Research Statement

Overview

I am a behavioral economist who specializes in examining how cognitive biases and social beliefs influence individual decision-making, our interactions with others, and our interpretation of our experiences. A hallmark of my approach is the integration of theoretical modeling with empirical studies to offer a well-rounded understanding of human behavior. My work incorporates ideas from psychology—such as errors in projection, context-dependent preferences, and mental simulation—which I explore across a wide variety of settings and applications. I highlight ways that non-traditional preferences and biases shape economic behavior and how these often have nuanced implications for behavior. For instance, my research has practical applications in various domains: it offers strategies for firms to minimize turnover through better workload allocation, suggests methods for organizations to enhance charitable giving, and provides insights for advertisers to optimize product marketing. I discuss my major lines of research and some of the implications below. Although my primary goal is to further our understanding of errors in beliefs that shape decision-making in economic settings, the insights that my papers provide are easily applied across a variety of microeconomics fields.

My work to date follows three major themes. The first theme of my research focuses on how a person's preferences are constructed—rather than static or innate—and the importance of context in decision-making. The conventional model in economics suggests that a person's choices should not vary with innocuous changes in the environment in which the choices are made or if choices are presented in informationally equivalent ways. Guided by work in psychology (see, e.g., Slovic and Lichtenstein 2006) I have reexamined this view using both theory and experiments. The second theme involves how reference-dependent preferences themselves are constructed, and how sensations of gain or loss can shape memories and subsequent beliefs. Regarding the former, I present novel evidence that preferences over risky lotteries are domain specific. Regarding the latter, I present a previously unexplored interplay between the psychological factors of reference-dependence and attribution bias. Each paper contributes to the behavioral economics literature in distinct yet complementary ways. The first paper dives deep into the theoretical underpinnings of how individuals learn from experience, particularly focusing on the role of attribution errors when learning is shaped by reference-dependent preferences. It lays out a formal model that captures how misattribution can distort the learning process over time, eventually leading individuals to form biased beliefs about the intrinsic value of prospects. The second paper builds on this theoretical framework but shifts the lens to the empirical domain. There, I present findings from two real-effort experiments designed to test whether individuals wrongly attribute sensations of elation or disappointment to the intrinsic (dis)utility of working on a task. The third theme involves how we understand other people. Thinking about others is central in economic models, but decades of research in psychology suggests systematic ways in which we may fail to accurately portray others in our mind's eye. Across three papers (and ongoing future work) I have explored how (i) neurobiological factors shape our decision-making about others; (ii) our predictions about others are shaped by our own internal states; and (iii) how our inferences about others are shaped by our (mis)interpretations of their actions.

I. Constructed Preferences and Context Dependence

In **Pavlovian Processes in Consumer Choice** (*The American Economic Review*), we study whether the form in which items are displayed at the time of decision affects the dollar value that subjects place on them. Using a Becker-DeGroot auction under three different conditions— (i) text dis-

plays, (ii) image displays, and (iii) displays of the actual items—we find that subjects' willingness-to-pay is 40–61 percent larger in the real than in the image and text displays. Furthermore, follow-up experiments suggest the presence of the real item triggers preprogrammed consummatory Pavlovian processes that promote behaviors that lead to contact with appetitive items whenever they are available. Our results speak to practical marketing concerns of firms. For instance, we help to explain companies' efforts to find the right packaging and display for their products. In particular, our results support the ongoing movement in retail to display real products to consumers and allow for more sensory interaction between consumers and products.

In **A Model of Relative Thinking** (*The Review of Economic Studies*), we take a significant step toward formalizing the intuitive but underexplored concept of "range-based relative thinking." While the notion that people evaluate options relative to the context is not new (see, e.g., Parducci 1965; Mellers and Cooke 1994; Bordalo et. al 2012; Kőszegi and Szeidl 2013), we present a novel formulation—inspired by neuroscience (e.g., Louie, Khaw, and Glimcher 2013) to capture this behavior. We depart from existing literature by introducing a new set of assumptions about how people weigh different dimensions of a choice. Specifically, our main assumption is that individuals give less importance to a consumption dimension when the outcomes along that dimension have greater variability in the available choice set. This assumption suggests novel economic implications. For example, our model yields predictions about labor choices under income uncertainty, suggesting that a worker will be less motivated to exert effort for a fixed return when the range of potential incomes is broader. This finding could have meaningful implications for understanding behavior in labor markets, particularly in gig economy settings where income is highly variable. In addition to its predictions, a selling point of our framework is designed to be versatile. It can be applied in deterministic settings, but also extends naturally to scenarios involving uncertainty. This flexibility allows the model to be used in a wide range of economic contexts, making it a valuable tool for both empirical researchers and policymakers. Finally and critically, we validate the model's predictions through experimental evidence, thereby bridging the gap between theory and empirical observations. The results corroborate the model's predictions about how the range of available options impacts people's sensitivity to fixed differences within that range.

