UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

WORKING PAPERS SES

Compressed Beliefs

Christian Zihlmann

N. 544
I.2026

Faculté des sciences économiques et sociales et du management
Wirtschafts- und sozialwissenschaftliche Fakultät

Compressed Beliefs*

Christian Zihlmann§

January 22, 2026

Abstract

Subjective beliefs are central to economic inference, and incentive-compatible belief elicitation mechanisms are widely assumed to identify these latent objects. This paper shows that elicited belief reports causally depend on an uninformative cognitive default induced by the elicitation design. From the lab to sports betting to official inflation expectations, reported beliefs are highly malleable, even under theoretically and behaviorally compatible incentives. I propose experimentally varying the cognitive default during belief elicitation. This exogenous variation allows the construction of inferred beliefs that are stable across elicitation designs and empirically outperform incentivized reports in predicting realized outcomes and participants' own behavior.

Keywords: Belief Elicitation, Subjective Beliefs, Probabilistic Beliefs, Cognitive Default, Field Evidence

JEL Classification Codes: C81, C90, D81, D83

*I thank David Danz, Holger Herz, Frederik Schwerter, Georg Weizsäcker, and Florian Zimmermann for helpful comments. Financial support from the Swiss National Bank, and the Swiss National Science Foundation, grant no. 222695, is gratefully acknowledged. Confidentiality declaration: This paper uses propriety data of the German central bank, the "Bundesbank-Online-Panel-Households". The results published and the related observations and analysis may not correspond to results or analysis of the data producers.

§Zihlmann: University of Fribourg, Switzerland (email: christian.zihlmann@unifr.ch).

"[W]hat we observe is not nature in itself
but nature exposed to our method of questioning."
Heisenberg (1989, p.25)

1 Introduction

Probabilistic beliefs about how the uncertain future might unfold—how likely it is that the economy will grow, a relationship will last, or a job will be secure—shape some of the most important decisions in life. For this reason, economists have devoted substantial attention to studying such subjective beliefs (see Benjamin, 2019, for a survey).

Yet unlike behavior, subjective beliefs are not directly observable but remain locked in the mind. As researchers, we rely on belief elicitation methods to make these latent objects measurable. But what if the very act of elicitation contaminates what we observe?

This paper shows that elicited belief reports are systematically malleable, even under state-of-the-art theoretically and behaviorally compatible incentives. Specifically, reported beliefs causally depend on an objectively uninformative cognitive default.

The consequences are twofold. (i) Reported beliefs reflect the researcher's implicit or explicit choice in designing the belief elicitation task, which is inconsistent with Manski (2004)'s exogeneity criterion: valid inference requires that the act of measurement does not alter what is being measured. (ii) The common identifying assumption in the belief elicitation literature—that incentive-compatible elicitation mechanisms reveal agents' subjective beliefs—is challenged.[1] Reported beliefs need not coincide with latent subjective beliefs, and incentives alone are insufficient to guarantee their identification.

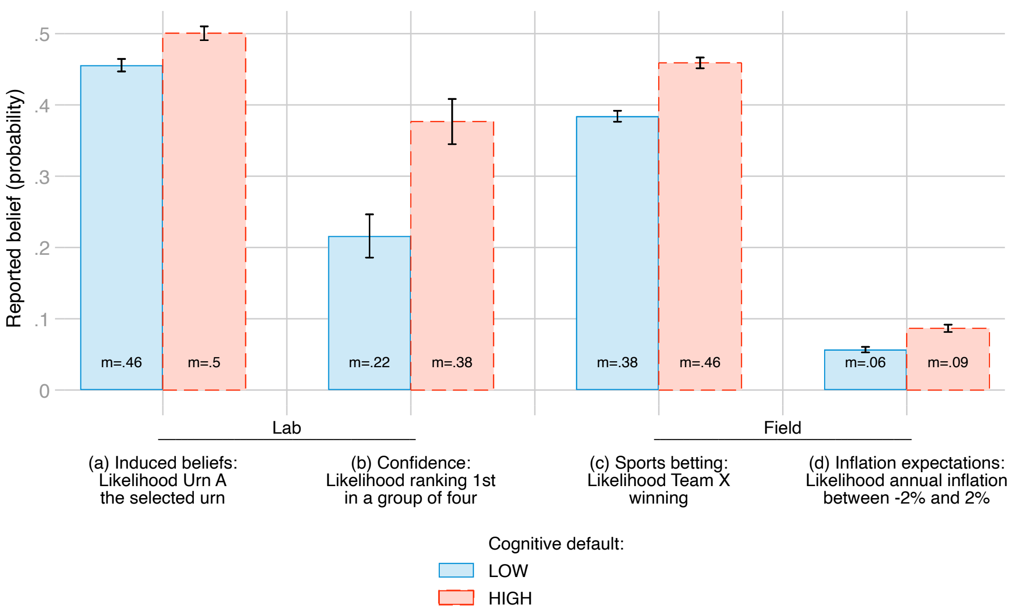
To fix ideas, suppose the observed *reported belief* is a convex combination of a latent subjective *root belief* and a cognitive default induced by the elicitation design. Specifically, assume the cognitive default is represented by an uninformative ignorance prior, which assigns equal probability mass to each category the state space was divided into (e.g., 50-50 in the binary case). A parameter $\alpha \in [0, 1]$ determines the weight placed on the default. This simple model predicts that reported beliefs are contaminated: they are sensitive to variation in the cognitive default, an artifact of the elicitation design.

I demonstrate that exogenously manipulating the cognitive default indeed systematically affects properly incentivized belief reports in different subject pools and across four domains. Reported beliefs are contaminated by the cognitive default in both the

[1] The concern is not whether participants truthfully report their beliefs; I believe appropriate incentives are necessary to address that issue. Rather, even under incentive compatibility and intended truthful reporting, artifacts of the elicitation design itself causally shape belief reports.

lab (the canonical balls-and-urns paradigm and confidence in self-placement) and unique field data (sports bettors' bets and official inflation expectations), see Figure 1. Consider Panel (b), for instance, measuring participants' confidence in self-placement. Incentives are proper, behaviorally compatible, and identical for both groups. Yet, a researcher who elicited an individual's confidence in ranking first versus not in a randomly composed group of four (creating a HIGH cognitive default of 50%) would reach a strikingly different conclusion than a researcher who elicited the probability distribution over all four ranks (creating a LOW cognitive default of 25%). In the HIGH condition, participants report an average likelihood of 38% of ranking first, being highly overconfident about their placement in the group. The average participant in the LOW condition, however, is not overconfident: the average probability of ranking first is statistically indifferent from the objective likelihood of 25%. Both elicitation designs have been used in the literature, yet they lead to starkly different—and even qualitatively opposing—conclusions.

Figure 1: Reported beliefs causally depend on the cognitive default



Note: All panels display the mean of reported beliefs separated by conditions. The LOW condition (solid blue) has a lower cognitive default than the HIGH treatment condition (dashed red). Whiskers indicate robust standard errors in cross-sectional data and in panel data, they are clustered on individual level. Panel (a): $N = 104$, $n = 1248$; Panel (b): $N = 101$; Panel (c): $N = 415$, $n = 1660$; Panel (d): $N = 2477$. In all four panels, the difference in reported beliefs between HIGH and LOW is statistically significant with $p < .001$.

Contamination by the cognitive default means that incentivized belief reports should not be taken at face value since they are a function of an objectively irrelevant object. This turns reported beliefs into a compressed version of subjective root beliefs: the reported belief is an asymptotically biased estimator of the root belief. To draw valid inferences about subjective probabilistic beliefs, we must first undo this contamination.

This paper's second insight is that the very sensitivity of reported beliefs to the cognitive default can be exploited for identification by using experimentation when eliciting beliefs. By exogenously manipulating the cognitive default, we can quantify the magnitude of malleability in belief reports and assess the extent the reported beliefs reveal the latent object they aim to measure, the subjective root beliefs.

Building on this insight, I show that we can construct a better estimator of subjective root beliefs under plausible assumptions, the *inferred belief*. The inferred belief is a more stable object, as it remains robust across elicitation designs that induce different cognitive defaults. Theoretically, unlike the reported belief, the inferred belief is an asymptotically unbiased estimator of the root belief conditional on plausible assumptions. It thus allows for valid inference. Moreover, the inferred belief is at least as good as the reported belief in predicting the root belief regarding linear and quadratic loss.

Empirically, I continue to show that across the four data sets, the inferred belief satisfies an important qualitative criterion: it is stable across elicitation designs that govern the objectively irrelevant cognitive default. The inferred belief also more accurately represents the presumed root belief than the incentivized reported belief does: it is a better predictor of the actual realized state of the world, representing external consistency; and of the individuals' own deterministic beliefs or actions, representing internal consistency.

Revisiting the confidence data discussed earlier, the inferred belief suggests that participants are actually underconfident in their self-placement. This is in contrast to the conclusions drawn from the incentivized belief reports, but it aligns closely with a coarse elicitation of confidence in self-placement, both qualitatively and quantitatively.

The implications of compressed beliefs are manifold. First and foremost, the construct validity of elicited beliefs is violated (Snowberg and Yariv, 2025). Our tool does not (only) measure what it should; it is sensitive to irrelevant manipulations of the elicitation design.

The contamination of reported beliefs by the cognitive default is a potential confounder for well-known errors in probabilistic reasoning. For example, compression towards the cognitive default may partially be responsible for phenomena such as underestimating the probability of likely events and overestimating that of rare events. Therefore, before investigating whether (root) beliefs deviate from normative benchmarks, or whether and how they differ across groups, it is essential to first account for the distortion introduced by the very process of eliciting those beliefs. Compression effects can mask the true pattern of root beliefs, leading to biased inferences.

This paper relates to Danz, Vesterlund and Wilson (2022) ("DVW2022"), Enke and Graeber (2023) ("EG2023") and Ba, Bohren and Imas (2024) ("BBI2024"). DVW2022 show that information on incentives can lead to compressed belief reports. Hence, one way to minimize compression effects is to use behaviorally compatible incentives that do not convey information on the quantitative effects of incentives, a sensible recommendation I adhered to. BBI2024 propose a two-stage model that reconciles under- and overreaction to information. In the first stage, individuals form root beliefs; in the second, noisy cognition causes them to rely partly on a cognitive default given by the ignorance prior, which compresses and attenuates reported beliefs. EG2023 link compressed beliefs to cognitive noise, which is the individual's awareness of being uncertain regarding the probabilistic answer to a given question. EG2023 show that compression towards the cognitive default may partially explain a large set of documented anomalies in probabilistic reasoning, such as base rate insensitivity and conservatism.

My contribution is twofold. (i) Across domains and subject pools, I causally document the consequences of reported beliefs compressing toward the cognitive default—their malleability even under proper incentives. (ii) I propose a method to recover root beliefs from incentivized but contaminated belief reports. Unlike reported beliefs, these inferred beliefs are independent of the specific elicitation design and, under plausible assumptions, asymptotically unbiased in estimating root beliefs.

The paper also relates to research originating in psychology and decision theory, putting forward the idea that probability judgments depend on how the state space is described or represented (Tversky and Koehler, 1994; Fox and Rottenstreich, 2003; Fox and Clemen, 2005; Clemen and Ulu, 2008; Sonnemann, Camerer, Fox and Langer, 2013; Prava, Clemen, Hobbs and Kenney, 2016). Benjamin, Moore and Rabin (2017) study beliefs about random samples and discuss that the partitioning may confound inference. Motivated by these findings, this paper asks whether latent subjective beliefs can be uniquely identified from reported beliefs under incentive-compatible elicitation. It provides causal evidence from laboratory and field settings that reliable identification generally fails, highlights the consequences for economic inference, and proposes a candidate method to recover the latent object. Benjamin (2019) argues that many errors in probabilistic reasoning may be confounded by compression effects, pointing out the need that we must first undo the effects of compression to study other belief biases. This study contributes to this objective by introducing a simple and practical method for inferring root beliefs, which in turn facilitates the study of errors in probabilistic reasoning.

I also relate to the literature on eliciting beliefs and subjective expectations about probabilistic events (Schotter and Trevino, 2014; Schlag, Tremewan and Van der Weele, 2015; Manski, 2018; Charness, Gneezy and Rasocha, 2021; Healy and Leo, 2024). The focus has been on theoretical incentive compatibility and the study of proper scoring rules (Brier et al., 1950; Hossain and Okui, 2013; Holt and Smith, 2016; Wilson and Vespa, 2018). Horse races between different scoring rules are commonly studied (Huck and Weizsäcker, 2002; Rutström and Wilcox, 2009; Andersen, Fountain, Harrison and Rutström, 2014; Trautmann and van de Kuilen, 2015). Recently, scholars have begun to investigate whether scoring rules are also behaviorally compatible (Danz et al., 2022).

