Cover Letter

My name is Bnaya Dreyfuss and I am writing in application to the NBER Digitization Tutorial 2023. I am currently a 3rd year Phd student at the Harvard Economics department. My main advisors are Benjamin Enke, Jesse Shapiro and David Yang. At Harvard, I took classes satisfying field requirements for public economics, IO, and behavioral economics. Before joining Harvard I completed my B.A. and M.A. in economics at the Hebrew University of Jerusalem.

I am interested in applying behavioral insights and models to important policy-relevant contexts. In particular, these applications involve automation and Human-Algorithm interaction. I have been actively pursuing this kind of work for the past few years.

The first line of work I have been interested in uses behavioral economics to explain frictions in technology adoption. My interest in this topic started when I joined a group of economists and engineers from Hebrew University and Cornell, on an interdisciplinary research project examining reference dependence and human-robot interaction (HRI). With robots entering the workforce and automation playing an increasingly defining role in the economy, we can expect situations in which people's wage is affected by the performance of their robot co-workers. The project explores the role of expectations in these new work environments. In the first paper (Kshirsagar et. al 2019), which garnered considerable media attention (e.g., Forbes), we adapted an existing human-human experimental paradigm (Gill and Prowse, 2012) to a human-robot context. We documented a "discouragement effect" where robots' performance had a significant negative effect on humans' effort, consistent with the model's predictions. In a follow-up project we compare a collaborative (rather than competitive) design and show that this effect disappears and perhaps reverses as we switch to an economic setting with additive compensation scheme.

The second topic I have been working on applies behavioral economics school-assignment algorithms. The first project started as my MA thesis at Hebrew University, and has recently been published at AEJ:Micro (Dreyfuss et al., 2022a). It seeks to explain recent puzzling evidence showing that a large share of school applicants do not report their true preferences in strategy-proof matching mechanisms. Evidence for such behavior has been documented, using observed choices (Shorrer and Sóvágó, 2019), surveys (Larroucau et al., 2021) and lab experiments (Rees-Jones and Skowronek, 2018). Our project uses a model of expectations-based reference dependent preferences (Kőszegi and Rabin, 2006, 2007 and 2009) to show that such puzzling behavior might be partly intentional: if applicants anticipate feeling disappointed when not being matched with their top choices and factor it into their decision-making, they might realize that by reporting truthfully they would be setting themselves up for future disappointment, which can be averted by avoiding "reach" (i.e., prestigious) programs, and reporting high-probability "safe" study programs at the top of their list instead. Drawing on experimental data from Li (2017), we show that the behavior patterns found in his experiment are indeed remarkably well-predicted by a class of models of expectations-based loss aversion.

In a follow-up project (Dreyfuss et al., 2022b), we explore how we can redesign these assignment mechanisms to increase the number of applicants applying to the most prestigious study programs. We theoretically show that when the number of applicants is sufficiently large, an assignment mechanism where candidates respond to admission offers from programs (instead of ranking them in advance),

completely eliminates the loss-aversion channel, and thus this assignment mechanism remains strategy proof even for extremely loss-averse applicants. The data strongly support our hypotheses.

In addition to robots and algorithms, I am studying frictions in the adoption of artificial intelligence (AI) technologies. In some tasks, AI has long overtaken the best human minds: examples include chess, driving, or image processing. In healthcare, AIs have recently allowed us to reach unparalleled diagnosis accuracy rates, for some classes of diseases; yet, some evidence points to a substantial lag in adoption of these technologies. Our proposed project seeks to explore the role of trust in algorithms as a potentially crucial barrier to adoption. Trust is a complex and dynamic phenomenon, too often considered as a black box; we wish to investigate one of its mechanisms: how it evolves with evaluation of past performance.

We start from an intuitive observation that people seem to forgive humans much more easily than machines, for the same mistakes. As we observe a machine making errors, we often seem prompt to view it as "broken" and to stop trusting it. In contrast, trust in humans seems more resilient: it usually takes a history of unreliability for us to stop trusting them completely. Therefore, the main goal of this project is to understand and quantify gaps in trust resilience, in the ultimate hope of redesigning AI-based recommendation systems in a way that would help reduce those gaps.

At present, we are in the process of running pilots and finalizing our experimental design, and we are hoping to run our first pilot on a population of physicians in the next few weeks.

I am excited by the opportunity to work alongside top students and researchers at the NBER Digitization Tutorial. As I hope I have demonstrated above, important policy-relevant applications of behavioral economics in the context of automation and Human-Algorithm interactions are at the core of my research agenda. I hope that my participation in the workshop will help me identify new sets of questions and topics to explore, and potentially will ignite exciting collaborations with other students and researchers.

Sincerely,

Bnaya Dreyfuss

Department of Economics, Harvard University

Email: bdreyfuss@g.harvard.edu

References:

Dreyfuss, Bnaya, Ori Heffetz, and Matthew Rabin. 2022. "Expectations-Based Loss Aversion May Help Explain Seemingly Dominated Choices in Strategy-Proof Mechanisms." American Economic Journal: Microeconomics, 14 (4): 515-55.

Dreyfuss, Bnaya, Ofer Glicksohn, Ori Heffetz, and Assaf Romm. 2022. "Deferred Acceptance with News Utility." NBER WP 30635.

Gill, D. and Prowse, V. (2012). A Structural Analysis of Disappointment Aversion in a Real Effort Competition. American Economic Review, 102(1):469–503.

Kőszegi, B. and Rabin, M. (2006). A Model of Reference-Dependent Preferences. The Quarterly Journal of Economics, 121(4):1133–1165.

Kőszegi, B. and Rabin, M. (2007). Reference-Dependent Risk Attitudes. American Economic Review, 97(4):1047–1073.

Kőszegi, B. and Rabin, M. (2009). Reference-Dependent Consumption Plans. American Economic Review, 99(3):909–36.

Kshirsagar A., Dreyfuss B., Ishai, G., Heffetz, O. and Hoffman, G. (2019). Monetaryincentive competition between Humand and Robots: Experimental Results. Proc. of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI'19), IEEE.

Larroucau, T. Martínez, M. Neilson, C. and Rios, I. (2021). Application Mistakes and Information Frictions in College Admissions. Working paper.

Li, S. (2017). Obviously Strategy-Proof Mechanisms. American Economic Review, 107(11):3257–87.

Rees-Jones, A. and Skowronek, S. (2018). An Experimental Investigation of Preference Misrepresentation in the Residency Match. Proceedings of the National Academy of Sciences, 115(45): 11471-11476.

Shorrer, R. and Sóvágó, S. (2019). Obvious Mistakes in a Strategically Simple College-Admissions Environment. Working paper.