Cambridge University Press
978-0-521-06356-2 - The Foundations of Causal Decision Theory
James M. Joyce
Excerpt
More information

Introduction
A Chance to Reconsider

So, you have decided to read this book. Are you sure that's wise? Reading a tome entitled *The Foundations of Causal Decision Theory* is a sure sign of interest in the study of rational choice, which is a very unfortunate interest to have. You probably have heard all the jokes about decision theorists who cannot make wise choices. (How many does it take to screw in a light bulb? Who knows; you can't draw up a decision matrix in the dark. What do you call a decision theorist with a bank balance? A meal ticket.) Unfortunately, there is more than a grain of truth in these jests. It is widely known, at least among people who have no interest in rational choice theory, that about half of those who study decision making are patsies who can be easily exploited. Here are some facts of which you may be unaware: Most business schools keep a few decision theorists on staff (and overpay them grossly) to make it easier to win money at the weekly poker game. Ninety percent of all swampland is owned by people who have read Leonard Savage's *The Foundations of Statistics*. In most banks one can get a loan without putting up collateral merely by showing that one has a product that appeals to *Econometrica* subscribers. No one has ever paid the full sticker price for a car who did not know that Daniel Bernoulli invented the concept of utility. What P. T. Barnum really said was that there is an *expected utility* theorist born every minute.

You've never heard of any of this? I'm not a bit surprised. It has been common practice in all cultures at all times to identify people with an interest in decision theory at an early age, and to keep them in the dark about their plight so as not to spoil the chances of easy exploitation. I am only being permitted to reveal this information now because it has recently been discovered that it does not matter whether or not a person realizes that he is a sucker. The tendency to make foolish choices has long been known to run in families. Until recently this was thought to be a matter of "nurture" rather than "nature." The teaching of sound decision-theoretic principles was believed to engender unsound decision making practices in one out of every two cases. Parents leaving copies of *The Theory of Games and*

Cambridge University Press
978-0-521-06356-2 - The Foundations of Causal Decision Theory
James M. Joyce
Excerpt
More information

Economic Behavior open for impressionable eyes to see, cautioning against playing the lottery, delivering stern lectures on the importance of not crying over spilt milk, and the like, were believed somehow to cause their children to buy retail, overestimate their odds in games of chance, and develop a love of decision theory.

Surprisingly, it turns out that being a sucker is congenital. There is a "patsy gene"! It causes its carriers both to have an interest in decision theory and to make lousy choices. If you have this gene, you can resign yourself to a life of misery and exploitation, since that is sure to be your lot. You will never get your money's worth. You will always be an easy mark for unscrupulous operators bent on exploitation. And, if your ship should happen to come in, you can be sure that someone will talk you out of it at the pier. The silver lining in this otherwise dark cloud is that you can enjoy the study of decision theory in peace. It will not make you any better at choosing actions, but it will not make you any worse either.

I am sure you are wondering how to tell whether or not you have the gene. At the moment there is no reliable biological test, so we are going to have to make inferences from your behavior. People with no interest in the theory of rational decision making always lack the gene. You are not one of these, however, since a noncarrier would never even pick up a book entitled *The Foundations of Causal Decision Theory*. Among people like you, who have some interest in rational choice theory, 50 percent have the gene and 50 percent do not. A history of poor decision making is a sure indicator of having the gene. About 49 percent of decision theorists fall into this class. But, this probably is not you. I suspect that you have made fairly wise decisions up to now. So, the good news is that the odds are 50 : 1 that you are not a carrier. The bad news is that you may be one of the unlucky 1 percent who have the virulent, late-onset form of sucker disease. These are the people who get talked out of their life savings on the telephone, who spend themselves into penury giving money to television preachers, and so on. How can you tell whether this is to be your lot?

As it turns out you are doing the one thing that is known to provide a foolproof test. Research has shown that this very book provides a completely reliable way of determining whether or not a person has the sucker gene. Anyone who chooses to read as far as page 3 has the gene; anyone who stops on page 2 or before lacks it. Research also shows that people who turn to page 3 tend to enjoy reading the book (if only for the fun of detecting all the fallacies), and reading it does not have any deleterious side effects. You are on page 2 now. Think carefully about what you should do next.