This paper suggests that studying the design of the belief elicitation task deserves more attention. Despite theoretically and behaviorally incentive-compatible elicitation, I document systematic shifts in belief reports induced solely by manipulations of the cognitive default. This evidence suggests that task design is a first-order component of behaviorally compatible belief elicitation (Danz, Vesterlund and Wilson, 2024).

The findings also have direct implications for interpreting and designing survey-based expectations. For example, probabilistic inflation expectations are elicited by major central banks, such as the Federal Reserve, the Bank of England, and the European Central Bank. Differences in inflation expectations across time and countries may arise mechanically from design choices rather than underlying shifts in expectations, limiting interpretation and comparability in practice.

Section 2 briefly presents the conceptual framework and derives the hypotheses. The four subsequent sections each provide evidence from a different domain of probabilistic beliefs. In Section 3, I provide evidence from the classical ball-and-urns paradigm (Bayesian likelihoods). Section 4 deals with confidence in self-placement. Section 5 presents field evidence from sports betting, and Section 6 considers official inflation expectations elicited in a representative panel by the German central bank, the Bundesbank.

2 Conceptual Framework

Before moving to the experimental evidence, it is helpful to briefly discuss the underlying conceptual framework and the hypotheses derived from it.

2.1 Reported beliefs

Consider a model in which an agent i reports probabilistic beliefs *as if* they are a mixture of their subjective belief θ_i and a default likelihood d:

$$\tilde{\theta}_i(\theta_i, \alpha_i, d) = (1 - \alpha_i) \cdot \theta_i + \alpha_i \cdot d, \qquad 0 \leq \alpha_i \leq 1, \ 0 < d < 1 \tag{1}$$

where $\tilde{\theta}_i \in [0, 1]$ is the reported likelihood of a probabilistic event. This *reported belief* $\tilde{\theta}_i$ is observed by the researcher and ideally properly incentivized.

Let $\theta_i \in [0, 1]$ be an agent's latent subjective belief about a probabilistic event, it is the belief that people hold in their heads. This latent object of interest may be inaccessible even to the agent itself, for instance due to noisy cognition (see EG2023). Suppose this *root belief* θ_i exists also in absence of its elicitation, it is free of any distortion induced by the elicitation procedure. I remain agnostic how these root beliefs are determined. They may follow objective rules of probability, or be distorted by errors in probabilistic reasoning. Root beliefs represent the latent object we are interested in—we may precisely want to investigate whether root beliefs deviate from objective probabilities.

Let $d \in (0, 1)$ be a scalar that denotes the cognitive default. While multiple factors may simultaneously determine d, I continue to assume that it reflects the ignorance prior that assigns uniform mass to all categories the states of the world were divided into for elicitation. The most prominent case is likely the binary category—the probability that an event happens or not—which yields a cognitive default of 50-50.[2]

Let $\alpha_i \in [0, 1]$ denote the weight an agents' belief report is contaminated by the default probability d. An agent who is not relying on the default but reports their root belief would be characterized with $\alpha_i = 0$. An $\alpha_i > 0$ implies that reported beliefs are a compressed version of the subjective root belief: reported beliefs $\tilde{\theta}$ are too insensitive to variation in root beliefs θ, and at the same time, overly sensitive to variation in the objectively uninformative default d. The reliance on the cognitive default can stem from

[2]Formally, let Ω be the set of possible states of the world. A partition k_A of Ω is a set of mutually exclusive events A, $A \subseteq \Omega$, the state space was divided into. Partitions jointly cover the state space Ω in its entirety. Then, $d = \frac{k_A}{K}$, where K is the total number of partitions, and k_A is the number of partitions that contain the event A in question. For instance, in the binary case when we ask for the likelihood that event A occurs vs. not, $K = 2$ and $k_A = 1$ and hence $d = \frac{1}{2}$.

various factors. I remain agnostic about the sources of the distortion.[3]

Using the reported belief $\tilde{\theta}$ for inference means that we implicitly impose the assumption that all agents place zero weight on the cognitive default, i.e. $\alpha_i = 0$ for all i. Equation 1 can be rewritten as $\tilde{\theta}_i = \theta_i + \alpha_i(d - \theta_i)$, and it becomes clear that the first moment of the population mean $\tilde{\theta}$ is asymptotically biased whenever $\mathbb{E}[\alpha] \neq 0$.[4] Because of that, using reported beliefs as an outcome variable in OLS leads to attenuation bias: $\tilde{\beta}_1 \xrightarrow{plim} (1 - \alpha)\beta_1$.

Prediction 1 is straightforward and highlights the consequences of using the biased estimator, the incentivized reported belief $\tilde{\theta}$. I also refer to Figure 9 in the Appendix for a visualization of Prediction 1 with simulated data. It illustrates, for example, that compressed beliefs can generate overweighting of rare events and underweighting of likely events. Formal statements and proofs are relegated to Appendix A.

Prediction 1 (Reported belief $\tilde{\theta}$)**.**
The mean of reported beliefs depends on the cognitive default d and is asymptotically biased towards $\alpha(d - \theta)$.

2.2 Inferred beliefs

When reported beliefs are a function of the cognitive default, they become dependent on the ignorance prior, itself a function of the elicitation design, specifically the researcher's choice of how to divide the state space into categories. It is precisely this endogeneity that we can leverage to our advantage by using experimentation.

Suppose we run a randomized experiment that exogenously varies the location of the cognitive default d. Let $\tilde{\theta}_i(LOW)$ and $\tilde{\theta}_i(HIGH)$ denote the potential outcomes for agent i under the two experimental groups, the LOW or HIGH cognitive default group. Due to randomization, we expect the mean subjective root belief to be the same in both groups, so $\theta(LOW) = \theta(HIGH) = \theta$. Therefore, the latent object of interest is stable across the two experimental groups. If proper incentives successfully reveal the underlying

[3]Belief compression can result from incentives (Offerman, Sonnemans, Van de Kuilen and Wakker, 2009; Hossain and Okui, 2013; Schlag and van der Weele, 2013), information about incentives (DVW2022, Danz et al., 2024), complexity or cognitive noise (EG2023, BBI2024, Ambuehl and Li, 2018; Khaw, Li and Woodford, 2021; Oprea, 2024; Enke, Graeber, Oprea and Yang, 2025).

[4]The reported belief is an unbiased estimator of θ in only two cases. The first case is when the default d coincides with the root belief θ. In principle, one could design d to equal θ. However, this would require prior knowledge of the latent θ, rendering the task of belief elicitation futile. The second case is when no agent relies on the cognitive default d at all, i.e., $\forall i,\ \alpha_i = 0$. As I will show, this assumption is unrealistic and can be empirically rejected in all four data sets.

root belief, no differences in belief reports across the two experimental groups should be expected—the identifying assumption in belief elicitation. Under this assumption, measuring the same latent object using identical incentives with objectively irrelevant variations of elicitation designs should not lead to systematically different belief reports. However, if beliefs are reported as if they follow Equation 1, reported beliefs in the two groups should differ because $d(LOW) < d(HIGH)$ and thus $\tilde{\theta}(LOW) < \tilde{\theta}(HIGH)$.

The average treatment effect in reported beliefs is identifiable under the standard identification assumptions of randomized controlled trials, $ATE = \mathbb{E}\left[\tilde{\theta}(HIGH) - \tilde{\theta}(LOW)\right]$. Before proceeding, an implicit assumption of the model is critical and worth discussing explicitly here. Equation 1 treats α as orthogonal to the cognitive default d and with it, the experimental group T_i. Formally, we need mean independence to proceed, so $E[\alpha_i|d] = E[\alpha_i]$.[5]

The ATE in reported beliefs allows us to identify the expectation $\mathbb{E}[\alpha]$, and with it, the average magnitude of compression in belief reports. Intuitively, varying the default d varies reported beliefs only because of the location shift in the default d itself. Comparing the location shift in d to the change in $\tilde{\theta}$ allows us to infer $\mathbb{E}[\alpha]$. In a finite sample, exogenously varying the cognitive default d allows us to estimate $\widehat{ATE}$ which in turn helps us to recover $\hat{\alpha} = \frac{\widehat{ATE}}{d(LOW)-d(HIGH)}$.[6] See Appendix A.1 for a more detailed elaboration. An important qualitative test is that the obtained $\hat{\alpha} \in [0, 1]$. We will see that this holds true in all four data sets.

Having access to $\hat{\alpha}$ is immensely helpful. First, it helps us to grasp the extent to which belief reports reveal underlying subjective root beliefs. Thus, we can assess the malleability of reported beliefs. Second, conditional that reported beliefs follow the specification in Equation 1 and the assumptions mentioned earlier, it helps us to construct an estimator of root beliefs that is free of any bias on the aggregate. Impose $\alpha_i = \hat{\alpha}$ for all i, and compute the *inferred belief* $\hat{\theta}_i$ as follows:

$$\hat{\theta}_i := \frac{\tilde{\theta}_i - \hat{\alpha} \cdot d}{1 - \hat{\alpha}} = \frac{(1 - \alpha_i)\theta_i + (\alpha_i - \hat{\alpha})d}{1 - \hat{\alpha}}. \tag{2}$$

On individual level, the inferred belief is not a perfect estimator of the subjective root belief θ_i. It will induce some error—not every individual's α_i is represented well by the

[5] EG2023 and BBI2024 endogeneize α to reflect noisy cognition. Also in these models, α is orthogonal to d, and the evidence in Enke and Graeber (2019) and BBI2024 supports this assumption. My evidence reported in the subsequent sections is also consistent with this assumption.

[6] A simple OLS regression of reported beliefs on a constant (absorbing the term $(1 - \alpha)\theta$) and the cognitive default d that varies exogenously by experimental group estimates $\hat{\alpha}$ through the β_1 coefficient.

sample average.[7]

However, $\hat{\theta}$ matches the first moment of θ asymptotically, conditional on the assumptions of the model described in Equation 1. Thus, $\hat{\theta}$ is an unbiased estimator of θ, unlike the incentivized reported belief $\tilde{\theta}$. Moreover, the inferred belief is independent of the cognitive default and hence, stable across design variations of the elicitation task. Also, the coefficient of $\hat{\beta}_1$ converges to β_1—we can therefore estimate true differences in root beliefs using the inferred beliefs. In addition to unbiasedness, I also show that linear and squared loss are weakly lower for the inferred belief $\hat{\theta}$ compared to the reported belief $\tilde{\theta}$.[8] Formal proofs of Prediction 2 are relegated to Appendix A.

Prediction 2 (Inferred belief $\hat{\theta}$).

a) The mean of the inferred belief is independent of the ignorance prior d and asymptotically unbiased.

b) The inferred belief is at least as good as the reported belief in estimating root beliefs regarding linear and quadratic loss.

2.3 Discussion

Equation 1 adopts a linear specification. The literature often documents inverse S-shaped patterns between reported and objective beliefs (see Benjamin, 2019, for a review): evidence suggests that when objective probabilities approach the extremes (close to 0 or 1), individuals' reported beliefs tend to deviate less from those objective probabilities. This inverse S-shape could arise from root beliefs truly following an inverse S-shape relative to objective benchmarks, which would be unproblematic. Alternatively, it could arise

[7] An obvious alternative is to estimate α_i for each individual. Note that this may be very costly in practice, since it requires a within-subject experiment—something that is probably out-of-scope for many use cases such as central banks eliciting inflation expectations. Moreover, within-subject designs come along with additional assumptions on exclusions restrictions. Finally, whether such inferred beliefs are a better estimator than using reported beliefs simply boils down to the error term in measuring α_i, which is an empirical rather than theoretical question.

