Gershman, Samuel. What Makes Us Smart: The Computational Logic of Human Cognition. Princeton University Press, 2021. 205 pages. Paperback, $35.00.

An enigma of human cognition explored by "What Makes Us Smart: The Computational Logic of Human Cognition" is the ability for people to be rational yet prone to simple errors. Drawing on statistical and computational logic, Samuel Gershman advances its supremacy as an effective information-processing system subject to constraints on data and resources. Human cognition is governed by two fundamental principles in response to these limitations, which include inductive bias and approximation bias. While these biases are essential to intelligence, they inadvertently engender cognitive externalities, which Gershman claims are perceived as failures of the human mind. However, by analyzing situations associated with these cognitive shortcomings under an empirical and multidisciplinary lens, Gershman concludes that the apparent errors are actually indicative of an effective information-processing system at work.

A key focus of this book involves assessing the role of inductive bias in decision-making and error propagation, where inductive biases are described as constraints on hypotheses before observing data, or in statistical parlance, employing prior beliefs to bias posterior beliefs. Gershman claims this is necessary to deal with a theoretically infinite confluence of data, and appropriately notes that "people need very little data to make strong inferences about physical and mental causality," a sign of the effectiveness of this bias (p. 26). After exploring perceptual errors that are caused by this phenomenon observed in illusions, Gershman details the specific inductive biases people have, which include causality, compositionality, and object-orientation. He then turns attention to apparent errors that this bias contributes to and details why they evince rationality. For example, he introduces the concept of information cascades, where "the judgments of other people typically provide useful information, so using that information should lead to higher accuracy, on average" (p. 39). This runs counter to the notion of conformity as an

irrational phenomenon. Another instance of supposed inductive bias errors supporting rationality involves the positive test strategy as an optimal information acquisition method, where under set assumptions, the tendency to select information in accordance with one's hypotheses tends to be maximally informative. He then relates this to revising beliefs when there is disconfirmatory evidence, which is an interesting yet dense exploration into the Bayesian statistical underpinnings of hypothesis updating. Lastly, Gershman draws on evidence from multiple domains in subsequent chapters to further describe the nature of inductive biases. One interesting feature is his multidisciplinary approach in relating indifference curves in economics to the cognitive thought processes regarding preference inconsistency. His later commentary on intrinsic motivation as related to preference construction is an especially interesting analysis into how incentives can be demotivating, which is important for business and policy decisions.

After spending much of the book considering the nature of inductive bias, Gershman then focuses on approximation bias in the remaining chapters. Unlike the other cognitive principle, approximation biases are constraints on the thought process of cognitive systems based on limited resources, such as time, memory, and energy. To begin, he empirically assesses the cost of computation by introducing the concept of efficient information coding from an information theory standpoint. This computational analysis of efficient optimization can be dense for some readers, but his detailed considerations are admirable. His analysis of efficient coding regarding social and economic judgments is perhaps more relatable, where he illustrates the relative nature of judgments by noting the seemingly irrational phenomenon of people being risk-averse for gains and risk-seeking for losses. Gershman's focus on language design is also an effective example of approximation bias, where language seeks to balance informativeness and effort. He observes the importance of ambiguity in language, as it reduces effort when context provides

sufficient information about meanings. One interesting feature of this section is his comparison of different language constructions on the basis of this linguistic trade-off, where one extreme maximizes ambiguity at the expense of expressivity, while the other exhibits the converse. Lastly, Gershman explores the role of randomness in the context of approximation biases, where he makes several interesting claims. One claim is that the mind approximates Bayesian inference by sampling, similar to the Monte Carlo technique that he adequately describes. Another key point he makes here is the dilemma between exploration and exploitation, where an individual aims to maximize a reward over some time by gathering enough information yet exploring enough options. Gershman successfully conveys his point of randomness in cognition here.

This book focuses on the computational nature of human cognition and effectively asserts that phenomena ascribed as cognitive errors are in fact the result of sophisticated processes. By referencing multiple empirical studies, and providing extensive interdisciplinary analyses, Gershman adequately bolsters the claim of the mind's superiority. While this book covers many domains, some familiarity with statistical and computational concepts is recommended.

Mark Rapala
Marquette University
Milwaukee, Wisconsin