Belief Uncertainty

Marina Agranov* Wanying Huang†

September 12, 2026

Abstract

We experimentally study whether individuals form precise beliefs implied by the frequentist interpretation of probability. Despite being given a fully specified data-generating process involving one million independent draws, about half of reported beliefs are unsure, reflecting uncertainty about what the realized frequency will be. Although these beliefs remain centered around the underlying probability, they are far more dispersed than the law of large numbers predicts. This uncertainty is systematic, persistent across questions, and responds to information but is not eliminated by it. Information affects subjects with sure and unsure initial beliefs differently. Greater belief uncertainty predicts lower willingness to take risks. Our findings suggest that eliciting belief uncertainty, rather than only point beliefs, reveals systematic features of belief formation missed by standard elicitation methods. It also distinguishes two sources of belief uncertainty: uncertainty already present in initial beliefs and uncertainty generated by new information, both of which shape updated beliefs.

*Division of the Humanities and Social Sciences, Caltech and NBER. Email: marina.agranov@gmail.com. Agranov gratefully acknowledges the support of NSF grant SES-2214040.

†Department of Economics, Monash University. Email: kate.huang@monash.edu

"Our beliefs are like wild animals, they can be coaxed and cajoled, but never truly tamed." – Ariel Sabar

1 Introduction

Economists have a long tradition of modeling individuals' beliefs with probability distributions. For a binary state, such a belief distribution reduces to a single number, often referred to as a prior. For example, a commuter holds a prior belief that there is a 50% chance of rain tomorrow; an investor puts the odds of a recession at one in three; and a patient judges a treatment to carry a 5% risk of complications. Across these different contexts, uncertainty is represented in the same way: by a single probability that captures the decision maker (DM)'s belief. This raises a natural question: how do individuals form their prior beliefs in the first place?

Philosophers and mathematicians propose an answer to this question using the frequentist interpretation of probability.[1] According to this interpretation, beliefs are probabilistic assessments formed from observed data. When samples are large and representative, it follows from the celebrated law of large numbers (LLN) that observed frequencies converge to their true probabilities. A DM who observes large samples of independent draws and forms beliefs in the frequentist sense should therefore report a precise belief: they should not only express the belief as the corresponding long-run frequency but also be *sure* of it. Understanding whether people recognize this implication of the LLN is important as it is central to the statistical reasoning that underpins many modern institutions. Clinical trials estimate treatment effects from large samples, public opinion polls approximate population preferences, and insurance and macroeconomic policy rely on the predictability of aggregate outcomes despite individual-level uncertainty.

In this paper, we investigate whether people form the precise beliefs implied by the LLN and how information affects those beliefs. To do so, we put the frequentist approach under the microscope by designing an experiment that allows participants to report a belief distribution rather than a single point belief. In particular, subjects are presented with scenarios involving a simple statistical process: N independent draws of a Bernoulli random variable that equals one with probability p, generating an empirical frequency of ones, $\hat{X}_N$, on the unit interval. Subjects are informed of the value of p and told that the draws are independent and sampled with replacement. The central object of our study is each subject's belief distribution over $\hat{X}_N$, which we refer to as the subject's *initial belief.*

To elicit these beliefs, we ask subjects to allocate one hundred probability points across a set of intervals that partition the unit interval. The number of points allocated to each interval represents, in percentage points, the subject's initial belief that $\hat{X}_N$ falls

[1] See, e.g., Von Mises (1981); Hájek (2002).

within that interval. Subjects are incentivized to report these beliefs truthfully using a BDM-style mechanism (Becker et al., 1964). Because the sample size is extremely large, i.e., $N = 1{,}000{,}000$, the objective distribution of $\hat{X}_N$ is effectively degenerate at p. A subject who understands the LLN should therefore assign all one hundred points to the interval containing p. In contrast, a subject who does not understand the LLN may either assign all one hundred points to an incorrect interval or distribute them across multiple intervals. We refer to the latter pattern as *belief uncertainty* and call the corresponding belief *unsure*; otherwise, the belief is *sure*.

Our first finding is that belief uncertainty is prevalent. Even in a controlled environment in which the data-generating process is fully specified and involves one million independent draws, subjects exhibit belief uncertainty in about half of all cases. Nevertheless, their reported belief distributions remain centered around the truth, although they are much more dispersed than the LLN would predict. This suggests that subjects understand, to some extent, the central tendency implied by the LLN, but underestimate the degree of concentration it predicts. Additional analyses show that this aggregate belief uncertainty is systematic and persistent at the individual level: for a majority of subjects, it recurs across questions and is associated with performance on related tasks. Together, these findings suggest that for many subjects, knowing the underlying probability is insufficient to form a precise belief about the corresponding empirical frequency. This departure from the LLN benchmark has important implications for standard belief-updating experiments, which typically assume that informing subjects of the probability p induces a precise prior belief exactly at p.

Having documented the prevalence of belief uncertainty, we next examine how it responds to new information and how it affects subjects' choices under risk. We operationalize these two questions in the experiment as follows. After subjects report their initial beliefs about $\hat{X}_N$, they receive additional information about the same N draws and report their *updated beliefs*. Specifically, each draw is subjected to an informative test of known precision, which yields either a positive or negative signal. Because the signals are informative, a positive signal is more likely when the draw equals one, whereas a negative signal is more likely when the draw equals zero. The test's precision measures how accurately the resulting signal reflects the underlying draw. After all N draws are tested, subjects report their updated beliefs about a *conditional* empirical frequency: the proportion of draws equal to one among those that receive a positive (negative) signal. Finally, one draw is randomly selected from the corresponding subset, and subjects choose between a safe option and a risky option that pays off if they correctly predict whether the selected draw equals one or zero. We repeat this procedure for different values of the underlying probability p and the precision of the test.

Our second main finding is that belief uncertainty responds systematically to information, but is not eliminated by informative signals. At the aggregate level, receiving signals increases both the prevalence of belief uncertainty and the average dispersion of reported beliefs, as measured by entropy. This aggregate pattern, however, masks important heterogeneity: information seems to affect subjects with initially sure and initially unsure beliefs in opposite ways. In particular, some initially sure subjects become unsure after receiving signals, although their updated beliefs exhibit relatively little uncertainty. In contrast, initially unsure subjects report less uncertainty after receiving signals but often remain unsure. For both groups, high-precision signals reduce uncertainty in updated beliefs, whereas contradicting signals increase it only among initially sure subjects.

Interestingly, this heterogeneity in subjects' responses to information does not appear to reflect fundamentally different updating rules. Both initially sure and initially unsure subjects update in the correct direction and exhibit the familiar inverse-S pattern: relative to the Bayesian benchmark, their updated beliefs are compressed toward 50/50. Although this compression appears stronger among initially unsure subjects, the difference largely reflects compression already present in their initial beliefs. Once we account for this initial compression, the additional compression induced by the signals is statistically indistinguishable between the two groups. Thus, differences in updated beliefs appear to be driven mainly by differences in initial beliefs, rather than by differences in how people update.

To better understand the origin of belief uncertainty, we further elicit subjects' beliefs about how a group of experts—statisticians and mathematicians—would answer the same updating questions. Strikingly, subjects who report uncertain updated beliefs are much more likely to believe either that these questions have no unique correct answer or that the correct answer is itself a belief distribution rather than a point estimate. This pattern suggests that subjects' reported belief distributions reflect genuine uncertainty about the correct answer rather than reporting noise around an underlying precise belief.[2]

Our third and final finding is that belief uncertainty plays an important role in shaping choices under risk. Holding fixed subjects' average updated beliefs about the payoff-relevant state, greater belief uncertainty is associated with a lower likelihood of choosing the risky bet. This relationship persists after controlling for subjects' attitudes toward risk, ambiguity, and compound lotteries, all of which are elicited in auxiliary tasks at the end of the experiment. In other words, failing to account for belief uncertainty would make subjects appear more risk averse than their independently elicited risk parameters suggest. One possible interpretation is that belief uncertainty operates like ambiguity in the decision problem.[3] Unlike standard ambiguity-elicitation methods, which typically

[2]If reporting noise were driving this pattern, subjects should still believe that there is a unique point-valued correct answer, even if they are uncertain about what that answer is.

[3]See, for example, Gilboa and Schmeidler (1989), Schmeidler (1989), and Klibanoff et al. (2005).

measure attitudes toward ambiguity using tasks *unrelated* to the decision of interest—such as the classic urn-and-balls paradigm (Ellsberg, 1961; Halevy, 2007; Abdellaoui et al., 2015)—our experiment elicits belief uncertainty directly within each updating problem. It therefore provides a measure of uncertainty that is tightly linked to the risky decision at hand.

To assess the robustness of our findings, we conduct two additional treatments. The first examines whether eliciting subjects' initial beliefs affects the prevalence of uncertainty in their updated beliefs. The second examines whether and how reported belief uncertainty depends on the coarseness of the partition of the unit interval. Across both treatments, our results remain quantitatively similar and qualitatively unchanged, suggesting that they are not driven by these design choices.

We contribute to the literature on belief formation and updating in three ways. First, we show that belief uncertainty is not an artifact of noisy reporting but reflects a stable feature of how a large group of individuals encodes uncertainty. It emerges even in a simple and controlled environment, persists at the subject level, and responds systematically to information.

Second, we identify two conceptually distinct sources of belief uncertainty: one arising from belief formation itself and the other from the arrival of new information. The first, which is the most common in our data, is evident among the subjects who report uncertain beliefs from the outset and continue to do so after receiving information, albeit with reduced dispersion. This suggests that these subjects entertain a distribution over possible empirical frequencies throughout the updating process. The second source is evident among a small fraction of subjects who initially report precise beliefs that are fully consistent with the LLN, but later become uncertain after observing informative signals. For these subjects, the arrival of information itself introduces belief uncertainty rather than simply shifting an otherwise precise belief in a predictable direction.

Third, our results provide a new perspective on the origins of the well-documented compression of subjective beliefs. Although subjects with uncertain initial beliefs exhibit substantially more compressed average updated beliefs than those with precise initial beliefs, this gap does not appear to be driven by differences in how the two groups respond to information. Instead, the two groups update in broadly similar ways. The greater compression among initially unsure subjects is largely inherited from their initial beliefs, which are already compressed before any additional information is observed. Thus, an important source of deviations from Bayesian benchmarks may lie not in the updating process itself, but in how individuals encode uncertainty before updating.

The remainder of the paper proceeds as follows. The rest of this section reviews the related literature. Section 2 describes the experimental design. Section 3 presents the main findings. Section 3.4 reports evidence from two additional treatments that assess the robustness of our results. Section 4 discusses the interpretation of the findings and concludes.

Related Literature

A large experimental literature studies how individuals form and report beliefs. See Schotter and Trevino (2014); Healy and Leo (2024) for surveys of elicitation methods and Benjamin (2019) for a survey of probabilistic-reasoning errors. Much of this literature has focused on documenting systematic departures from Bayesian belief formation and updating, such as base-rate neglect, conservatism, and confirmation bias. More recent work emphasizes the roles of complexity (Oprea, 2020; Agranov and Reshidi, 2026), cognitive noise (Enke and Graeber, 2023), and decision confidence (Hu, 2023). Across this literature, however, beliefs are typically elicited as point probabilities—that is, as a single number representing an individual's assessment of the likelihood of a particular event occurring. We depart from this practice by eliciting the probabilities subjects assign to multiple events corresponding to different ranges of an empirical frequency, thereby recovering subjects' belief distributions.

