Pre-analysis plan: Misguided Inference and Asymmetric Correction

Christoph Drobner* A. Yeşim Orhun†

August 18, 2026

1 Background

This study investigates how overconfidence and underconfidence about own ability shape inferences about the economic environment when outcomes depend on both ability and the environment. We focus on a setting where individuals can switch between environments. We test the hypothesis that overconfident individuals are more likely than underconfident individuals to select environments they perceive as more ability-contingent, and that this selection generates asymmetric belief correction (i.e., overconfident participants correct their biases more than underconfident participants do).

2 Experimental Design

The experiment has two stages. Before they begin, participants answer a short questionnaire that includes an attention check and read experimental instructions followed by nine comprehension check questions (initial comprehension check). The attention check and comprehension questions serve both screening and education (see Screening)

2.1 Stage 1: quiz and prior performance belief

Participants complete a 12-question multiple-choice logic quiz (4 options, 10 seconds per question). The quiz contains 6 *shared* questions taken by all participants and 6 *condition-specific* questions whose difficulty is determined by the participant's randomly assigned quiz

*Central European University; drobnerc@ceu.edu.
†University of Michigan; aorhun@umich.edu.

difficulty condition (EASY or DIFFICULT). We define γ^* as the percentage of a reference pool the participant beats on the 6 shared questions. The reference pool consists of 100 participants who completed the same 6 shared questions in a pilot.[1] The participant beats a reference participant whose score is strictly lower, and ties are broken by a coin flip, so $\gamma^* = \Pr(\text{higher}) + \frac{1}{2}\Pr(\text{tie})$ on a 0–100 scale.

Before the belief question, participants read how the comparison works, including the coin-flip tie break. They then observe their score on the entire 12-question quiz and report their belief $\gamma_0 \in [0, 100]$ about what percentage of the 100 reference-pool participants they outperformed on the 6 shared questions. Participants also rate how sure they are of that answer on a continuous scale with five labeled anchors, from "No clue at all" to "Absolutely certain." The point answer and the confidence rating map into a Beta distribution over γ^*, which participants review as a histogram of implied probabilities and confirm or adjust before continuing. Only γ_0 enters the analysis; the confidence rating is recorded for descriptive use.

2.2 Stage 2: manager-selection task

Participants read instructions about the manager-selection task and answer nine comprehension questions that largely re-test the same concepts as in the initial comprehension check but using exact numbers that are now relevant to the participant. We refer to this set of nine questions as the *repeat-comprehension check*. If the participant provides an incorrect answer, the correct answer is shown and the participant can proceed by selecting it.

In each of the two rounds, participants receive tokens from one of the two managers, Alex or Sam. Each round displays 50 tokens. Each token draw is either green or red; a green token pays \$0.05 and a red token pays nothing. At the end of the experiment, one of the two rounds is randomly selected and the participant is paid the cumulative token earnings from that round. Of the two managers, Alex and Sam, one is the Performance Manager and the other is the Random Manager. Participants do not know who is which, but can learn at the end of the experiment if they wish.

The *Performance Manager* compares the participant's quiz score on the 6 shared questions to that of one randomly drawn participant from the reference pool on each of 50 independent draws per round. It pays a green token if the participant's score is higher, red if it is lower, and resolves ties by a coin flip. This produces a green token with probability γ^*.

The *Random Manager* produces a green token with disclosed probability γ_R on each of 50 draws per round, implemented as a wheel of fortune with 100 fields, γ_R green and $100 - \gamma_R$

[1] The pilot study used the same attention check screening criteria as the main study and was run before the main study to initialize the reference score distribution.

red, spun independently for each token draw. By design, $\gamma_R = \gamma^*$: the disclosed rate of the Random Manager equals the participant's true performance rate. The value of γ_R is disclosed to participants, but they are not told that $\gamma_R = \gamma^*$.

