

Evidence,  
Decision and  
Causality

Arif Ahmed



## EVIDENCE, DECISION AND CAUSALITY

Most philosophers agree that causal knowledge is essential to decision-making: agents should choose from the available options those that probably *cause* the outcomes that they want. This book argues against this theory and in favour of Evidential or Bayesian Decision Theory, which emphasizes the symptomatic value of options over their causal role. It examines a variety of settings, including economic theory, quantum mechanics and philosophical thought-experiments, where causal knowledge seems to make a practical difference. The arguments make novel use of machinery from other areas of philosophical inquiry, including first-person epistemology and the free-will debate. The book also illustrates the applicability of decision theory itself to questions about the direction of time and the special epistemic status of agents.

ARIF AHMED is Senior Lecturer in Philosophy at the University of Cambridge.



# EVIDENCE, DECISION AND CAUSALITY

ARIF AHMED

*University of Cambridge*



**CAMBRIDGE**  
UNIVERSITY PRESS

# CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9781107020894](http://www.cambridge.org/9781107020894)

© Arif Ahmed 2014

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2014

Printed in the United Kingdom by Clays, St Ives plc

*A catalogue record for this publication is available from the British Library*

*Library of Congress Cataloguing in Publication data*

Ahmed, Arif.

Evidence, decision, and causality / Arif Ahmed, University of Cambridge.

pages cm

Includes bibliographical references and index.

ISBN 978-1-107-02089-4 (hardback : alk. paper)

1. Decision making. 2. Evidence. 3. Causation. I. Title.

BD184.A36 2014

122 – dc23 2014020933

ISBN 978-1-107-02089-4 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

# Contents

<i>Preface</i>	<i>page</i> vii
Introduction	I
0.1 Causalism and evidentialism	I
0.2 Evidence	3
0.3 Causality	4
0.4 Decision theory	5
0.5 Evidential Decision Theory and Causal Decision Theory	7
0.6 Predestination	9
0.7 Why it matters	13
1 Savage	16
1.1 Simplifications and idealizations	16
1.2 States, events, outcomes and acts	17
1.3 Rational preference and the Savage axioms	20
1.4 From preference to probability	23
1.5 Utility and the representation theorem	26
1.6 Dominance and fatalism	30
2 EDT and CDT	35
2.1 Preference over news items	36
2.2 The Jeffrey–Bolker axioms	38
2.3 Credence	40
2.4 Evidential Decision Theory	43
2.5 Newcomb’s problem	46
2.6 K-partitions	48
2.7 Causal Decision Theory	52
2.8 Matters arising	54
3 Causalist objections to CDT	60
3.1 Egan–Gibbard	61
3.2 The Piaf maxim	68
3.3 Objective Decision Theory	74

4	Realistic cases	81
4.1	Remedial cases	82
4.2	Medical cases	89
4.3	The Tickle Defence	91
4.4	Economic cases	99
4.5	Psychological cases	102
4.6	<i>Prisoners' Dilemma</i>	108
5	Deterministic cases	120
5.1	<i>Betting on the Past</i>	120
5.2	<i>Betting on the Laws</i>	130
5.3	Objections	135
5.4	The openness of the past	141
6	Quantum-mechanical cases	146
6.1	The device	146
6.2	Identical settings	148
6.3	Non-identical settings	149
6.4	QM versus CDT	155
6.5	Objections	157
7	The standard Newcomb case	166
7.1	Deflationary responses	166
7.2	The Discontinuous Strategy	170
7.3	The case for one-boxing	180
7.4	The case for two-boxing	194
7.5	Conclusion	211
8	'The ultimate contingency'	213
8.1	Dualism and the Ramsey Thesis	215
8.2	Arguments for the Ramsey Thesis	217
8.3	Dynamic inconsistency and Dutch books	226
8.4	Anti-dualism	233
	<i>References</i>	236
	<i>Index</i>	247

## *Preface*

Causality is a pointless superstition. These days it would take more than one book to persuade anyone of that. This book focuses on the ‘pointless’ bit, not the ‘superstition’ bit. I take for granted that there are causal relations and ask what doing so is good for. More narrowly still, I ask whether causal belief plays a special role in decision. My argument that it does not consists largely of schemes for extracting money from the people who think it does. I conduct the argument using the framework of decision theory. Decision theory is convenient for this purpose because it (a) quantifies causal beliefs and (b) isolates their role in practical deliberation.

Because of this aim the book couldn’t serve as a stand-alone introduction to decision theory. Anyone who took it that way would find it distorted and lacunary. Distorted because of the focus on the dominance principle at the expense of almost everything else that Savage’s axioms entail; lacunary because of the complete absence from the story of any approach outside the Ramsey–Savage expected-utility paradigm.