Our experimental design is most closely related to earlier work in psychology and behavioral science documenting systematic errors in reasoning about random samples. For example, Kahneman and Tversky (1972) elicit beliefs about empirical frequencies for samples of sizes 10, 100, and 1,000. They find that participants are largely insensitive to sample size and, consistent with our findings, report subjective sampling distributions that are too diffuse relative to their objective counterparts.[4] These early studies, however, examine relatively small samples and rely on unincentivized belief reports. A notable exception is Benjamin, Moore, and Rabin (2017), who elicit incentivized belief reports about the proportion of heads in samples of 10, 1,000, and one million fair-coin flips. They find that this insensitivity to sample size persists even with incentives and much larger samples and refer to it as *nonbelief in the LLN*.[5] We build on this literature on beliefs about random samples by varying the underlying probability and tracking individual-level belief distributions from initial formation through updating. In particular, we identify how belief uncertainty responds to information and affects choices.

Methodologically, our work contributes to the literature on incentivized methods for eliciting full subjective distributions rather than point beliefs (Harrison et al., 2015, 2017;

[4]Similar patterns have also been found using samples of fewer than ten observations, see, e.g., Wheeler and Beach (1968); Peterson et al. (1968).

[5]To formalize this phenomenon, Benjamin, Rabin, and Raymond (2016) propose a two-stage model of belief formation. More specifically, individuals first draw a subjective rate β from a full-support distribution centered at the true rate and then form beliefs as though observations were generated according to β. Their model predicts that subjective sampling distributions are more dispersed than their objective counterparts and remain dispersed even as the sample size grows arbitrarily large.

Eyting and Schmidt, 2021).[6] In a related study, Harrison and Swarthout (2021) repeatedly elicit incentivized belief distributions over the unknown composition of an urn after subjects observe successive samples of draws. In contrast, our subjects know the data-generating probability and form beliefs about the realized empirical frequency, both before and after conditioning on additional information. We adapt a BDM-style mechanism to incentivize truthful reporting of these beliefs. Because subjects assign probabilities to predefined intervals of the empirical frequency, our work also relates to the literature on partition dependence, which shows that reported beliefs may vary with how the outcome space is partitioned (Tversky and Koehler, 1994; Fox and Rottenstreich, 2003; Suleymanov, 2025).

At a conceptual level, our paper is related to the broader theoretical literature on belief updating under ambiguity (Gilboa and Schmeidler, 1989). In multiple-priors models, ambiguity is represented by a set of priors rather than a single prior, raising the question of how this set should change when information arrives. Full Bayesian updating applies Bayes' rule separately to each prior (Jaffray, 1992; Pacheco Pires, 2002; Gilboa and Marinacci, 2011), whereas maximum-likelihood updating retains only those priors that assign the highest likelihood to the observed information before updating them (Gilboa and Marinacci, 2011). Partial Bayesian updating provides an intermediate approach: it retains priors whose likelihood exceeds an event-dependent threshold, nesting these two updating rules as limiting cases (Kovach, 2021; Hill, 2022). Although our elicited belief distributions are not formally sets of priors, our finding that belief uncertainty persists after information arrives echoes a central feature of these models: uncertainty need not be resolved through updating.

Finally, our work contributes to the literature on second-order beliefs and compound lotteries. The theoretical literature has proposed models that allow choices to depend not only on the mean probability of an outcome but also on uncertainty about that probability (Klibanoff et al., 2005; Nau, 2006; Seo, 2009; Ergin and Gul, 2009). In our setting, if one were to interpret the empirical frequency as a first-order belief, then our elicited distribution over that frequency can be viewed as a second-order belief. Beyond documenting the persistence of this second-order uncertainty, we find that it is systematically related to choice: greater belief uncertainty is associated with less risk taking, in line with models in which uncertainty about probabilities makes risky options less attractive to ambiguity-averse decision makers.

[6] Recent evidence also shows that both cognitive biases and elicitation design can affect higher moments of reported belief distributions (Holm et al., 2025).

2 Experimental Design

Structure of the Experiment. The experiment consists of three parts. In Part I, we familiarize participants with the types of scenarios they will encounter in the experiment and the procedure used to report belief distributions. In this part, we elicit only their initial belief distributions across several scenarios, as described below. In Part II, participants encounter new scenarios and again report their initial belief distributions. They then receive additional information that they may use to update their beliefs. We subsequently elicit their updated belief distributions and ask them to make risky choices based on the information received in each scenario. Finally, in Part III, we administer a series of control tasks and collect auxiliary measures of participants' preferences.

Participants first receive general instructions describing the overall structure of the study. Before starting each part, they also receive more specific instructions tailored to that part. Throughout the instruction period, participants complete comprehension questions designed to assess their understanding of the experimental procedures. Only participants who answer these questions correctly are allowed to proceed. These precautions are intended to mitigate inattention and ensure that participants understand both the tasks and how their payments depend on their choices. The complete instructions and screenshots are provided in the Online Appendix.

Part I. Part I consists of three rounds. In each round, participants are presented with a different scenario.[7] To illustrate, consider a deck of 100 cards containing 20 red cards and 80 black cards (see Figure 1). The computer randomly draws one card from the deck with replacement, records its color, and returns it to the deck before making the next draw. This process is repeated $N = 1{,}000{,}000$ times. Participants are then asked to report their beliefs about the percentage of the N draws for which the card is red, which we denote by $\hat{X}_N$. We refer to these reported beliefs as participants' *initial* belief distributions. We use the term "initial" rather than "prior" to distinguish these distributions over $\hat{X}_N$ from the conventional point belief about the probability of red on a single draw. They are initial in the sense that participants form them solely from the description of the repeated-draw process, before receiving any additional information.

To elicit these belief distributions, participants distribute 100 probability points across 11 statements, as illustrated in Figure 2. Each statement corresponds to an interval of possible values for the percentage of red draws $\hat{X}_N$. Specifically, statement i says that "the percentage of times (out of 1 million) in which the drawn card was red is between a_i% and b_i%," where $a_i \in \{0, 5, 15, 25, \ldots, 85, 95\}$ and $b_i \in \{4, 14, 24, 34, \ldots, 94, 100\}$. Here, a_i and b_i are the inclusive lower and upper bounds of statement i. Participants

[7] The exact specification of the scenarios varies across rounds; section 2 of the Online Appendix presents all scenarios.

Figure 1: Example of a Scenario

There is a deck with 100 cards: **20** are **RED** and **80** are **BLACK**.

The computer repeats the following process **1 million times**:

- The deck is well-shuffled.
- ONE card is randomly selected from the deck; each card is equally likely to be selected.
- The computer records the color of the selected card and returns it to the deck.

Figure 2: Eliciting Belief Distributions

For each column, indicate how likely you believe the statement in that column is CORRECT.

The percentage of times (out of 1 million) in which the selected card was **RED** is between:

0% and 4%	5% and 14%	15% and 24%	25% and 34%	35% and 44%	45% and 54%	55% and 64%	65% and 74%	75% and 84%	85% and 94%	95% and 100%
	10	**80**	**10**							

Your current sum is: 100

Next

indicate how likely they think each statement is to be true by assigning $A_i \in [0, 100]$ probability points to the corresponding column i. Since the 11 statements are mutually exclusive and collectively exhaustive, the reported numbers must sum to 100 for the belief distribution to be well defined. Any column left blank is automatically assigned a value of zero.

To make the task less repetitive, we vary both the objects used in the scenarios, such as decks of cards, urns of balls, and bags of poker chips, and the composition of these objects across rounds in Parts I and II. To further ensure that participants pay attention to these differences, they answer two simple comprehension questions about each scenario before reporting their beliefs. These questions are inspired by the characterization assessment method studied by Bernheim et al. (2026).[8] For example, in the scenario described above, participants are asked how many red cards are in the deck and how many times the computer draws a card and records its color.

Part II. Part II consists of six rounds. In each round, participants are presented with a new scenario and answer three questions. The first question is the same as in Part

[8] The idea is to provide additional individual-level evidence about whether the decision-maker understands key aspects of the problem. In our case, these key aspects are the composition of the objects and the computer's sampling procedure. These simple questions vary across scenarios.

I: participants are asked to report their initial beliefs about $\hat{X}_N$. After they report these beliefs, an informative test with precision α is applied to each of the N draws. For each draw, the test result matches the color of the drawn card with probability α.[9] Thus, each test result provides an informative but imperfect signal about the color of the corresponding drawn card. The second question then asks participants to report their beliefs about the percentage of red draws *among the subset that receive a positive test result*.[10] For example, participants report their beliefs about the percentage of drawn cards that are actually red among those that receive a likely-red test result. They report these beliefs using a table similar to the one used in Part I. We refer to these elicited beliefs as participants' *updated* belief distributions.

For the third question, the computer randomly selects one draw from the subset that receives a positive test result. Participants then choose among three bets: a safe bet that pays \$2 for sure, a risky-red bet that pays \$5 if the card in the selected draw is red, and a risky-black bet that pays \$5 if that card is black.

We take two additional steps to mitigate confusion, fatigue, and potential inattention among subjects. First, in two randomly selected rounds of Part II, participants also answer two simple comprehension questions similar to those used in Part I. These questions appear after the scenario is shown but before participants report their initial beliefs. Second, we include short "brain breaks" after every three rounds. During these breaks, participants are asked to locate a camouflaged animal hidden in a nature image.

Part III. In Part III, participants complete a series of additional control tasks and auxiliary measures. These tasks are designed to help us understand why some participants report nondegenerate belief distributions and whether this belief uncertainty is related to their preferences, such as attitudes toward risk and ambiguity, or to probabilistic sophistication more generally. We briefly describe each task below.

(i) *Single-Best-Guess Question:* For each round in Part I, we ask participants to report their beliefs about $\hat{X}_N$ again, but this time as a *single best guess* rather than a distribution of probability points across the 11 statements. That is, participants report their best estimate of $\hat{X}_N$ as a number between 0 and 100.

(ii) *Small-Sample Question:* For one randomly selected round in Part I, we ask participants to report their beliefs again, but with the number of computer draws reduced from $N = 1{,}000{,}000$ to $N = 100$.

[9] The test is applied independently across draws, so the N draws generate N corresponding test results.

[10] In the experiment, positive test results are described in intuitive terms. For example, in the card scenario, participants are told that the test result reports "likely red" with probability α if the drawn card is red, and "likely black" with probability α if the drawn card is black. To reduce potential confusion, we relabel the test results across scenarios so that the test result always corresponds to the outcome whose empirical frequency is elicited in the first question.

(iii) *Statistician Question:* For one randomly selected round in Part II, we remind participants of the scenario and the updated belief distribution they reported for the subset of draws with a positive test result, i.e., their answer to the second question in Part II. We then ask what they think a group of mathematicians and statisticians from an elite scientific institution would report in response to the same question. Participants choose one of the following three options: (i) all members of the group would allocate all 100 probability points to a single statement; (ii) all members would distribute the points across the same two or more statements; or (iii) the experts would disagree and submit different reports.

(iv) *Risk, Ambiguity, and Compound-Lottery Preferences:* We use standard multiple-price lists to measure participants' attitudes toward risk, ambiguity, and the reduction of compound lotteries. Following the technique of Gillen et al. (2019), we elicit each measure twice to reduce measurement error.

(v) *Probability Matching, Compound-Lottery Reduction, Cognitive Reflection, and Overconfidence:* We also elicit participants' tendency to engage in probability matching (Humphreys, 1939; Loomes, 1998; Rubinstein, 2002), their ability to reduce compound lotteries, their cognitive reflection test (CRT) scores (Frederick, 2005), and measures of overconfidence (Moore and Healy, 2008; Dean et al., 2023).

Parameters. Across Parts I and II, we use seven different object compositions. We denote the percentage of target objects in a scenario by $x \in \{10, 20, 30, 50, 70, 80, 90\}$. We also use two levels of test precision, $\alpha \in \{0.6, 0.9\}$. We refer to the case with $\alpha = 0.9$ as the *high-precision* environment and the case with $\alpha = 0.6$ as the *low-precision* environment.

For each participant, three scenarios with different compositions are randomly selected for Part I. The same three compositions are used for the single-best-guess questions in Part III. Three of the four remaining compositions are then randomly selected for Part II. To examine whether participants respond differently to high- versus low-precision information, each Part II scenario is presented twice, once under each level of test precision.