You have made a wise choice! Of course, I am sorry for you that you made it because it indicates that your life is going to be miserable. On the bright side, your life would have been miserable whether or not you turned the page, and you will enjoy reading the book (at least a little, for the fallacies). Too bad about the gene, but you cannot change your basic biological makeup and you may as well get as much pleasure as you can out of the bad hand you have been dealt. No use crying over spilt milk, as they say.

You have just faced a *Newcomb problem*. These are choice situations in which one option (e.g., not turning to page 3) reliably indicates the presence of some desirable state of affairs (lacking the patsy gene) *without doing anything to bring that state about*, while another option (turning the page) reliably indicates the presence of some undesirable state of affairs (having the gene) but, again, *without doing anything to bring that state about*. What makes the problem interesting is that this second, less propitious option has benefits not associated with the first (e.g., the pleasure of finding the fallacies). Let us call the first option the *auspicious act* since it serves as a sign or augury of favorable results. The second option will be the *efficacious act* since it plays a causal role in helping to secure the small side benefit. As we shall see later, there are a variety of choice problems with this curious structure.

There are two schools of thought on the issue of how rational agents should behave when faced with Newcomb problems. Proponents of *evidential decision theory* feel that the auspicious option should always be selected. Actions, they believe, ought to be evaluated in terms of the *evidence* they provide for thinking that desirable outcomes will ensue. Defenders of *causal decision theory*, on the other hand, claim that acts are best assessed on the basis of their ability to *causally promote* desirable outcomes. A rational agent, on their view, will always perform an act that is *maximally efficacious* in bringing about desirable consequences. Thus, while an evidential decision theorist would have encouraged you *not* to turn to page 3 of this book (and presumably would not have turned it herself), a causal decision theorist would have advised the opposite.

Both these theories characterize rational desire and choice in terms of *subjective expected utility maximization*. The common ground here is the Bayesian doctrine that the strengths of a rational agent's beliefs can always be measured by a *subjective probability function* P defined over *states of the world*, and the view, which comes down to us from Daniel Bernoulli, David Ricardo, John Stuart Mill, and others, that an agent's desires can be described in terms of a real-valued *utility func-*

tion u defined over *outcomes*. Following Richard Jeffrey, proponents of evidential decision theory maintain that the utility of an act A is best identified with its *news value*, which is given by

Jeffrey's Equation. $\mathrm{V}(A) = \sum_S \mathrm{P}(S/A)\mathrm{u}(\mathrm{O}[A, S])$

where S ranges over a set of mutually exclusive and jointly exhaustive "states of the world," where $\mathrm{P}(S/A) = \mathrm{P}(A \,\&\, S)/\mathrm{P}(A)$ is the decision maker's *subjective conditional probability* for the state S given A, and where $\mathrm{u}(\mathrm{O}[A, S])$ is the utility of the outcome that A would produce if it were performed when S obtained. $\mathrm{V}(A)$ captures the sense in which A provides the decision maker with evidence for thinking that desirable outcomes will ensue.

Advocates of causal decision theory think that actions should be chosen on the basis of their efficacy value, which is defined using

Stalnaker's Equation. $\mathrm{U}(A) = \sum_S \mathrm{P}(S \backslash A)\mathrm{u}(\mathrm{O}[A, S])$

Here $\mathrm{P}(S\backslash A)$ is a probability that is supposed to capture the decision maker's beliefs about the extent to which the act A is likely to causally promote the state S. (We will discuss the definition of $\mathrm{P}(\bullet\backslash A)$ at length in Chapters 5 and 6.)

In most cases there is no conflict between the evidential and causal approaches to decision making because acts usually indicate good results by *causing them*, which makes U-maximization and V-maximization equivalent. In Newcomb problems, however, indicating and causing come apart, and auspiciousness is no longer a reliable mark of efficacy. There has been a great deal of discussion of the differences between causal and evidential decision theories, and the broad (albeit not universal) consensus is that causal decision theory gets the answers right in situations where the two approaches disagree. It seems clear, for example, that denying oneself the pleasure of finding the fallacies in this book merely to give oneself evidence that one is not a congenital sucker is irrational since one gains nothing at all by doing it.