[8] The superiority in prediction accuracy in the horse race between the reported beliefs and the inferred beliefs boils down to whether agents are better characterized by the sample-average $\hat{\alpha}$ or by imposing $\alpha_i = 0$ for all i. The winner of this horse race depends on the distribution of α_i solely. I show that it is always the case that the majority of $i's$ are better characterized by $\hat{\alpha}$ than 0 if α_i is uniformly or uni-modal distributed. A higher bar is linear and quadratic loss: $\hat{\theta}$ strictly outperforms $\tilde{\theta}$ in both MAE and MSE when α_i is uni-modally distributed. Assume $\alpha_i \sim Beta(a, b)$, with $a, b \geq 1$, which allows for considerable flexibility in the distributional shape. While we may infer a belief that is closer to θ for the majority of observations, we may correct in the wrong direction for a few other observations. For example, for participants who do not suffer from reliance on the ignorance prior, with $\alpha_i = 0$. A higher bar is thus assessing mean absolute error ("MAE") and mean squared error ("MSE")—particularly the MSE is sensitive to such wrong corrections, since errors are squared. I show in Appendix A that the two estimators perform equally well if α_i is distributed uniformly regarding MAE and MSE. As soon as we move towards a uni-modal distribution, $\hat{\theta}$ strictly outperforms $\tilde{\theta}$ in both MAE and MSE.

from belief compression itself—that is, a tendency to rely less on the cognitive default during belief reporting when root beliefs are extreme. In what follows, I adopt the first interpretation and assume that reported beliefs are compressed towards the cognitive default in a linear fashion, independent of the location of the root belief. Formally, let the weight α be mean independent of the latent root belief, so $\mathbb{E}[\alpha_i|\theta_i] = \mathbb{E}[\alpha_i]$. This makes Equation 1 a deliberately stylized representation. Linear formulations, such as for example the widely used neo-additive weighting function, are often used for their tractability and interpretability, not because they capture all nuances of real-world belief formation. Indeed, this simplicity can be a limitation: the model may well approximate belief reporting over a broad range of probabilities but fail near the extremes, where more flexible functional forms may be necessary. Thus, the framework here is intentionally linear and minimalist, designed to highlight broad patterns rather than provide a fully structural account.

A second assumption is that the cognitive default is well-represented by the ignorance prior, assumed to be common across agents. The cognitive default representing a uniform probability mass across categories is consistent with previous theoretical notions and empirical evidence supporting the view that the ignorance prior serves as an empirically relevant cognitive default (Enke and Graeber (2019), BBI2024).

Several diagnostic checks can be implemented to assess whether Equation 1 and its underlying assumptions are approximately valid. First, manipulating $d(T)$ should shift the mean of the reported belief $\tilde{\theta}$, but not its variance. Second, the inferred belief $\hat{\theta}$ should be independent of the treatment condition $d(T)$. Third, regressing $\tilde{\theta}$ on $\hat{\theta}$, the treatment $d(T)$, and their interaction should yield a stable slope, that is, the interaction term should be insignificant. Finally, regressing $\tilde{\theta}$ on the treatment and a proxy for θ, along with their interaction, should also show no significant interaction term, indicating stable slopes across treatment conditions. In the remainder of the paper, I will revisit those checks.

2.4 Hypotheses

Based on the framework presented, I formulate the following two hypotheses to be tested in the experiments reported in the next sections. Hypothesis 1 was pre-registered for all four experiments, and Hypothesis 2 for all experiments except the one reported in Section 5; see each study's pre-registration link for details.

Hypothesis 1. *The reported belief $\tilde{\theta}$ is on average higher when the cognitive default d is exogenously larger.*

Hypothesis 2. *The inferred belief $\hat{\theta}$ is better aligning with the presumed root belief, and its performance regarding linear and squared loss is at least as good as the performance of the reported belief $\tilde{\theta}$.*

3 Bayesian Beliefs

The experimental design and the hypothesis presented in this section were pre-registered prior to data collection on aspredicted.org (ID 213187), and approved by the IRB of the University of Fribourg, Switzerland, Ref. 2024-06-05.

3.1 The Experiment

The first study employs a workhorse paradigm of the literature, the ball-and-urn task, which is frequently used to induce probabilistic beliefs (Schlag et al., 2015). An advantage of the ball-and-urn task is that the objective data-generating process is known—the implied normative benchmark is well defined and adheres to Bayes' Rule.

There are two urns, A and B, both containing 10 balls, either red or blue. The computer selects one of the two urns by a pre-defined distribution (the base rate). It remains unknown which urn was selected, but the computer randomly draws a ball (the signal) from the selected urn. The key parameters in this task is the base rate ($b \in 20, 40, 60, 80$), which was implemented as a fair 10-sided die roll, and the signal diagnosticity ($q \in 30, 70$) of the ball drawn of the selected urn. The die roll and the random draw of the signal were randomized by the computer before the first session took place.

Participants are then asked to state a probabilistic guess that Urn A is the selected urn. Once a likelihood for Urn A was entered (but not yet confirmed) by participants, the computer instantly and automatically showed the corresponding probability that Urn A was not selected. See Figure 2 for screen shots.

Each participant completed the ball-and-urn task under behaviorally compatible incentives: they were incentivized by a binarized scoring rule that would earn them either CHF 8 or nothing (Hossain and Okui, 2013). Yet, only qualitative information was provided, since this has been shown to minimize distorted reporting because of (information on) incentives, see DVW2022.[9] After participants submitted their probabilistic belief in a problem, I elicited their cognitive uncertainty ("CU") using the same wording as EG2023.[10] The elicitation of CU was not incentivized.

3.1.1 The two treatment conditions

The two different conditions exogenously varied the cognitive default, as shown in Figure 2. Participants assigned to the HIGH condition faced two urns, Urn A and Urn B. The LOW condition is identical except that the base rate probability mass previously assigned to Urn B is now divided into two equal components, Urn B and Urn C, as shown in Figure 2b.[11] Objectively, the likelihood that Urn A was selected is exactly the same in both treatment conditions—the base rate as well as the signal diagnosticity are exactly identical in both conditions. Importantly, also incentives are identical in both treatment conditions: only the belief report on Urn A was incentivized. Yet, in condition LOW, the cognitive default of Urn A being the selected urn is $d_{LOW} = \frac{1}{3}$ compared to $d_{HIGH} = \frac{1}{2}$ in the HIGH condition.

3.1.2 Procedures

The experiment was conducted in February and March 2025 on-site at FriLab, the laboratory of the University of Fribourg, Switzerland. The average payout including the show-up fee was CHF 23, and the average duration was about 50 minutes. A total of 105 participants participated. One participant will be excluded, adhering to the pre-registered exclusion criteria, because they reported a belief that perfectly matched the statistically correct likelihood for all problems.

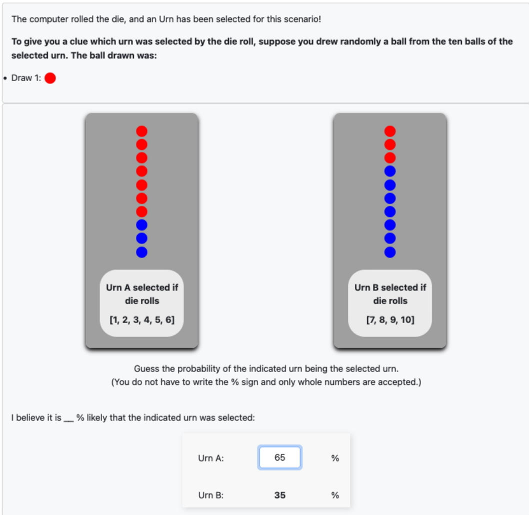
[9] Participants were truthfully informed that *The payment rule is designed so that you can secure the largest expected earnings by reporting your most-accurate guess. The precise payment rule details are available on request.*

[10] *Your decision on the previous screen indicates that you believe there is an x % chance that Urn A was selected. How certain are you that the statistically correct likelihood that Urn A was selected is actually somewhere between $(x-1)$ % and $(x+1)$ %?* Participants could move a slider with no default position from 0 (very uncertain) to 100 (very certain). Participants received an explanation that one can compute a statistically correct likelihood, using the laws of probability based on Bayes's Rule, that does not rely on information that participants do not have.

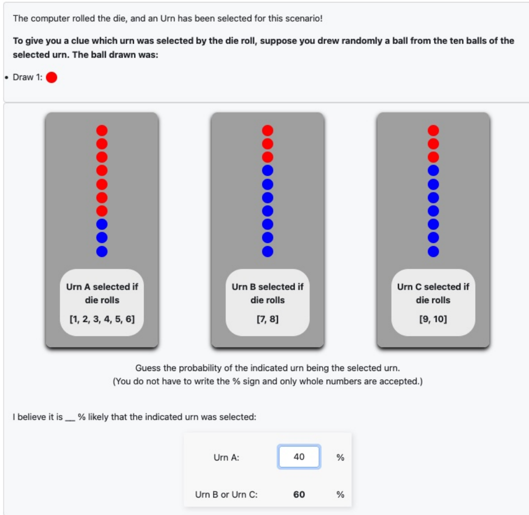
[11] Each participant faced six different ball-and-urn problems, once in condition LOW, and once in condition HIGH. In total, participants completed 12 ball-and-urn problems. Whether participants faced first six times the condition LOW and then HIGH, or vice versa, was randomly determined. The order of the six different ball-and-urn problems were randomly determined for each participant within each condition.

Figure 2: Experimental design ball-and-urns

(a) HIGH cognitive default



(b) LOW cognitive default



Note: Both panels show the same ball-and-urn problem, once in condition LOW (right panel) and once in HIGH (left panel). Participants are incentivized to guess the likelihood that Urn A being the selected urn. The left panel displays condition HIGH, in which there were two urns, generating an cognitive default that Urn A is the selected Urn of 50%. The right panel displays condition LOW, in which Urn A is exactly identical, but the former Urn B was divided into two identical sub-urns B and C. While this does not vary the likelihood that Urn A being the selected Urn, it does vary the cognitive default to 33%. The corresponding likelihood that Urn A was not selected was computed automatically and dynamically by the computer.

3.1.3 Replicating EG2023

Before turning to the results, I begin with benchmarking my data to EG2023 and test whether their key findings replicate in this subject pool.

One can estimate Equation 1 by regressing the reported beliefs on the Bayesian beliefs, the self-reported measure of cognitive uncertainty, and their interaction. This assumes that the measure of self-reported cognitive uncertainty represents to some degree the total size of compression towards the default α_i.[12] For cognitively uncertain participants, who presumably rely more strongly on the cognitive default, we should find (i) higher intercepts (ii) and lower sensitivity to the induced Bayesian belief.

I replicate this key insight of EG2023: the higher a participant's self-reported cognitive uncertainty, the stronger their reported belief is contaminated by the cognitive default d, and the lower the reported belief's sensitivity to the induced Bayesian belief. Hence, cognitive uncertainty is associated with α and predicts the degree of belief compression towards the cognitive default. See Table 5 in the Appendix for more details.

Second, self-reported cognitive uncertainty is independent of the cognitive default,

[12] Note that i needs to be aware of their cognitive noise in order to report uncertainty.

see Table 6 in the Appendix. This also aligns with the model in EG2023 and evidence reported in Enke and Graeber (2019). On average, cognitive uncertainty does not differ in the two treatment conditions.

Third, cognitive uncertainty is uni-modally distributed, refer to Figure 11 in the Appendix. The distribution matters for the inferred belief's performance regarding MAE and MSE, see Prediction 2 formalized in Proposition 3, which assumes that the distribution of α must not be bi-modally distributed.

To sum up, my data is fully consistent with EG2023: compressed beliefs are associated with cognitive noise.

3.2 Results

3.2.1 Reported beliefs

Result 3.1. *The incentivized reported belief depends on the cognitive default: When exposed to the HIGH default, participants report a higher likelihood for Urn A than when exposed to the LOW default.*

Evidence for Result 3.1 is displayed in the left panel of Figure 3a, which illustrates the key prediction of the model described in Equation 1 and formulated as Prediction 1: reported beliefs are a function of the cognitive default.