Incentives. All tasks in our experiment are incentivized. Participants' payments consist of several parts. First, all participants receive a \$9 completion fee. Second, one of the nine rounds across Parts I and II is randomly selected for a bonus of up to \$10. This bonus constitutes the main incentive scheme in the experiment, which we describe in detail below. Third, participants can earn an additional bonus of up to \$21 based on two randomly selected tasks from Part III.[11] Finally, 10% of participants are eligible for an

[11] For the attitude-elicitation tasks, the single-best-guess task, and the small-sample task, 10% of participants are randomly selected to be eligible for payment, and their bonus is based on one randomly

additional \$1 bonus for correctly answering one randomly selected simple comprehension question from Parts I and II.

We now describe the payment rule for the randomly selected round from Parts I and II. If a round from Part I is selected, one of the 11 statements is randomly chosen, and the participant's bonus is determined using the standard BDM mechanism for that statement.[12] This amounts to implementing 11 BDM mechanisms in parallel, one for each event described by a statement. More specifically, conditional on the selected statement, whether the participant receives the \$10 bonus depends on their reported belief about the corresponding event and on whether the realized frequency falls within the range specified by the statement. Thus, participants are incentivized to truthfully report their beliefs about the 11 events, which together form a belief distribution over the empirical frequency.[13]

If a round from Part II is selected, one of the three questions in that round is randomly chosen for payment. If one of the two belief-distribution questions is selected, payment is again determined by applying the BDM mechanism to one of its 11 statements, selected at random. If the betting question is selected, participants are paid according to their chosen option: either the safe payment or the realized outcome of the risky bet.

Implementation. The experiment was programmed in oTree (Chen et al., 2016) and conducted on Prolific between March and May 2026 with U.S.-based adults. We excluded participants who failed either of the two sets of comprehension questions at the beginning of Part I or Part II.[14] A total of 304 subjects completed the study. The study was approved by Caltech (IRB IR24-1455) and preregistered on AsPredicted (#281802). The experiment lasted approximately 50 minutes. Participants earned \$14.20 on average, including the \$9 completion fee.

Discussion of the Experimental Design. A few features of our experimental design are worth discussing. First, note that our BDM mechanism is a variant of the standard single-draw BDM mechanism, in which the event of interest is typically the realization of a binary random variable $\omega \in \{0, 1\}$, such as whether a drawn card is red or black. In our setting, however, the object of interest is the empirical frequency of such a binary variable

selected task among these tasks; for the remaining tasks in Part III, all participants are eligible for payment based on one randomly selected task.

[12]We use the BDM mechanism as it is an incentive-compatible method for eliciting truthful responses regardless of participants' risk attitudes (Becker et al., 1964).

[13]Following Danz et al. (2022), we also inform participants that they have no incentive to misreport their beliefs if they wish to maximize their subjective expected payoff in the experiment. This technique has become standard practice in the belief-elicitation literature, as it helps participants understand the payment method and reinforces the incentive compatibility of truthful reporting.

[14]For each set of comprehension questions, participants had two attempts. We excluded those who failed both attempts for either set.

$\hat{X}_N = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(\omega_i = 1)$, which takes values in $[0, 1]$. We therefore adapt the standard BDM mechanism to our frequentist setting so that the payment rule remains incentive-compatible for eliciting truthful beliefs about events of the form $\hat{X}_N \in [a_i\%, b_i\%]$.

Second, implementing a BDM-like mechanism to elicit belief distributions requires specifying a finite partition across which participants allocate probability points. Our 11-bin partition coincides with that used in Kahneman and Tversky (1972) and Benjamin et al. (2017). To examine whether our results are sensitive to this design choice, we conduct an additional robustness treatment that is otherwise identical to the main treatment but uses a finer, 20-bin partition to elicit participants' belief distributions (see Section 3.4). Although we find some evidence of partition dependence, consistent with previous findings (Tversky and Koehler, 1994; Fox and Rottenstreich, 2003; Benjamin, 2019), our main findings are similar across the 11- and 20-bin treatments.

Third, one may be concerned that eliciting initial belief distributions in each scenario could make participants' uncertainty about these beliefs more salient, which could, in turn, affect their subsequent belief updating and risky choices. To examine this possibility, we conduct another robustness treatment that is identical to the main treatment except that participants' initial beliefs are not elicited in Part II (see Section 3.4). We find no meaningful differences in behavior between the two treatments, suggesting that eliciting initial beliefs does not drive our main results.

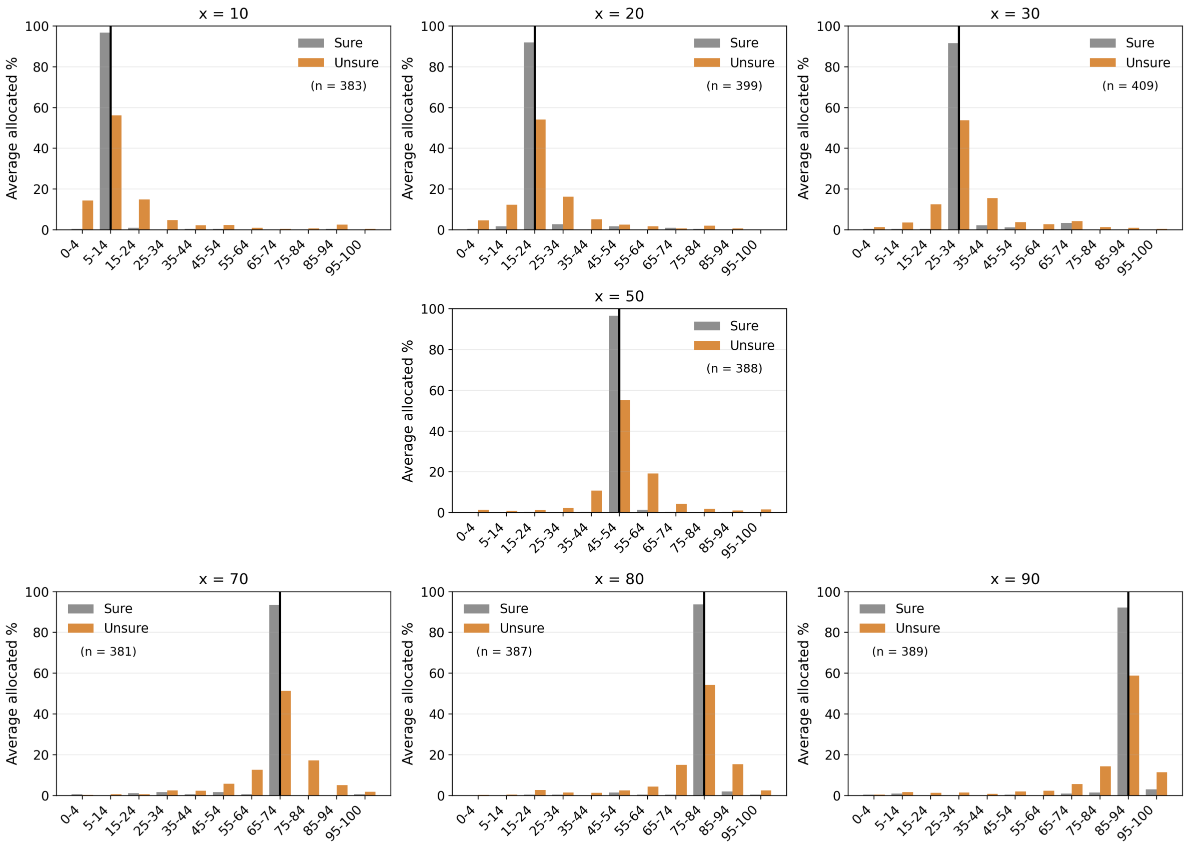
3 Results

This section proceeds as follows. First, we focus on the initial beliefs elicited in Parts I and II, where subjects are presented with scenarios involving one million draws (Section 3.1). Second, we examine how subjects incorporate additional informative signals and describe the resulting updated belief distributions, focusing in particular on changes in belief uncertainty (Section 3.2). Third, we study how these updated beliefs affect subjects' choices under risk (Section 3.3).

Data Analysis Approach. Recall that, in our experiment, subjects report a belief distribution $\hat{p}$ represented by an 11-dimensional vector $\hat{p} = (\hat{p}_1, \dots, \hat{p}_{11})$ whose entries sum to one. Each component $\hat{p}_i \in [0, 1]$ denotes the probability assigned to the interval event $\{\hat{X}_N \in [a_i\%, b_i\%]\}$. In the analysis that follows, we classify a degenerate belief distribution—that is, one that assigns probability one to a single interval event—as a *sure belief* and a nondegenerate belief distribution—that is, one that assigns positive probability to at least two interval events—as an *unsure belief.* We say that subjects exhibit *belief uncertainty* if they report an unsure belief.

We compute the following three outcomes of interest: (1) whether a subject re-

Figure 3: Average Initial Belief Distributions: Sure versus Unsure



Notes: The figure shows the aggregate probability assigned to each interval for initial beliefs elicited from Parts I and II. Each panel corresponds to a prior composition, with the vertical line marking the interval containing the true prior composition.

ports an unsure belief; (2) the reported mean belief, $\bar{m} = \sum_{i=1}^{11} \hat{p}_i \cdot m_i$, where m_i is the midpoint of the interval $[a_i, b_i]$; and (3) the Shannon entropy (with base 2) of $\hat{p}$, $H(\hat{p}) = -\sum_{i=1}^{11} \hat{p}_i \log_2 \hat{p}_i$. The entropy measure captures the amount of uncertainty in the reported belief distribution, with higher values corresponding to greater belief uncertainty.[15]

3.1 Initial Belief Uncertainty

We find that initial belief uncertainty is both prevalent and substantial. Pooling across Parts I and II, 51.2% of initial belief reports are unsure. Among these reports, mean entropy is 1.23 bits ($s.d. = 0.74$). This pattern is also consistent across underlying compositions.[16] In Figure 3, we present the empirical distribution of reported initial beliefs

[15] A sure belief has an entropy of zero, whereas every unsure belief has positive entropy. To give a sense of magnitudes, uncertainty about the outcome of a fair coin flip has an entropy of 1 bit, while a belief distribution assigning probability 0.5 to one interval and 0.25 to each of two others has an entropy of 1.5 bits.

[16] See Table A.1 in the Appendix for summary statistics by prior composition.

separately for each composition x, distinguishing between sure and unsure reports.[17] As seen in the figure, sure reports are highly concentrated around $x/100$, whereas unsure reports remain centered at this benchmark but are substantially more dispersed. This suggests that subjects who report unsure initial beliefs understand, on average, the central tendency implied by the LLN but underestimate the degree of concentration implied by the large sample size.

To see whether subjects' initial belief uncertainty is sensitive to sample size, we compare their Part I belief reports with those from the corresponding small-sample question in Part III. Recall that in this control question, the number of draws is reduced from one million to one hundred.[18] We find that unsure reports are more frequent in the small-sample environment than in the large-sample environment (62.5% versus 54.3%; paired t-test: $t = -2.774$, $p = 0.006$). Nonetheless, among subjects who report unsure beliefs in both environments ($n = 136$, 44.7%), mean entropy is similar across the two environments (1.29 versus 1.30 bits; paired t-test: $t = -0.222$, $p = 0.824$). Thus, a smaller sample increases the frequency of unsure reports but leaves the amount of uncertainty reported largely unchanged. This limited sensitivity echoes earlier evidence that subjective sampling distributions are insufficiently sensitive to sample size (Kahneman and Tversky, 1972; Benjamin et al., 2017).

We conduct two additional analyses to ensure that the observed prevalence of unsure initial beliefs does not simply reflect confusion about the task, inattention, or elicitation noise. First, we examine subjects' answers to the simple comprehension questions, which were administered immediately before belief elicitation. We find that subjects answered these questions correctly more than 95% of the time, regardless of whether their initial belief report was sure or unsure.[19] This high accuracy suggests that subjects understood the process and attended to the key features of each scenario. Thus, belief uncertainty is unlikely to be driven by task misunderstanding or inattention.

