Updated Pre-Analysis Plan

1 Overview

Participants complete a repeated belief-updating task with the same probabilistic structure across conditions. In the **Control (Abstract)** condition, the task is framed as an urn problem. In the **Narrative** condition, the identical likelihood structure is embedded in a short story about two suspects of theft. Signals in each round are either agnostic or non-diagnostic. Diagnostic signals are denoted $s \in \{-1, +1\}$, with negative and positive sign indicating respectively a signal against and or in support of the true state of the world. Non-diagnostic signals are denoted $s = 0$. In the **follow-up study**, the same narrative is used but the non-diagnostic signals are replaced by blank information (no message, still $s = 0$).

2 Primary Estimand and Structural Model

Our **primary analysis** uses the structural misperception model. Let $\theta \in \{0, 1\}$ denote the true state and $s_t \in \{-1, 0, +1\}$ the actual signal in round t with known state-dependent likelihoods (as induced by the design). Participants hold an individual prior π_{t-1} (their reported belief at $t-1$). We assume:

1. **Perception step (systematic bias).** The actual signal s_t is perceived as $\tilde{s}_t \in \{-1, 0, +1\}$ according to a misperception matrix $\mathbf{\Pi}$ with elements $\Pi_{ij} = \Pr(\tilde{s} = j \mid s = i)$, rows summing to one. $\mathbf{\Pi} = \mathbf{I}$ implies perfect perception (no bias).

2. **Bayesian updating on perceived evidence.** The latent posterior $\pi_t^\star$ is the Bayesian posterior obtained by updating π_{t-1} with the perceived signal $\tilde{s}_t$ using the experiment's true likelihood ratio.

3. **Reporting noise (unsystematic error).** The reported belief $b_t \in [0, 1]$ equals $\pi_t^\star$ plus zero-mean noise $\mathcal{N}(0, \sigma^2)$.

Identification is driven by the fact that $\mathbf{\Pi}$ captures directional misreadings of signal content (systematic bias), whereas σ^2 absorbs symmetric over/under-reactions and random imprecision.

2.1 Main estimand.

The components of $\mathbf{\Pi}$, with emphasis on the **non-diagnostic row**, i.e., $\Pi_{0,-1}$ and $\Pi_{0,+1}$, quantify the tendency to treat non-diagnostic evidence as if informative. Larger off-diagonal mass indicates stronger misuse of non-diagnostic signals.

Primary hypotheses (Main Experiment).

- **H1 (Narrative misuse of non-diagnostic signals).** In the Narrative condition, the off-diagonal mass in the non-diagnostic row is significantly greater than zero: $\Pi_{0,-1} + \Pi_{0,+1} > 0$.
- **H1′ (Benchmarking to Control).** The off-diagonal mass in the non-diagnostic row is larger in Narrative than in Control.

2.1.1 Follow-Up Study: Blank Information in Narrative

In the follow-up, the narrative framing remains identical but non-diagnostic signals are absent (blank line). We estimate the same structural model, extended to include a "blank" category $s = 0$ that carries no state-contingent likelihood information.

Primary hypotheses (Follow-Up).

- **H2 (No bias from blank information).** Blank information is not misperceived as diagnostic: the off-diagonal mass from the blank row is null, i.e., $\Pi_{0,-1} = \Pi_{0,+1} = 0$ and $\Pi_{0,0} = 1$.[1]

2.2 Estimation and Inference

We fit the model by maximum likelihood. Inference is based on participant-level bootstrap (10,000 replicates) for $\mathbf{\Pi}$ and σ^2, reporting percentile CIs. All tests in H1/H1′ and H2 are pre-specified as one-sided in the direction stated.

3 Secondary and Robustness Analyses

For continuity with our prior pre-analysis plan, we will reproduce two reduced-form diagnostics in the **Appendix**. First, we regress the absolute difference between reported beliefs and the Bayesian benchmark on the Treatment (Narrative indicator) variable.

Second, we estimate a dynamic regression model, relating log-odds to (i) lagged log-odds, (ii) diagnostic evidence Δz_t, incrementing through rounds, and (iii) dummy variables measuring participants' tendency to confirmation bias for both diagnostic and non-diagnostic signals.

[1] Equivalently, any movement in reported beliefs following a blank is attributed to σ^2 (unsystematic noise), not to systematic misperception.

The Bayesian benchmark implies $\rho = 1$, $\beta = \ln\big(\theta/(1-\gamma-\theta)\big) = 0.4055$, and all dummy coefficients equal to zero, so departures map to conservatism/overreaction ($\beta \lessgtr 0.4055$) and confirmation/misuse of non-diagnostic signals (non-zero dummy coefficients). We keep the same practical details (probability truncation to $[0.01, 0.99]$ to define log-odds; standard errors clustered by subject) and run both diagnostics separately by treatment.

In the follow-up (blank information), we rerun the same regressions, replacing the non-diagnostic indicators with blank-information indicators. These checks are ancillary to the primary structural analysis and are reported in the **Appendix**.

3.1 Data Handling and Exclusions

The study is conducted in the lab; consequently, we do not anticipate meaningful attrition or inattention. Participants are asked to perform comprehension checks at the beginning of the session.

Extreme-belief zone (dominance due to rounding). With a quadratic scoring rule rounded to integers, payoffs exhibit a plateau near 0 and 1. Let

$$S(b, y) = \text{round}\big(100\,[1-(b-y)^2]\big), \quad y \in \{0, 1\}.$$

Once $b \geq 0.93$, the payoff when the realized state is $y = 1$ already rounds to the maximum (100). Pushing b above 0.93 cannot increase this payoff, but it does strictly decrease $S(b, 0)$ if the state is $y = 0$. Hence any $b > 0.93$ is weakly dominated by reporting $b = 0.93$. Symmetrically, for $b \leq 0.07$, the payoff when $y = 0$ is already at its maximum, and any $b < 0.07$ is weakly dominated by reporting $b = 0.07$. Therefore, "making beliefs more extreme" inside $[0, 0.07] \cup [0.93, 1]$ can only reduce expected payoff.
To address this, we pre-specify an extreme-belief zone

$$\mathcal{E} \;=\; [0,\, 0.07] \;\cup\; [0.93,\, 1].$$

Our primary structural analysis includes all observations. As a pre-registered robustness, we will re-estimate all models excluding observations where the incoming belief (the pre-signal belief at round $t-1$) lies in $\mathcal{E}$, since updates within $\mathcal{E}$ are not interpretable as meaningful intensification. We will report (i) the share of excluded observations, (ii) trimmed vs. untrimmed estimates side-by-side, and (iii) any differences in inference.

Transform conventions. For the reduced-form log-odds regressions, probabilities are truncated to [0.01, 0.99] to define the log-odds transform.

Commitment. The structural model described above is the **primary** analysis for both the main experiment and the follow-up. All reduced-form analyses from the previous PAP will be executed and reported in the **Appendix**.