The difficulty is that from the theoretical point of view causal decision theory is something of a mess; it lacks an appropriate foundation. Evidential decision theory, in contrast, is the model of what a decision theory should be as far as foundational matters are concerned (or so it will be argued). The standard method for justifying any version of expected utility theory involves proving a *representation theorem* that shows that an agent whose beliefs and preferences satisfy certain axiomatically specified constraints will automatically behave as if she is maximizing expected utility as the theory defines it. Such a theorem ensures that the theory's concept of expected utility makes sense and that it can be applied across a broad range of decision situations. It is essential that a decision theory have a representation theorem before it can be taken seriously. As we shall see in Chapter 4, Ethan Bolker has proved a powerful and elegant representation theorem for evidential decision theory that sets the standard by which all other representation theorems should be judged. No similarly compelling result has yet been obtained to serve as a foundation for causal decision theory.

This leaves us in a difficult position. Our best account of rational decision making – the one that seems to give the right answers in Newcomb problems – lacks the minimum theoretical foundation necessary to justify its use, whereas the account that has an adequate theoretical underpinning – the one for which an acceptable representation theorem can be proved – sometimes gives wrong answers. Thus, decision theorists appear to be faced with a choice between an ill-founded theory with true consequences and a well-founded theory with false consequences.

Fortunately, these foundational differences are not as serious as they appear. I will show how to express Jeffrey's Equation and Stalnaker's Equation as instances of a general *conditional expected utility theory* whose defining equation is

$$\mathrm{V}(X\|A) = \sum_S \frac{\mathrm{P}(S \,\&\, X\|A)}{\mathrm{P}(X\|A)} u(\mathrm{O}[A, S])$$

where X is any proposition expressible as a disjunction of states, $\mathrm{P}(Y\|A)$ is the probability that an agent associates with a proposition Y when she supposes that she will perform A, and $\mathrm{u}(\mathrm{O}[A, S])$ is her utility for the outcome that would ensue if S were to obtain when A was performed. $\mathrm{V}(X\|A)$ is the news value associated with X *on the supposition that A is performed*. I will argue that any tenable theory of rational choice must postulate expected utilities that obey some version of this equation and that it should ask people to maximize, not the unconditional utilities of their acts, but the utilities of their acts *conditional on the supposition that they are performed*. In other words, a rational agent should always choose to perform an act A such that $\mathrm{V}(A\|A)$ is greater that $\mathrm{V}(B\|B)$ for any alternative B.

To see that this characterization of prudential rationality is broad enough to encompass both the evidential and causal approaches, notice first that by substituting the ordinary conditional probability $\mathrm{P}(\bullet/A)$ for $\mathrm{P}(\bullet\|A)$ one obtains an "evidentialist" conditional utility theory whose defining equation is

$$\mathrm{V}(X/A) = \sum_S \frac{\mathrm{P}(S \,\&\, X/A)}{\mathrm{P}(X/A)} \mathrm{u}(\mathrm{O}[A, S])$$

Similarly, substituting the causal probability $\mathrm{P}(\bullet\backslash A)$ for $\mathrm{P}(\bullet\|A)$ yields a notion of conditional expected utility appropriate to causal decision theory:

$$\mathrm{V}(X\backslash A) = \sum_S \frac{\mathrm{P}(S \,\&\, X\backslash A)}{\mathrm{P}(X\backslash A)} \mathrm{u}(\mathrm{O}[A, S])$$

Since these equations reduce to Jeffrey's Equation and Stalnaker's Equation, respectively, when $X = A$, evidential and causal decision theory can each be understood as different versions of a generalized conditional decision theory when the relevant conditional probabilities are properly interpreted.