Second, to provide further evidence that unsure reports capture genuine belief uncertainty, we study the relationship between unsure initial belief reports and subjects' responses in the single-best-guess task. Intuitively, if subjects generate their single-best guesses by drawing from their reported belief distributions, those who report unsure beliefs should be more likely to give guesses that differ from the true composition.[20] Our

[17] The number of observations varies slightly across compositions because, as described in the experimental design, scenario compositions in Parts I and II are randomly drawn.

[18] As a benchmark, for each underlying composition, we compute the entropy of the objective distribution over the 11 intervals induced by Binomial$(100, x/100)$ and then average across the compositions observed in the sample. The resulting average entropy is 0.92 bits. For $N = 1{,}000{,}000$, the corresponding objective entropy over the 11 intervals is essentially zero.

[19] Across Parts I and II, the correlation between reporting an unsure belief and answering both comprehension questions correctly is small and insignificant at the round level ($r = -0.008$, $p = 0.767$) and at the subject level ($r = -0.040$, $p = 0.487$).

[20] Overall, 28% of single-best-guess reports differ from the true composition. Figure A.1 in the Ap-

data is consistent with this prediction: at the subject-question level, the correlation between unsure reports in Part I and incorrect single-best guesses is 0.31 ($p < 0.001$); at the subject level, the correlation between the number of unsure reports and the number of incorrect single-best guesses is 0.36 ($p < 0.001$). Thus, unsure reports appear to capture systematic variation in beliefs that carries over to the single-best-guess task, rather than task-specific elicitation noise.

We conclude this section by noting that uncertainty in initial beliefs is persistent at the individual level. In Part I, belief uncertainty recurs across questions for a majority of subjects: 54% report unsure beliefs in two or more of the three Part I questions. This persistence is particularly pronounced at the extremes: about 44% of subjects report unsure beliefs in all three questions, and 79% of these subjects continue to do so in all six initial-belief questions in Part II. In contrast, of the 39% of subjects who report sure beliefs in all three Part I questions, 92% continue to report sure beliefs throughout Part II. At the subject level, the fraction of unsure initial beliefs is also strongly positively correlated across parts ($r = 0.85$, $p < 0.001$). This persistence motivates the subsequent analysis, in which we examine whether subjects with sure and unsure initial beliefs respond differently to the information they receive.

Result 1: *Despite the concentration implied by the LLN, initial belief uncertainty is prevalent even in a simple setting in which the data-generating process is known and the sample size is extremely large. Moreover, the tendency to report uncertain initial beliefs is an individual-specific feature, present among roughly half of subjects.*

3.2 How Information Affects Belief Uncertainty

We next study how receiving new information affects belief uncertainty and whether this effect depends on the type of information received. We distinguish signals along two dimensions: whether they confirm or contradict the prior composition and whether they are of high or low precision.

At the aggregate level, receiving information makes subjects more likely to report unsure beliefs: on average, the share of unsure reports rises by 6 percentage points from initial to updated beliefs (clustered s.e. $= 0.016$, $p < 0.001$).[21] This increase is larger following low-precision than high-precision signals but does not differ significantly between confirming and contradicting signals.[22] To examine whether these aggregate patterns differ by initial belief type, Table 1 reports OLS regressions estimated separately

pendix presents histograms of these reports.

[21]To account for subject heterogeneity, we cluster standard errors at the individual level.

[22]The low- versus high-precision gap is 3.2 percentage points ($\beta = -0.032$, clustered s.e. $= 0.015$, $p = 0.032$), while the confirming-versus-contradicting difference is small and statistically insignificant ($\beta = 0.024$, clustered s.e. $= 0.021$, $p = 0.244$).

Table 1: How Belief Uncertainty Responds to Different Information

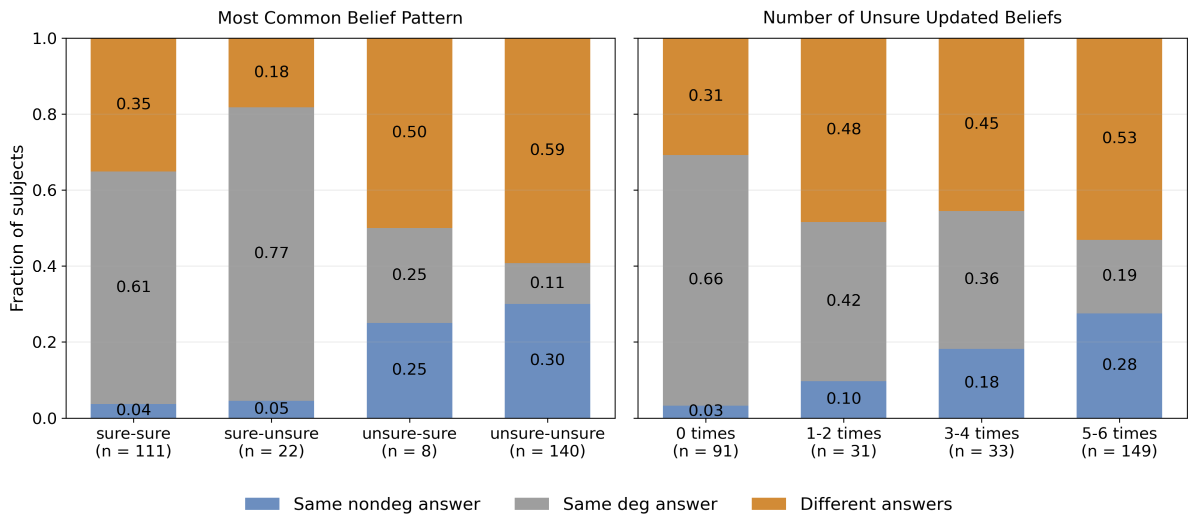
	Change in unsure-belief indicator			Change in entropy		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Sure Initial Beliefs						
Constant	0.221***	0.183***	0.232***	0.228***	0.162***	0.274***
	(0.026)	(0.030)	(0.029)	(0.031)	(0.029)	(0.040)
Contradicting signal		0.069**			0.128***	
		(0.030)			(0.036)	
High-precision signal			−0.022			−0.090***
			(0.020)			(0.026)
Observations	905	767	905	905	767	905
Subjects	181	169	181	181	169	181
Panel B. Unsure Initial Beliefs						
Constant	−0.099***	−0.091***	−0.075***	−0.046**	−0.052*	0.043
	(0.016)	(0.019)	(0.016)	(0.022)	(0.030)	(0.028)
Contradicting signal		−0.006			0.041	
		(0.022)			(0.034)	
High-precision signal			−0.048***			−0.180***
			(0.017)			(0.033)
Observations	919	809	919	919	809	919
Subjects	176	175	176	176	175	176

Notes: Each column reports an OLS regression, with standard errors clustered at the subject level. The dependent variables in columns (1)–(3) and (4)–(6) are the change in the unsure-belief indicator and the change in entropy from the initial to the updated report, respectively. In columns (2) and (5), scenarios with prior composition 50 are excluded because contradicting signals are undefined in those scenarios. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

for sure and unsure initial beliefs. For each measure of belief uncertainty, we regress the change from the initial to the updated belief on a constant and indicators for whether the signal contradicts the prior composition and whether it is highly precise. We estimate these regressions separately by initial belief type, with standard errors clustered at the subject level.

We find that the aggregate results mask substantial heterogeneity by initial belief type. First, the aggregate increase in belief uncertainty appears to be driven by subjects with sure initial beliefs, particularly following signals that contradict the prior composition. By contrast, for subjects with unsure initial beliefs, receiving information reduces their tendency to report unsure beliefs. Despite this reduction, the latter group remains more uncertain than the former after receiving information: the mean entropy of their updated beliefs is 1.16 bits, more than five times the mean of 0.23 bits among subjects with sure initial beliefs. Second, for both groups, higher-precision signals reduce belief uncertainty but do not eliminate it: even after receiving a signal with 90% precision, 88% of subjects with unsure initial beliefs continue to report unsure updated beliefs, with a mean entropy of 1.05 bits. Thus, these results show that the effect of information on belief uncertainty is not uniform: it increases uncertainty among subjects with sure initial

Figure 4: Subjects' Responses to the Statistician Question



beliefs and only partially reduces it among those with unsure initial beliefs, even when signals are highly precise.

Result 2a: *At the aggregate level, receiving information increases the prevalence of unsure beliefs. This pattern masks opposing responses by initial belief type: information increases belief uncertainty among initially sure subjects but only partially reduces it among initially unsure subjects. High-precision signals reduce uncertainty in both groups, while contradicting signals increase uncertainty only among initially sure subjects.*

Statistician Question. To better understand why subjects report unsure beliefs after receiving information, we examine their responses to the statistician question. Recall that in this task, subjects indicate whether they believe that a group of experts answering the same belief-elicitation question would (i) all report the same degenerate belief distribution, (ii) all report the same nondegenerate belief distribution, or (iii) disagree with one another. Figure 4 plots the fraction of subjects selecting each response, separately by their own belief-reporting behavior.

In the left panel of Figure 4, we classify subjects according to their modal initial-to-updated belief pattern across the six Part II rounds.[23] Most subjects fall into one of two common belief-transition patterns. For about 46% of subjects, the modal pattern is unsure-to-unsure; these subjects are also the least likely to believe that all experts would report the same degenerate belief distribution. For another 37%, the modal pattern is sure-to-sure; these subjects are also most likely to correctly believe that all experts

[23]We exclude the 23 subjects with no unique modal pattern (i.e., a tie between two or more of the four initial-to-updated belief patterns across their six Part II rounds).

would report the same degenerate belief distribution. The right panel of Figure 4 shows the same relationship using the number of unsure updated beliefs each subject reports across the six Part II rounds. As this number increases, the fraction who believe that all experts would report the same degenerate belief distribution decreases. Consistent with this pattern, we find that the frequency of unsure updated beliefs is positively correlated with the belief that experts would not unanimously report a degenerate belief ($r = 0.42$, $p < 0.001$).[24]

Taken together, these patterns suggest that subjects' own tendency toward persistent belief uncertainty reflects their genuine belief that either the correct distribution is nondegenerate or no unique correct distribution exists, rather than merely noise in their belief reports.

Belief Updating. We now examine whether subjects with sure and unsure initial beliefs update differently relative to the Bayesian benchmark.[25] In Figure 5, we plot the average reported mean belief against the 45-degree line, separately by initial-belief type. We make three observations.

First, subjects who report unsure initial beliefs in Part II (right panel) already exhibit the familiar inverse-S shape even before receiving any information. That is, they overestimate small probabilities and underestimate large ones, so their mean beliefs are compressed toward the center. After observing signals, this compression becomes even more pronounced in their updated beliefs. Second, subjects with sure initial beliefs (left panel) start close to the 45-degree line, with the inverse-S compression emerging mainly after they observe signals. Third, despite these differences, the change from initial to updated beliefs appears comparable across panels, suggesting that the two groups respond to information in broadly similar ways.

To test these patterns more formally, Table 2 reports the results of three regressions, with standard errors clustered at the subject level. Column (1) regresses updated mean beliefs on the Bayesian posterior benchmark, an indicator for unsure initial beliefs, and their interaction. The negative and statistically significant interaction coefficient confirms the greater compression among subjects with unsure initial beliefs. Column (2) uses the same specification for initial beliefs with the prior benchmark and shows that this difference predates the observation of signals. Column (3) regresses changes in mean beliefs on the corresponding Bayesian change implied by the signal, again including the indicator and interaction. Consistent with the figure, we find little evidence that the

[24]We find further evidence that subjects whose initial or updated mean beliefs deviate more from the Bayesian benchmark are less likely to believe that all experts would unanimously report a degenerate belief. See Table A.2 in the Appendix.

