Tom Griffiths  
Professor, Information Technology, Consciousness and Culture  
Princeton University

Abstract:

Anybody trying to understand human decision-making faces a challenge: the two dominant perspectives on how human minds work are fundamentally at odds with one another. The classic approach of assuming that people are rational agents has the advantage of making it possible to derive generalizable models for any new circumstance, but has been shown to give a poor characterization of actual human behavior. By contrast, the “heuristics and biases” perspective that has replaced it provides explanations for the ways in which people deviate from rationality, but doesn’t give us an overarching framework that we can use to make predictions about how people will act in new situations. In this talk I will present a framework that reconciles these two perspectives, based on the idea of “resource rationality” — that human decision-making is impaired by limited cognitive resources, but that people nonetheless make rational use of the resources that are available to them. This framework can be used to capture some classic results from the heuristics and biases literature and provides a way to automatically derive new models of behavior with similar generality to classic approaches to rationality.