Perhaps an overlooked contribution of this paper is presented in Online Appendix D, where we offer the reader a guide to unifying our approach to that of Kőszegi and Szeidl (2013). These two models *seem* contradictory: Kőszegi and Szeidl suggest that a wider range along an attribute *increases* rather than diminishes its weight in the choice process. Our approach to unification is simple: in cases where people's attention is directed to the relevant dimensions, we suggest relative thinking dominates. In complex situations where dimensions may be neglected, the bigger-range-increases-incremental-weight hypothesis might be the dominant force. Our Appendix D offers a formal model which accommodates both forces in a tractable-yet-compelling manner.

II. Reference-Dependent Preferences

In **Learning with Misattribution of Reference Dependence** (*Journal of Economic Theory*), we explore how individuals learn from personal experience. We focus on how that learning can be distorted by an error of attribution whereby a person wrongly attributes feelings stemming from reference-dependent evaluations to her intrinsic tastes. The paper fills a significant gap in existing literature: while research has documented that deviations from expectations shape our experienced utility, little attention has been paid to how people might learn in environments with such preferences. We posit that people may misattribute these sensations when forming beliefs about a product, service, or job. The paper presents a dynamic model in which an individual's beliefs about the expected value of a prospect are influenced by such misattribution. This model produces

a variety of belief dynamics, including contrast effects and overly-volatile beliefs, which help to explain observed errors in decision-making. It provides both short-run and long-run results, illustrating how misattribution can lead to abandonment of learning or the formation of pessimistic beliefs. This latter result is not obvious in our setting. Although we focus on a setting where outcomes are truly i.i.d., belief convergence is not an immediate result of the law of large numbers because encoded outcomes are serially correlated: prior outcomes shift the agent's reference point, thereby influencing the current encoded outcome. We show that reference-dependence alone is not the driver of long-run distortions; rather loss aversion is a key determinant. The paper concludes by emphasizing the role of "expectations management" as a strategy to influence a person's subsequent beliefs, a finding that has direct applications in marketing and human resources.

Although beyond the scope of the main model, we present an important extension in Online Appendix C, where we consider settings under which the agent's actions influence the distribution of data she receives. We assume that (in each period) the agent takes an action to maximize her expected utility according to her true utility function conditional on her erroneous beliefs. This seemingly benign extension introduces the concept of a *biased-belief personal equilibrium*, as it extends Kőszegi and Rabin (2006) to the case where the agent holds erroneous expectations.

In **Reference Dependence and Attribution Bias: Evidence from Real-Effort Experiments** (*American Economic Journal: Microeconomics*), we conduct real-effort experiments to empirically investigate the role of expectations-based reference dependence in shaping willingness to work. The paper is framed around a key question: do people correctly understand that their reference-dependent experiences—of either elation or disappointment—will not persist in the future? The experiments manipulate initial expectations about the task and then measure how those expectations affect future willingness to work. The results confirm that individuals indeed misattribute their reference-dependent sensations to the intrinsic utility of a task. In particular, when individuals work on a task that exceeds their expectations, hours later they overestimate its intrinsic value and are more willing to work on it. Conversely, if the task falls short of expectations, they underestimate its value and are less willing to engage with it later. In an Appendix, we highlight that—perhaps surprisingly—this basic result is not predicted by Kőszegi and Rabin (2006; 2007). Our evidence is important in two ways. First, it provides support for expectations-based reference points. Second, it enriches our understanding of how expectations shape labor supply decisions and has implications for task design and incentives in organizational settings.

More recent work seeks to bridge the gap between my first and second thematic areas. In **Risk and Loss Attitudes Vary Across Domains** (*Working Paper*) I explore how people assess risks in different settings. While previous research has emphasized the ways individuals perceive and respond to risks over monetary outcomes, our everyday lives brim with non-monetary uncertainties that also shape our choices. Are our risk attitudes consistent across monetary and non-monetary contexts? This paper offers experimental evidence suggesting context-dependent disparity in risk perception between monetary and non-monetary domains. My findings are threefold. Firstly, participants showed a higher degree of loss aversion when facing non-monetary outcomes compared to monetary ones. Secondly, I observe less variation in preferences across participants in the non-monetary domain than in the monetary one. Lastly, preferences exhibited noticeable curvature within the non-monetary domain but relatively little in the monetary domain. These findings suggest the importance of examining context when thinking about risk.