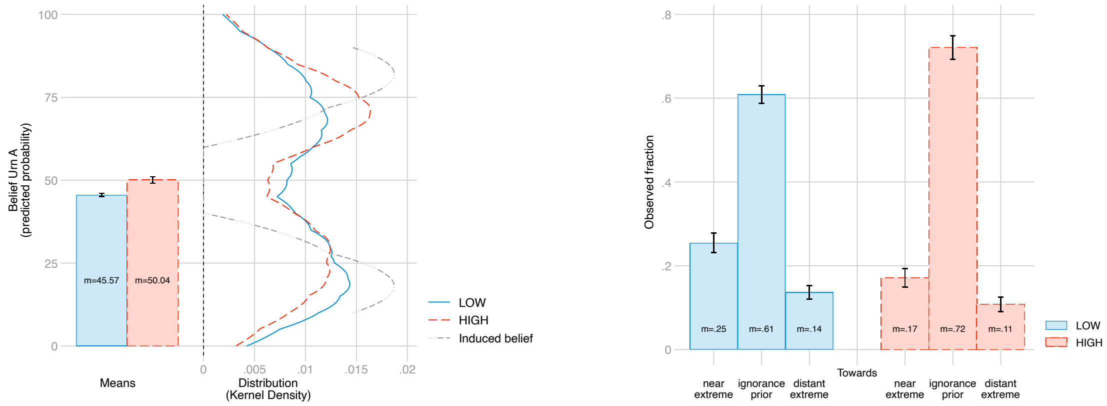
The cognitive default causally and substantially influences reported beliefs, even tough they were elicited with a proper scoring rule in a behaviorally compatible way. The average participant, over all problems, believes that Urn A is the selected urn with a likelihood of 45% when facing three urns (condition LOW with a cognitive default of $d_{LOW} = \frac{1}{3}$). This likelihood rises to 50% in condition HIGH in which participants were confronted with two urns, generating a default of $d_{HIGH} = \frac{1}{2}$. The difference is highly significant ($p < .001$), and remains at that significance level when controlling for problem and time fixed-effects (see Table 7 in the Appendix).

Figure 3a also displays the distributions. Visually, we observe that reported beliefs are too compressed: compared to the induced Bayesian belief, there is too much mass in the center of the probability range—reported beliefs exhibit too little variance. Moreover, the entire distribution in the LOW condition is shifted downward compared to the HIGH condition. It places less mass on higher probability values and more mass around lower probabilities; reported beliefs are skewed towards the cognitive default. As predicted by the model, the variance does not differ by treatment conditions (Levene's test, $p = .764$).

Figure 3: Compressed beliefs in the lab

(a) Sensitivity of reported beliefs to variation in the cognitive default

(b) Direction of the deviation of reported beliefs from induced Bayesian beliefs



Note: The left panel displays the mean of reported beliefs in the LOW treatment condition (solid blue) and the HIGH treatment condition (dashed red), along with whiskers indicating robust standard errors clustered on individual level. It also plots the distribution of the two reported beliefs along the distribution of the induced Bayesian beliefs (short-dashed gray). All distributions are kernel density estimates using the optimal default bandwidth. For reported beliefs that are not equal to the Bayesian belief, the right panel displays the frequency of the nature of the deviation. A reported belief is compressed towards the cognitive default (given by ignorance prior) if the belief lies between the Bayesian belief and the ignorance prior, i.e. $\tilde{\theta} \in (\theta, d]$ when $\theta < d$ or $\tilde{\theta} \in [d, \theta)$ when $\theta > d$. Beliefs that are not compressed towards the ignorance prior are classified as moving towards the nearest extreme if $\tilde{\theta} \in [0, \theta)$ when $\theta < d$ or $\tilde{\theta} \in (\theta, 1]$ when $\theta > d$; and as moving towards the distant extreme if $\tilde{\theta} \in (d, 1]$ when $\theta < d$ or $\tilde{\theta} \in [0, d)$ when $\theta > d$. Whiskers indicate robust standard errors clustered on individual level.

To see how reported beliefs differ from Bayesian beliefs, it is helpful to analyze the direction of the deviation. Compression towards the cognitive default as outlined in Equation 1 posits that reported beliefs lie between the root belief θ and the cognitive default d. Accordingly, reported beliefs are classified as moving towards the cognitive default if $\tilde{\theta} \in (\theta, d]$ when $\theta < d$ or $\tilde{\theta} \in [d, \theta)$ when $\theta > d$. Reported beliefs that are not in between θ and the default d are classified as moving towards the near extreme if the reported belief falls between θ and the nearer end of the scale, and as moving towards the distant extreme if beliefs move beyond d, toward the far end of the scale opposite to θ. Take for instance a Bayesian belief of 22%. Reported beliefs that fall between 0% and 22% would be classified as moving to the near extreme, reports that fall between 22% and the cognitive default of 33.3% as moving towards the cognitive default, and reports larger than 33.3% as moving towards the distant extreme point. Those classification rules align with DVW2022 and mimic their analysis.

Figure 3b illustrates the results. Reported beliefs that are moving towards the cognitive default are by far the most frequent category: In both conditions, individuals are significantly more likely to report a belief that is consistent with compression effects than not ($p < .001$).[13] In both conditions, the largest fraction of reported beliefs are consistent with compression towards the cognitive default.

3.2.2 Inferred beliefs

The exogenous variation in d helps us to estimate the magnitude to which reported beliefs are malleable, without the need to know or presume the latent θ. Using the exogenous variation in the cognitive default d alone, as outlined in Section 2, leads to an estimate of $\hat{\alpha} = .26$. This implies that the reported belief comprises only three-quarters of the root belief. One quarter represents a systematic distortion that depends on the elicitation task. I continue by computing the inferred belief as specified in Equation 2.

An important test is whether the inferred belief is still a function of the objectively irrelevant cognitive default, as the reported belief is. This may be the case if the model in Equation 1 is not an accurate enough way of describing reported beliefs. Table 10 in the Appendix shows that while the observed reported belief is sensitive to the cognitive default, the inferred belief is not: inferred beliefs do not depend on the treatment condition ($p = .897$), as predicted by Prediction 2.[14] Thus, inferred beliefs are not contaminated by design artifacts of the elicitation task.

Figure 4a analyzes whether the reported belief or the inferred belief lie closer to the induced Bayesian belief. Observations are classified into "Yes" if the inferred belief's absolute distance to θ is lower or equal than the reported belief's absolute distance. This will be the case for any agent i who is better characterized by $\alpha_i = \hat{\alpha}$ than by $\alpha_i = 0$. Figure 4a shows that for roughly 80 percent of observations, the inferred belief is closer to the induced Bayesian belief than the reported belief.[15] That fraction is about equal in both treatment conditions, and a test of the equality of proportions rejects the null for both conditions ($p < .001$).

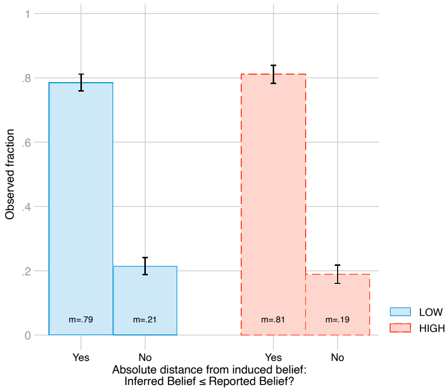
[13] Note that comparisons across the two conditions is not helpful since the ranges of the three categories vary due to a different d, and hence, mechanically, may produce different results.

[14] Also, regressing reported beliefs on the treatment condition, inferred beliefs, and its interaction reveals that the slope is stable across conditions ($p = .402$), suggesting α to work uniformly across conditions.

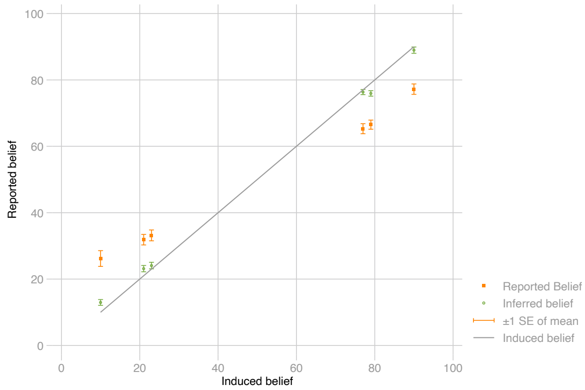
[15] Rather than just a binary classification, we can also assess the magnitude of the linear and squared loss—MAE and MSE—of the inferred and reported belief in predicting the latent truth θ. Table 11 in the Appendix reports the results and shows that the inferred belief does not perform statistically different than the reported belief, for both linear and squared loss, being in line with Prediction 2. Actually, when using an inclusion criterion commonly used in the literature when comparing losses (see Danz et al., 2024; Hossain and Okui, 2013), the betweenness criterion that excludes observations that are likely caused by simple mistakes (for example, due to confusing the probability of Urn A vs not Urn A and hence, reporting 24% instead of 74%), the inferred belief performs better than the reported belief and at a statistically significant level, both in linear and squared loss.

Figure 4: Reported vs. inferred beliefs

(a) Whether inferred belief is closer to the induced Bayesian belief



(b) Means



Note: The left panels shows, separate by treatment condition, whether the inferred belief is closer to the Bayesian belief compared to the reported belief. The right panel shows, for each problem, the mean reported belief for treatment condition HIGH, the standard design used in the literature with two urns, represented in yellow squares. The inferred belief is represented by a green circle. Whiskers indicate standard errors. For better visualization, the Bayesian beliefs in problems 2 and 3, and in problems 4 and 5, were adjusted by ±1% to ensure they are sufficiently separated. The solid gray 45-degree line represents the points where reported beliefs equal the Bayesian beliefs.

Result 3.2. *The inferred belief is closer to the induced Bayesian belief than the reported belief.*

Turning toward the implications, Figure 4b mimics a common visualization in the literature on errors in probabilistic reasoning: the means of belief reports are plotted against the Bayesian belief. The standard elicitation design used in the literature is with two urns, so the condition HIGH is plotted in yellow squares. It shows large deviations from induced Bayesian beliefs, and the direction is consistent with the previous literature (Benjamin, 2019). Figure 4b thus visualizes the insensitivity of reported beliefs to the correct likelihood.

For low probability ranges (or, alternatively, when the root beliefs are lower than the cognitive default), people appear to believe that the event is more likely than it actually is. The Bayesian likelihood is overestimated. For higher probability ranges (or, alternatively, once the root belief is larger than the cognitive default), the opposite happens: people underestimate the Bayesian likelihood. As seen previously, such a pattern is well-documented in the literature, and often attributed to people overestimating the likelihood of rare events and underestimating the likelihood of probable events.

However, the inferred belief—a theoretically unbiased estimator of the latent root belief—is much closer to the induced Bayesian belief (green circles). The overweighting of small probabilities and underweighting of large probabilities partially vanishes.

Thus, this evidence offers an alternative explanation for the pattern attributed to biased probabilistic reasoning: reported beliefs are attenuated relative to the induced Bayesian beliefs because of compression towards the cognitive default d. Compressed beliefs may be partially responsible for well-known errors in probabilistic reasoning, aligning with the arguments and evidence put forward in EG2023 and BBI2024.

3.3 Implications: Grether decomposition

To further examine the implication of compression effects, I turn to Grether decompositions (Grether, 1980) that investigate deviations from objective Bayesian beliefs by generating a linear relationship between people's beliefs π, the objective likelihood ratio, and the objective prior odds, see Equation 3. $p()$ refers to objective correct probabilities, $\pi()$ refers to a person's belief—in this exercise, either the reported belief $\tilde{\theta}$ or the inferred belief $\hat{\theta}$.

$$\ln\left(\frac{\pi(A|S)}{\pi(\neg A|S)}\right) = \gamma \ln\left(\frac{p(S|A)}{p(S|\neg A)}\right) + \delta \ln\left(\frac{p(A)}{p(\neg A)}\right). \tag{3}$$

Those decompositions shall measure the sensitivity to both the likelihood ratio, captured by the parameter γ, and the base rate, captured by δ. A Bayesian would express full sensitivity to both the likelihood ratio and the prior odds, and hence, $\gamma = \delta = 1$. Overreaction is present if $\gamma, \delta > 1$. The canonical finding is underreaction, often referred to as underinference from signals and base rate neglect, and identified when $\gamma, \delta < 1$.

However, if people report compressed beliefs as in Equation 1, then mechanically we would observe that both parameters $\gamma, \delta < 1$. Intuitively, because participants report a belief that responds to variation to the induced Bayesian belief θ only with weight $(1-\alpha)$, compression effects automatically generate insensitivity to both the likelihood ratio and the base rate. Suppose a subject that fully relies on the cognitive default $d = .5$. Then, regardless of the Bayesian belief θ, $\tilde{\theta}$ will always be 50%, and hence, $\gamma, \delta = 0$.

