Career Concerns with Costly Agent Assessment and Misspecified Learning: Theory and Applications

Federico Echenique^{*} Anqi Li[†]

This paper studies a model of career concerns with misspecified learning, focusing on how misperceptions of ability and effort productivity both influence and are shaped by players' strategic interactions.

The model features three players: an agent, an evaluator, and the market. The agent has an innate ability and can exert costly, hidden, efforts to influence their output. The evaluator can scrutinize the agent at a cost, with greater scrutiny lowering output errors and enhancing the informativeness of output as a signal of ability. Based on the realized output, the market rewards the agent for their ability and possibility their effort-driven output.

While the agent correctly perceives their mean ability and effort productivity, the evaluator and market misjudge the agent's mean ability, while forming perceptions of their effort productivity. Misperceived effort productivity influences the evaluator's scrutiny level, which in turn affects the agent's effort choice and the resulting output distribution. In response to the observed output, the evaluator and market adjust their perceptions of effort productivity to better explain the data. This feedback loop between perception, scrutiny, and effort persists until a steady state is reached.

We formalize these interactions using a Berk-Nash equilibrium framework. A Berk-Nash equilibrium comprises the evaluator's evaluation policy, the agent's effort choice, and the evaluator and market's perception of the agent's effort productivity. In equilibrium, the evaluation policy is optimal given the perceived productivity, and the perception minimizes the Kullback-Leibler divergence between perceived and actual output distributions generated by the evaluation policy and effort. This equilibrium concept has a misspecified Bayesian learning foundation, where the evaluator and

^{*}Department of Economics, University of California, Berkeley, fede@econ.berkeley.edu.

[†]Department of Economics, University of Waterloo, angellianqi@gmail.com.

market hold misspecified prior beliefs about effort productivity and update their beliefs over time based on realized outputs. The limiting posterior beliefs as time approaches infinity correspond to the locally asymptotically stable equilibria of our game (in the sense of Lyapunov).

Our main results highlight how ability misspecifications propagate through the feedback loop involving evaluation policies and perceived effort productivity. Depending on whether the evaluator overestimates or underestimates the agent's mean ability, the perceived effort productivity acts as either a substitute or a complementary explanation for the discrepancy between perceived and actual output data. In the case of complementarity, our game often generates multiple equilibria, including some that are unstable (in the sense of Lyapunov) and others featuring significant distortions, even when the initial misspecification is small. In these scenarios, we use local asymptotic stability as the equilibria. We identify conditions that robustly reduce the distortions across all locally asymptotically equilibria. These include: slight corrections of the initial misspecification, or slight increases in scrutiny levels across all perception levels, driven by changes in the exogenous environment.

Our findings reveal new mechanisms that drive differential outcomes among groups of individuals who are identical in reality but are subject to systematically biased beliefs about their innate ability. When applied to contexts like discrimination in education and labor markets, our model predicts that gender and racial majorities tend to be over-rewarded for their ability and under-rewarded for their effort (relative to their intrinsic contributions), whereas the opposite holds for minorities. Minorities may out-earn majorities, but only due to inflated perceptions of their effort productivity and the increased scrutiny they face.

Additionally, our results offer actionable insights into reducing distortions and improving agents' actual outcomes and welfare. They also highlight the consequences of different affirmative action policies. Within our framework, we demonstrate that imposing a color-sighted, impartial evaluation requirement effectively halts the belief cascades that lead to multiple, "bad," equilibria in cases of complementarity. In contrast, color-blind evaluation fails to achieve the same effect. These findings underscore the importance of tailored policy interventions to mitigate biases and foster equitable outcomes.