From preferences to choices and back again: evidence for human inconsistency and its implications

Christopher Lucas
Carnegie Mellon University

Charles Kemp
Carnegie Mellon University

Thomas Griffiths
University of California, Berkeley

Abstract: People's choices can be predicted given information about their preferences. Learning people's preferences is the inverse problem of inferring preferences from choices. Given the apparent relationship between choice prediction and preference learning, it is natural to ask whether the two are mutually consistent. Given weak assumptions, we show that no single, consistent model of the relationship between choices and preferences can explain both the choices people make and their inferences about others' preferences. This finding implies that people make systematic errors in learning about others' preferences, and indicates that some accounts of preference learning, e.g., those based on simulating choices as well as unconstrained rational models, are inadequate. We also consider alternative assumptions, which allow consistent models but require a new interpretation of decoy effects in multiattribute choice.