

Behavioral Economics

Lecture 7: Overconfidence and ABCs

Daniel Martin

danielmartin@ucsb.edu

ABCs and the Grether model

- ▶ In the Grether-style quasi-Bayesian model, the worker combines prior beliefs and AI information, but may *misweight* them

$$\log \left(\frac{\mu_R(s_1)}{\mu_R(s_2)} \right) = \alpha + \beta_1 \log \left(\frac{\Pr(R|s_1)}{\Pr(R|s_2)} \right) + \beta_2 \log \left(\frac{\mu(s_1)}{\mu(s_2)} \right)$$

- ▶ Interpretation:
 - ▶ α allows a baseline pull toward one state
 - ▶ β_1 governs the weight put on AI evidence
 - ▶ β_2 governs the weight put on one's own prior
- ▶ Overconfident workers: too much weight on own prior $\rightarrow \beta_2 > 1$

Side note

- ▶ This is different from Jin et al. (2022), where overconfidence meant $\beta_1 > 1$ (overconfident about reading of report)
- ▶ But overconfident prior here is likely the outcome of an overconfident belief about ability to guess age
- ▶ So the overconfident prior is the outcome a downstream updating process where the worker over-weighted their signal ($\beta_1 > 1$) when forming the prior

ABCs and the non-Bayesian models

- ▶ In the non-Bayesian interpretation, workers do not process the AI as a fully specified likelihood
- ▶ They move *part of the way* from their own prior toward an AI-implied anchor

$$\mu_R = \varepsilon_R \mu + (1 - \varepsilon_R) \rho_R \quad (\text{Contraction Rule})$$

$$\mu_R = w \mu + (1 - w) \pi_R \quad (\text{weighted Inertial Updating})$$

- ▶ ρ_R or π_R (or ϕ_R in the case of weighted subjective Inertial Updating) captures what the worker takes the AI message to mean
- ▶ ε_R or w captures *reliance on one's own view*
- ▶ Overconfidence: reliance on self is too high $\rightarrow \varepsilon_R$ or w is too high

Two possible explanations

Grether/QB view

- ▶ Worker is *trying to combine evidence correctly*
- ▶ But the evidence is *misweighted*
- ▶ Bias lives in distorted treatment of prior odds and signal odds

Non-Bayesian view

- ▶ Worker uses a *partial-adjustment rule*
- ▶ The AI rec is what the worker adjusts too
- ▶ Bias lives in *how far* the worker moves towards that rec

Follow-up questions:

1. Can both approaches explain the behavioral patterns in the ABCs paper?
2. Or is there a new analysis that would separate theories using the ABCs data?

Hard to separate here because don't elicit priors, as in Agarwal et al. (2023) and Kovach et al. (2026)

References I

- Agarwal, N., Moehring, A., Rajpurkar, P. & Salz, T. (2023), Combining human expertise with artificial intelligence: Experimental evidence from radiology, NBER Working Paper 31422, National Bureau of Economic Research. Revised November 2025.
- Jin, G. Z., Luca, M. & Martin, D. (2022), 'Complex disclosure', *Management Science* **68**(5), 3236–3261.
- Kovach, M., Martin, D. & Tserenjigmid, G. (2026), 'Learning from an unknown dgp'. Working paper.