III. Thinking About Others

In **A Neurocomputational Model of Altruistic Choice and Its Implications** (*Neuron*), we propose

a neurocomputational model and empirically validate it with behavioral and fMRI data. Previous research focused on altruism through the lens of strategic considerations or psychological theories; this work integrates neuroscientific insights to explain how altruistic choices are made. The model posits that altruistic choices are determined by calculating an overall value based on two specific attributes: the benefits for oneself and the benefits for others. It employs a multi-attribute version of the Drift-Diffusion Model (DDM) to account for the noise in value signal computations and to make predictions about both choice and reaction time. Specifically, we explain when and why generous choices are slower or faster than selfish choices, and why they produce greater response in TPJ and vmPFC brain regions without invoking competition between automatic and deliberative processes or reward value for generosity. We then directly explore the neural substrates of this decision-making process, identifying specific brain regions responsible for computing these attribute values and their integration. Through rigorous fMRI data analysis and simulations, we confirm our theoretical predictions and offers novel insights into individual variations in altruistic behavior. We offer a perhaps surprising punchline to those familiar the intuitive two-system approach to altruistic choice (see, e.g., Kahneman 2011): generous decisions are sometimes unintended mistakes resulting from the noisy choice process rather than genuinely pro-social preferences.

In **Failures in Forecasting: An Experiment on Interpersonal Projection Bias** (*Management Science*), we investigate people's ability to predict about others. Specifically, we explore predictions about others' willingness to work (WTW) on a real-effort task. Our results reveal a pronounced "interpersonal projection bias" in these forecasts. We show that a predictor's current state—specifically, their level of fatigue—alters their predictions as if they believed others shared their tiredness. Our experiment involved two groups, "workers" and "predictors"; the latter guessed the WTW of the former. Tired predictors significantly underestimated the WTW of fresh workers, and conversely, fresh predictors overestimated the WTW of tired workers, with distortions ranging from 21% to 50%. We also provide a novel comparison between inter- and intrapersonal projection bias, finding that the former may actually be more pronounced. This work contributes to the existing literature by empirically measuring and dissecting "projection bias" in settings where thinking about others is crucial. Our findings can be readily applied to managerial decision-making and organizational behavior. For example, they caution that managers or decision-makers could make suboptimal choices in task allocation and incentive design if they do not account for projection bias, especially in contexts involving effort provision and fatigue.

In **Heterogeneous Tastes and Social (Mis)Learning** (*Working Paper*), we explore another domain for errors in understanding others: accounting for when others have different tastes from us. More specifically, we examine the degree to which people successfully engage in social learning when there is naturally occurring heterogeneity in tastes. To do so, we employed a three-stage field-in-the-lab experiment with sequential observational learning in which participants faced uncertainty about the nominal value of a gift card. Despite the complexities of the learning environment, we find support for social learning, and participants' inferences obey some simple comparative statics predicted by rational models. But we find significant and systematic departures from rational social learning. Our evidence can be broadly categorized as *under-inference* from behavior that ought to provide a strong signal about the underlying state and *over-inference* from behavior that provides a weak signal. We provide a framework for understanding such patterns as arising from errors in the beliefs about others. The paper is novel both in its basic exploration of social learning under varied tastes and also its approach to exploring natural (rather than induced) tastes. **IV.**

Ongoing Work

I conclude with a brief mention of ongoing experimental work. In **The Person or the Situation? Projection Bias and Inference about Others** (joint with Tristan Gagnon-Bartsch and Jeongbin Kim) we experimentally explore whether misinference about others leads to stereotyping. In **Autonomy and Adherence to Expert Advice** (joint with Xavier Giné, Aprajit Mahajan, and Enrique Seira) we use a Lego-building task—completed for speed—to explore whether and why expert advice is followed more when a person chooses to adopt it versus has it imposed upon them. Finally, in **Misunderstanding What Makes Other People Happy** (joint with Matthew Rabin) we explore a novel experimental paradigm to recover participants' beliefs about the correlates of others' happiness and provide simple non-parametric tests for biases in belief dynamics.

External References

BORDALO, P., GENNAIOLI, N., AND SHLEIFER, A. (2012). Salience Theory of Choice Under Risk. *Quarterly Journal of Economics*, 127(3), 1243–1285.

KAHNEMAN, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

KŐSZEGI, B., AND RABIN, M. (2006). A Model of Reference-Dependent Preferences. *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 SZEIDL, A. (2013). A Model of Focusing in Economic Choice. *Quarterly Journal of Economics*, 128(1), 53–104.

LOUIE, K., KHAW, M. W., AND GLIMCHER, P. W. (2013). Normalization is a general neural mechanism for context-dependent decision making. *Proceedings of the National Academy of Sciences*, 110(15), 6139–6144.

MELLERS, B. A., AND COOKE, A. D. J. (1994). Trade-Offs Depend on Attribute Range. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5), 1055–1067.

PARDUCCI, A. (1965). Category Judgment: A Range-Frequency Model. *Psychological Review*, 72(6), 407–418.

SLOVIC, P. AND LICHTENSTEIN, S. (2006). *The Construction of Preference.* Cambridge University Press.