The two managers therefore produce tokens from the same underlying distribution. This eliminates payoff differences across managers and ensures that any cross-group difference in belief updating reflects belief-driven manager selection rather than differences in objective token exposure. Manager types are assigned once and remain fixed across the two rounds; token observations are therefore perceived as informative about whether Alex or Sam is the Performance Manager, conditional on the participant's current performance belief.

In each round, participants observe a summary of the 50 token outcomes and report (i) an updated performance belief $\gamma_t \in [0, 100]$ and (ii) their *manager-type belief* $\theta_t \in [0, 100]$, the percent chance that the current manager is the Performance Manager rather than the Random Manager. Because $\gamma_R = \gamma^*$, the two managers are objectively indistinguishable, so θ_t reflects only *perceived* manager type, inferred from a (possibly) biased performance belief.

Before round 1, participants are randomly assigned a manager. At the end of round 1, participants decide whether to stay with their current manager or switch to the other manager for round 2. All participants are told that their choice will be honored 75% of the time, which reflects a 50/50 experimental randomization between two *manager assignment regime* conditions: either the *active-choice* arm, in which the round-2 manager is always the one the participant chooses; or the *control* arm, in which the participant is randomly assigned a manager. Because 75% is the average across the two arms, this statement is accurate for the sample and identical across arms, so the choice is made under a common expectation. This design elicits the manager selection from everyone while randomly varying whether it is carried out, which isolates the causal effect of selection on belief correction.

2.3 Belief elicitation

The manager-type belief θ_t is elicited on a 0–100 scale in response to the question: "Given these tokens, how likely is [current manager] to be the Performance Manager versus the Random Manager?" Because exactly one of the two managers is the Performance Manager and the other is the Random Manager, this value also fixes the probability assigned to the other manager; the two sum to 100. The performance belief γ_t is elicited via a single slider on the 0–100 scale, in response to the question: "Given these tokens, what percentage of the 100 other participants do you now expect to beat on the shared quiz questions?" It therefore concerns the same object as γ_0, on the same scale, but without the confidence rating.

The performance belief γ_t and the manager-type belief θ_t are elicited on the same screen,

and their left-right placement is randomized across participants (see Randomization). All belief reports are incentivized with the binarized scoring rule of Hossain and Okui (2013). After the manager-selection task, one belief is selected uniformly at random from the five elicited reports (the performance beliefs γ_0, γ_1, γ_2 and the manager-type beliefs θ_1, θ_2) and paid out under the rule, with a maximum bonus of \$1.

2.4 Payments

Participants receive a flat \$3.50 completion fee, the token bonus in a randomly chosen round (up to \$2.50), and the belief bonus in a randomly chosen belief question (up to \$1). The experiment is expected to take 20 minutes.

2.5 Other features

Basic demographic information (age and gender) is collected at the beginning of the study, before the quiz and the treatment assignment. Near the end of the study, participants complete one trial of the Ebbinghaus visual-illusion task, which we use as a bot screen. The study also carries a honeypot field, invisible to human participants but visible to software that fills in the page, which we record as a second bot screen. At the end of the study, the participants are given the option to learn how their quiz performance compared and what types of managers Alex and Sam each were.

3 Screening and randomization

We plan to collect data from 2,000 US participants on Prolific who complete the entire experiment, randomized in equal numbers to the active-choice and control regimes.

3.1 Screening

The study requires a desktop or laptop: participants on other devices are blocked at entry. Two additional rules remove participants, both administered before any experimental condition is assigned.

First, a directed-response attention check appears on the first page after the consent form: participants are asked to select a specified option from an occupation list. Failure terminates the study immediately.

Second, a comprehension gate applies to the initial comprehension check, described below. Nine questions are asked in total. At the end of the first pass, participants who made at

most two errors proceed, after a page showing the correct answer for each question missed. With three or more errors, the participant re-reads the whole instruction block and answers the questions again. At the end of the second pass, two or more errors terminate the study; participants who made at most one error proceed, after a page showing the correct answer for the question missed.

A participant whose identifier already carries a completed or terminated record cannot restart the study.

