expectation for each joint probability is $\sum_U \eta_U * P(Col_U \cap \sigma_U) = 1/6$. However, by varying the concentration parameter, we can vary how much the subjects believe that the urn they are sampling looks like the mean urn with 1/6 in all colors. If the concentration parameter is low $\alpha < 1$, then individuals believe that more dissimilar distributions are more plausible a priori, that is, those that have more extreme proportions of urns and letters. Hence, we can say that they expect the sampling to be very informative. On the other hand, if the concentration parameter is higher $\alpha > 1$, then they believe that the urns that have probabilities that resemble more like the mean are more probable, i.e. the sampling is expected to be less informative. When $\alpha = 1$ we have the special case of the uniform distribution, where they think that all possible urns are equally likely.

Another advantage of assuming a Dirichlet distribution is that it can be used as a conjugate prior to the multinomial distribution, which reflects the likelihood of sampling with replacement from an urn. Such conjugacy allows a direct interpretation of the observations, if for example subjects observe more A-Red balls, then its component in the posterior increases, placing more probability mass on those urns with more A-Red balls. The primary focus of the analysis in this study will be on the scenario where $\alpha = 1$, that is, uniform priors, as this assumption on the priors is commonly employed in the literature, for example, in the classic analysis of the Ellsberg paradox. However, we perform an additional analysis for two specific values of $\alpha = 1.5$ and $\alpha = 0.15$ to understand how robust the results are to the prior specification.

By deriving the conditional probability of the signal $P(\sigma_u|Col_i)$, this work is able to address two issues that have been previously raised in the DfD-DfE literature. First, the importance of prior beliefs to estimate the differences between experience and description (Aydogan, 2021). Second, the comparison of individual attitudes and beliefs should only take into account the actual information the subjects faced, not the underlying distribution on the urn as highlighted by Abdellaoui et al. (2011). To estimate $P(\sigma_u|Col)$, we need to assume a prior distribution, and its updating with the sampled information reflects the information seen by the subject, not the underlying distribution of the urn.

4 Results

4.1 Data

We recruited 400 subjects from prolific, comprising 194 males and 197 females[5]. The average bonus won was \$7.8. Individuals were assigned to the Decision from Description or Decision from Experience arms. Table 2 shows the composition of each group by demographic characteristic.

	All	DfD	DfE
Age			
Mean	38.98	38.35	39.61
SD	12.76	13.17	12.32
Education in %			
High school	33.08	32.18	34.17
Higher education (Bachelor)	49.00	48.02	50.25
Post-graduate education	17.66	19.80	15.58
Gender in %			
Female	49.00	48.02	50.25
Male	48.26	50.50	46.23
Non-binary	1.99	0.99	3.02

Table 2: Descriptive statistics for the complete sample

As stated in the pre-registration, the analysis was only performed using those individuals who

[5]The rest of the subjects reported they were non-binary or preferred not to reveal their gender

passed both comprehension checks, both questions can be found in B. The sample is then reduced to 295 subjects of which 142 subjects were assigned to the DfE arm and 153 subjects were assigned to the DfD arm. Table 3 shows the composition of the subjects who passed both comprehension checks and who will be used in the analysis.

	All	DfD	DfE
Age			
Mean	37.83	37.54	38.15
SD	12.36	12.89	11.79
Education in %			
High school	34.58	33.33	35.92
Higher education (Bachelor)	48.47	47.06	50.00
Post-graduate education	16.95	19.61	14.08
Gender in %			
Female	50.17	50.98	49.30
Male	47.12	48.37	45.77
Non-binary	2.37	0.65	4.23

Table 3: Descriptive statistics excluding individuals who did not pass the comprehension check

When probabilities were recovered, there were five individuals whose probabilities did not lie inside the unit interval, all of them in the DfE arm. Attitudes and a-neutral probabilities were recalculated for these individuals, forcing their beliefs to lie in the unit interval. The analyses without these individuals do not differ qualitatively and are found in the Appendix G. The analyses without probabilities being in the unit interval do not differ in magnitude from the analysis presented here.

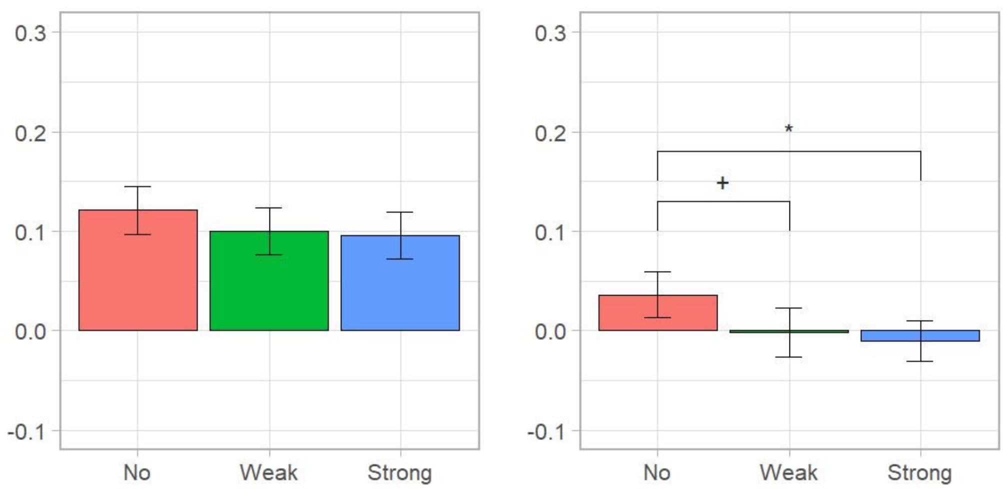
4.2 Attitudes

Table 4 presents the summary statistics of ambiguity attitude indexes per treatment arm. The average of aversion across signal strengths was higher in the DfD arm than in the DfE arm, while the inverse is true for insensitivity. Furthermore, the table shows that comparing the average of each urn between the two arms shows the same pattern of higher aversion in DfD than in DfE, and that the average for the weak and strong signals for DfE is close to neutrality, something that becomes apparent in Figure 5 and Figure 6. Nevertheless, for a-insensitivity it is higher in all urns but for the strong signal, where no difference was found. The comparison in aversion is consistent with Ert and Trautmann (2014), where the subjects preferred the ambiguous urn when they were given the opportunity to sample before. The results in a-insensitivity highlight the importance of the accuracy of the signal. A strong signal allows subjects to better differentiate between the probabilities of events no matter in which condition. This equality in the strong signal hints that a change in information might change attitudes.

	DfE		DfD			
	Mean	SD	Mean	SD	t	p
Aversion						
Average	0.01	0.23	0.11	0.28	-5.14	0.00
No	0.04	0.27	0.12	0.30	-2.54	0.01
Weak	-0.00	0.29	0.10	0.29	-2.96	0.00
Strong	-0.01	0.24	0.10	0.29	-3.45	0.00
Insensitivity						
Average	0.36	0.22	0.28	0.25	3.98	0.00
No	0.40	0.29	0.29	0.31	2.97	0.00
Weak	0.39	0.30	0.30	0.29	2.60	0.01
Strong	0.29	0.25	0.25	0.29	1.26	0.21

Table 4: Comparison of means between urns in the DfE and DfD arms

Within-treatment ANOVA analysis reveals that aversion is not different between signal strengths for the DfD arm ($F(2, 304) = 1.84, p = 0.16$), while in the DfE arm, differences in signal strength impact differences in aversion ($F(2, 282) = 3.11, p = 0.05$). Post-Hoc analysis reveals that the aversion is higher in the non-information signal compared to the strong signal ($M_{S-N} = -0.5, t = -2.40, p_{bonf} = 0.03$)[6] and in turn that the former is higher than the weak signal at the 10% level ($M_{W-N} = -0.4, t = -1.89, p_{bonf} = 0.09$). In summary, the more informative the signals, the lower the aversion in DfE. No differences were found for those in DfD. This finding suggests that the informativeness of the contents of the urn has an effect on how much sampling reduces aversion when the contents of the urn are not known beforehand.