What we gain by moving to conditional decision theory is a unified framework within which both evidential and causal decision theory can be expressed in formally similar terms. In the penultimate chapter of this book I prove a representation result for conditional decision theory that generalizes Bolker's theorem. Since this new theorem provides an equally secure foundation for both versions of decision theory there will no longer be any reason to prefer one to the other on purely formal grounds. In the first instance, this should comfort and encourage the causal decision theorists because they no longer need to worry about the foundational deficiencies of the approach. Evidential decision theorists gain something too, though, for one of the main morals I wish to draw is that there is a deep sense in which Jeffrey's theory is exactly right: *all value is a kind of news value* even if not all kinds of news value are relevant to the choice of actions.

The plan of the book is as follows: Chapters 1–3 comprise a general introduction to expected utility theory that is meant to prepare readers for the discussion of evidential and causal decision theory that takes place in Chapters 4–7. Chapter 1 provides a quasi-historical introduction to expected utility theory as it applies to casino gambling, the case where it works best. It is meant primarily for readers who are coming to the subject for the first time. Those who already know a little bit about expected utility theory will miss nothing important by skipping the first chapter entirely. The second chapter clarifies the concept of a decision problem. My treatment here is slightly non-

Cambridge University Press
978-0-521-06356-2 - The Foundations of Causal Decision Theory
James M. Joyce
Excerpt
More information

standard since I follow Jeffrey in supposing that the components of decision problems are always *propositions* and yet still maintain a distinction among actions, states, and outcomes. Chapter 3 is an extended critical discussion of the influential formulation of expected utility theory that appears in Leonard Savage's *The Foundations of Statistics*. I shall argue that Savage's theory is not ultimately acceptable as a foundation for expected utility theory. This is an important conclusion in the present context because causal decision theorists have tended to assume that some appropriately modified version of Savage's theory would supply an adequate formal underpinning for their account. Since this is not the case, the need for a representation theorem for causal decision theory becomes all the more pressing.

Chapter 4 begins the heart of the book. In it I present Jeffrey's "evidentialist" version of decision theory and sketch Bolker's representation theorem for it. While Bolker's theorem will be seen to avoid nearly all the pitfalls that beset Savage's approach, it will be argued, nevertheless, that it is not entirely acceptable because it does not sufficiently constrain rational beliefs. A better version of Bolker's result will then be proved, one that, for the first time, obtains *unique* representations within Jeffrey's framework. This turns out to be a very important advance.

Chapter 5 treats the topic of causal decision theory. First I will argue that its proponents are correct in thinking that an adequate solution to Newcomb's problem requires an account of rational decision making that portrays agents as having beliefs about causal connections that are not ultimately reflected in their ordinary conditional subjective probabilities. The nature of these "causal" beliefs will be discussed at length. In the course of this investigation it will become apparent that all the various formulations of causal decision theory suffer from a common problem of "partition dependence"; they apply only when the decision situation is described in a very specific and detailed way. This makes it difficult to apply the theory to real-life decisions, and it greatly complicates the foundational challenges that it faces. The lesson will be that causal expected utility must assume the form of a conditional decision theory if the problem of partition dependence is to be solved.

The topic of Chapter 6 is the concept of *supposition* that underlies the notion of conditional expected utility. It is well known that there are at least two ways to suppose that a proposition is true: One can suppose it *indicatively* by provisionally adding the proposition to one's stock of knowledge, or one can suppose it *subjunctively* by imagining a possible circumstance in which the proposition is true that otherwise deviates minimally from the way that things actually are. I do not think the relationship between these two notions has yet been adequately understood, and the goal of Chapter 6 is to shed some light on the issue and, more generally, to clarify the concept of a supposition itself. In passing, I will show how the infamous "problem of old evidence" can be partially solved (or, better, how part of the problem can be completely solved), and a generalization of Jeffrey conditionalization to Rényi–Popper measures will be presented.

In Chapter 7 we examine a number of the representation results that have been proposed as foundations for causal decision theory. I shall argue that none of them is fully acceptable. I then go on to prove a representation theorem for conditional decision theory along the lines of Bolker's Theorem for evidential decision theory. This theorem will be seen to provide a common theoretical underpinning for both causal and evidential decision theories.