3.2 Randomization

The quiz-difficulty condition (EASY vs. DIFFICULT) is block randomized within strata defined by the participant's score on the 6 shared questions $(0, 1, \ldots, 6)$ crossed with gender.[2] This gives $7 \times 2 = 14$ cells.

The manager assignment regime (active-choice vs. control) is block randomized within cells defined by the cross of four pre-treatment variables, all fixed at the moment the round-1 beliefs are submitted:

1. round-1 performance-belief bias $\gamma_1 - \gamma^*$, in four bins with cutpoints -15, 0, 15;
2. no round-1 performance-belief updating, $\mathbf{1}[\gamma_1 = \gamma_0]$;
3. gender;
4. error on the repeat-comprehension check,
   $\mathbf{1}$[one or more first-try errors on the repeat-comprehension check].

This gives $4 \times 2 \times 2 \times 2 = 32$ cells. The manager assignment regime affects round 2, and its randomization occurs after the round-1 belief report, so every stratifying variable is predetermined with respect to the treated round.

The remaining randomizations are order and placement randomizations that prevent position confounds from influencing results. We independently randomize: (i) color assignment to each manager name; (ii) which manager is the Performance Manager; (iii) the left-right placement of the manager-type belief elicitation relative to the performance belief elicitation on the same page; and (iv) the left-right placement of the Yes and No buttons on the end-of-study opt-in screen.

We will check for color, placement, and name-preference effects. If any are significant, we will control for them in our analyses.

[2] Gender is binarized as male = 1; female, non-binary, and prefer not to say = 0.

4 Hypotheses

H1 (Misguided selection). Overconfident participants are more likely than underconfident participants to select the environment they perceive to be ability-contingent (the Performance Manager), while underconfident participants are correspondingly more likely to select the environment they perceive to be random (the Random Manager).

H2 (Asymmetric correction). Through endogenous learning, overconfident participants reduce the bias in their beliefs about their own performance more than underconfident participants do.

5 Empirical analysis

All analyses use the sample of participants who complete the entire experiment. Throughout, standard errors are heteroskedasticity-robust (HC1), and every regression includes a full set of saturated stratum fixed effects: for H1 these are the 14 quiz-condition strata; for H2 these are the 32 regime cells defined above.

Notation.

- γ^*: the participant's percentile rank on the 6 shared questions against the reference pool. $\gamma_R = \gamma^*$: the disclosed Random-Manager rate.
- γ_0: the prior performance belief, reported after the quiz. γ_1, γ_2: the performance beliefs reported after rounds 1 and 2.
- $\text{OC}_i = \mathbf{1}[\gamma_{0i} > \gamma_i^*]$: overconfidence, from the predetermined prior belief γ_0. Underconfidence (UC) is the complement.
- Easy_i: the randomly assigned easy quiz condition.
- θ_1: the belief that the round-1 manager is the Performance Manager, reported after observing tokens in round 1.
- θ_1^{chosen}: the belief that the manager *chosen* for round 2 is the Performance Manager. It equals θ_1 when the participant decides to stay with the round-1 manager and $100 - \theta_1$ when the participant decides to switch, so it is defined from the stated choice.
- θ_1^{faced}: the belief that the manager *faced* in round 2 is the Performance Manager. It equals θ_{1i} when the round-2 manager is the round-1 manager and $100 - \theta_{1i}$ otherwise.

- $\text{Active}_i \in \{0, 1\}$: the arm indicator, 1 in the active-choice arm and 0 in the control arm.
- $\lambda_{s(i)}$: stratum fixed effects.

5.1 H1: misguided selection

The manager selection motive is expected pay. A participant expects the Performance Manager to pay green at their current believed rate γ_1 and the Random Manager at the disclosed rate γ_R, so they prefer the Performance Manager when $\gamma_1 > \gamma_R$. Because $\gamma_R = \gamma^*$, this preference holds exactly when $\gamma_1 > \gamma^*$. We classify participants on the prior belief γ_0 rather than γ_1 because γ_0 is predetermined with respect to the round-1 tokens and is the belief the quiz-difficulty instrument shifts. The selection is misguided because the belief driving it is biased.