Anyway, many excellent introductions to the subject are already available at various levels of mathematical sophistication. For instance, Peterson 2009 is written at a mathematically elementary level, Gilboa 2009 and Kreps 1988 are more difficult and Fishburn 1970 is mathematically fairly advanced. *This* book presupposes no mathematical knowledge beyond completely elementary set theory and probability.

But it probably *could* function as a philosophical companion to any such introduction. Its central topic is perhaps the main debate in the philosophical foundations of decision theory. That subject has a bearing on more traditional preoccupations of metaphysics, including causality itself, the asymmetry of time and the nature of self-knowledge. Pursuing it also forces us to touch upon live issues in psychology, economics, the theory of voting and the foundations of quantum mechanics. So I hope in these chapters to illustrate why the philosophy of decision theory is important, or at least interesting, even if you are not a philosopher of decision theory.

I owe the reader an apology for the number and the length of the footnotes. They largely involve (i) references, (ii) numerical calculations and (iii) objections and clarifications that are unlikely to interest most readers but likely to interest some. In cases (ii) and (iii), I felt that including these items in the main text would break up the main thread of the argument, which you should be able to follow without reading any of them.

Whilst writing this book I have had the benefit of conversations with, and/or written comments from, the following people: Helen Beebe, Sharon Berry, Simon Blackburn, Rachael Briggs, Lucy Campbell, Adam Caulton, John Collins, Tom Dougherty, Adam Elga, Luke Fenton-Glynn, Alison Fernandes, Alexander Greenberg, Alan Hájek, Caspar Hare, Jane Heal, Hykel Hosni, Jennan Ismael, Leon Leontyev, Isaac Levi, Hanti Lin, Penelope Mackie, John Maier, Adam Morton, Daniel Nolan, Harold Noonan, Huw Price, Paolo Santorio, Wolfgang Schwarz, Shyane Siriwardena, Julia Staffel and Paul Weirich. I thank them all. I also thank two referees from the Press for their extremely helpful written comments, Dr John Gaunt for meticulous copyediting, and Alexander Greenberg and Shyane Siriwardena for assistance with the index. And I thank my editors at CUP, Sarah Green and Hilary Gaskin, for their efficiency, patience and goodwill.

I am also grateful to audiences at the Aristotelian Society, the University of Auckland, the Australian National University, Birmingham University, Bristol University, the University of Cambridge, Columbia University, the Institute of Philosophy, Monash University, the University of Nottingham, Princeton University, the Scuola Normale Superiore di Pisa, the University of Sheffield, the University of Sussex and the University of Sydney, where I delivered talks on material that has ended up in the book.

I wrote some of this book whilst holding a Leverhulme Research Fellowship at the Faculty of Philosophy, University of Cambridge, and then at the Sydney Centre for the Foundations of Science, University of Sydney. I wrote some of it whilst holding a Distinguished Visiting Professorship at the Scuola Normale Superiore di Pisa. And I wrote some of it whilst holding a Visiting Fellowship at the Research School of Social Sciences in the Australian National University, Canberra. I am grateful to all of these institutions for their hospitality, and also to the Leverhulme Trust. I also thank the Faculty of Philosophy at the University of Cambridge, and Girton College, Cambridge, for granting me leave from teaching and administrative duties in order to finish the book.

Material from [section 3.1](#) has appeared in my paper ‘Push the button’, *Philosophy of Science* 79 (July 2012): 386–95, and is reproduced by permission

of the University of Chicago Press. Copyright 2012 by the Philosophy of Science Association. All rights reserved.

Material from [section 5.1](#) has appeared in my paper ‘Causal Decision Theory and the fixity of the past’, published online in the *British Journal for the Philosophy of Science* on 16 September 2013 by Oxford University Press on behalf of the British Society for the Philosophy of Science.

Material from [section 5.2](#) has appeared in my ‘Causal Decision Theory: a counterexample’, in *Philosophical Review*, 122, 2: 289–306. Copyright 2013, Cornell University. Reprinted by permission of the present publisher, Duke University Press, [www.dukepress.edu](http://www.dukepress.edu).

Material from s. 7.2 is here reproduced from my paper ‘Infallibility in the Newcomb Problem’, published online in *Erkenntnis* on 11 April 2014, copyright Springer Science+Business Media Dordrecht 2014, and appears here with kind permission from Springer Science+Business Media.

Material from [section 7.3.3](#) has appeared in A. Ahmed and H. Price, ‘Arntzenius on “Why Ain’cha rich?”’, *Erkenntnis* 77: 15–30, copyright 2012 Springer Science+Business Media B.V., and reappears here with kind permission from Springer Science+Business Media B.V.