Thus, the prediction is that compression effects account for some of the attenuation in $\gamma, \delta < 1$. Inferred beliefs, therefore, should yield larger values of γ, δ. To test this predictions empirically, I run Grether regressions using reported beliefs and inferred beliefs. The results are displayed in Table 12 in the Appendix. When using reported beliefs, both coefficients are significantly smaller than 1 ($\gamma_{\tilde{\theta}} = .66, \delta_{\tilde{\theta}} = .60$, both $p < .001$). We would conclude that people underinfer from signals and exhibit base rate neglect.

When using inferred beliefs that are free from compression towards the cognitive default, we would come to a more nuanced conclusion: both coefficients are higher in magnitude, indicating less insensitivity to the likelihood ratio and base rate, and statistically indistinguishable from 1 ($\gamma_{\hat{\theta}} = .87$ with $p = .246$,$\delta_{\hat{\theta}} = 1.04$ with $p = .715$).

Therefore, compression towards the cognitive default can generate attenuated γ, δ-parameters in Grether regressions.

4 Confidence in Self-Placement

The experimental design and the hypothesis presented in this section were pre-registered prior to data collection on aspredicted.org (ID 213187), and approved by the IRB of the University of Fribourg, Switzerland, Ref. 2024-06-05.

4.1 The Experiment

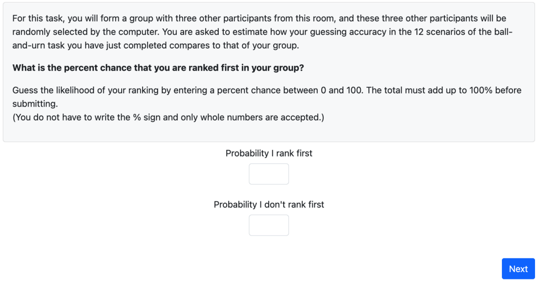
The experiment discussed here followed as a separate part after participants completed all 12 ball-and-urn rounds, but before seeing their payoff and realizations. Participants were incentivized to guess the probability that they would rank first out of four randomly selected participants in the same session. This mimics a standard design used to assess confidence (Niederle and Vesterlund, 2007; Exley and Nielsen, 2024). The task asks participants to rank themselves relative to their group, which measures confidence in self-placement.

Participants were randomly divided into two groups. The elicitation design for group HIGH was: "What is the percent chance that you are ranked first in your group?" Participants needed to enter a probability for ranking first versus not ranking first that needed to sum up to 100%, creating a default of $d_{HIGH} = 50\%$. In group LOW, participants were asked "What is the percent chance that you are ranked first, second, third, or fourth in your group?" They needed to enter a probability for each rank, with the requirement that probabilities sum up to 100%. This creates a default in LOW of $d_{LOW} = 25\%$, see Figure 5. Participants were properly incentivized in a behaviorally compatible way. Importantly, in both groups, only the belief that they would rank first was payoff-relevant, and participants were explicitly informed about this.

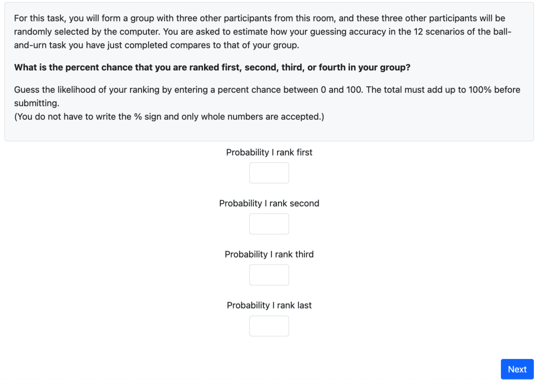
After eliciting probabilistic beliefs of ranking first, participants were asked to pick

Figure 5: Experimental design

(a) HIGH cognitive default



(b) LOW cognitive default



Note: The left panel displays the condition with a HIGH cognitive default, the right panel the condition with a LOW cognitive default. Participants were randomly assigned to conditions. Only the guess for ranking first was incentivized to hold incentives constant across the two conditions.

their modal rank: "What do you think is your most likely rank within your randomly selected group?" This coarse elicitation was not incentivized.

4.2 Results

4.2.1 Reported beliefs

Figure 6 displays participants' confidence in ranking first. Exogenously varying the cognitive default has a substantial and highly significant effect on expressed confidence between the two groups (t test: $p < .001$).

Result 4.1. *The incentivized reported belief to rank first depends on the cognitive default: Participants in HIGH express more confidence in self-placement than in LOW.*

Both elicitation designs were used in the literature, and it matters. A researcher who uses the binary partition to elicit belief reports about ranking first versus not ranking first would conclude that, on average, participants are overconfident. Specifically, the average participant in HIGH reports the likelihood to rank first to be 37.7%, which is substantially and significantly above the rational benchmark of 25% (t test against theoretical value of .25: $p < .001$).

A researcher eliciting beliefs over all ranks would conclude differently: The average participant in LOW reports a 21.6% likelihood of ranking first—a significant difference from the 37.7% in the HIGH condition ($p < .001$).[16] Yet, the average likelihood of 21.6% in

[16] Note that as expected (due to random assignment), the participants' performance and with it, the

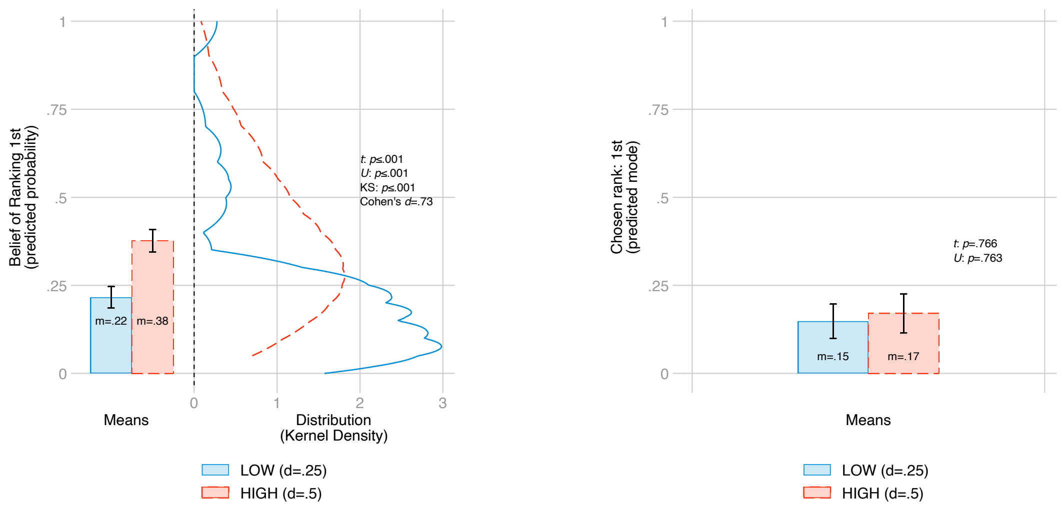
LOW does not significantly differ from the benchmark of 25% (t test against theoretical value of .25: $p = .270$). Thus, we would conclude that LOW participants are neither under- nor overconfident.

We should not take reported beliefs at face value, as we have seen earlier. Neither reported belief, whether elicited in condition LOW or HIGH, accurately reflects the root belief according to the model on compressed beliefs.

Figure 6: Confidence in self-placement

(a) Probability of ranking first LOW vs. HIGH

(b) Most likely rank: 1st (no treatment groups)



Note: Figure 6a displays the means of the predicted probability of the ranking first by experimental group, along with kernel density estimates that show the distribution of the predicted probabilities. Condition LOW in solid blue faced $d_{LOW} = .25$, condition HIGH in dashed red $d_{HIGH} = .5$ Whiskers indicate standard errors. Reported statistics are the p value of a two-sided Welch's unequal variance t test, the two-sided Mann-Whitney U test, a Kolmogorov-Smirnov equality-of-distributions test, and the effect size is expressed as Cohen's d. Figure 6a displays the means of the coarse elicitation of ranking first that followed afterwards, in which participants needed to pick their most likely rank in their group of four. Whiskers indicate standard errors. $N = 101$.

4.2.2 Inferred beliefs

Using the exogenous variation in d to estimate $\hat{\alpha}$ suggests that the reported belief of the average participant relies on the cognitive default with a weight of 64%. The larger share of the observed reported belief is simply a systematic distortion. Root beliefs only contribute 36% to the elicited belief reports. As previously, I infer the root beliefs of ranking first using $\hat{\alpha}$ as described in Equation 2.[17] A first diagnostic check is fine: the

likelihood to rank first, does not differ across the two groups. Also, the variance of reported beliefs does not differ across the two conditions ($p = .470$).

[17] Tests reveal that α does not seem to depend on the experimental group, and hence, d. The magnitude of CU does not depend on which elicitation design participants faced ($p = .722$), and regressing the

inferred belief's mean is independent of the treatment condition and with it, the cognitive default ($p \approx 1$), and so is the variance ($p = .469$).

The inferred belief of ranking first suggests that, on average, participants assign a 15.5% likelihood to ranking first—much lower than their reported beliefs suggest. This indicates that participants are actually underconfident regarding their relative standing, aligning with previous literature using a coarse elicitation and documenting underplacement in difficult tasks (Hoelzl and Rustichini, 2005; Moore and Healy, 2008). The qualitative conclusion of the inferred belief, namely that the average participant is underconfident in their self-placement, also aligns with the model's qualitative prediction: if $d_{LOW} = 25\%$ and the average participant's reported belief is below 25%, Equation 1 predicts that the average root belief must be lower than 25% since reported beliefs move towards the cognitive default. In other words, reported beliefs must somewhere lie between the root belief and the cognitive default.

The subjective root belief θ_i remains latent, but we can use an individual proxy for it. After the probabilistic elicitation, participants were tasked with a coarse elicitation of guessing their most likely rank in their group of four. The mode rank may seem less informative, yet, it is argued that a coarse elicitation may better reflect participants' subjective belief because it is more natural to answer (Haaland, Roth and Wohlfart, 2023; Healy and Leo, 2024). In any case, it helps us to assess how strongly the probabilistic belief elicitation aligns with the coarse belief elicitation. There were no treatment conditions in the modal elicitation, and hence, we should not expect any treatment differences—under the assumption that there is no spill-over from the previous probabilistic belief elicitation, supported by evidence showing there are no significant differences among the two experimental groups in the coarse elicitation.

Result 4.2. *The inferred belief is internally more consistent: it weakly outperforms the reported belief in predicting the coarse elicitation of ranking first.*

Figure 6b shows that on average, 15.8% of participants indicated that they believe they will most likely rank first in their group of four. This coarse elicitation is qualitatively (and quantitatively) aligning with the inferred belief's conclusion: participants exhibit underconfidence in relative placement. The inferred beliefs much better align with the modal elicitation. The tables in Appendix C provide further evidence for this: the inferred

reported beliefs on a constant and on the mode belief as a proxy for θ, I find that the slope (the term $1 - \alpha$) is not different between the two groups ($p = .636$). Also, the regressing $\tilde{\theta}$ on $\hat{\theta}$, the treatment condition and their interaction reveals a non-different slope ($p = .440$) as predicted by the model.

belief weakly outperforms the incentivized reported belief in predicting the coarse belief regarding linear and quadratic loss and AIC and BIC, aligning with Prediction 2.

5 Sports Betting

The experimental design and the hypothesis presented in this section were pre-registered prior to data collection on aspredicted.org (ID 110583), and approved by the IRB of the University of Fribourg, Switzerland, Ref. 2022-10-02.

5.1 The Experiment

I was working with a non-profit organization ("the Organizer") based in Switzerland that had organized a parimutuel prediction tournament for every FIFA World Cup for more than 30 years. In this type of betting, commonly known as pool betting, bettors wager against each other, and the Organizer essentially acts as the matchmaker. The regular prediction tournament is not of particular interest for this research, so I briefly discuss it more profoundly in Appendix D.1.

For the 2022 edition, the Organizer included an additional betting game which was designed to be a natural field experiment. There was a separate betting slip for this additional game, and it was announced in the rules book, as described in Table 1. Bettors could participate voluntarily and at no cost. Bettors were not aware that they were participating in a study.

In this additional betting game, bettors had to make a probabilistic bet on the outcomes of four group-stage fixtures, as shown in Table 1. There are three possible and mutually exclusive outcomes in group-stage matches: Home win, Draw, Away win.

