Learning with Misspecified Models: the case of overestimation

Jimena Galindo

October 25, 2023

Abstract

I design a framework and a laboratory experiment that allows for the comparison of multiple theories of misspecified learning. I focus on a framework with endogenous information and a data-generating process ruled by two fundamentals: an ego-relevant parameter and a state. Within this framework I study three forces that can lead to misspecified beliefs: initial misspecifications, learning traps and biased updating. I find that biased updating is the main driver of misspecified beliefs in the lab. In addition, I vary the degree of ego-relevance of the parameter by introducing a stereotype treatment. The data is consistent with biased updating in both cases but for potentially different reasons: when learning about themselves, subjects attribute successes to their own ability and failures to luck. Instead, in the stereotype treatment, they compensate for initial negative biases by over-attributing positive signals to the ability of others. This translates into similar observed choices but different dynamics in beliefs.

1 Introduction

A growing body of literature in economics explores how people develop incorrect beliefs about fundamentals. Most of this research centers on scenarios where agents passively observe the world, and incorrectly integrate the received information into their beliefs.[1] However, many real-world situations cast agents as active participants in the generation of information. In these cases, the information they observe is influenced by their actions and subsequent behavior is in turn determined by how they incorporate the information into their beliefs.

As an example consider a student who needs to decide how much effort to put into studying for an exam. Their decision will depend on two factors: their belief about their intrinsic ability, and their belief about how difficult the exam will be. The outcome they observe will be affected by how much they decide to study. Imagine that the student puts in a moderate amount of effort and gets a surprisingly good grade. Did they get a good grade because they are smarter than they thought? Or because the exam was easier than they had anticipated? Their future exam preparation strategy will depend fundamentally on which line of reasoning they take. This feedback loop is referred to as an *endogenous information process* and is at the center of the forces I study.

To understand what the main forces at play are, I compare a set of theories that model learning in settings with endogenous information, and which can rationalize the persistence of misspecified beliefs. I develop a unifying framework that nests multiple theories of learning and generates testable predictions for each of them. Then I cast this framework in a laboratory experiment and test the predictions to identify which of the theories are consistent with the behavior I observe in the lab. The experiment features an agent who needs to learn two parameters: one that pertains their own characteristics (an *ego-relevant* parameter), and an exogenously determined state—in the context of the student preparing for exams, the two parameters would be their intrinsic ability and the difficulty of the exam.

Misspecified beliefs about an ego-relevant parameter are often referred to as over-

[1] See Benjamin [2019] for a review of the literature on errors of probabilistic thinking.

confidence (or underconfidence) and have been documented by behavioral scientists[2] and economists[3] in a variety of settings. Oster et al. [2013] find that subjects who are at high risk of having Huntington's disease overestimate their probability of being healthy and make retirement decisions as if they were healthy. Hoffman and Burks [2020] show that workers overestimate the quality of their match to their current employment and are unlikely to look for other opportunities. Camerer and Lovallo [1999] find that entrepreneurs are overconfident about the quality of their enterprise, which leads to excessive entry and early exit from markets.

In all of these examples, holding an incorrect belief about a fundamental leads to suboptimal choices with potentially high costs. In spite of the abundance of evidence, the scope of the existing research is limited in terms of the frameworks it considers. Most of the experimental evidence documenting the bias is collected in settings where subject are passive learners.[4] They observe a noisy signal and report their beliefs over subsequent rounds.[5] Although these studies provide important insights, they are not flexible enough to incorporate the richer theories that have been proposed more recently. In particular, they do not allow the study of endogenous learning or for learning about multiple parameters at once.[6]

In my experiment, I move away from the standard framework of passive learning to analyze a richer set of learning mechanisms. In particular, the interaction between the two parameters together with an endogenous information process gives rise to three possible mechanisms that allow for the persistence of incorrect beliefs: the presence of learning traps, incorrect initial beliefs, and misattribution bias. The theories that I consider incorporate different combinations of these mechanisms.

When the setting features learning traps, even an agent who incorporates all information

[2]Kelley and Michela [1980] provides a review of the psychology literature.