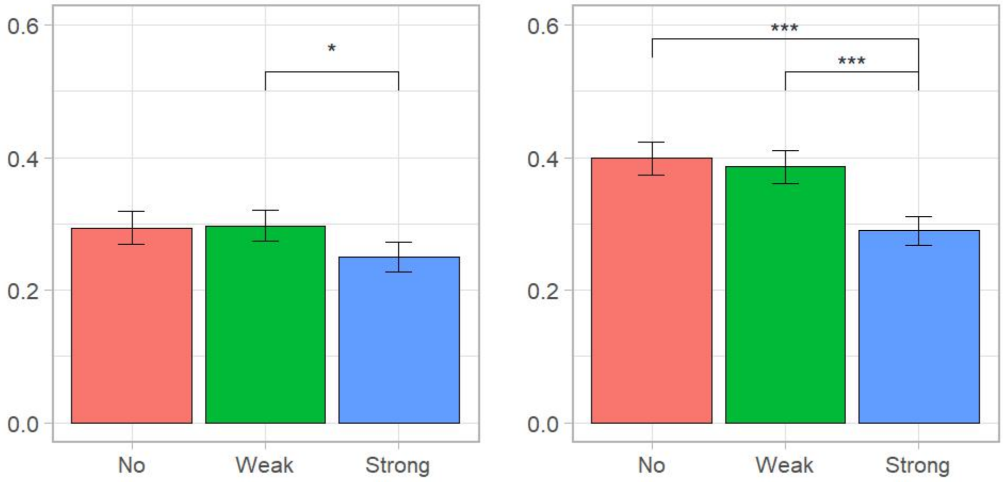
(a) Decision from Description arm

(b) Decision from Experience arm

Figure 5: Mean of the index of aversion by urn by urn, $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$ describe significance level of the difference of means derived from the post-hoc analysis

[6] p_{bonf} refers to the p-values once the Bonferroni correction is applied

The pattern of differences within treatment for insensitivity differs from that found in aversion. The within-treatment ANOVA analysis for the DfD arm reveals that we can only reject the null hypothesis at the 10% level ($F(2, 304) = 2.83, p = 0.06$). Post hoc analysis reveals that the insensitivity in the Strong signal is lower than in the Weak signal ($M_{S-W} = -0.05, t = -2.30, p_{bonf} = 0.03$). On the other hand, for DfE, a difference is found between signals ($F(2, 282) = 10.58, p < 0.01$), and the post hoc pairwise analysis reveals that the no information signal has a higher insensitivity than both the weak signal ($M_{S-N} = -0.11, t = -4.31, p_{bonf} < 0.01$) and the strong signal ($M_{S-W} = -0.10, t = -3.85, p_{bonf} < 0.01$). The complete Pairwise Post-hoc comparisons made are in Table 9 in Appendix C.



(a) Decision from Description arm

(b) Decision from Experience arm

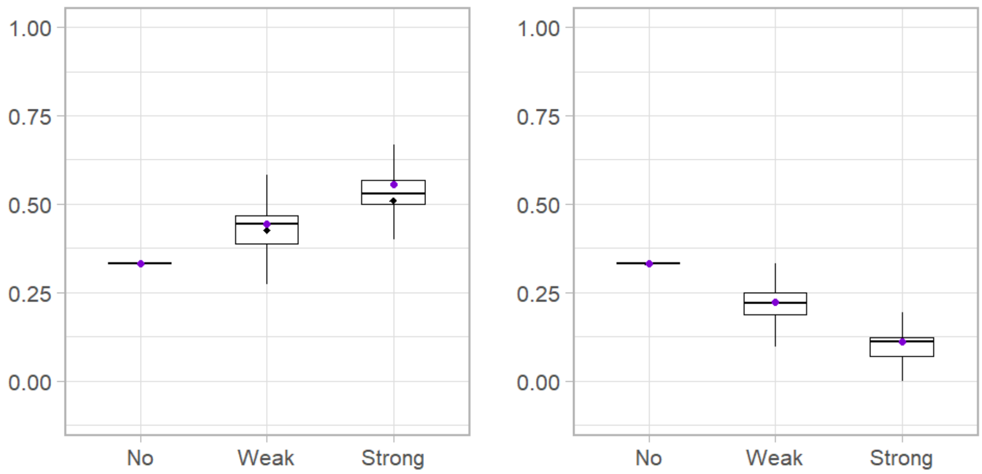
Figure 6: The mean of the index of a-insensitivity by urn, $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$ describe the significance level of the difference of means derived from the post hoc analysis

The effects of varying signal informativeness on aversion and a-insensitivity are interesting. It shows that the difference between experience and description is not only about the magnitude of the indexes, but also in how attitudes are affected by informativeness. In aversion this difference of effect of informativeness on attitudes is starker, for subjects in the DfE arm it is just enough for the signal to be informative to change attitudes, while no change is observed in DfD. For changes in the a-insensitivity it is not enough for a signal to be informative of the state space as it was in aversion; the signal must be strong, as the weak signal is not statistically different in sensitivity to probabilities as the no information signal. In the DfE arm, people exhibited less insensitivity to uncertainties regarding the composition of the strong signal urn than in the weak and no-information urns, respectively. Thus, not only is the magnitude of index affected when assigned to Experience or Description, but also how indices vary as signals become more informative.

4.3 Beliefs

As described in Section 3.2, a-neutral probabilities were estimated for the individuals in both arms. As seen in Section 2 Table 1 the values of the conditional probabilities of $P(A|Red) - P(B|Blue)$ and $P(A|Blue) - P(B|Red)$ follow the same pattern, which means that the Bayesian posterior is the same for both of these probabilities. As the pattern increases for $P(A|Red) - P(B|Blue)$ the signal is more informative we will call this positive relation, as the pattern decreases for $P(A|Blue) - P(B|Red)$ we call it a negative relation. This notation is handy for presenting descriptive statistics in Figure 7 for DfD and Figure 8 for DfE.

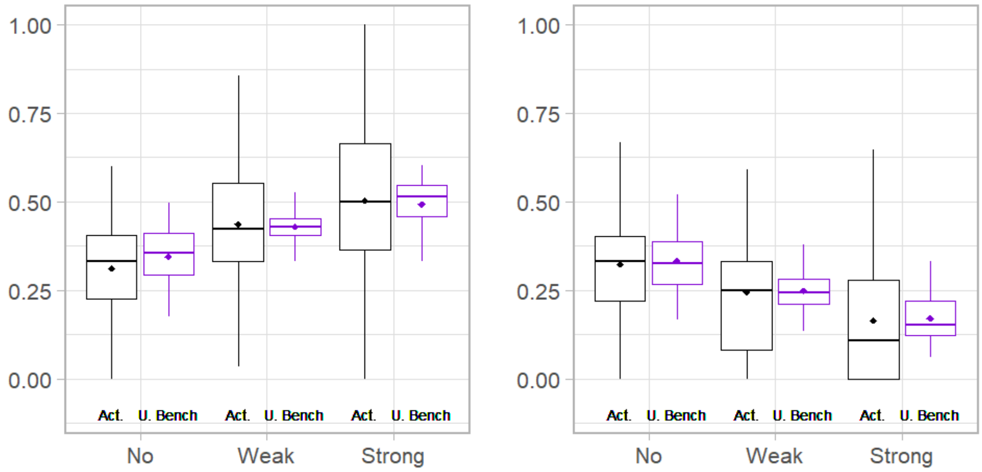
For individuals in the DfD arm, the main distribution statistics are summarized in Figure 7. The black points in this figure are the measure sample posteriors and the purple points represent the Bayesian benchmark. As an example, for the no information sample, the Bayesian posterior is 1/3 for both Positive and Negative relation signals. For the non-informative signal, both the Bayesian posterior and the measured posteriors juxtapose, meaning that both are very close. For weak and strong signals, the main pattern seems to be that the sample mean posterior (in black) are lower than the benchmark (in purple). In fact, for the strong urn more than half of individuals underestimate the probability with respect to the benchmark.