Bettors were randomized into two experimental groups[18], and I exogenously varied the cognitive default through how the state space of these possible outcomes was categorized. For two out of the four matches, the state space was divided into three categories, with each possible outcome being assigned a separate category (treatment condition HIGH cognitive default). For the other two of the four matches, the three outcomes were divided into two categories only, with two outcomes being combined into a single category

[18]The randomization was performed on the computer, where the organization and I randomized the mailing list into two different groups. The betting slip for the regular tournament was fully identical for both groups, and only the betting slips for the experimental task differed.

(treatment condition LOW cognitive default).[19] The experimental groups differed across the matches whether they face condition HIGH or LOW.

To incentivize bettors properly, I again use theoretically and behaviorally compatible incentives: while bettors were de facto incentivized with a proper scoring rule, they were provided with qualitative information about the incentives only. Thus, bettors were truthfully instructed that reporting accurate beliefs would maximize their expected profit. Quantitative information about the payment rule was, of course, available upon request.[20] Bettors were selected quasi-randomly for payout, see Appendix D.2 for more information.

A total of 420 unique bettors participated in the 2022 edition of the tournament. All but one bettor chose to participate in the special game, too, and completed the corresponding betting slip. However, four bettors reported a probability of over 100% for at least one match, so they were excluded from the analysis.[21] This yields a final sample size of 415 bettors. 205 bettors submitted the betting slips of group 1, and 210 bettors participated in group 2, with no statistically significant difference (χ^2: $p = .477$). The 13 winning ranks that were eligible for payout were shared by 17 bettors. Table 15 in the Appendix displays the payouts of the special game by rank and bettor. The average winning bettor earned a prize money of CHF 103 (approx. $120).

5.2 Results

5.2.1 Reported Beliefs

Result 5.1. *Reported beliefs depend on the cognitive default: Bettors in HIGH systematically report a higher likelihood than bettors in LOW.*

Result 5.1 is supported by Figure 7. Bettors in LOW and are displayed in solid blue. Bettors in condition HIGH are displayed in red dashed lines. Visually, we observe that the distribution of $\tilde{\theta}_{LOW}$ is skewed toward the cognitive default of 50%—assigning equal probability to each of the two categories—in all four matches. In contrast, the distribution of $\tilde{\theta}_{HIGH}$ is more skewed towards its cognitive default of two-thirds.

For all four matches, bettors in HIGH predict a significantly higher average likelihood that the outcome occurs than bettors in LOW, see Figure 7.

[19]There would be many more potential configurations of the state space partitioning. For example, we could further split up "France wins" into "France wins with 1 goal difference" and "France wins with 2 goal differences" and so on, essentially dividing the event "France wins" into further sub-events. Indeed, many large bookmakers do offer precisely such bets.

[20]Three bettors requested this information from the Organizer.

[21]Bettors were informed that probabilities need to add up to 100% to make a betting slip valid.

Table 1: The experimental design

Instructions on the betting slip. *For each of these four matches, guess the probability of each outcome occurring. Note that the sum of the stated probabilities must add up to 100% for each match. For more information, please refer to the rules book.*

Experimental group 1

How probable is it, that ...		*How probable is it, that ...*	
France wins or draw	%	Portugal wins	%
Denmark wins	%	Uruguay wins or draw	%
	100%		**100%**
How probable is it, that ...		*How probable is it, that ...*	
Spain wins	%	Serbia wins	%
Draw	%	Draw	%
Germany wins	%	Switzerland wins	%
	100%		**100%**

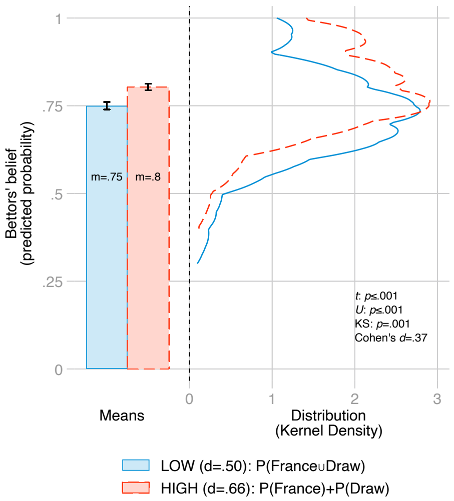
Experimental group 2

How probable is it, that ...		*How probable is it, that ...*	
France wins	%	Portugal wins	%
Draw	%	Draw	%
Denmark wins	%	Uruguay wins	%
	100%		**100%**
How probable is it, that ...		*How probable is it, that ...*	
Spain wins	%	Serbia wins or draw	%
Germany wins or draw	%	Switzerland wins	%
	100%		**100%**

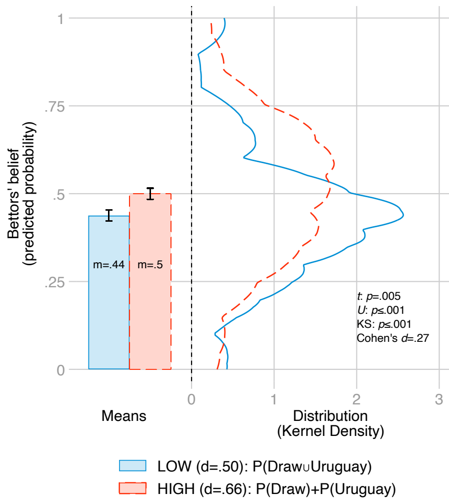
Instructions in the rules book. [Translated from German.]
Special jackpot "The End": For the last edition of the Toto, we have implemented something special. The organizers have received prize money for an additional game, which gives you the chance to win an extra prize for free. The entry form can be found below the standard betting slip of the Toto. Participation is voluntary.
Rules. *Guess the probability of occurrence for the indicated match outcomes for the four matches in the group stage. Your goal is to guess the probabilities as accurately as possible (whole numbers only). Note that the sum of the stated probabilities must add up to 100% for each match.*
Scoring and Payout. *If you rank 33rd, 66th, 99th, 133rd, 166th, 199th, 233rd, 266th, 299th, 333rd, 366th, 399th, or 433rd at the end of the regular Toto, you will receive a payout of CHF 200 multiplied by your guessing accuracy percentage (possible guessing accuracies range from 0% to 100%). The guessing accuracy is calculated in such a way that it pays off for you to guess as accurately as possible and reveal your true guess. The precise payment rule and the calculation formula are available upon request. If two or more participants land on the same rank, the prize money of CHF 200 will be evenly split and then multiplied with your guessing accuracy.*
In short: *The more accurately you guess, the higher your expected payout.*

Figure 7: Compressed beliefs in sports betting

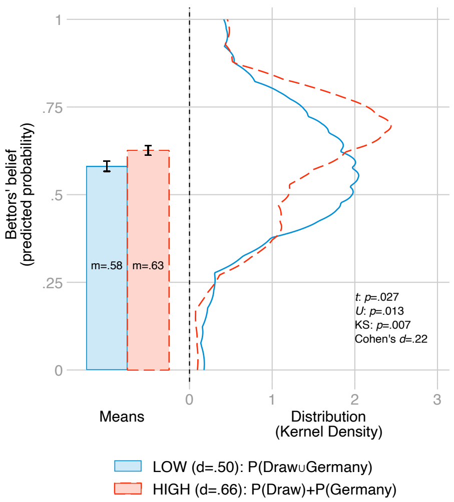
(a) France vs. Denmark



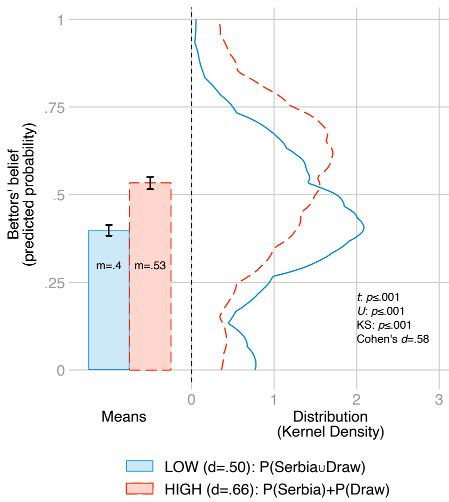
(b) Portugal vs. Uruguay



(c) Spain vs. Germany



(d) Serbia vs. Switzerland



Note: For each of the four matches, the figure displays the means of the predicted probability of the match outcome by experimental group, along with kernel density estimates that show the distribution of the predicted probabilities. Condition HIGH faced three partitions, and the reported probability is the sum of the two events $P(A) + P(B)$, resulting in a cognitive default of two-thirds. Group LOW faced two partitions, and the outcome $P(A) \cup P(B)$ was combined into a single partition since it was described as a union event, resulting in an cognitive default of 50%. Reported statistics are the p value of a two-sided Welch's unequal variance t test, the two-sided Mann-Whitney U test, a Kolmogorov-Smirnov equality-of-distributions test, and the effect size is expressed as Cohen's d. $N = 415$.

Table 16 in Appendix D.3 exploits the panel structure of the data set and displays panel regression results, confirming the findings reported here: bettors in HIGH report on average that the respective event in question is by 7.5 percentage points or about 14% more likely than bettors in LOW. Regardless of the model employed and the controls included, this is effect is highly statistically significant with $p < .001$.

The average reported probability over the four matches in condition HIGH is $\tilde{\theta}_{HIGH} = 61.59\%$. In condition LOW, the mean likelihood was judged to be $\tilde{\theta}_{LOW} = 54.11\%$. Therefore, solving for $\hat{\alpha}$ gives us a distortion parameter for the average bettor of $\hat{\alpha} = .4492 \approx .45$. That is, only roughly 55% of the variation in root beliefs manifests itself in observed reported beliefs due to the insensitivity generated by belief compression.[22]

Taken together, Figure 7 and Table 16 provide external validity for compressed beliefs: also in the field, reported beliefs causally depend on the induced cognitive default.

5.2.2 Inferred Beliefs

I continue to recover participants' inferred belief $\hat{\theta}_i$ as specified in Equation 2. Again, the inferred belief $\hat{\theta}$ is independent of the treatment conditions ($p = .493$) and thus not a function of the researcher's choice of the elicitation design that determines the cognitive default.[23] For a descriptive comparison of the reported and inferred belief, see Figure 12 in the Appendix.

Naturally, we do not observe the latent root belief, and to assess performance of the two estimators $\hat{\theta}$ and $\tilde{\theta}$, we must thus assume a proxy for the root belief θ. Candidates proposed in the literature are i) the realized state of the world, representing the external validity of beliefs, and ii), the individual's behavior, representing the internal validity of beliefs (see Schlag et al., 2015, for a discussion).[24]

A nice feature of the dataset is that I observe the behavior of the bettors in the regular prediction tournament and can therefore assess the *internal validity* of their probabilistic

[22]Tests reveal that α does not depend on the experimental group and neither on the treatment condition. Regressing the reported beliefs on a constant (estimating the term αd in Equation 1), and on betting behavior in the regular tournament interacted with the experimental group, I find that the slope (estimating the term $1-\alpha$ in Equation 1) is not different between the two experimental groups ($p = .544$). Employing the same regression but replacing the experimental group dummy with a treatment dummy reveals a similar result ($p = .713$).

[23]The inferred belief is also independent of the experimental group ($p = .367$).

[24]The first approach tests whether bettors correctly predict the actual realization of the state of the world, the task for which they were incentivized for. A potential criticism of using the actual outcomes as a proxy for root beliefs is that we do not necessarily know whether bettors correctly predict the actual match outcomes. In other words, we do not know whether betters on average hold correct beliefs. Potentially, root beliefs may be subject to errors and biases, too.

beliefs. The Organizer provided me with bettors' wagers on the four matches in the regular tournament, where bettors had to predict the score of each match.[25] Therefore, they had to make a discriminating choice among the three possible outcomes—a coarser belief elicitation. I find this betting behavior to be independent of the experimental condition in all four matches, assessed with a χ^2 test (p values: $p_1 = .808$, $p_2 = .914$, $p_3 = .317$, $p_4 = .575$). For each bettor and each match, I compute linear and squared loss of the two estimands $\tilde{\theta}$ and $\hat{\theta}$, evaluated against bettors' own betting behavior and the actual outcomes of the match fixtures.