[25]Here, the Bayesian benchmark refers to the Bayesian prior point belief implied by the scenario for initial beliefs, and the Bayesian posterior point belief after observing the signal for updated beliefs.

Figure 5: Observed Beliefs versus Bayesian Beliefs

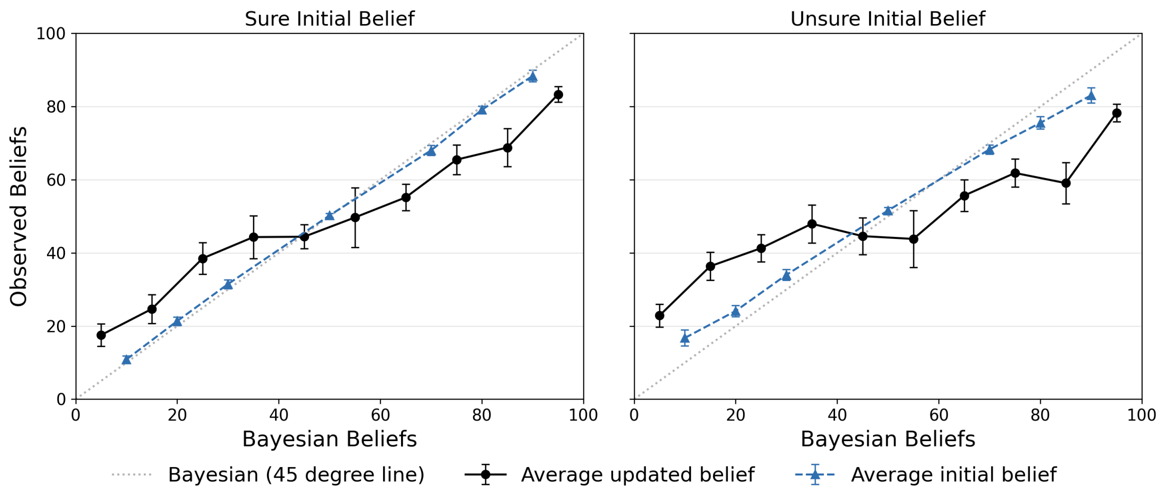


Table 2: Beliefs Relative to the Bayesian Benchmark

	Mean belief		Change in mean belief
	(1) Updated	(2) Initial	(3) Updated − Initial
Bayesian benchmark	0.657***	0.973***	0.676***
	(0.024)	(0.009)	(0.050)
Dummy for unsure initial belief	7.412***	5.817***	0.353
	(1.984)	(1.139)	(1.161)
Bayesian benchmark × dummy	−0.155***	−0.102***	−0.038
	(0.033)	(0.022)	(0.069)
Constant	17.747***	1.219***	0.203
	(1.478)	(0.375)	(0.874)
Observations	1,824	1,824	1,824
Subjects	304	304	304
R^2	0.456	0.903	0.348

Notes: Each column reports an OLS regression, with standard errors clustered at the subject level. All beliefs are expressed in percentage points. The dependent variables in columns (1)–(3) are the updated mean belief, initial mean belief, and their difference, respectively. The corresponding Bayesian benchmarks are the posterior, prior, and posterior-minus-prior change implied by the signal. The unsure-initial-belief dummy equals one for a nondegenerate initial belief distribution and zero otherwise. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

two groups differ in their responsiveness to new information. Thus, these results suggest that the greater compression in updated beliefs among unsure subjects primarily reflects differences in their initial beliefs, rather than differences in their responses to new information.

Result 2b: *Differences in updated beliefs are driven mainly by differences in initial beliefs, not by how people update. Subjects who are initially unsure already hold compressed beliefs, and this compression carries through to their updated beliefs. Moreover, the persistence of unsure beliefs seems to reflect subjects' genuine belief that either the*

Table 3: How Belief Uncertainty Affects Risky Behavior

	Indicator for choosing a risky bet			
	(1)	(2)	(3)	(4)
Updated belief entropy	-0.091^{***}	-0.130^{***}	-0.124^{***}	-0.126^{***}
	(0.017)	(0.036)	(0.036)	(0.036)
Updated mean belief		0.006^{***}	0.006^{***}	0.006^{***}
		(0.000)	(0.000)	(0.000)
Updated belief entropy × updated mean belief		0.001^{**}	0.001^{***}	0.002^{***}
		(0.001)	(0.001)	(0.001)
Risk CE				0.002^{**}
				(0.001)
Ambiguity CE (normalized)				0.000
				(0.001)
Compound CE (normalized)				0.002^{**}
				(0.001)
Constant	0.788^{***}	0.372^{***}	0.189	0.142
	(0.021)	(0.038)	(0.168)	(0.168)
Additional controls	No	No	Yes	Yes
Observations	1,824	1,824	1,824	1,788
Subjects	304	304	304	298

Notes: Each column reports a linear random-effects regression with a subject-level random intercept; standard errors are in parentheses. The dependent variable indicates whether a risky bet is chosen. The mean updated belief is folded so that higher values always correspond to a stronger belief in favor of the risky option. Columns (3) and (4) control for subjects' CRT scores, absolute and relative overconfidence, probability matching, compound-lottery-reduction ability, and the Part II simple-question score. Column (4) additionally includes CEs from the 20-ball risk, ambiguity, and compound-lottery urn questions and excludes subjects with negative CE values. Ambiguity CE and Compound CE are normalized relative to Risk CE. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

correct distribution is nondegenerate or no unique correct distribution exists.

3.3 Choice Under Risk

We next examine how subjects' updated beliefs and the uncertainty surrounding them relate to their choices under risk. Across all Part II rounds, subjects choose one of the two risky options 72.5% of the time and the safe option 27.5% of the time. Subjects choose the safe option more often when their updated mean belief is close to 50%. As their beliefs shift toward either state, they increasingly choose the risky option corresponding to that state (see Figure A.2 in the Appendix). Thus, their choices generally align with their updated mean beliefs.

Table 3 examines how belief uncertainty relates to risky behavior. Column (1) regresses the indicator for choosing a risky bet on the entropy of the updated belief distribution. A one-bit increase in entropy is associated with a 9-percentage-point decrease in the probability of choosing a risky bet. Column (2) adds subjects' average updated belief, folded so that higher values indicate a stronger belief in favor of one of the risky options, and its interaction with entropy. As expected, a stronger mean belief in favor of a risky option is associated with a higher probability of choosing a risky bet. The positive interaction term indicates that the negative effect of belief uncertainty on risky behavior is attenuated when subjects hold stronger beliefs in favor of the risky bet.

Columns (3) and (4) add controls for individual characteristics and task performance, followed by preferences toward risk, ambiguity, and compound lotteries. The coefficients on entropy and its interaction with mean beliefs remain similar. These findings are also robust to alternative measures of these preferences (see Table A.3 in the Appendix). Thus, the entropy of subjects' updated beliefs seems to capture variation in risky choice beyond that explained by these elicited preferences and characteristics.

Finally, to assess the extent to which belief uncertainty discourages risky choice, we compare subjects' observed choices with those predicted by a standard expected-utility (EU) model with CARA utility. For each subject, we estimate their risk aversion parameter using the certainty equivalent elicited from a multiple-price list in Part III.[26] We then combine the estimated CARA parameter with the subject's round-specific updated mean belief to predict whether an EU maximizer would choose a risky bet. The EU model matches observed choices in about 71% of rounds. However, it predicts risky choice in 92.3% of rounds, compared with only 72.5% observed in the data. Thus, a standard expected-utility model that does not account for belief uncertainty would substantially *overpredict* risk-taking.

Result 3: *Subjects with greater uncertainty in their updated beliefs are less likely to choose a risky bet. This holds even after controlling for elicited preferences toward risk, ambiguity, and compound lotteries. As a result, without accounting for belief uncertainty, subjects may appear to be more risk-averse than they truly are.*

3.4 Robustness Treatments

We conduct two additional treatments to assess the robustness of our main findings. In the first treatment (T1), we do not elicit subjects' initial beliefs in Part II. This allows us to test whether eliciting these beliefs in the main treatment makes belief uncertainty more salient and thereby affects subsequent belief updating or risky choice. In the second treatment (T2), we replace the 11-bin elicitation table with a finer 20-bin partition to examine whether our results depend on the number of bins.[27] Apart from these treatment-specific changes, all other features of the design remain the same. We recruited an additional 147 subjects for T1 in April 2026 and 152 subjects for T2 in May 2026, following the same implementation protocol as in the main treatment.

[26] We use the risky 20-ball urn for this estimation. The estimated CARA parameter has 25th, 50th, and 75th percentiles of 0.010, 0.030, and 0.052, respectively.

[27] Specifically, we split each 10-percentage-point interior bin, such as 5–14%, into two five-percentage-point bins, 5–9% and 10–14%, while leaving the endpoint bins unchanged.

First, comparing T1 with the main treatment, we find no evidence that eliciting initial belief distributions in Part II affects subjects' subsequent beliefs or choices. When initial beliefs are not elicited, both the frequency of unsure updated beliefs and the entropy of those reports are statistically indistinguishable from their counterparts in the main treatment. Likewise, as in the main treatment, greater updated-belief uncertainty is associated with a lower likelihood of choosing a risky option, and this relationship remains statistically significant after controlling for other individual characteristics and preferences.[28]

Second, comparing T2 with the main treatment, we find that the finer partition increases the frequency of unsure initial beliefs by 11.2 percentage points but has no statistically significant effect on updated beliefs. To compare the dispersion of unsure reports across partitions, we normalize entropy by the maximum attainable entropy under each partition, thereby placing dispersion on a common scale.[29] Conditional on an unsure report, normalized entropy is statistically indistinguishable across treatments for both initial and updated beliefs.[30] Thus, the finer partition makes subjects more likely to report initial belief uncertainty but does not change the dispersion of their beliefs conditional on an unsure report.

This comparison also allows us to examine *partition dependence*, a phenomenon documented in the literature whereby reported beliefs vary with the partition of the outcome space. To see if the finer partition changes how subjects allocate probability mass in our setting, we first classify each initial-belief report as sure or unsure using the partition under which it was elicited. For reports classified as unsure, we map each T2 report from the 20-bin partition back onto the coarser 11-bin partition used in the main treatment. We then calculate the probability mass assigned to the interval containing the true composition and average it across these subjects' reports in Parts I and II. The resulting average is 62.9% in T2, compared with 54.8% in the main treatment (Welch t-test, $p < 0.001$; see also Figure A.3 in the Appendix). Thus, in our setting, partition dependence takes the form of a shift toward the interval containing the truth. This is consistent with the evidence that splitting an event into a larger number of bins can increase the total probability assigned to that event (Tversky and Koehler, 1994; Benjamin, 2019).

Finally, when we replicate the main analyses using the T2 data, the rest of our results remain qualitatively similar under the finer partition (see Appendix A.1). Overall, the results from T1 and T2 suggest that our main findings are robust to both design choices.

[28]See Tables A.4 and A.6 in the Appendix.

[29]For a reported belief distribution p, normalized entropy is defined as $H(p)/\log_2 n$, where n is the number of bins. A value of 0 corresponds to a degenerate belief distribution, whereas a value of 1 corresponds to a uniform distribution over all n bins.

[30]See Table A.5 in the Appendix.

4 Discussion

Our experiment begins from a simple question: when individuals know the probability governing a stochastic process, do they form the precise beliefs about long-run frequencies implied by the law of large numbers? For roughly half of the subjects in our experiment, the answer is no. Even with a fully specified data-generating process and one million independent draws, these subjects assign positive probability to multiple possible empirical frequencies.[31] At the same time, their beliefs are centered around the underlying probability and respond systematically to changes in the environment. Thus, their behavior does not suggest a complete failure to understand the stochastic process. Rather, it points to a distinction between knowing the probability governing individual realizations and being certain about the empirical frequency generated by those realizations.