(a) Positive relation signals (b) Negative relation signals

Figure 7: Comparison actual posterior (Black) vs Bayesian benchmark (Purple) in DfD arm

In the context of DfE the comparison with Bayesian benchmarks is not as straightforward. As subjects sample as much as they wish and see different sequences of information, the benchmark is not a single point. Furthermore, the benchmark posterior distribution is contingent on the parameter assumed for the priors. Figure 8 shows a comparison between the distributions of the actual posterior,denoted *Act.*, and the distribution of the Bayesian posterior derived assuming uniform priors ($\alpha = 1$), denoted *U. bench*. It shows that the measured posteriors have a more variable distribution than the uniform benchmark. Figure 8 shows comparable means in both positive and negative relation signals between benchmark and the actual posteriors. However, in the negative relation signals the median of the strong signal is higher, indicating more overreaction, whereas in the positive relation overreaction is less evident. This underlines the importance of the analysis of the under- and overreaction using Eq. 3 to derive inferences about the subjects' behavior. Figure 12 for the comparison of the benchmarks with the values of $\alpha = 1.5$ and $\alpha = 0.15$ is available in Appendix D.



(a) Comparison for positive relation signals (b) Comparison for negative relation signals

Figure 8: Comparison actual posterior (Black) vs Bayesian benchmark for $\alpha = 1$ (Purple) in DfE arm

4.3.1 Uniform Prior Analysis

Table 5 shows the parameter that captures the weight individuals give to the information from the signal (weight of information) for both arms and for the pooled models. The hypothesis that $\rho = 1$ for Model 1 is rejected ($t = -8.32, p < 0.01$), but not for Model 2 ($t = -1.58, p = 0.11$). Thus, subjects in DfD assign less weight to information than their theoretical Bayesian counterparts would. Interestingly, the estimation of the mean in the DfD arm is lower than for Ambuehl and Li (2018), but similar to what they find in the case of DfE. Such differences might be explained as we do not assume that attitudes toward inferential risk are the same as those of simple risk, contrary to what Ambuehl and Li do[7].

Interestingly, in both Model 1 and Model 2 the estimation of the constant is not remote from the value of the logodds of the prior $ln(\frac{1/3}{2/3}) \approx -0.69$. In fact, in the DfD arm we have $t =$

[7]See page 32 paragraph 1 of Ambuehl and Li (2018) for the complete assumption of the authors

	DfD	DfE	Pool	
	(1)	(2)	(3)	(4)
Weight of Info. (ρ)	0.81^{***}	0.92^{***}	0.81^{***}	0.81^{***}
	(0.02)	(0.05)	(0.03)	(0.03)
W. Info x DfE			0.11^{*}	0.12^{*}
			(0.05)	(0.05)
DfE			0.02	0.02
			(0.03)	(0.03)
Male				0.01
				(0.03)
Bachelor				−0.00
				(0.03)
Post-Graduate				0.01
				(0.04)
Constant	-0.70^{***}	-0.68^{***}	-0.70^{***}	-0.70^{***}
	(0.01)	(0.03)	(0.02)	(0.03)
s_idios	0.52	0.94	0.74	0.74
s_id	0.00	0.00	0.00	0.00
R^2	0.48	0.22	0.31	0.31
Adj. R^2	0.48	0.22	0.31	0.31
Num. obs.	1324	1146	2470	2461

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$. Random Effects estimation

Table 5: Reactiveness to information for the DfD and DfE arms and pooling both conditions

$-0.47, p = 0.63$, and in the DfE arm $t = 0.68, p = 0.49$. Note that this experiment was not designed to capture base rate neglect, as prior probabilities where symmetrical and their log-odds did not change through the experiment. However, it gives some evidence of the adequacy of the estimated model.

In order to infer whether there were differences between the estimated response to information, for this, the data for both arms were pooled and the linear model in Equation 4 shown in the analysis section was estimated. From Model 3 it can be concluded that individuals in the Description arm underreact to information and that the weight of information of individuals in the Experience condition is different and higher. This means that the updating of subjects in the DfE arm is closer to the Bayesian benchmark than those in the DfD arm. Thus, if we assume uniform priors, we find that there is a learning gap. An additional fact found in Table 5 is that the coefficient of DfE in Model 3 and Model 4 is not significant. Thus, the only effect of experience was via the different information contained in the signals.

4.3.2 Comparison with different priors

As discussed in Section 3.2.2, the value of the initial parameter α for the prior could matter for our analysis. It is intuitive to think that how much subjects weight information depends on how informative they thought the urns were from the beginning. If a subject had a prior belief that extreme distributions were more plausible, $\alpha \leq 0.16$ for example, and got to sample from the strong urn, then their initial beliefs would likely not move much, offering an alternative explanation for the result in the previous section. Although there is no reason to believe that individuals hold such type of initial beliefs, exploring how this assumed initial parameter affects the estimation of the weight given to information is important to assess the robustness of the conclusions.

	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 0.15$
Weight of Info. (ρ)	0.92^{***}	1.02^{***}	0.76^{***}
	(0.05)	(0.06)	(0.04)
Constant	-0.68^{***}	-0.68^{***}	-0.67^{***}
	(0.03)	(0.03)	(0.03)
s_idios	0.94	0.93	0.95
s_id	0.00	0.00	0.00
R^2	0.22	0.22	0.20
Adj. R^2	0.22	0.22	0.20
Num. obs.	1146	1146	1146

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$. Random Effects estimation

Table 6: Reactiveness to information for different initial parameters α for the Experience arm

In this section, we will refer to three cases and compare them, our initial case $\alpha = 1$, with a uniform distribution across all possible urns, $\alpha = 0.15$ where urns with more extreme distributions of colors and letters are more likely, and $\alpha = 1.5$ where the urn with a uniform proportion of letters in each color is given the most probability. Table 6 shows the estimate of the average weight of information for each assumed concentration parameter α. As expected, if individuals expect ex ante that the extreme urns are very probable then they expect the signals to be very informative and, thus, their ratio of likelihood of signals would be higher. With $\alpha = 0.15$ the parameter $\rho < 1$ ($t = -5.33, p < 0.01$) implies that individuals would have conservatism as in the case of Description. Thus, in order to have results closer to those in the literature (Grether, 1992; Ambuehl and Li, 2018), a starting belief of highly informative urns must be assumed for the experience arm. In contrast, a parameter of $\alpha = 1.5$ produces a parameter close to 1, thus closer to the Bayesian benchmark.

Table 7 again shows the estimated differences between experience and description. No differential impact of experience is observed when assuming a parameter of $\alpha = 0.15$. In contrast, the coefficient is significant and higher in the model employing $\alpha = 1.5$. Again, as in Table 5, the only effect of experience comes from its effect on information weighting. Consequently, the effect of experience is not invariant to the type of initial beliefs assumed. The parameters $\alpha \geq 1$ produce a differential effect of experience in updating. The assumption that individuals anticipate signals to be highly informative from the beginning does not produce a differential effect in updating from

	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 0.15$
Weight of Info. (ρ)	0.81^{***}	0.81^{***}	0.81^{***}
	(0.03)	(0.03)	(0.03)
W. Info x DfE	0.11^{*}	0.21^{***}	-0.05
	(0.05)	(0.06)	(0.05)
DfE	0.02	0.02	0.03
	(0.03)	(0.03)	(0.03)
Constant	-0.70^{***}	-0.70^{***}	-0.70^{***}
	(0.02)	(0.02)	(0.02)
s_idios	0.74	0.74	0.75
s_id	0.00	0.00	0.00
R^2	0.31	0.32	0.30
Adj. R^2	0.31	0.31	0.30
Num. obs.	2470	2470	2470

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$. Random Effects estimation

Table 7: Pooled models of reactiveness to information for different parameters α for the Experience arm

experience.