[3]See Benjamin [2019] for a review of the literature in economics.

[4]Götte and Kozakiewicz [2022] and Ozyilmaz [2022] are exceptions that study settings with endogenous information processes.

[5]Bracha and Brown [2012] and Möbius et al. [2022] are some examples.

[6]Coutts et al. [2020] studies an environment with an ego-relevant parameter and an exogenous state but does not incorporate the endogenous information process.

correctly, may fall into learning traps as outlined by Hestermann and Yaouanq [2021]. These traps are characterized by a combination of an incorrect belief and an optimal action which produce information that confirms the incorrect belief. Once an agent falls into a trap, the belief will be stable and even with a correctly specified model of learning, they will not be able to abandon their misspecified beliefs. If, the agent is dogmatic about their initial belief, Heidhues et al. [2018] show that they will inevitably fall into a trap and thus will be able to rationalize and sustain their initial misspecification belief.[7]

Ba [2023] moves away from dogmatism and endows the agents with a mechanism through which they can abandon incorrect beliefs; this allows them to avoid falling into learning traps. To do so agents perform Bayesian hypothesis tests that evaluate which is the more likely parameter out of two possibilities. By doing this, she characterizes the set of situations in which, even agents who consider alternative paradigms, may become trapped.[8]

Lastly, misattribution bias is the more classical explanation and has been widely studied in behavioral science.[9] Agents who suffer from misattribution bias will attribute successes to their own ability—the ego-relevant parameter—and failures to bad luck—the state. Under this model of learning, even an agent who initially has a correct initial belief may become overconfident if they observe a sequence of successes. In this case, the main driver of the bias is not an initial misspecification or the presence of learning traps, it is the updating procedure itself.[10]

These theories provide the main building blocks for a simplified framework that can be directly implemented in a laboratory experiment. In the experiment subjects make choices and receive feedback that depends on their own ability, an exogenous parameter and the choice they made. I track their choices as well as their beliefs about their own ability. The goal is to identify which of the 3 forces—the presence of learning traps, misspecified initial

[7] Götte and Kozakiewicz [2022] study the case of agents with dogmatic initial beliefs in a laboratory experiment.

[8] A similar mechanism is proposed by Schwartzstein and Sunderam [2021] in a setting with persuasion.

[9] See Kelley and Michela [1980] for a review.

[10] A more general framework that can be used to model this bias has also been proposed by Brunnermeier et al. [2005] and empirically studied by Bracha and Brown [2012].

beliefs, or misatribution bias—better explains the observed behavior. To determine the fit of the models, I compare the behavior predicted by each theory to the benchmark given by the fully Bayesian updating procedure.

I also study whether the learning mechanism is inherently linked to the ego-relevance of the parameters or if it is a more general phenomenon. I vary the degree of ego-relevance by introducing a treatment in which subjects learn about the ability of another participant. In this treatment, the participants know only the gender and nationality of the other, and thus can induce stereotypes—a different type of misspecification.

If correct learning about the parameters happens at higher rates in the stereotype treatment, it would suggest that the bias is intrinsically linked to the ego-relevance of the parameter. In contrast, if similar biased behavior arises in both treatments, it is more likely that the main driver of these types of misspecified beliefs is the updating procedure itself or the endogenous information process.

Although some agents do fall into learning traps, I find that the behavior of most subjects is better explained by misattribution bias: good news are treated as signaling high ability, while bad news are attributed to a low state. I also find that misattribution is no more prevalent in the ego-relevant than in the stereotype condition. This suggests that the main driver of the misspecification is the updating procedure; however, the underlying mechanism by which the bias is generated may be different in both treatments: While in the ego-relevant condition subjects prefer to hold themselves in high esteem, in the stereotype condition updating seems to be driven by some sort of bias overcorrection—when subjects realize that they underestimated the ability of another participant based on their gender and nationality, they compensate by overestimating their ability.

Finally, I estimate the structural parameters of the models to study model-heterogeneity in the sample. I find that even at an individual level the behavior is better explained by a general model of misattribution bias for most subjects. There is a smaller group of subjects that can be better explained through dogmatic beliefs and hypothesis testing and none of

them behave in line with the fully Bayesian benchmark.