This distinction has implications for how belief updating is typically studied. In many experiments, the researcher specifies a prior probability and then studies how subjects respond to new information. The specified probability is generally treated as the subject's prior belief, so deviations of the resulting posterior from the Bayesian benchmark are attributed to the updating process. Our results suggest that these two objects—the probability supplied by the experimenter and the belief represented by the subject—need not coincide. Subjects with unsure initial beliefs already exhibit substantial compression toward 50/50 before receiving any additional information. Although their updated beliefs are also more compressed than those of subjects with sure initial beliefs, the difference largely reflects this initial compression. Once we examine the change from initial to updated beliefs, the two groups respond to information in broadly similar ways.

More generally, this suggests that some departures from Bayesian posterior beliefs may originate before updating begins. A posterior belief combines at least two conceptually distinct stages: how the decision maker represents the initial probability and how that representation changes in response to information. Observing only the final posterior makes it difficult to distinguish between the two. Our results illustrate that this distinction can matter empirically. A subject may appear to underreact to information relative to a Bayesian benchmark not only because of how she processes the signal, but also because the belief from which she begins already differs from the point prior imposed by the researcher. Eliciting belief distributions therefore provides a way to separate distortions in initial belief formation from distortions introduced during updating.

Our findings also highlight a second role of information: it changes not only the location of beliefs but also the uncertainty surrounding them. This reveals two conceptually distinct sources of belief uncertainty in our data. For the largest group of unsure subjects,

[31] In other words, some individuals retain uncertainty about the mapping from an underlying probability to the distribution of its empirical frequencies.

uncertainty is already present before any additional information arrives, reflecting how individuals represent a stochastic process even when its underlying probability is known; informative signals tend to reduce this uncertainty, particularly when they are highly precise, but rarely eliminate it. For another, smaller group, initial beliefs are precise and consistent with the LLN, but uncertainty emerges when subjects confront the additional inference required by the signal, especially when the signal contradicts the initial composition. Thus, information need not simply move a precise prior to a precise posterior or monotonically resolve uncertainty: it can reduce pre-existing belief uncertainty for some individuals while generating new uncertainty for others. Distinguishing between these two sources would be impossible if beliefs were elicited only as point estimates.

Our choice results suggest that this additional dimension of beliefs also has behavioral consequences. Holding fixed the average updated belief about the payoff-relevant state, subjects are less likely to choose a risky option when their reported belief distribution is more dispersed. This relationship remains after controlling for independently elicited attitudes toward risk, ambiguity, and compound lotteries. Thus, two individuals with the same average probability assessment may make different choices depending on how certain they are about that assessment. One possible interpretation is that belief uncertainty can make an objectively risky environment subjectively resemble an ambiguous one. Although the underlying probability, sampling process, and precision of the test are all known in our experiment, subjects may remain unsure which probability to use when evaluating the risky option. In this sense, whether a decision feels risky or ambiguous may depend not only on the information available in the environment but also on how the decision maker represents that information.

Our setting is deliberately simple, and this simplicity is useful for interpreting the broader relevance of the results. Subjects know the underlying probability, know that observations are independently drawn with replacement, and face an extremely large sample. These conditions leave relatively little objective reason for uncertainty about empirical frequencies. Nevertheless, substantial belief uncertainty remains. Outside the laboratory, individuals typically face additional uncertainty about the underlying data-generating process, the relevance and independence of observations, and the reliability of information. Our results therefore raise the possibility that uncertainty about one's own probability assessment may be important in a much broader class of decisions. At the same time, the precise amount of uncertainty we elicit depends to some extent on how the outcome space is partitioned, and our experiment does not identify the cognitive mechanism generating these belief distributions. The robustness treatments nevertheless show that the main patterns—the persistence of belief uncertainty, its response to information, and its relationship with choice—are not driven by the particular elicitation procedure.

Taken together, our findings suggest that representing an individual's belief by a single probability can conceal an economically meaningful dimension of uncertainty. Individuals may differ not only in what probability they assign to an event, but also in how certain they are that this is the probability they should use. Measuring this uncertainty helps distinguish belief formation from belief updating, reveals how information can change the precision as well as the location of beliefs, and improves our understanding of how beliefs translate into choice.

References

Abdellaoui, M., P. Klibanoff, and L. Placido (2015): "Experiments on compound risk in relation to simple risk and to ambiguity," *Management Science*, 61, 1306–1322.

Agranov, M. and P. Reshidi (2026): "Disentangling Suboptimal Updating: Task Difficulty, Structure, and Sequencing," *Quantitative Economics*.

Becker, G. M., M. H. DeGroot, and J. Marschak (1964): "Measuring utility by a single-response sequential method," *Behavioral science*, 9, 226–232.

Benjamin, D. J. (2019): "Chapter 2 - Errors in probabilistic reasoning and judgment biases," in *Handbook of Behavioral Economics - Foundations and Applications 2*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, North-Holland, vol. 2 of *Handbook of Behavioral Economics: Applications and Foundations 1*, 69–186.

Benjamin, D. J., D. A. Moore, and M. Rabin (2017): "Biased beliefs about random samples: Evidence from two integrated experiments," Tech. rep., National Bureau of Economic Research.

Benjamin, D. J., M. Rabin, and C. Raymond (2016): "A Model of Nonbelief in the Law of Large Numbers," *Journal of the European Economic Association*, 14, 515–544.

Bernheim, D., A. J. Lucia, K. Nielsen, and C. Sprenger (2026): "When Are Decisions Improvable? An Evaluation of Diagnostic Methods," *working paper.*

Chen, D. L., M. Schonger, and C. Wickens (2016): "oTree—An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 9, 88–97.

Danz, D., L. Vesterlund, and A. J. Wilson (2022): "Belief elicitation and behavioral incentive compatibility," *American Economic Review*, 112, 2851–2883.

Dean, M., P. Ortoleva, E. Snowberg, and C. Camerer (2023): "Econographics," *Journal of Political Economy Microeconomics*, 1(1), 1–224.

Ellsberg, D. (1961): "Risk, ambiguity, and the Savage axioms," *The quarterly journal of economics*, 75, 643–669.

Enke, B. and T. Graeber (2023): "Cognitive uncertainty," *The Quarterly Journal of Economics*, 138, 2021–2067.

Ergin, H. and F. Gul (2009): "A theory of subjective compound lotteries," *Journal of Economic Theory*, 144, 899–929.

Eyting, M. and P. Schmidt (2021): "Belief elicitation with multiple point predictions," *European Economic Review*, 135, 103700.

Fox, C. R. and Y. Rottenstreich (2003): "Partition priming in judgment under uncertainty," *Psychological Science*, 14, 195–200.

Frederick, S. (2005): "Cognitive reflection and decision making," *Journal of Economic perspectives*, 19, 25–42.

Gilboa, I. and M. Marinacci (2011): "Ambiguity and the Bayesian Paradigm," Manuscript prepared for *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress of the Econometric Society*, edited by Daron Acemoglu, Manuel Arellano, and Eddie Dekel.

Gilboa, I. and D. Schmeidler (1989): "Maxmin expected utility with non-unique prior," *Journal of mathematical economics*, 18, 141–153.

Gillen, B., E. Snowberg, and L. Yariv (2019): "Experimenting with measurement error: Techniques with applications to the Caltech cohort study," *Journal of Political Economy*, 127, 1826–1863.

Hájek, A. (2002): "Interpretations of Probability," in *The Stanford Encyclopedia of Philosophy*, ed. by E. N. Zalta, Metaphysics Research Lab, Stanford University.

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

Harrison, G. W., J. Martínez-Correa, J. T. Swarthout, and E. R. Ulm (2015): "Eliciting subjective probability distributions with binary lotteries," *Economics Letters*, 127, 68–71.

——— (2017): "Scoring rules for subjective probability distributions," *Journal of Economic Behavior & Organization*, 134, 430–448.

Harrison, G. W. and J. T. Swarthout (2021): "Belief distributions, Bayes rule and Bayesian overconfidence," Tech. rep., Working Paper.

Healy, P. J. and G. Leo (2024): "Belief elicitation: a user's guide," *Handbook of Experimental Methodology, ed. by L. Yariv and E. Snowberg, Elsevier*, 1.

Hill, B. (2022): "Updating confidence in beliefs," *Journal of Economic Theory*, 199, 105209.

Holm, H. J., M. Samahita, R. van Veldhuizen, and E. Wengström (2025): "Anchoring and subjective belief distributions," *Journal of Economic Behavior & Organization*, 240.

Hu, E. H. (2023): "Confidence in inference," Tech. rep., Working paper.

Humphreys, L. (1939): "Acquisition and extinction of verbal expectations in a situation analogous to conditioning," *Journal of Experimental Psychology*, 25(3), 294–301.

Jaffray, J.-Y. (1992): "Bayesian updating and belief functions," *IEEE transactions on systems, man, and cybernetics*, 22, 1144–1152.

Kahneman, D. and A. Tversky (1972): "Subjective probability: A judgment of representativeness," *Cognitive psychology*, 3, 430–454.

Klibanoff, P., M. Marinacci, and S. Mukerji (2005): "A smooth model of decision making under ambiguity," *Econometrica*, 73, 1849–1892.

Kovach, M. (2021): "Ambiguity and partial Bayesian updating," *arXiv preprint arXiv:2102.11429.*

Loomes, G. (1998): "Probabilities vs Money: A Test of some Fundamental Assumptions About Rational Decision Making," *The Economic Journal*, 108(447), 477–489.

Moore, D. and P. J. Healy (2008): "The Trouble with Overconfidence," *Psychological Rev*, 115, 502–517.

Nau, R. F. (2006): "Uncertainty aversion with second-order utilities and probabilities," *Management science*, 52, 136–145.

Oprea, R. (2020): "What makes a rule complex?" *American economic review*, 110, 3913–3951.

Pacheco Pires, C. (2002): "A rule for updating ambiguous beliefs," *Theory and Decision*, 53, 137–152.

Peterson, C. R., W. M. DuCharme, and W. Edwards (1968): "Sampling distributions and probability revisions." *Journal of Experimental Psychology*, 76, 236.

RUBINSTEIN, A. (2002): "Irrational diversification in multiple decision problems," *European Economic Review*, 46(8), 1369–1378.

SCHMEIDLER, D. (1989): "Subjective Probability and Expected Utility without Additivity," *Econometrica*, 57, 571–587.

SCHOTTER, A. AND I. TREVINO (2014): "Belief elicitation in the laboratory," *Annu. Rev. Econ.*, 6, 103–128.

SEO, K. (2009): "Ambiguity and second-order belief," *Econometrica*, 77, 1575–1605.

SULEYMANOV, E. (2025): "Entropy Regularized Belief Reporting," *arXiv preprint arXiv:2506.22649*.

TVERSKY, A. AND D. KOEHLER (1994): "Support theory: A nonextensional representation of subjective probability," *Psychological Review*, 101.

VON MISES, R. (1981): *Probability, statistics, and truth*, Courier Corporation.

WHEELER, G. AND L. R. BEACH (1968): "Subjective sampling distributions and conservatism," *Organizational Behavior and Human Performance*, 3, 36–46.