4.4 Summary of results

		Between arms			Within arms	
		DfD		DfE	DfD	DfE
Ambiguity Attitudes	Aversion	No	$\underset{*}{>}$	No	No $\underset{n.s.}{=}$ Weak $\underset{n.s.}{=}$ Strong	No $\underset{*}{>}$ Strong
		Weak	$\underset{**}{>}$	Weak		No $\underset{+}{>}$ Weak
		Strong	$\underset{**}{>}$	Strong		Weak $\underset{n.s.}{=}$ Strong
	Insensitivity	No	$\underset{**}{>}$	No	No $\underset{n.s.}{=}$ Strong	No $\underset{***}{>}$ Strong
		Weak	$\underset{**}{>}$	Weak	No $\underset{n.s.}{=}$ Weak	No $\underset{n.s.}{=}$ Weak
		Strong	$\underset{n.s.}{=}$	Strong	Weak $\underset{*}{>}$ Strong	Weak $\underset{***}{>}$ Strong
		DFD		DFE		
Weight of information	Uniform prior		$\underset{*}{<}$			
	$\alpha = 1.5$		$\underset{***}{<}$			
	$\alpha = 0.15$		$\underset{n.s.}{=}$			

Table 8: Summary of results of between arm and within arms analyses.

Table 8 shows a summary of the results. In summary we have found that for attitudes the gap consists of two phenomena, first between arms aversion is worse for all signal strengths in DfD than in DfE. As for insensitivitivy, it is worse in DfE than in DfD for the non-informative and weak signal. Second, within-arms the DfE the non-informative signal has higher aversion than the weak and strong signal, while no difference is found for DfD. The aforementioned difference suggests that the informativeness of the signal only affects aversion in DfE. In both DfD and DfE, in contrast to aversion, only the strongest signals decrease insensitivity, suggesting that it is not enough to be informative but that it needs a certain threshold of informativeness to increase how individuals discriminate between probabilities.

In our main analysis, we found that for beliefs, the individuals in the DfE arm give more weight to information than in the DfD arm, resulting in a closer level to the Bayesian benchmark. However, these results do depend on the priors on assuming uniform priors, or priors that do not expect urns with dissimilar proportions of letter-color pairs. We consider these differences to be suggestive evidence for a learning gap when learning from signals, although further exploration is needed of the role of priors.

5 Discussion

Returning to the example of the doctor, our results tell us that there is a difference when she is relying on her experience or the well-established diagnostic guidelines. In her diagnosis, she gives a higher weighting to the evidence when she relies on her experience. Second, the results of aversion tell us that when she relies on her experience, she does not hesitate that much on her diagnosis, especially if the symptons are strongly associated with certain disease. But her diagnosis might be less precise, leading to less decisive action, unless her association with certain diagnostic paths is strong.

The results of this study highlight the potential effects of uncertainty on previous findings. In contrast with previous DfD-DfE literature, we use a new measure in which the aversion and insensitivity measures are orthogonal to each other. An advantage of the method is that aversion and independence are totally orthogonal (Baillon et al. (2021) Theorem 12). This is not the case for the curvature parameters for the probability weighting functions of the Goldstein-Einhorn family. This is crucial because the comparison of aversion between description has a different sign than the comparison of a-insensitivity, as aversion is higher in Description and a-insensitivity is higher in Experience on average.

This work is able to identify that the strength of the association between the PR-event and the information, characterized here by the noise of the underlying signal, has a more pronounced effect on subjects in the experience condition. The subjects in the Experience condition never know the actual distribution of joint colors and letters, so as it becomes clearer that the letters better predict the color, their aversion to betting on the colors decreases. In contrast, subjects in the description condition have all the information available to them from the beginning, so the degree of association between the letters and colors does not have the same effect on their aversion.

For insensitivity, the pattern is not the same. Insensitivity is only lower for the strong signal urn. A similar but less evident pattern occurs in the description condition, contrasting with the lower aversion observed in both the weak and strong signal urns. Therefore, unless the contrast between colors in the signal is very evident, subjects may struggle to discriminate between probabilities. This suggests that while subjects might be better at updating their beliefs in an experience-based context, they may still face challenges in distinguishing between different probabilities. If we accept the conventional wisdom that Decision from Experience is easier for subjects to process, then this may indicate how uncertainty influences the integration of sampled information.

For the risk case, our findings show that risk attitudes are not uniform between simple risk and more complex risk, in agreement with (Armantier and Treich, 2016; Halevy, 2007). Moreover, the comparison of attitudes between the Experience arm and the Description arm supports previous results that there is a difference between attitudes to compound risks, of which conditional probabilities can be seen as a special case, and ambiguity (Abdellaoui et al., 2015). Therefore, our results for attitudes underscore the need to account for this difference in risk attitudes, which is a departure from previous studies in the literature of pitfalls in belief updating (Ambuehl and Li, 2018).

It is also important to specify the type of uncertainty that the subjects face in this study. Before making their choices, subjects know which letter is on the sampled ball, which means some uncertainty has already been resolved. This is a key difference from the work of Epstein and Halevy (2024) where no signal has been drawn and the elicitation is done for both signals. Therefore, in this study, we refer to the signals as uncertain signals, but we do not address the attitudes of subjects toward signal uncertainty directly, as their definition of attitudes towards signal ambiguity requires the violation of the martingale property of Bayesian updating, and thus the elicitation of conditional probabilities of both signals. In this experiment, since there is no uncertainty about the color composition of the urn, what introduces uncertainty is the addition of the information of the letter, as their relationship is unknown, a setting with parallels to those explored by Kops and Pasichnichenko (2023). Thus, what I capture is the posterior ambiguity attitude generated by knowing the realization of an uncertain signal whose conditional distribution on color is unknown.

The analysis of beliefs reveals that for the description subjects in the Decision from Description arm, the conclusions are similar to those in the literature with only a binary state space. Subjects exhibit underreaction in their update of the probabilities of PR-events (Holt and Smith, 2009; Ambuehl and Li, 2018; Coutts, 2019). This finding contradicts the findings of Ba et al. (2024), as they find overreaction when there are more than two state spaces. There could be multiple sources of that difference, for once they compare over and underreaction to a expected value, while our states are not collapsible to a single number. We assume a log-odds functional form, which Ba et al. avoids. We do not use the initial beliefs of individuals as they do; instead, we correct for their ambiguity attitudes.

In the Decision from Experience arm, subjects exhibit a heightened response to information only when they do not anticipate the signal to be highly informative prior to sampling. These priors may be considered the most plausible according to the literature; as evidenced by (Aydogan et al., 2024), who estimated the prior values of a beta distribution in a model with two states, finding that the priors are approximately uniform and that individuals assign greater weight to central values than to extreme values, analogously to our $\alpha \geq 1$. Our results are therefore similar to those reported by (Moreno and Rosokha, 2016), where it is observed that new information is given greater weight under conditions of ambiguity, although in their study, the unknown element was the distribution of the states rather than the relation between signal and PR-events. However, in contrast with these two studies, in this study the state space is more complex as it contains more than two possible states, thus completely discarding extreme distributions is not possible. Hence, this study follows Aydogan (2021) in showing the importance of priors in showing gaps in the DfD-DfE literature; however, since we are using the belief hedge method, in this study they are not needed to measure attitudes, but for determining if there are learning (from signal) gaps.

A possible solution to this is to measure prior beliefs before sampling and estimating parameters for initial priors. However, such a measure should have been done before sampling, hence it would be necessary to elicit conditional behavior on both signals. Thus we would be comparing contingent elicitation's of the prior with contingents of the posterior, this comparison might affect the results as some literature argues contingent thinking adds bias (Aina et al., 2024).

Both in the Description and Experience some of the retrieved a-neutral probabilities are equal to zero. This phenomenon occurs more in the Experience arm (as seen in Figure 18 of Appendix E) and for the Strong signal. Apparently, once the contrast is so high, individuals act as if the event with negative correlation with the signal had a very low probability, even thought subjects have been told at least one of the balls in each color was marked A or B. Due to the log-odds form assumed for the reaction to information, any data point with zero probability cannot be used. This might have affected downwards the estimation of reaction to information.