In what follows I first discuss the theoretical framework and the predictions of each of the theories. Then I introduce a unifying example and my hypotheses. In section 4 I describe the experimental design and in section 5 I present the data and the results. Section 6 outlines the estimation of the parameters and the model fit analysis.

References

Cuimin Ba. Robust misspecified models and paradigm shifts. 2023.

Daniel J. Benjamin. *Errors in probabilistic reasoning and judgment biases*, pages 69–186. 2019. doi: 10.1016/bs.hesbe.2018.11.002.

Anat Bracha and Donald J. Brown. Affective decision making: A theory of optimism bias. *Games and Economic Behavior*, 75:67–80, 5 2012. ISSN 0899-8256. doi: 10.1016/J.GEB.2011.11.004.

Markus K Brunnermeier, Jonathan A Parker, Andrew Abel, Roland Bénabou, An-Drew Caplin, Larry Epstein, Ana Fernandes, Christian Gol-Lier, Lars Hansen, David Laibson, Augustin Landier, Erzo Luttmer, Sendhil Mullainathan, Filippos Papakonstantinou, Wolfgang Pesendorfer, Larry Samuelson, and Robert Shimer. Optimal expectations. *The American Economic Review*, 95:1092–1118, 2005.

Colin Camerer and Dan Lovallo. Overconfidence and excess entry: An experimental approach. *American Economic Review*, 89:306–318, 3 1999. ISSN 0002-8282. doi: 10.1257/aer.89.1.306.

Alexander Coutts, Leonie Gerhards, Zahra Murad, Kai Barron, Thomas Buser, Tingting Ding, Han Koh, Yves Le Yaouanq, Robin Lumsdaine, Cesar Mantilla, Luis Santos Pinto, Giorgia Romagnoli, Adam Sanjurjo, Marcello Sartarelli, Peter Schwardmann, Sebastian

Schweighofer-Kodritsch, Séverine Toussaert, Joël Van Der Weele, and Georg Weizsäcker. What to blame? self-serving attribution bias with multi-dimensional uncertainty. 2020.

Lorenz Götte and Marta Kozakiewicz. Experimental evidence on misguided learning *. 2022.

Paul Heidhues, Botond Kőszegi, and Philipp Strack. Unrealistic expectations and misguided learning. *Econometrica*, 86:1159–1214, 2018. ISSN 0012-9682. doi: 10.3982/ecta14084.

Nina Hestermann and Yves Le Yaouanq. Experimentation with self-serving attribution biases. *American Economic Journal: Microeconomics*, 13:198–237, 2021. ISSN 19457685. doi: 10.1257/mic.20180326.

Mitchell Hoffman and Stephen V. Burks. Worker overconfidence: Field evidence and implications for employee turnover and firm profits. *Quantitative Economics*, 11:315–348, 2020. ISSN 1759-7323. doi: 10.3982/QE834.

Harold H Kelley and John L Michela. Attribution theory and research. 1980. URL www.annualreviews.org.

Markus M. Möbius, Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. Managing self-confidence: Theory and experimental evidence. *Management Science*, 68:7793–7817, 11 2022. ISSN 0025-1909. doi: 10.1287/mnsc.2021.4294. URL https://pubsonline.informs.org/doi/10.1287/mnsc.2021.4294.

Emily Oster, Ira Shoulson, and E. Ray Dorsey. Optimal expectations and limited medical testing: Evidence from huntington disease. *American Economic Review*, 103:804–30, 4 2013. ISSN 0002-8282. doi: 10.1257/AER.103.2.804.

Hakan Ozyilmaz. Mental models and endogenous learning. 2022. URL https://drive.google.com/file/d/1sNnaXye8p2bZ3Zzd36dYF0Dza47E26PG/view.

Joshua Schwartzstein and Adi Sunderam. Using models to persuade. *American Economic Review*, 111:276–323, 1 2021. ISSN 19447981. doi: 10.1257/aer.20191074.