A Omitted Tables and Figures

Table A.1: Summary Statistics of Reported Initial Beliefs

	Unsure beliefs			Sure beliefs	Total
Prior composition	Share	Mean belief (s.d.)	Mean entropy (s.d.)	Mean belief (s.d.)	Observations
10%	51.17%	16.76 (15.60)	1.11 (0.69)	10.87 (6.96)	383
20%	53.63%	24.13 (11.46)	1.29 (0.72)	21.36 (8.10)	399
30%	56.48%	34.00 (11.43)	1.29 (0.75)	31.47 (8.12)	409
50%	45.36%	51.59 (6.16)	1.37 (0.84)	50.24 (3.95)	388
70%	51.97%	68.30 (8.90)	1.28 (0.75)	68.03 (9.55)	381
80%	49.61%	75.55 (11.94)	1.20 (0.69)	79.12 (7.13)	387
90%	50.13%	83.08 (14.94)	1.08 (0.71)	88.39 (10.91)	389

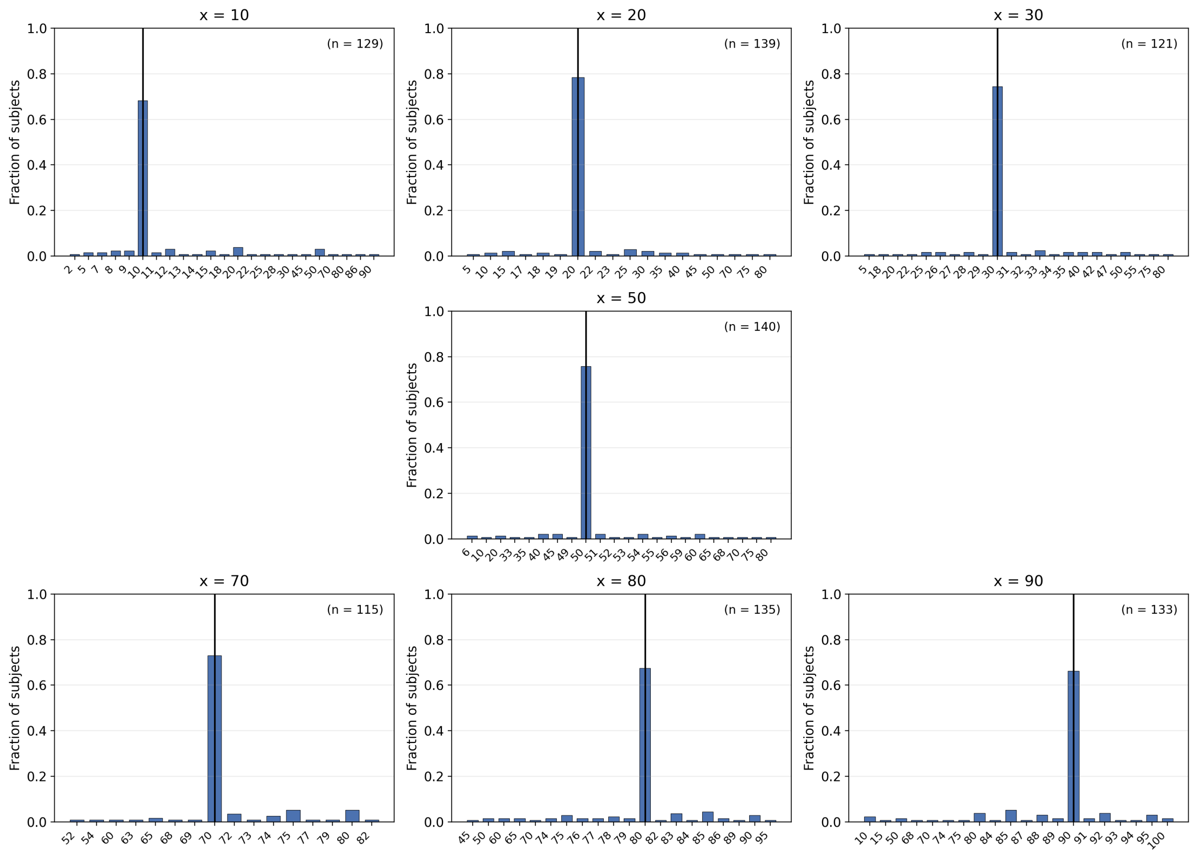
Notes: The table reports summary statistics for initial beliefs pooled across Parts I and II. Prior compositions and belief means are expressed in percentage points, and entropy is measured in bits. Standard deviations are reported in parentheses.

Table A.2: Statistician Question and Deviations from the Bayesian Benchmark

	Indicator for "same degenerate answer"	
	(1)	(2)
Mean absolute updated-belief deviation	-0.0076^{**}	
	(0.0030)	
Mean absolute initial-belief deviation		-0.0174^{***}
		(0.0030)
Constant	0.5160^{***}	0.4655^{***}
	(0.0626)	(0.0306)
Observations	304	304
R^2	0.020	0.103

Notes: Each column reports an OLS regression, with standard errors in parentheses. The dependent variable equals one if the subject reports that all experts would provide the same degenerate answer in the statistician question and zero otherwise. In column (1), the regressor is the subject's mean absolute deviation between their updated mean beliefs and the Bayesian posterior benchmarks, averaged across the six Part II rounds. In column (2), the regressor is the corresponding mean absolute deviation between their initial mean beliefs and the Bayesian prior benchmarks, averaged across the three Part I rounds. All beliefs are expressed in percentage points. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure A.1: Distribution of Single-Best-Guess Reports by Prior Composition



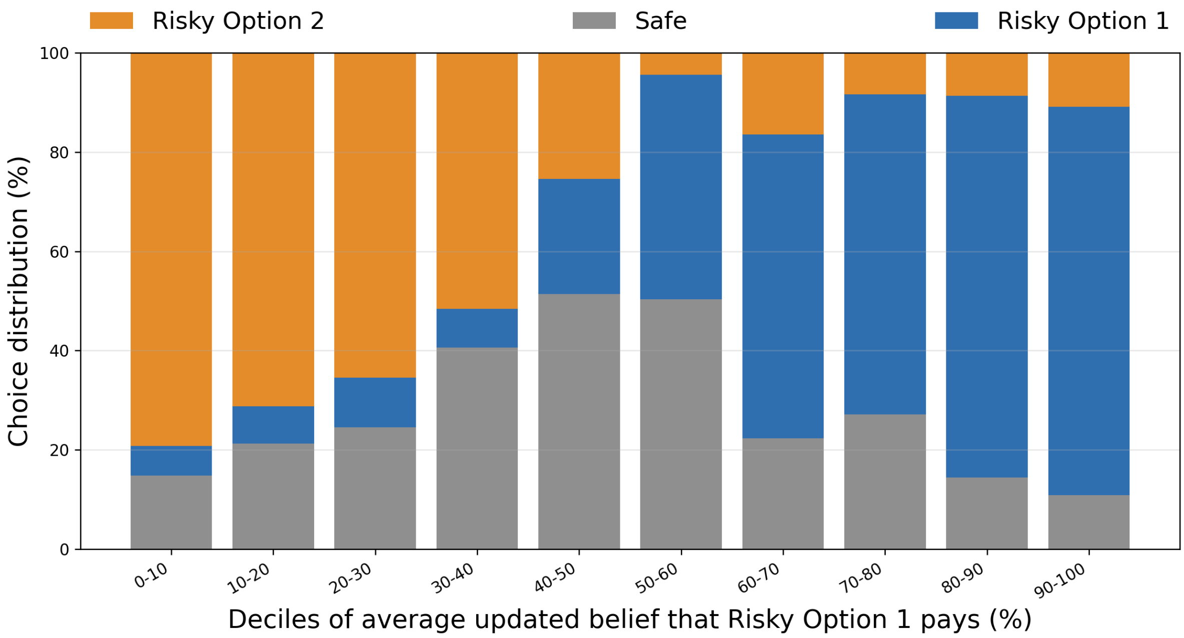
Notes: The figure shows the histograms of subjects' initial beliefs when asked to report a single best guess of the empirical frequency $\hat{X}_N$ as a number between 0 and 100. Each panel corresponds to a prior composition, with the vertical line marking the true prior composition.

Table A.3: Risky Choice with Alternative Certainty-Equivalent Measures

	Indicator for choosing a risky bet		
	(1) 20-ball urn CE	(2) 30-ball urn CE	(3) Average CE
Updated belief entropy	-0.126^{***}	-0.126^{***}	-0.131^{***}
	(0.036)	(0.036)	(0.036)
Updated mean belief	0.006^{***}	0.006^{***}	0.006^{***}
	(0.000)	(0.000)	(0.000)
Updated belief entropy × updated mean belief	0.002^{***}	0.001^{***}	0.002^{***}
	(0.001)	(0.001)	(0.001)
Risk CE	0.002^{**}	0.002^{**}	0.002^{**}
	(0.001)	(0.001)	(0.001)
Ambiguity CE (normalized)	0.000	0.000	−0.000
	(0.001)	(0.001)	(0.001)
Compound CE (normalized)	0.002^{**}	0.002^{**}	0.003^{**}
	(0.001)	(0.001)	(0.001)
Constant	0.142	0.203	0.181
	(0.168)	(0.172)	(0.168)
Additional controls	Yes	Yes	Yes
Observations	1,788	1,806	1,782
Subjects	298	301	297

Notes: Each column reports a linear random-effects regression with a subject-level random intercept; standard errors are in parentheses. The dependent variable indicates whether a risky bet is chosen. The updated mean belief is folded so that higher values correspond to a stronger belief in favor of the risky option. All specifications control for subjects' CRT scores, absolute and relative overconfidence, probability matching, compound-lottery-reduction ability, and the Part II simple-question score. Column (1) uses certainty equivalents (CEs) from the 20-ball urn tasks and restricts the sample to subjects with positive risk, ambiguity, and compound CEs. Column (2) uses the corresponding 30-ball urn CEs and applies the analogous restriction. Column (3) uses the averages of the 20- and 30-ball urn CEs and restricts the sample to subjects with positive CEs in both sets of tasks. Ambiguity and compound CEs are normalized relative to the corresponding risk CE. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure A.2: Choices in Risky and Safe Bets



Notes: The figure groups subject-round observations by the decile of the updated mean belief that risky option 1 will pay. For each decile, it reports the shares choosing the safe bet, risky option 1, and risky option 2. The sample pools all six Part II rounds from the main treatment.

A.1 Robustness Treatments

Table A.4: T1 versus Main Treatment

	Dependent variable		
	(1) Risky choice	(2) Unsure updated belief	(3) Updated-belief entropy
Indicator for T1	−0.020	−0.012	0.068
	(0.029)	(0.045)	(0.081)
Prior × precision × signal FE	Yes	Yes	Yes
Observations	2,706	2,706	1,513
Subjects	451	451	307

Notes: Each column reports an OLS regression, with standard errors clustered at the subject level in parentheses. The dependent variables in columns (1)–(3) are an indicator for choosing a risky bet, an indicator for reporting an unsure updated belief, and updated-belief entropy, respectively. The sample pools the main treatment ($n = 304$) and T1 ($n = 147$). The indicator for T1 equals one for subjects in T1 and zero for subjects in the main treatment. All specifications include prior composition × signal precision × signal realization fixed effects. Column (3) is restricted to rounds with a nondegenerate updated belief. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.5: T2 versus Main Treatment

	Dependent variable			
	Initial belief		Updated belief	
	(1) Unsure	(2) Normalized entropy	(3) Unsure	(4) Normalized entropy
Indicator for T2	0.112^{**}	−0.003	0.058	0.006
	(0.044)	(0.025)	(0.044)	(0.025)
Prior FE	Yes	Yes	No	No
Prior × precision × signal FE	No	No	Yes	Yes
Observations	4,104	2,256	2,736	1,594
Subjects	456	309	456	321

Notes: Each column reports an OLS regression, with standard errors clustered at the subject level in parentheses. The dependent variables in columns (1) and (3) are indicators for reporting an unsure belief. Columns (2) and (4) are restricted to unsure reports and use entropy normalized by the maximum attainable entropy under each partition, $H/\log_2 n$, where $n = 11$ in the main treatment and $n = 20$ in T2. The sample pools the main treatment ($n = 304$) and T2 ($n = 152$). The indicator for T2 equals one for subjects in T2 and zero for subjects in the main treatment. Columns (1) and (2) pool initial beliefs from Parts I and II and include prior-composition fixed effects. Columns (3) and (4) use updated beliefs from Part II and include prior composition × signal precision × signal realization fixed effects. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