6 Conclusion

This work adapted the DfD-DfE frameworks to study learning from noisy signals. Until now, most studies on learning and belief updating were unable to study ambiguity attitudes simultaneously, as they will confound beliefs. Using a recently introduced method, we obtained the required separation and were able to study unconfounded beliefs and ambiguity attitudes. We found the following DfD-DfE gaps. First, subjects in Description have more aversion and are more sensitive to probabilities, both attitudes decrease with informativeness for subjects in DfE, but only insensitivity does in DfD. The other gap is a learning gap, we find that, assuming that subjects do not hold extreme priors, in Experience subjects react more to information than those in Description. These findings are significant because they show a reversion of the widespread pattern in the literature of underreaction to noisy signals.

References

Abdellaoui, M., Klibanoff, P., and Placido, L. (2015). Experiments on Compound Risk in Relation to Simple Risk and to Ambiguity. *Management Science*, 61(6):1306–1322. Publisher: INFORMS.

Abdellaoui, M., L'Haridon, O., and Paraschiv, C. (2011). Experienced vs. Described Uncertainty: Do We Need Two Prospect Theory Specifications? *Management Science*, 57(10):1879–1895.

Aina, C., Amelio, A., and Brutt, K. (2024). Contingent Belief Updating.

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

Araujo, F., Chabris, C., and Meyer, M. (2024). The Effect of On-the-Job Experience on Base-Rate Neglect: Evidence from Medical Professionals. *Google Docs*.

Armantier, O. and Treich, N. (2016). The Rich Domain of Risk. *Management Science*, 62(7):1954–1969. Publisher: INFORMS.

Augenblick, N., Lazarus, E., and Thaler, M. (2024). Overinference from Weak Signals and Underinference from Strong Signals. arXiv:2109.09871 [econ, q-fin].

Aydogan, I. (2021). Prior Beliefs and Ambiguity Attitudes in Decision from Experience. *Management Science*, page mnsc.2020.3841.

Aydogan, I., Baillon, A., Kemel, E., and Li, C. (2024). How much do we learn? Measuring symmetric and asymmetric deviations from Bayesian updating through choices. *Working Paper*.

Ba, C., Bohren, J. A., and Imas, A. (2024). Over- and Underreaction to Information: the Role of Complexity in Belief-Updating. *SSRN Electronic Journal*.

Baillon, A., Bleichrodt, H., Li, C., and Wakker, P. P. (2021). Belief hedges: Measuring ambiguity for all events and all models. *Journal of Economic Theory*, 198:105353.

Baillon, A., Huang, Z., Selim, A., and Wakker, P. P. (2018). Measuring Ambiguity Attitudes for All (Natural) Events. *Econometrica*, 86(5):1839–1858.

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

Bohren, J. A., Hascher, J., Imas, A., Ungeheuer, M., and Weber, M. (2024). A Cognitive Foundation for Perceiving Uncertainty.

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

Crawford, V. P. and Sobel, J. (1982). Strategic Information Transmission. *Econometrica*, 50(6):1431–1451. Publisher: [Wiley, Econometric Society].

Cubitt, R., Kopsacheilis, O., and Starmer, C. (2022). An inquiry into the nature and causes of the Description - Experience gap. *Journal of Risk and Uncertainty*, 65(2):105–137.

Dimmock, S. G., Kouwenberg, R., and Wakker, P. P. (2016). Ambiguity attitudes in a large representative sample. *Management Science*, 62(5):1363–1380. Publisher: INFORMS.

Eil, D. and Rao, J. M. (2011). The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself. *American Economic Journal: Microeconomics*, 3(2):114–138.

Epstein, L. G. and Halevy, Y. (2024). Hard-to-Interpret Signals. *Journal of the European Economic Association*, 22(1):393–427.

Ert, E. and Trautmann, S. T. (2014). Sampling experience reverses preferences for ambiguity. *Journal of Risk and Uncertainty*, 49(1):31–42.

Esponda, I., Vespa, E., and Yuksel, S. (2024). Mental Models and Learning: The Case of Base-Rate Neglect. *American Economic Review*, 114(3):752–782.

Gigerenzer, G. (1996). The psychology of good judgment: frequency formats and simple algorithms. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, 16(3):273–280.

Goldstein, W. M. and Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review*, 94(2):236–254. Place: US Publisher: American Psychological Association.

Grether, D. M. (1992). Testing bayes rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior & Organization*, 17(1):31–57.

Halevy, Y. (2007). Ellsberg Revisited: An Experimental Study. *Econometrica*, 75(2):503–536.

Hertwig, R. (2015). Decisions from Experience. In *The Wiley Blackwell Handbook of Judgment and Decision Making*, pages 239–267. John Wiley & Sons, Ltd.

Hertwig, R. and Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences*, 13(12):517–523.

Holt, C. A. and Smith, A. M. (2009). An update on Bayesian updating. *Journal of Economic Behavior & Organization*, 69(2):125–134.

Kops, C. and Pasichnichenko, I. (2023). Testing negative value of information and ambiguity aversion. *Journal of Economic Theory*, 213:105730.

Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N., and Wills, A. J. (2016). Attention and associative learning in humans: An integrative review. *Psychological Bulletin*, 142(10):1111–1140. Place: US Publisher: American Psychological Association.

Lejarraga, T. and Hertwig, R. (2021). How experimental methods shaped views on human competence and rationality. *Psychological Bulletin*, 147(6):535–564.

Li, C., Turmunkh, U., and Wakker, P. P. (2019). Trust as a decision under ambiguity. *Experimental Economics*, 22(1):51–75.

Liang, Y. (2019). Learning from Unknown Information Sources. *SSRN Electronic Journal*.

Milgrom, P. R. (1981). Good News and Bad News: Representation Theorems and Applications. *The Bell Journal of Economics*, 12(2):380–391. Publisher: [RAND Corporation, Wiley].

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

Möbius, M. M., Niederle, M., Niehaus, P., and Rosenblat, T. S. (2022). Managing Self-Confidence: Theory and Experimental Evidence. *Management Science*, 68(11):7793–7817. Publisher: INFORMS.

Oprea, R. and Vieider, F. M. (2024). Minding the gap: On the origins of probability weighting and the description-experience gap. *Working Paper*.

Wakker, P. P. (2010). *Prospect Theory: For Risk and Ambiguity*. Cambridge University Press.

Wulff, D. U., Mergenthaler-Canseco, M., and Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, 144(2):140–176.

