Planned Analysis for Narrative Willingness to Pay

In a previous experiment[1], we documented that subjects in the narrative framing act as if they use nondiagnostic signals to update their beliefs when they should not. In this new study, we investigate a related question: do people attach monetary value to those nondiagnostic signals because they perceive them as informative?

We test whether nondiagnostic signals ("Info") increase the mean willingness to pay (WTP) for a three-signal bundle specifically under the Narrative framing. Our primary estimand is the 2×2 interaction (Narrative $\times$ Info), i.e.,

$$\Delta \;=\; \big[E(WTP \mid N, I) - E(WTP \mid N, \neg I)\big] \;-\; \big[E(WTP \mid C, I) - E(WTP \mid C, \neg I)\big].$$

A positive Δ supports the hypothesis.

Belief composition and robustness. We acknowledge that any comparison of mean WTP across treatments is potentially confounded by differences in participants' prior beliefs, since the theoretical value of information depends nonlinearly on the prior and peaks around $p = 0.5$. This "belief-composition" issue cannot be fully resolved ex-ante, because we make no assumptions about the shape or distribution of beliefs across treatments. Our strategy is therefore exploratory and robustness-oriented rather than prescriptive. We will report treatment effects under several complementary approaches—conditioning on the stated prior, residualizing WTP by its theoretical benchmark, reweighting to balance the prior distribution, and examining relative deviations—without committing to a single specification as the uniquely correct one. We will also inspect the data for extreme priors (close to 0 or 1) where both WTP and EVSI are near zero and relative differences become unstable; analyses may be repeated excluding such observations. This flexible approach ensures transparency and guards against over-interpreting belief-driven variation as treatment effects.

Following this approach, we define a tentative analysis aimed at answering our research question. The following model can be estimated using a pooled OLS with errors clustered at the participant level

$$WTP_{it} - WTP^*_{it} = \beta_0 + \beta_1 I_i + \beta_2 N_i + \beta_3 (I_i \times N_i) + \delta_{s\times t} + \epsilon_{it}$$

where WTP_{it} is the WTP indicated by participant i in round t, WTP^*_{it} is the theoretical

[1] Pre-registered trial

WTP based on their belief at round $t-1$, I is a dummy equal to one if the subject is assigned to a treatment with nondiagnostic signals, N a dummy equal to one if the subject is assigned to a treatment with the story presented in narrative form, $\delta_{s\times t}$ are dummies accounting for story and round specific effects, while ϵ_{it} is the idiosyncratic error.
We aim to test the statistical significance of the diff-in-diff coefficient β_3.

Although our primary analysis uses all five rounds per story to maximize precision, repeated interaction with the pricing/learning environment can generate artifacts that might bias the results. For example, after participants observe a random price and a purchase/non-purchase outcome, they may inflate their WTP to secure information acquisition, anchor on past prices, or respond to cumulative spend and reference points. Receiving (or not) a signal also changes the posterior, and therefore the marginal value of the next signal. This path dependence can propagate into later rounds even after we subtract the normative benchmark based on reported beliefs. Although these problems are common to all experimental conditions, we plan to add a first-round-only robustness check. Round 1 provides an experience-free snapshot, thereby offering a cleaner test of treatment differences. Concordant estimates between the all-rounds specification and the first-round replication strengthen internal validity. Discrepancies, instead, would be informative about learning/anchoring or path-dependence rather than belief composition per se.