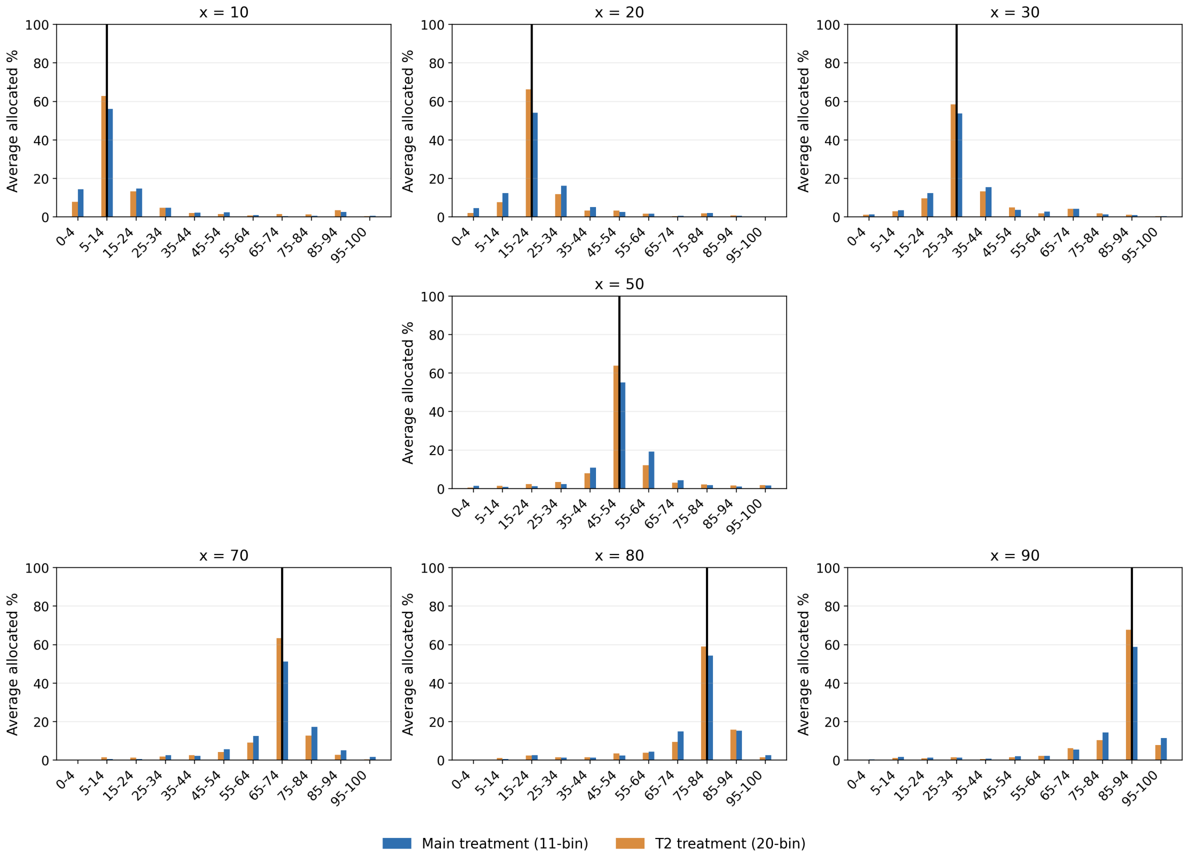
To assess the robustness of our main findings, we replicate the key analyses from the main text separately for each additional treatment. For T1, Table A.6 replicates Table 3. For T2, Figure A.4 replicates Figure 3, Table A.7 replicates Table 1, and Table A.8 replicates Table 3.

Table A.6: Belief Uncertainty and Risky Choice (T1)

	Indicator for choosing a risky bet			
	(1)	(2)	(3)	(4)
Updated belief entropy	-0.083^{***}	-0.089^{*}	-0.083	-0.096^{*}
	(0.025)	(0.053)	(0.053)	(0.054)
Updated mean belief		0.007^{***}	0.007^{***}	0.006^{***}
		(0.001)	(0.001)	(0.001)
Updated belief entropy $\times$ updated mean belief		0.001	0.001	0.001
		(0.001)	(0.001)	(0.001)
Risk CE				0.000
				(0.001)
Ambiguity CE (normalized)				0.000
				(0.001)
Compound CE (normalized)				-0.000
				(0.001)
Constant	0.764^{***}	0.290^{***}	0.632^{***}	0.651^{***}
	(0.029)	(0.054)	(0.236)	(0.253)
Additional controls	No	No	Yes	Yes
Observations	882	882	882	858
Subjects	147	147	147	143

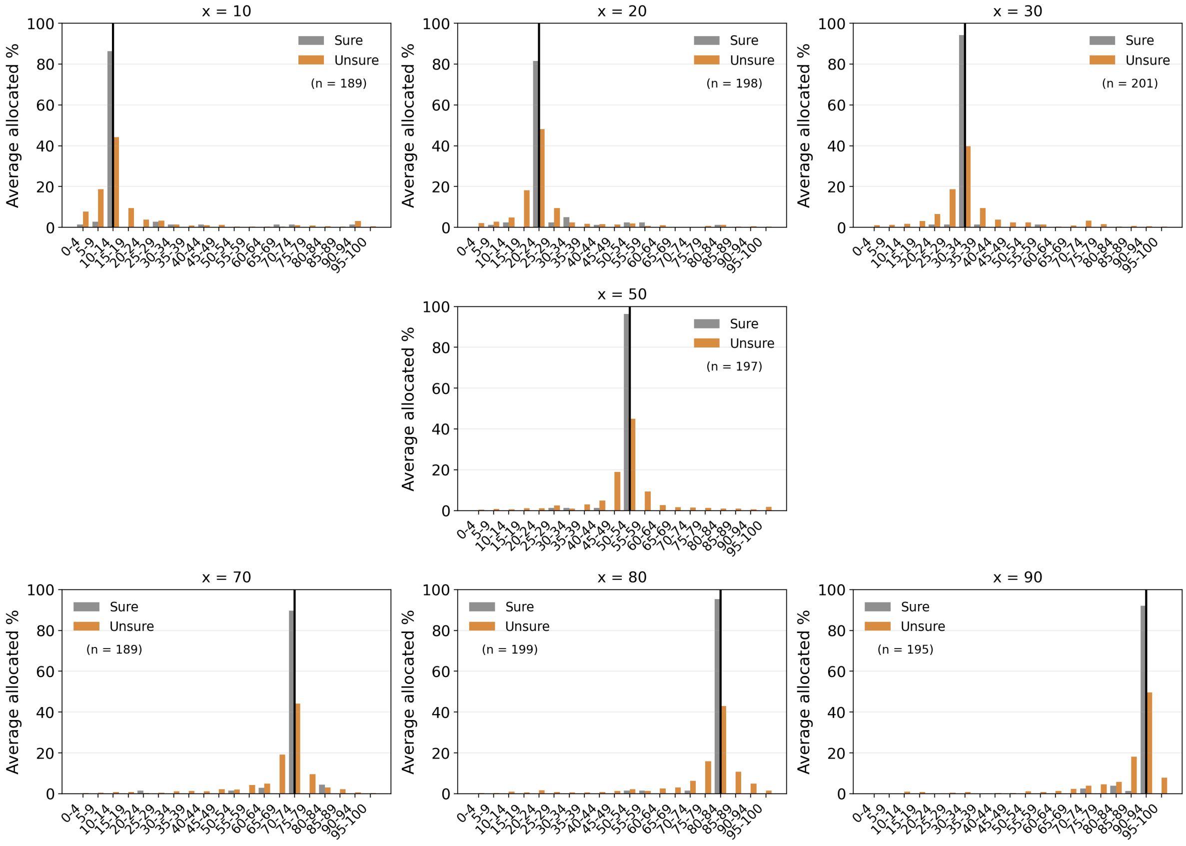
Notes: Each column reports a linear random-effects regression with a subject-level random intercept; standard errors are reported in parentheses. The dependent variable indicates whether a risky bet is chosen. The updated mean belief is defined so that higher values indicate a stronger belief in favor of the risky option. Columns (3) and (4) control for subjects' CRT scores, absolute and relative overconfidence, probability matching, compound-lottery-reduction ability, and the Part II simple-question score. Column (4) additionally includes certainty-equivalent (CE) measures from the 20-ball risk, ambiguity, and compound-lottery urn tasks and excludes subjects with negative CE values. The ambiguity and compound-lottery CEs are normalized relative to the risk CE. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure A.3: Partition Dependence



Notes: The figure shows the aggregate probability assigned to each interval for non-degenerate initial beliefs elicited from Parts I and II, comparing T2 (20-bin reports aggregated to the main treatment's 11-bin partition, $n = 152$) with the main treatment ($n = 304$). Each panel corresponds to a prior composition, with the vertical line marking the interval containing the true prior composition.

Figure A.4: Average Initial Belief Distributions: Sure versus Unsure (T2)



Notes: The figure shows the aggregate probability assigned to each interval for initial beliefs elicited from Parts I and II in T2. Each panel corresponds to a prior composition, with the vertical line marking the interval containing the true prior composition.

Table A.7: How Belief Uncertainty Responds to Different Information (T2)

	Change in unsure-belief indicator			Change in entropy		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Sure Initial Beliefs						
Constant	0.111***	0.116***	0.125***	0.131***	0.128***	0.157***
	(0.030)	(0.036)	(0.034)	(0.036)	(0.041)	(0.047)
Contradicting signal		0.006			0.031	
		(0.036)			(0.049)	
High-precision signal			−0.029			−0.053
			(0.023)			(0.042)
Observations	343	286	343	343	286	343
Subjects	70	67	70	70	67	70
Panel B. Unsure Initial Beliefs						
Constant	−0.072***	−0.074***	−0.054***	−0.045	−0.086*	0.030
	(0.017)	(0.020)	(0.017)	(0.038)	(0.047)	(0.044)
Contradicting signal		−0.002			0.089	
		(0.021)			(0.058)	
High-precision signal			−0.036**			−0.148***
			(0.018)			(0.037)
Observations	569	494	569	569	494	569
Subjects	109	106	109	109	106	109

Notes: Each column reports OLS estimates, with standard errors clustered at the subject level in parentheses. The dependent variables in columns (1)–(3) and (4)–(6) are, respectively, the change in the unsure-belief indicator and the change in entropy from the initial to the updated report. Columns (2) and (5) exclude scenarios with prior composition 50 because contradicting signals are undefined in those scenarios. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.8: Belief Uncertainty and Risky Choice (T2)

	Indicator for choosing a risky bet			
	(1)	(2)	(3)	(4)
Updated belief entropy	-0.073^{***}	-0.083^{**}	-0.073^{**}	-0.080^{**}
	(0.019)	(0.037)	(0.037)	(0.038)
Updated mean belief		0.007^{***}	0.007^{***}	0.006^{***}
		(0.001)	(0.001)	(0.001)
Updated belief entropy $\times$ updated mean belief		0.001	0.001	0.001
		(0.001)	(0.001)	(0.001)
Risk CE				0.001
				(0.001)
Ambiguity CE (normalized)				-0.000
				(0.001)
Compound CE (normalized)				0.001
				(0.001)
Constant	0.800^{***}	0.340^{***}	0.276	0.251
	(0.030)	(0.051)	(0.206)	(0.210)
Additional controls	No	No	Yes	Yes
Observations	912	912	912	882
Subjects	152	152	152	147

Notes: Each column reports a linear random-effects regression with a subject-level random intercept; standard errors are reported in parentheses. The dependent variable indicates whether a risky bet is chosen. Belief distributions in T2 are elicited using the 20-bin partition. The updated mean belief is defined so that higher values indicate a stronger belief in favor of the risky option. Columns (3) and (4) control for subjects' CRT scores, absolute and relative overconfidence, probability matching, compound-lottery-reduction ability, and the Part II simple-question score. Column (4) additionally includes certainty-equivalent (CE) measures from the 20-ball risk, ambiguity, and compound-lottery urn tasks and excludes subjects with negative CE values. The ambiguity and compound-lottery CEs are normalized relative to the risk CE. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.