A Bisection Method

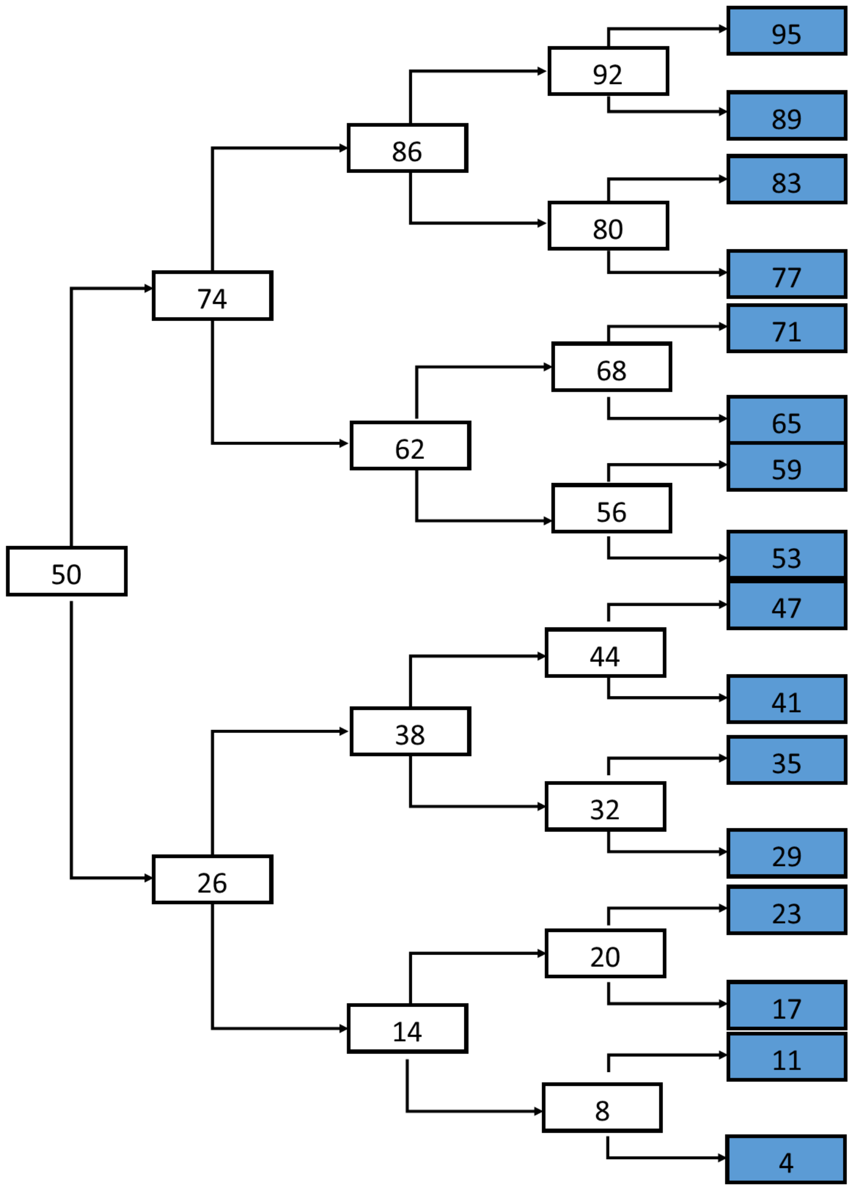


Figure 9: Sequence of lottery choices in the bisection method.

In this experiment, we implement the bisection method in the following way. Individuals first faced a choice, as seen in Figure 2. The choice was between betting on the color or betting on the lottery. Everyone started with a choice of 50% for single-color choices, and how the chances of winning changed in the lottery choice is illustrated in Figure 9. The mechanic was as follows if subjects chose to bet on the color, then in the sequence of they would move up, if they chose the lottery, then they would move down. This sequence would be repeated for 4 choices, the resulting matching probabilities are shown in blue.

B Comprehension Questions

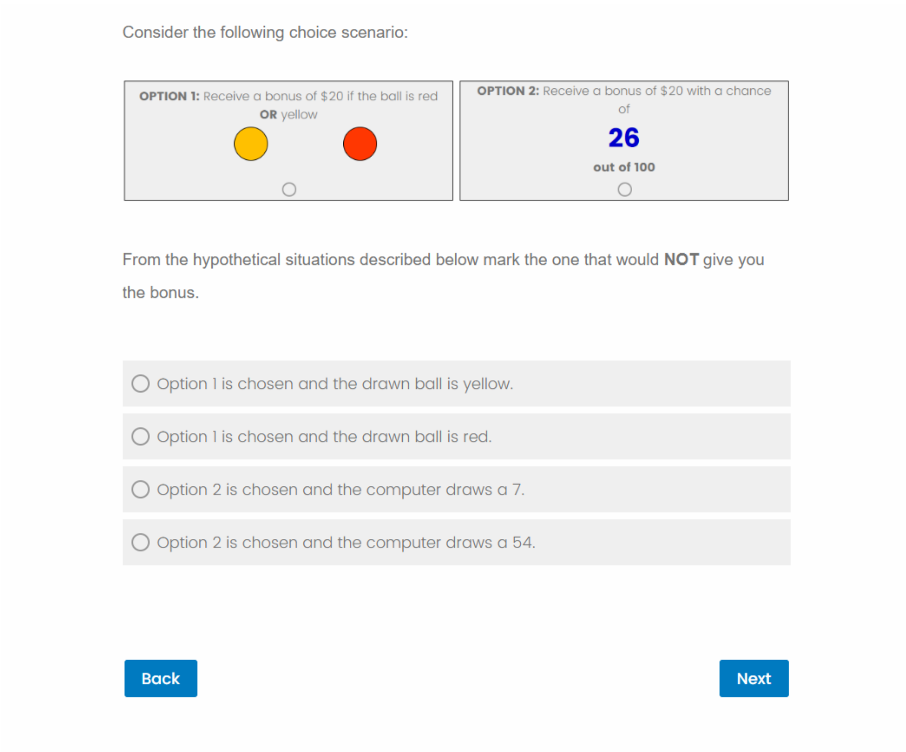


Figure 10: Comprehension question about payment mechanism

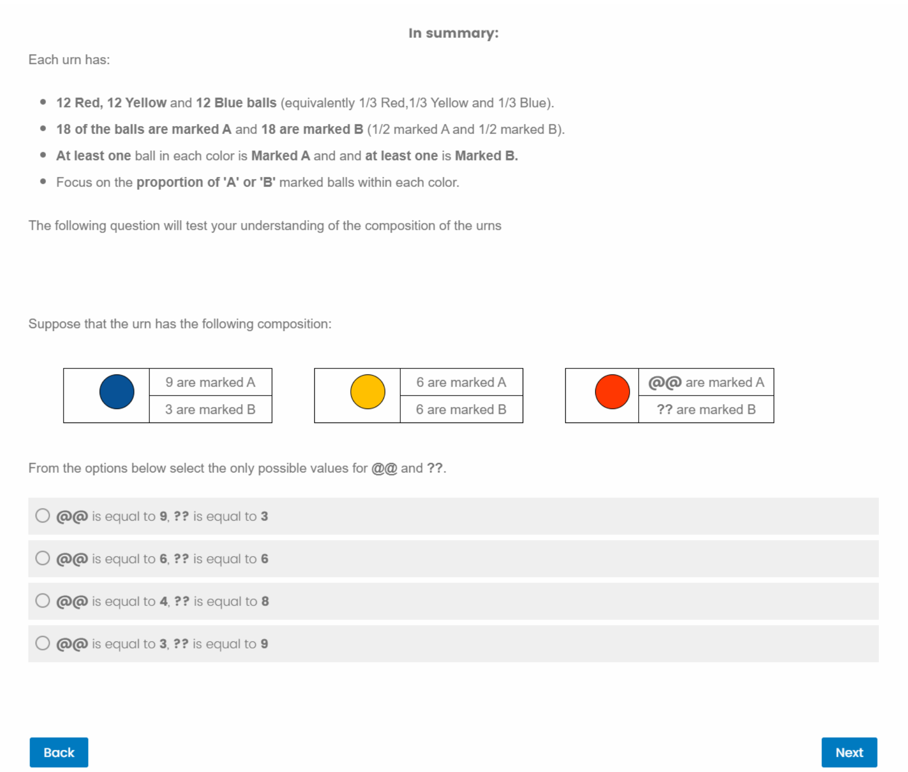


Figure 11: Comprehension question about content of the urn

C Post-Hoc analysis

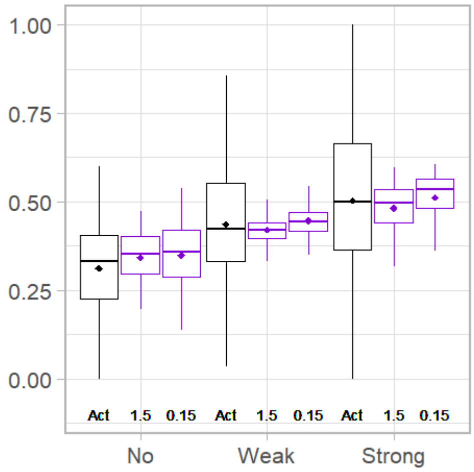
	Difference	t	p	p-corrected
Weak-N.inf	-0.04	-1.89	0.03	0.09
Strong-N.inf	-0.05	-2.40	0.01	0.03
Strong-Weak	-0.01	-0.45	0.33	0.98

Table 9: Mean comparison by signal accuracy in aversion for the experience condition