The primary outcome is $S_i = \mathbf{1}[\theta_{1i}^{\text{chosen}} > 50]$: the participant chooses the manager they believe is the Performance Manager. The secondary outcome is $\theta_{1i}^{\text{chosen}} - 50$, the strength of that selection.

First stage. We first establish that the easy condition shifts overconfidence:

$$\text{OC}_i = \pi_0 + \pi_1 \,\text{Easy}_i + \lambda_{s(i)} + u_i.$$

We report π_1 and the first-stage F statistic, and predict $\pi_1 > 0$: the easy condition raises the prior belief γ_0 and thus the share of overconfident participants.

Main test. For each outcome $Y_i \in \{S_i,\ \theta_{1i}^{\text{chosen}} - 50\}$ we estimate

$$Y_i = \beta_0 + \beta_1 \,\text{OC}_i + \lambda_{s(i)} + \varepsilon_i$$

by OLS and by 2SLS instrumenting OC_i with Easy_i, four regressions in total. H1 predicts $\beta_1 > 0$ in all four.

5.2 H2: asymmetric correction

The outcome is the round-1 to round-2 performance-belief correction,

$$C_i = |\gamma_{1i} - \gamma_i^*| - |\gamma_{2i} - \gamma_i^*|,$$

one observation per participant, where a positive value indicates a performance belief that moved closer to the true rate over the informed round. We use the round-2 movement rather than the change over the whole task because the asymmetry predicted by H2 operates through the informed selection made at the end of round 1: only after that selection does the manager the participant samples reflect their perceived ability-contingency.

The design sets the Random-Manager rate equal to the participant's true rate, $\gamma_R = \gamma^*$, so the two managers emit the same green distribution. The signal is therefore identical whether a participant faces the Performance or the Random Manager, and identical across the active and control arms.

H2a: facing the believed-Performance manager (mechanism, control arm)

Within the control arm, the realized manager is independent of the stated choice, so whether the participant faces the manager they believe is the Performance Manager is exogenous. We establish that facing that manager lowers bias. Because $\gamma_R = \gamma^*$, the manager carries no additional signal about ability, so the effect is belief-driven: the participant revises more when they think the manager is informative. This establishes the causal step linking H1 to H2.

Let $D_i = \mathbf{1}[\theta_i^{\text{faced}} > 50]$, with ties dropped in the binary form. Restricting to the control arm, we estimate

$$C_i = \beta_0 + \beta_1 D_i + \delta \, |\gamma_{1i} - \gamma_i^*| + \lambda_{s(i)} + \varepsilon_i,$$

and a dose version replaces D_i with $(\theta_i^{\text{faced}} - 50)$. We control for the entering bias $|\gamma_{1i} - \gamma_i^*|$ to absorb variation in the room available to correct beyond the variation controlled by the stratum fixed effects. We predict $\beta_1 > 0$ in both forms, and we verify that D_i is unrelated to overconfidence and to entering bias within the control arm.

H2b: asymmetric correction from realized selection (main test)

As the primary test, we estimate

$$C_i = \beta_0 + \beta_1 \, \text{OC}_i + \beta_2 \, \text{Active}_i + \beta_3 \, (\text{OC}_i \times \text{Active}_i) + \delta \, |\gamma_{1i} - \gamma_i^*| + \lambda_{s(i)} + \varepsilon_i$$

among all completers. All controls are pre-treatment, so the coefficients on Active and the interaction retain their experimental interpretation. The coefficient of interest, β_3, compares the causal effect of the active regime between overconfident and underconfident participants. H2 predicts $\beta_3 > 0$.

5.3 Robustness

We re-estimate every H1, H2a, and H2b specification on a restricted sample that excludes any participant with one or more first-try errors on the repeat-comprehension check, together with any participant flagged as a bot or AI agent by the Ebbinghaus trial (any answer other than the larger circle) or the honeypot field (any value). This restriction sharpens the mechanism among participants who understood the instructions and answered them themselves.