Table 2: Performance of the inferred belief versus the reported belief in sports betting

	Mean of benchmark $L(\tilde{\theta})$	Δ in means $L(\tilde{\theta}) - L(\hat{\theta})$	% Improvement $(L(\tilde{\theta}) - L(\hat{\theta}))/L(\tilde{\theta})$
External validity:			
MAE	.389	.052	16.5
		(.003)	(.97)
		{.001}	{.001}
MSE	.201	.000	7.85
		(.002)	(1.45)
		{.917}	{.001}
Internal validity:			
MAE	.349	.060	21.4
		(.003)	(1.17)
		{.001}	{.001}
MSE	.165	.008	15.7
		(.002)	(1.72)
		{.001}	{.001}

Note: The table shows the difference in means as well as the percentage improvement of using the inferred belief versus the reported belief regarding two common losses, the mean absolute error (MAE) and the mean squared error (MSE). For external validity, the losses refer to predicting actual match outcomes. For internal validity, the losses refer to predicting one's own (deterministic) betting behavior in the tournament. Standard errors are shown in parentheses, and p values are shown in braces, obtained from a t test against the theoretical value of 0. $N = 415$.

Table 2 reports the results. Regarding external validity, the inferred belief outperforms the incentivized reported belief significantly in linear loss and performs equally well in quadratic loss. Regarding internal validity, the inferred belief outperforms the incentivized reported belief in linear as well as quadratic loss, and significantly so.

[25] A caveat to note here is that while this deterministic subjective belief is incentivized, the incentives in the regular tournament are unclear due to the complex rule set, and bettors may have behaved strategically. Actually, there is no deterministic proper scoring rule for prediction tournaments (Witkowski, Freeman, Vaughan, Pennock and Krause, 2022).

Result 5.2. *The inferred belief outperforms the reported belief not only in predicting actual match outcomes, but also in predicting bettors' own betting behavior.*

The magnitude is also substantial: For instance, the inferred belief leads, compared to the reported belief, to a 16.5% improvement in linear loss regarding external validity, and a 21% improvement in linear loss regarding internal validity. Result 5.2 is also confirmed by panel regression analysis, see Table 17 in the Appendix. The inferred belief is thus a bettor predictor of actual match outcomes as well as of bettor's betting behavior in the regular pool betting tournament than the reported belief. The inferred belief is supposedly better representing the root belief.

6 Inflation Expectations

I continue to analyze the domain of inflation expectations by using secondary data collected by the German central bank ("Bundesbank"). Before the Bundesbank granted me access to their data, the hypotheses were pre-registered on aspredicted.org (ID 137464). The data employed in this section is confidential and property of the Deutsche Bundesbank, and the data source shall be cited as the "Bundesbank-Online-Panel-Households" ("BOP-HH").

6.1 The survey and the experiment

The BOP-HH is a monthly representative survey by the Bundesbank, which measures German citizens' inflation expectations and perceptions of the price level (see Beckmann and Schmidt, 2020, for details regarding the survey and its elicitation). The sample size for each wave is around 2,500 to 5,000 individuals who voluntarily participate in the survey. The accuracy of the survey responses are not incentivized.

The BOP-HH elicits inflation expectations for the upcoming year in two different ways. First, respondents are asked to give a deterministic point forecast of the inflation rate in 12 months time.[26] Second, after answering this point forecast and a few questions in between, respondents are asked to make a probabilistic inflation forecast. They are confronted with a specific partitioning of the state space and must assign a probability of occurrence to each state, where the probabilities must add up to 100%.[27]

[26]The point forecast is elicited with the following question (translated from German): *"What do you think will be the approximate inflation rate over the next twelve months?"*

[27]Specifically, respondents are asked (translated from German): *"How likely do you think it is that the*

In wave 30 of the survey, conducted in June 2022, the Bundesbank implemented a survey experiment regarding the elicitation of the probabilistic inflation expectations: the state space of possible inflation rates from $-\infty$ to ∞ was differently partitioned among the respondent pool. Respondents were randomly assigned to two experimental groups. The baseline group LOW faces the standard design that the Bundesbank uses: the state space is divided into ten partitions, as visualized in Figure 13 in the Appendix. In contrast, in group HIGH, the state space was divided into 14 different partitions, as shown in Figure 14.

As a consequence, the potential event that next year's inflation will be between -2% and 2% receives a cognitive default of $d_{LOW} = \frac{2}{10}$ in group LOW. In contrast, in group HIGH, the cognitive default is with $d_{HIGH} = \frac{6}{14}$ more than twice as large.

In total, 2,963 individuals participated in wave 30 in the two treatment variations. Following the procedure by the Bundesbank, I exclude all individuals that report point estimates either below or above 12%. I remove from this sample size all individuals who dropped out during the survey, and who provide either no answer or a "don't know" answer to the questions concerning the inflation expectations or socio-demographic characteristics. I further exclude respondents who provide heavily inconsistent responses regarding inflation expectations: their binary response to the question whether they expect inflation or deflation does not align with their numerical inflation (or deflation) expectation, which is why I also exclude those respondents. This yields a final sample size of exactly 2,477 respondents. 1,226 respondents were assigned to the LOW group, and 1,251 respondents are in the HIGH group.

6.2 Results

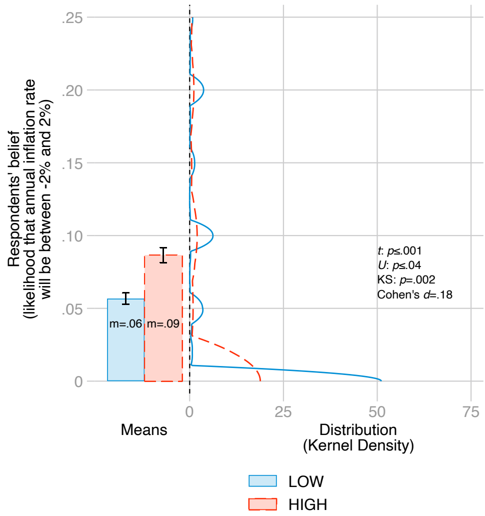
6.2.1 Reported Belief

Result 6.1. *Official inflation expectations depend on the cognitive default used for elicitation: Respondents in HIGH report an annual inflation rate between -2% and 2% to be more probable than respondents in LOW.*

Official inflation expectations depend causally on the survey design used by the central bank: In LOW, respondents on average believe that there is a 5.67% likelihood that the next year's inflation rate will be between -2% and 2%. Respondents in HIGH, however,

inflation rate will develop as follows over the next twelve months?" Respondents are then presented with the partitioning of the state space.

Figure 8: Official inflation expectations depend on the cognitive default



Note: The figure displays the mean likelihood that the annual inflation rate will be between -2% to 2% by experimental group, along with kernel density estimates that show the distribution of the predicted likelihoods. The state space that the inflation rate will be between -2% to 2% was divided into two categories in group LOW, yielding a cognitive default of $\frac{2}{10}$; and into six partitions in group HIGH, yielding a cognitive default of $\frac{6}{14}$. For visualization purposes, the graph censors likelihoods above .25. Reported statistics are the p value of a two-sided Welch's unequal variance t test, the two-sided Mann-Whitney U test, a Kolmogorov-Smirnov equality-of-distributions test, and the effect size is expressed as Cohen's d. $N = 2,477$.

believe that the exact same state will realize with a likelihood of 8.65%—an increase in reported probability of 2.98 percentage points, or 53% ($p < .001$), see Figure 8.

6.2.2 Inferred Belief

Next, I assess the performance of the inferred versus reported belief.[28] Considering internal validity, the BOP-HH survey also elicits a point estimate of annual inflation expectations. This prediction was elicited before respondents were asked about the probabilistic forecast, and with it, randomly assigned to the two experimental groups.[29] I proceed with using this point prediction to identify each individuals subjectively perceived most likely outcome, that is, the partition in which the individual respondent's point forecast falls. The results are presented in Table 18 in the Appendix. The inferred belief signifi-

[28]The population-average $\hat{\alpha} = 0.13$. The distortion parameter is substantially lower than in the sports betting experiment. A reason could be that at the time of the survey (June 2022), an annual inflation rate of -2% to 2% was quite unlikely: The inflation rate in Germany in June 2022 was 7.9%, and in June 2023—the rate that respondents needed to predict—at 6.4%, see https://www.destatis.de/EN/Press/2023/01/PE23_022_611.html.

[29]Therefore, we should not observe any treatment differences in point estimates, and we do not (t test: $p = .295$) and median (median test: $p = .751$).

cantly outperforms the reported belief on all losses with $p < .001$ in predicting the actual realization of the inflation rate, but also individual's own point forecasts.

Result 6.2. *The inferred belief outperforms the reported belief regarding external as well as internal validity.*

Table 3: Percentage of consistent respondents

Prediction	Reported Belief $\tilde{\theta}$	Inferred Belief $\hat{\theta}$	p value
Mean	77.47%	77.75%	.162
Median	74.61%	76.54%	.001
Mode	76.22%	77.03%	.001

Note: The table displays the frequency of consistent respondents for the reported belief and the inferred belief, and the associated p values from a test of proportions. A respondent is classified as consistent if the criterion is met, and as inconsistent otherwise (Engelberg et al., 2006): (1) Classify a respondent as consistent if the point prediction falls within the lower and upper bounds of the mean of the probabilistic forecast, the bounds of the mean are obtained by placing all of each partition's probability mass at the partition's lower and upper endpoint, respectively (mean); (2) Classify a respondent as consistent if the point estimate falls into the partition they assigned the median probability mass (median); (3) Classify a respondent as consistent if the point estimate falls into the partition they assigned the highest probability (mode).

Another common way to test internal validity is to assess the consistency between the point prediction and the probabilistic belief (see D'Acunto, Malmendier and Weber, 2023, for a review). Table 3 shows the results. In all three cases, the inferred belief outperforms the reported belief and increases the percentage of respondents whose probabilistic prediction is consistent with their point prediction.

7 Concluding Remarks

This paper demonstrates that reported beliefs are compressed towards a cognitive default. Even if incentivized with a proper scoring rule in a behavioral compatible way, reported beliefs are a function of the cognitive default, itself implied by the design of the elicitation task. This not only holds in a controlled laboratory setting with a classical subject pool, but also in a natural field experiment with sports bettors, as well as in a representative large-scale survey experiment collecting official inflation expectation data. Replication across these different populations and contexts increases confidence in the robustness and generalizability of the findings (Al-Ubaydli, List and Suskind, 2017).

The contamination of reported beliefs by the cognitive default is a potential confounder when drawing inferences. It may partially explain well-known errors in probabilistic reasoning, such as for example overestimating the likelihood of rare events and underestimating the likelihood of likely events.

Moreover, compression towards the cognitive default also implies that belief reports are malleable: two researchers who use a different elicitation design will obtain different reported beliefs, and potentially conclude differently, as shown in the confidence in self-placement data reported in this paper. Importantly, my results should not be interpreted as a failure of incentive-compatibility: in my view, incentives *are* necessary to incentivize truthful reporting conditional on a given elicitation design, yet they are insufficient to ensure that reported beliefs reliably identify an underlying latent subjective belief.

These findings have direct implications for the design and interpretation of elicited beliefs and survey-based expectations. Because reported beliefs depend on the design used for elicitation, belief reports and survey responses reflect both underlying root beliefs and elicitation-induced cognitive defaults.

As a result, changes in reported expectations over time or across surveys may arise mechanically from design choices rather than from shifts in beliefs. This suggests caution when comparing expectations (i) across surveys following survey redesigns and (ii) across countries using different survey designs. For the interpretation and comparability of economic data, this matters in practice. For instance, in 2025, the Fed, the European Central Bank, and the Bank of England all use different survey designs regarding the elicitation of probabilistic inflation expectations, all inducing different cognitive defaults.[30]

The paper advocates for a constructive solution: using experimentation in task design when eliciting probabilistic beliefs (within, not across, survey and experiment). Experimentally manipulating the cognitive default allows us to assess the extent to which our belief elicitation tool identifies the intended latent object it aims to measure, and to quantify the contamination of belief reports by an objectively irrelevant artifact—the cognitive default.

Deliberate variation in elicitation design also allows experimenters and survey designers to infer the latent root belief, which yields a measure of root beliefs that is robust to design choices. Theoretically, conditional on plausible assumptions, this inferred belief is an asymptotically unbiased estimator, as opposed to the observed belief reports. Empirically, in all four domains, the inferred belief indeed better represents the root belief than

[30] See the following sites, all last accessed on December 23, 2025: ECB Survey of Professional Forecasters; Fed Survey of Professional Forecasters; Bank of England Survey of External Forecasters; Fed Survey of Consumer Expectations.

the reported belief.

The result that belief reports are malleable and depend on artifacts of the elicitation design may also impose a challenge for meta-studies and replications. First, replication studies may not replicate (quantitatively, and potentially even qualitatively) the original study's effect if a slightly different elicitation design is used, so findings may not be robust. Second, comparability across studies may be limited, a challenge for meta-studies.