The book concludes with a short chapter that describes what has been accomplished and suggests some directions for future research.

By the way, I was kidding about the sucker gene. I decided to open the book with that example to weed out unsympathetic readers. It is good to know that you are still with me! Enjoy the fallacies!

1

Instrumental Rationality as Expected Utility Maximization

This chapter provides a brief quasi-historical introduction to *expected utility theory*, the most widely defended version of *normative decision theory*. The overarching goal of normative decision theory is to establish a general standard of rationality for the sort of *instrumental* (or "practical") reasoning that people employ when trying to choose means appropriate for achieving ends they desire. Expected utility theory champions *subjective expected utility maximization* as the hallmark of rationality in this means-ends sense.

We will examine the theory in the setting where it works best by applying it to the case of professional gamblers playing games of chance inside casinos. In this highly idealized situation, the end is always the maximization of one's own fortune, and the means is the ability to buy and sell wagers that offer monetary payoffs at known odds. Later chapters will consider more general contexts. Since the material here is presented in an elementary (and somewhat pedantic) way, those who already understand the concept of expected utility maximization and the rudiments of decision theory are encouraged to proceed directly to Chapter 2.

1.1 PASCAL AND THE "PROBLEM OF THE POINTS"

The *Port Royal Logic* of 1662 contains the first general statement of the central dogma of contemporary decision theory:

> In order to decide what we ought to do to obtain some good or avoid some harm, it is necessary to consider not only the good or harm in itself, but also the probability that it will or will not occur; and to view geometrically the proportion that all these things have when taken together.[1]

In modern terms, the suggestion here is that risky or uncertain prospects are best evaluated according to the *principle of mathematical expectation*, so that "our fear of some harm [or hope of some good]

[1] Arnauld and Nicole (1996, pp. 273–74).

ought to be proportional not only to the magnitude of the harm [or good], but also to [its] probability."[2]

This principle, and the theory of probability that underlies it, had been discovered in 1654 by Blaise Pascal, the greatest of the many great thinkers that Port Royal produced, during the course of a correspondence with Pierre de Fermat concerning a gambler's quandary now known as the *Problem of the Points.*[3] It had been posed to Pascal by a "reputed gamester," the Chevalier de Méré, who Pascal regarded as a fine fellow even though he suffered from the "great fault" of not being a mathematician. The question had to do with the fair division of a fixed pot of money among gamblers forced to abandon a winner-take-all game before anyone had won. Here is a simplified version of the problem that Pascal and Fermat considered (with dollars instead of "pistoles" as currency): Two gamblers, H and T, are playing a game in which a coin, known to be fair, is to be tossed five times and a $64 prize awarded to H or T depending on whether more heads or tails come up. Suppose that the first three tosses go head/tail/tail, and that the game is then interrupted, leaving the two gamblers with the task of finding an equitable way to dividing the $64. T, who has two of the three tails she needs to win, would surely feel cheated if the pot were split down the middle. H, on the other hand, would be justifiably upset if T got all the money since he still had a chance to win the game when it was stopped. Clearly, the fair division must give T something more than $32 but less than $64. The challenge for Pascal and Fermat was to find the right amount. Both men solved the instance of the Problem of the Points they were considering, but Pascal, in a great feat of mathematical genius, went on to treat the general case.

His solution had two parts. First, he invented the theory of probability more or less from scratch. Professional gamblers had long known that one could use nonnegative real numbers to measure the frequencies at which various events occur, and that these would give the odds at which various bets would be advantageous. Legend has it, for example, that the Chevalier de Méré made a lot of money laying even odds that he could roll at least one 6 in four tosses of a fair die. What the Chevalier realized, and his gullible opponents did not, was that the probability of this event was slightly more than one-half (about 0.518), and thus he was likely to win his bet more often than not. Unfortunately, probabilities were difficult to calculate, and gamblers were forced to find them empirically by observing the

2 Arnauld and Nicole (1996, pp. 274–75).
3 The best discussion of the Pascal/Fermat correspondence is found in Todhunter (1865/1949, pp. 7–21).