Research Statement

Normative decision theory aims to provide a formal account of instrumental rationality. This is the sort of rationality exhibited by agents who perform well in practical tasks, relative to their own goals and beliefs. Moral philosophy, by contrast, is often taken to be concerned with offering an account of a thicker sort of practical rationality exhibited by fully reasonable and socially concerned decision making agents. My research agenda is centered around exploring both sorts of rationality using the formal tools of mathematical decision theory, with the aim of yielding theoretical and practical insights of relevance to diverse fields including public policy and machine ethics.

A. Rational Choice

Given our status as planning creatures, many of the practical tasks we face are dynamic in nature, involving anticipated sequences of decisions and learning events. My early research has examined what practical rationality requires of agents facing such extended choice problems and has explored the significance of dynamic choice norms for various debates in Bayesian decision theory. Previous debates regarding such dynamic choice principles have been undertaken almost exclusively in the context of the variants of Bayesian decision theory most familiar to economists, e.g., the theories of Savage and von Neumann and Morgenstern. However, Bayesian philosophers have long recognized the greater elegance and generality of the approach pioneered by Richard Jeffrey, and so a key focus of my research has been to explore the significance of dynamic choice arguments in the context of Jeffrey-style Bayesianism. This project has resulted in several papers analyzing rational planning from a broadly Bayesian standpoint, including "Dynamic Consistency in the Logic of Decision" (*Philosophical Studies*), "Bradley Conditionals and Dynamic Choice" (*Synthese*, co-authored with Simon Huttegger), and "A Plan-Based Causal Decision Theory" (*Analysis*).

My fullest treatment of dynamic consistency norms in the context of Bayesian decision theories is presented in my essay, "Evidence, Causality, and Sequential Choice" (*Theory and Decision*), which yields an impossibility result: no decision theory that satisfies two plausible constraints can be dynamically consistent across the full range of standard dynamic choice problems. In a similar vein, my manuscript, "Deference and Decision" (co-authored with Daniel Herrmann, draft available upon request), asks whether any decision theory might at least satisfy a pair of highly intuitive deference principles that (roughly put) enjoin acquiescence to experts and evidence, respectively. The result is again negative but illuminating. These papers, however, like my earlier work, make the somewhat controversial assumption that agents can deliberate about what action to take whilst simultaneously holding definite credences regarding which action they will ultimately choose. In my paper, "Prediction Impairs Deliberation" (*Ergo*), I explore the implications of rejecting this assumption for a number of classic decision-theoretic puzzles.

While my investigations of rational choice have largely focused upon single-person decision theory, I have more recently turned my attention to strategic rationality in a joint project with Wolfgang Spohn and Mantas Radzvilas, "Dependency Equilibria: Extending Nash Equilibria to Entangled Belief Systems" (draft available upon request), where we rigorously investigate the oft overlooked implications of the causal vs evidential decision theory debate for non-cooperative game theory. Our paper extends standard justifications of the Nash equilibrium solution concept and shows that they apply to Spohn's more general *dependency equilibrium* concept, on the assumption that players choose their strategies according to the recommendations of evidential decision theory. These results have potential relevance for how we should think about not only human interactions but also the strategic behavior of artificial agents, whose algorithmic decision procedures may be especially prone to generating the kind of acausal correlational patterns that render the causal vs evidential decision theory debate significant.

B. Moral Choice

Within moral philosophy, Bayesian decision theory is often seen as tethered to a consequentialist outlook, leading deontologists to search for alternative guidelines for uncertain decision makers. My most recent line of research challenges this prejudice by exploring the extent to which Bayesian risk management procedures can be neatly integrated with traditional deontological moral principles, like the doctrines of double effect and doing and allowing. My essay, "A Causal Modeler's Guide to Double Effect Reasoning" (*Philosophy and Phenomenological Research*), initiates this project by developing a formal modeling framework within which moral decision problems are represented as normatively supplemented causal graphs and various consequentialist and double effect-inspired moral theories can be viewed as disagreeing over the inputs of a common decision rule that I dub *purified utility maximization.* (A current follow up project attempts to incorporate the distinction between doing and allowing harm into this picture as well.) From this modeling vantage point, it becomes evident that deontological moral theories of the relevant sorts, no less than consequentialist ones, can be naturally extended to accommodate empirical uncertainty on the part of moral agents in a fully Bayesian fashion, since purified utility maximization is easily restated as *expected* purified utility maximization.

This work has several important upshots beyond harmonizing two philosophically attractive strands of thought (Bayesianism and deontology) and thus defusing worrisome objections to both. First, the framework I have so far developed invites us to attend more carefully than we ordinarily might to the logical structure of potentially morally significant concepts like that of a *causal means.* A current project of mine in this vein employs recent work from computer science on *actual (token) causation* to refine and develop my earlier explication of causal means. Second, this work suggests new ways of approaching moral uncertainty. By situating rival moral theories within a common framework, the infamous problem of *intertheoretic value comparisons* can be rendered more tractable. Finally, supplying deontologists with formal decision procedures of the sort commonly taken for granted by consequentialists opens the door to broader application of deontological reasoning in settings where such procedures are prized (i.e., in fields like public policy and machine ethics). Whereas consequentialist decision procedures (whether under the guise of CBA, utilitarianism, etc.) run the risk of allowing prospective rewards to validate the use of morally tainted means, a rule like purified utility maximization aims to guide decision analysis in a manner that avoids achieving desired ends in illegitimate ways (e.g., via deception, exploiting the vulnerable, etc.), something that may prove especially critical in the training of potentially dangerous decision making algorithms for artificial agents.