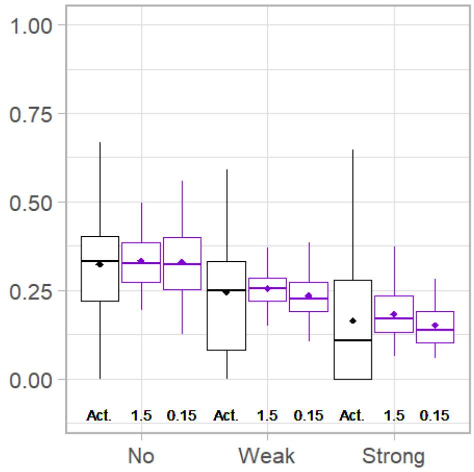
	Experience				**Description**			
	Difference	t	p	p-corrected	Difference	t	p	p-corrected
Weak-N.inf	-0.01	-0.46	0.32	0.97	0.00	0.15	0.56	1.00
Strong-N.inf	-0.11	-4.31	0.00	0.00	-0.04	-1.80	0.04	0.11
Strong-Weak	-0.10	-3.85	0.00	0.00	-0.05	-2.30	0.01	0.03

Table 10: Mean comparison by signal accuracy

D Beliefs

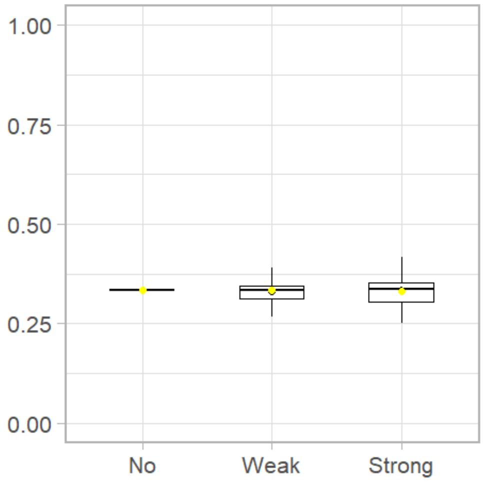


(a) Positive relation signals

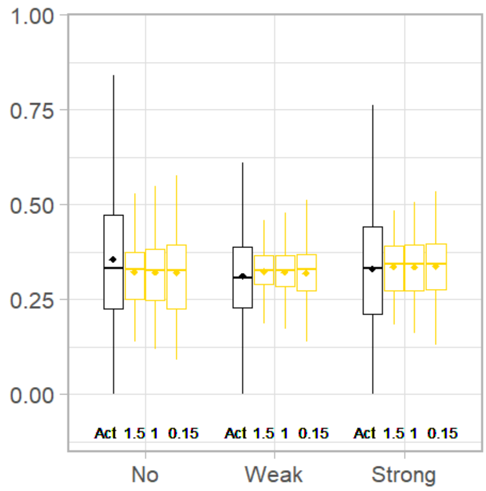


(b) Negative relation signals

Figure 12: Comparison actual posterior (Black) vs Bayesian benchmarks (Purple) with $\alpha = 0.15$ and $\alpha = 1.5$ in DfD arm

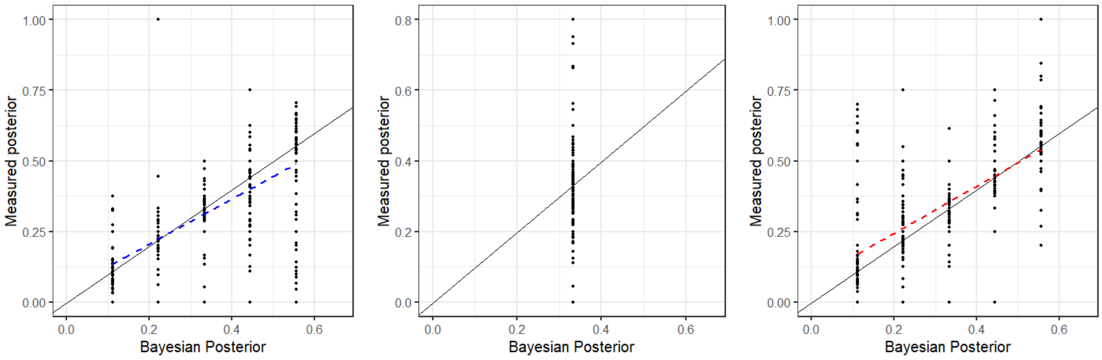


(a) Description



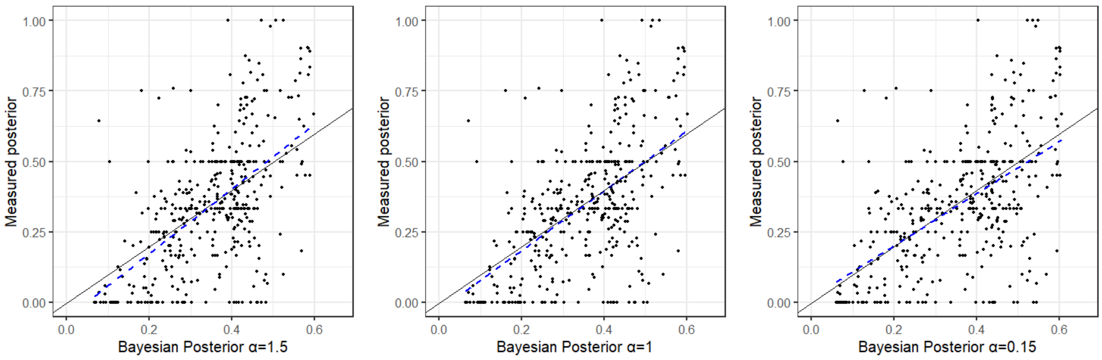
(b) Experience

Figure 13: Comparison Actual Posterior and Bayesian Posterior Yellow Signal



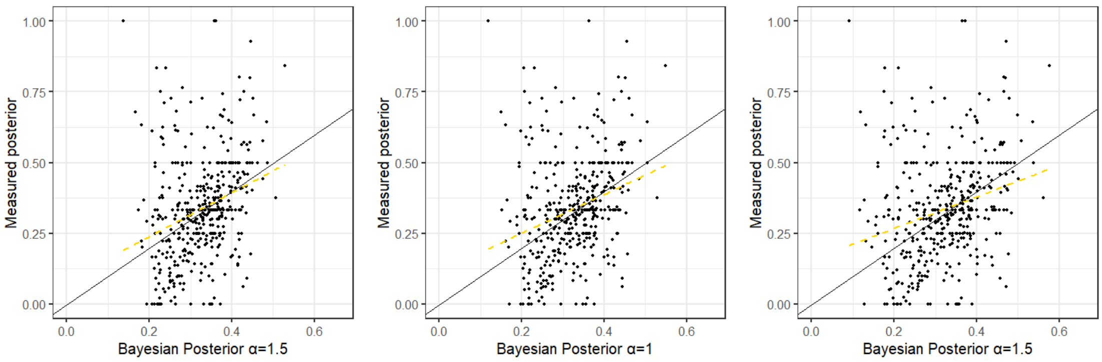
(a) Blue (b) yellow (c) Red

Figure 14: Plot points with different prior parameters Description



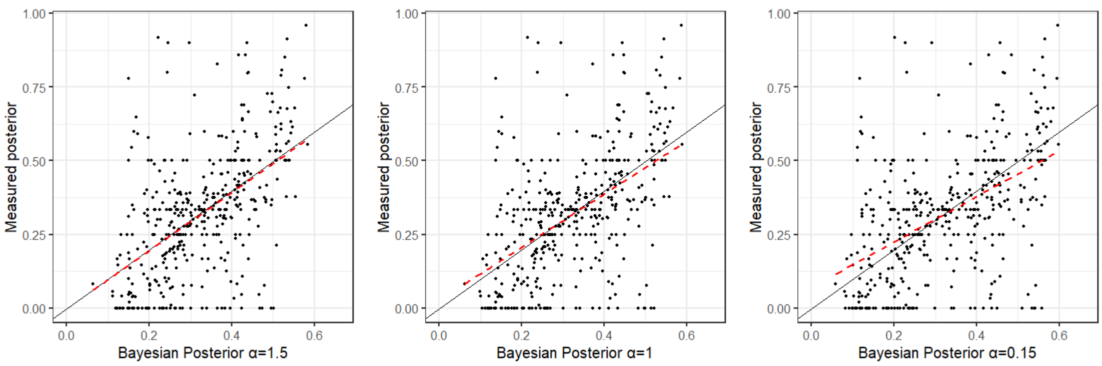
(a) $\alpha = 1.5$ (b) $\alpha = 1$ (c) $\alpha = 0.15$