Taken together, my results imply that the state-of-the-art belief elicitation method does not directly measure latent subjective beliefs, but that such beliefs can be imperfectly recovered once we account for cognitive defaults necessarily induced by the elicitation.

The proposed approach also has some limitations, and many open questions remain, to be addressed in future research. The first and perhaps most fundamental challenge is to know whether people actually hold probabilistic beliefs at all. If not, there may be little point in trying to elicit (or infer) probabilistic beliefs.

Another fairly fundamental question is whether decisions are also dependent on the cognitive default. The partitioning of choices may create a choice default similar to the cognitive default, such as 50-50 in a binary choice, and decisions may be influenced by that context. Scholars have begun to address this question theoretically (Ahn and Ergin, 2010) and empirically (Sonnemann et al., 2013; Enke et al., 2025). If behavior is also a function of a cognitive default created by the partitioning of choices, revealed preferences would also suffer from systematic bias, calling into question their validity as currently elicited since the construct validity is jeopardized (Snowberg and Yariv, 2025).

In terms of assumptions, a challenge is whether the model described in Equation 1 is actually a sufficiently accurate *as if* description of how agents report probabilistic beliefs. One can imagine that the linear model is sufficiently good at approximating a large range of root beliefs, but suffers from lack of precision at the extremes. Thus, if people hold extreme root beliefs, the approach presented here may be less accurate.

A more practical problem is that it is difficult to distinguish observations that are likely biased from those that are likely unbiased. Decompressing only the reported beliefs that are actually contaminated by the cognitive default would increase the accuracy of the inferred belief even further.

It would be interesting to see whether the contamination by the cognitive default is one reason why (reported) beliefs often diverge from behavior, and player's not best-responding to their reported beliefs (Costa-Gomes and Weizsäcker, 2008).

Finally, I believe the belief elicitation literature should devote more attention to study how belief reports depend on the design of the elicitation task. For instance, in what domains are compression effects large, and in what domains negligible? What does the compression of reported beliefs imply for the consensus finding in a certain strand of literature? These and many other questions are left for future research.

References

Ahn, David S and Haluk Ergin, "Framing contingencies," *Econometrica*, 2010, *78* (2), 655–695.

Al-Ubaydli, Omar, John A List, and Dana L Suskind, "What can we learn from experiments? Understanding the threats to the scalability of experimental results," *American Economic Review*, 2017, *107* (5), 282–286.

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

Andersen, Steffen, John Fountain, Glenn W Harrison, and E Elisabet Rutström, "Estimating subjective probabilities," *Journal of Risk and Uncertainty*, 2014, *48* (3), 207–229.

Ba, Cuimin, J Aislinn Bohren, and Alex Imas, "Over-and underreaction to information," *Available at SSRN 4274617*, 2024.

Beckmann, Elisabeth and Tobias Schmidt, "Bundesbank online pilot survey on consumer expectations 01/2020," Technical Report, Deutsche Bundesbank 2020.

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

—, Don A Moore, and Matthew Rabin, "Biased beliefs about random samples: Evidence from two integrated experiments," Technical Report, National Bureau of Economic Research 2017.

Brier, Glenn W et al., "Verification of forecasts expressed in terms of probability," *Monthly weather review*, 1950, *78* (1), 1–3.

Charness, Gary, Uri Gneezy, and Vlastimil Rasocha, "Experimental methods: Eliciting beliefs," *Journal of Economic Behavior & Organization*, 2021, *189*, 234–256.

Clemen, Robert T and Canan Ulu, "Interior additivity and subjective probability assessment of continuous variables," *Management Science*, 2008, *54* (4), 835–851.

Costa-Gomes, Miguel A and Georg Weizsäcker, "Stated beliefs and play in normal-form games," *The Review of Economic Studies*, 2008, *75* (3), 729–762.

D'Acunto, Francesco, Ulrike Malmendier, and Michael Weber, "What do the data tell us about inflation expectations?," in "Handbook of economic expectations," Elsevier, 2023, pp. 133–161.

Danz, David, Lise Vesterlund, and Alistair J Wilson, "Belief elicitation and behavioral incentive compatibility," *American Economic Review*, 2022, *112* (9), 2851–83.

— , — , and Alistair J. Wilson, "Evaluating Behavioral Incentive Compatibility: Insights from Experiments," *Journal of Economic Perspectives*, November 2024, *38* (4), 131–54.

Engelberg, Joseph, Charles F Manski, and Jared Williams, "Comparing the point predictions and subjective probability distributions of professional forecasters," 2006.

Enke, Benjamin and Thomas Graeber, "Cognitive Uncertainty," *NBER Working Paper*, 2019, (w26518).

— and — , "Cognitive uncertainty," *The Quarterly Journal of Economics*, 2023, *138* (4), 2021–2067.

— , — , Ryan Oprea, and Jeffrey Yang, "Behavioral Attenuation," Technical Report, Working Paper 2025.

Exley, Christine L. and Kirby Nielsen, "The Gender Gap in Confidence: Expected but Not Accounted For," *American Economic Review*, March 2024, *114* (3), 851–85.

Fox, Craig R and Robert T Clemen, "Subjective probability assessment in decision analysis: Partition dependence and bias toward the ignorance prior," *Management Science*, 2005, *51* (9), 1417–1432.

— and Yuval Rottenstreich, "Partition priming in judgment under uncertainty," *Psychological Science*, 2003, *14* (3), 195–200.

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

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart, "Designing Information Provision Experiments," *Journal of Economic Literature*, March 2023, *61* (1), 3–40.

Healy, Paul J and Greg Leo, “Belief elicitation: a user’s guide,” *Handbook of Experimental Methodology, ed. by L. Yariv and E. Snowberg, Elsevier*, 2024, *1.*

Heisenberg, Werner, *Physics and Philosophy: The Revolution in Modern Science*, London: Penguin Books, 1989.

Hoelzl, Erik and Aldo Rustichini, “Overconfident: Do you put your money on it?,” *The Economic Journal*, 2005, *115* (503), 305–318.

Holt, Charles A and Angela M Smith, “Belief elicitation with a synchronized lottery choice menu that is invariant to risk attitudes,” *American Economic Journal: Microeconomics*, 2016, *8* (1), 110–139.

Hossain, Tanjim and Ryo Okui, “The binarized scoring rule,” *Review of Economic Studies*, 2013, *80* (3), 984–1001.

Huck, Steffen and Georg Weizsäcker, “Do players correctly estimate what others do?: Evidence of conservatism in beliefs,” *Journal of Economic Behavior & Organization*, 2002, *47* (1), 71–85.

Johnson, N.L., S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Volume 2* Wiley Series in Probability and Statistics, Wiley, 1995.

Khaw, Mel Win, Ziang Li, and Michael Woodford, “Cognitive imprecision and small-stakes risk aversion,” *The review of economic studies*, 2021, *88* (4), 1979–2013.

Manski, Charles F, “Measuring expectations,” *Econometrica*, 2004, *72* (5), 1329–1376.

— , “Survey measurement of probabilistic macroeconomic expectations: progress and promise,” *NBER Macroeconomics Annual*, 2018, *32* (1), 411–471.

Moore, Don A and Paul J Healy, “The trouble with overconfidence.,” *Psychological review*, 2008, *115* (2), 502.

Niederle, Muriel and Lise Vesterlund, “Do women shy away from competition? Do men compete too much?,” *The quarterly journal of economics*, 2007, *122* (3), 1067–1101.

Offerman, Theo, Joep Sonnemans, Gijs Van de Kuilen, and Peter P Wakker, “A truth serum for non-bayesians: Correcting proper scoring rules for risk attitudes,” *The Review of Economic Studies*, 2009, *76* (4), 1461–1489.

Oprea, Ryan, "Decisions Under Risk are Decisions Under Complexity," *American Economic Review*, 2024, p. forthcoming.

Prava, Venkata R, Robert T Clemen, Benjamin F Hobbs, and Melissa A Kenney, "Partition dependence and carryover biases in subjective probability assessment surveys for continuous variables: model-based estimation and correction," *Decision Analysis*, 2016, *13* (1), 51–67.

Rutström, E Elisabet and Nathaniel T Wilcox, "Stated beliefs versus inferred beliefs: A methodological inquiry and experimental test," *Games and Economic Behavior*, 2009, *67* (2), 616–632.

Schlag, Karl H and J. J. van der Weele, "Eliciting probabilities, means, medians, variances and covariances without assuming risk neutrality," *Theoretical Economics Letters*, 2013, *3*, 38–42.

_ , James Tremewan, and Joël J Van der Weele, "A penny for your thoughts: A survey of methods for eliciting beliefs," *Experimental Economics*, 2015, *18*, 457–490.

Schotter, Andrew and Isabel Trevino, "Belief elicitation in the laboratory," *Annu. Rev. Econ.*, 2014, *6* (1), 103–128.

Snowberg, Erik and Leeat Yariv, "Evaluating Experimental Designs," Working Paper 34031, National Bureau of Economic Research July 2025.

Sonnemann, Ulrich, Colin F Camerer, Craig R Fox, and Thomas Langer, "How psychological framing affects economic market prices in the lab and field," *Proceedings of the National Academy of Sciences*, 2013, *110* (29), 11779–11784.

Trautmann, Stefan T and Gijs van de Kuilen, "Belief elicitation: A horse race among truth serums," *The Economic Journal*, 2015, *125* (589), 2116–2135.

Tversky, Amos and Derek J Koehler, "Support theory: a nonextensional representation of subjective probability.," *Psychological review*, 1994, *101* (4), 547.

Wilson, Alistair and Emanuel Vespa, "Paired-uniform scoring: Implementing a binarized scoring rule with non-mathematical language," Technical Report, Working Paper 2018.

Witkowski, Jens, Rupert Freeman, Jennifer Wortman Vaughan, David M Pennock, and Andreas Krause, "Incentive-compatible forecasting competitions," *Management Science*, 2022.

Working Paper 544 January 2026

Authors

Christian Zihlmann
Faculty of Management, Economics and Social Sciences,
Chair of Industrial Economics,
University of Fribourg, Boulevard de Pérolles 90, 1700 Fribourg, Switzerland;
Email: christian.zihlmann@unifr.ch.

Abstract

Subjective beliefs are central to economic inference, and incentive-compatible belief elicitation mechanisms are widely assumed to identify these latent objects. This paper shows that elicited belief reports causally depend on an uninformative cognitive default induced by the elicitation design. From the lab to sports betting to official inflation expectations, reported beliefs are highly malleable, even under theoretically and behaviorally compatible centives. I propose experimentally varying the cognitive default during belief elicitation. This exogenous variation allows the construction of inferred beliefs that are stable across elicitation designs and empirically outperform incentivized reports in predicting realized outcomes and participants' own behavior.

JEL Classification
C81;C90;D81:D83

Keywords
Belief Elicitation; Subjective Beliefs; Probabilistic Beliefs; Cognitive Default; Field Evidence

Working Papers SES collection

Last published

539 Dembetembe G., Raso Dominguéz X., Gugler P. : Engagements and disengagements of Multinational Enterprises with the United Nations Sustainable Development Goals: Where we are and where to go from here; 2024
540 Ducret R., EugsterN., Isakov D., Weisskopf JP. : The behavior of stock prices around the ex-day during a dividend shortage; 2025
541 Isakov D., Keane A., Schatt A., : The dividend policy of state-owned banks during an economic crisis; 2025
542 Dumas C.,Gautrain E.,Gosselin-Pali A., : Nutritional Benefits of Fostering: Evidence from Longitudinal Data in South Africa 2025
543 Dumas C., JAtiva X., Baumgartner S., : Antibiotic pollution and infant mortality in India: a research note 2025

Catalogue
https://folia.unifr.ch/unifr/collections/299490

DOI: https://doi.org/10.51363/unifr.ewp.fgx26v
ISSN: 3042-7517

Publisher
Université de Fribourg, Suisse, Faculté des sciences économiques et sociales et du management
Universität Freiburg, Schweiz, Wirtschafts- und sozialwissenschaftliche Fakultät
University of Fribourg, Switzerland, Faculty of Management, Economics and Social Sciences

Bd de Pérolles 90
CH-1700 Fribourg
Tél.: +41 (0) 26 300 82 00
decanat-ses@unifr.ch
www.unifr.ch/ses