Figure 15: Plot points with different prior parameters color Blue Experience



(a) $\alpha = 1.5$ (b) $\alpha = 1$ (c) $\alpha = 0.15$

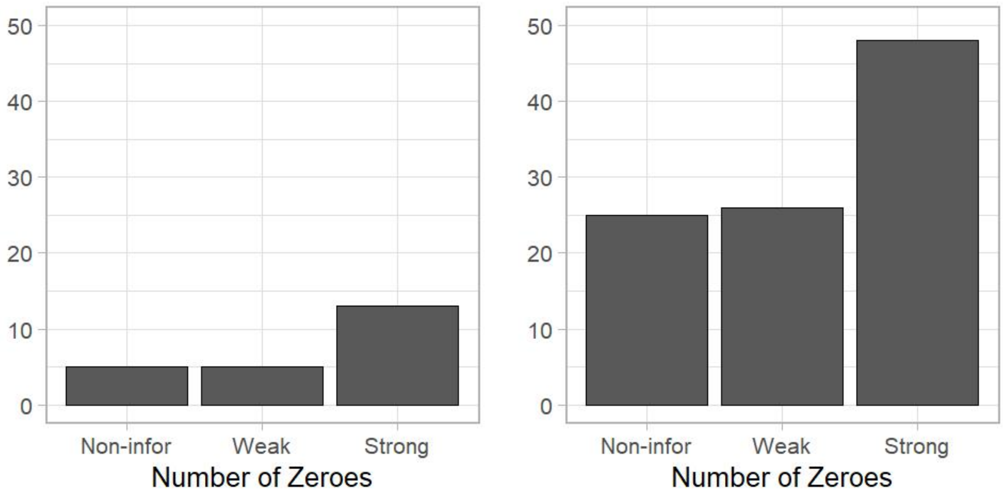
Figure 16: Plot points with different prior parameters color yellow



(a) $\alpha = 1.5$ (b) $\alpha = 1$ (c) $\alpha = 0.15$

Figure 17: Plot points with different prior parameters color Red Experience

E zeros in probabilities and Na's in logodds



(a) Description (b) Experience

Figure 18: Frequency of subjects with one zero in the probability estimation by Informativeness of the urn

F Further Exploration of the effect of priors

G Analysis without Individuals with probabilities outside the unit interval

Taking out the individuals whose probabilities lay outside the unit interval, we find that the differences in magnitudes are very small, and the comparisons between arms are still significative in the same way as they were in Section 4.2.

	Experience		Description			
	Mean	SD	Mean	SD	t	p
Aversion						
Average	0.02	0.22	0.11	0.28	-4.54	0.00
Non-informative	0.05	0.27	0.12	0.30	-2.24	0.03
Weak	0.01	0.28	0.10	0.29	-2.53	0.01
Strong	-0.00	0.23	0.10	0.29	-3.14	0.00
Insensitivity						
Average	0.36	0.22	0.28	0.25	4.01	0.00
Non-informative	0.40	0.30	0.29	0.31	3.04	0.00
Weak	0.39	0.30	0.30	0.29	2.55	0.01
Strong	0.29	0.25	0.25	0.29	1.31	0.19

Table 11: Comparison of means between urns in the Experience and Description arms

As there was no subject in the DfD arm whose probabilities were outside of the unit interval, within-treatment comparisons do not change for them. The within-treatment anova for the DfE arm does change in the case of aversion ($F(2, 274) = 2.81, p = 0.06$), while significant at the 5% level before it is only significant at the 10% level. However, for insensitivity, the results do not change as much ($F(2, 274) = 10.06, p < 0.01$). The post hoc comparison of the aversion index show that

	Difference	t	p	p-corrected
Weak-N.inf	-0.03	-1.56	0.06	0.18
Strong-N.inf	-0.05	-2.37	0.01	0.03
Strong-Weak	-0.01	-0.74	0.23	0.69

Table 12: Mean comparison by signal accuracy in aversion for the experience condition

the difference between the strong signal and the noninformative signal is still significant, however there is no difference between the weak and noninformative signal at the 10% level. However, for the a-insensitivity the differences are still significant for the comparison between the strong signal urn, and the non-informative and weak signals.

	Difference	t	p	p-corrected
Weak-N.inf	-0.02	-0.59	0.28	0.83
Strong-N.inf	-0.11	-4.26	0.00	0.00
Strong-Weak	-0.09	-3.64	0.00	0.00

Table 13: Mean comparison on insensitivity inde by signal accuracy

Now, looking at the results of our analysis for beliefs, taking out this subjects results in decrease of the estimate of weight of information and experience as shown in Table 14, and thus the coefficient is not significant at 5% level. Results in Table 15 show a similar pattern, there is no difference between the priors with, for priors $\alpha = 1.5$ and $\alpha = 0.15$ the only difference is that the coefficient estimate is lower, but the coefficient in $\alpha = 1.5$ is still significant

	DfD	DfE	Pool	
	(1)	(2)	(3)	(4)
Weight of Info. (ρ)	0.81^{***}	0.90^{***}	0.81^{***}	0.81^{***}
	(0.02)	(0.05)	(0.03)	(0.03)
W. Info x DfE			0.10^{+}	0.11^{*}
			(0.05)	(0.05)
DfE			0.02	0.02
			(0.03)	(0.03)
Male				0.01
				(0.03)
Bachelor				-0.00
				(0.03)
Post-Graduate				-0.00
				(0.04)
Constant	-0.70^{***}	-0.68^{***}	-0.70^{***}	-0.70^{***}
	(0.01)	(0.03)	(0.02)	(0.03)
s_idios	0.52	0.93	0.74	0.74
s_id	0.00	0.00	0.00	0.00
R^2	0.48	0.21	0.31	0.32
Adj. R^2	0.48	0.21	0.31	0.31
Num. obs.	1324	1119	2443	2434

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$. Random Effects estimation

Table 14: Reactiveness to information for the DfD and DfE arms and pooling both conditions

	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 0.15$
Weight of Info. (ρ)	0.81^{***}	0.81^{***}	0.81^{***}
	(0.03)	(0.03)	(0.03)
W. Info x DfE	0.10^{+}	0.19^{***}	-0.06
	(0.05)	(0.06)	(0.05)
DfE	0.02	0.02	0.03
	(0.03)	(0.03)	(0.03)
Constant	-0.70^{***}	-0.70^{***}	-0.70^{***}
	(0.02)	(0.02)	(0.02)
s_idios	0.74	0.74	0.74
s_id	0.00	0.00	0.00
R^2	0.31	0.32	0.30
Adj. R^2	0.31	0.31	0.30
Num. obs.	2443	2443	2443

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$. Random Effects estimation

Table 15: Pooled models of reactiveness to information for different parameters α for the Experience arm

H Fixed effects estimation reaction to information

	DfD	DfE	Pool	
	(1)	(2)	(3)	(4)
Weight of Info. (ρ)	0.81^{***}	0.90^{***}	0.81^{***}	0.81^{***}
	(0.02)	(0.05)	(0.03)	(0.03)
W. Info x DfE			0.10^{+}	0.11^{+}
			(0.06)	(0.06)
Controls	No	No	No	Yes
R^2	0.49	0.21	0.31	0.31
Adj. R^2	0.42	0.10	0.22	0.22
Num. obs.	1324	1146	2470	2461

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$ Fixed Effects estimation

Table 16: Reactiveness to information for the DfD and DfE arms and pooling both conditions

	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 0.15$
Weight of Info. (ρ)	0.81^{***}	0.81^{***}	0.81^{***}
	(0.03)	(0.03)	(0.03)
W. Info x DfE	0.10^{+}	0.20^{***}	-0.06
	(0.06)	(0.06)	(0.05)
R^2	0.31	0.31	0.30
Adj. R^2	0.22	0.22	0.21
Num. obs.	2470	2470	2470

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

Table 17: Pooled models of reactiveness to information for different parameters α for the Experience arm