Finally, and most of all, I am grateful to my family (Frisbee, Isla, Iona and Skye) for putting up with me during the writing of this book. My father, Dr G. M. Ahmed, died in early 2013 after a period of illness that he bore with characteristic stoicism and good humour. I should like to dedicate the book to my mother, Mrs S. Ahmed, and to the memory of my father.



# Introduction

## 0.1 Causalism and evidentialism

*Causalism* is the doctrine that rational choice must take account of causal information. Specifically it must attend to whether and how an agent's available acts are *causally relevant* to the outcomes that he desires or dreads. Nothing short of causal information will do the trick, and agents who ignore it will make the wrong decision in a variety of identifiable and realistic cases. This book argues against causalism.

*Evidentialism*, which it prefers, is the contrary view that only the *diagnostic* bearing of acts is of practical concern. It only matters to what extent this or that act is *evidence* of this or that outcome, regardless of whether the act causes the outcome or is merely symptomatic of it. Many cases where evidentialism and causalism make different practical recommendations are less realistic than causalism supposes. But anyway evidentialism is practically superior: we do better to follow the evidentialist recommendation than to follow the causalist one, *whenever* they diverge.

Two philosophical positions encourage evidentialism. The first is Humean scepticism about causality.<sup>1</sup> I strike the match and then it lights. However closely you examine that sequence you won't see the striking *making* the match light. However often you examine similar sequences you will only see the recurrence of that pattern. However widely you examine surrounding sequences you will only see recurrences of other patterns. All of this only shows that in such-and-such circumstances the striking of a match is a good *sign* of its lighting. So as far as we can see, distinct events are, as Hume said, entirely loose and separate. So causality does not impinge upon human experience in any other way than through the *evidential* relations between events (the patterns) that either constitute causality or are sustained by it. This is not to say that some deeper metaphysical relation of causality

<sup>1</sup> Hume 1949 [1738]: I.iii.

doesn't exist, but only that for practical purposes we needn't care what it is or whether it does.

The second motivation is Russellian eliminativism about causality.<sup>2</sup> Physics alone tells us what really causes what. There are no causes in physics. So there are no causes. To act as if there were is to act upon illusions. To act upon illusions is as counterproductive in the practical sphere as believing them in the theoretical sphere.

Causalism can seem attractive from the perspective of an anti-Humean tradition pre-dating Hume himself. According to it, Hume misses out half the story. In fact we are not just *observers*, but also *agents*. Therefore we must acknowledge aspects of reality that a mere spectator might have no reason to discern. In its most romantic and excitable version (Schopenhauer's), *only* the willing of an agent can penetrate the veil that conceals *all* objective truth from mere perception. But you needn't go that far to appreciate the insight, if it is one, that rational agents have reason as such to distinguish tighter *causal* relations between events from amongst the merely evidential relations of co-occurrence and statistical correlation that are equally available to Humeans and to the paralytic.<sup>3</sup> Still, many modern adherents of causalism make no reference to that tradition.

It is possible to question all of these standpoints. In connection with Humean scepticism, one might object that people can in fact *see* e.g. that one thing is pushing another.<sup>4</sup> In connection with Russellian eliminativism, one might deny both (a) that only physics can settle what causes what<sup>5</sup> and (b) that physics makes no room for causality.<sup>6</sup> And in connection with the 'anti-Humean' tradition, one might object that agency, being itself a causal notion, cannot reveal causality to anyone who would otherwise be innocent of it.<sup>7</sup>

But my aim in mentioning these metaphysical positions (for which plenty more might be said) wasn't to argue for evidentialism or causalism, but only to indicate what broader issues give additional interest to the dispute between them. My reasons for preferring evidentialism don't depend on these sweeping metaphysical arguments, but rather on attention to individual cases where the choice between these doctrines makes a difference.

<sup>2</sup> Russell 1913.

<sup>3</sup> The tradition includes Berkeley (1980 [1710]: sections 25–9); Reid (2001 [1792]); Schopenhauer (1995 [1819]: Book II); Collingwood (1940: 291, 322); Von Wright (1973: 108–13); and Menzies and Price (1993: 191–2).

<sup>4</sup> Michotte 1963.    <sup>5</sup> Cartwright 1979.    <sup>6</sup> Hitchcock 2007.

<sup>7</sup> Ahmed 2007 discusses some forms of this objection.

So as to convey what that involves, the rest of this Introduction sets out in informal terms: (0.2) the kind of evidential relation and (0.3) the kind of causal relations that will matter here; (0.4) the formal framework in which I propose to conduct the debate; (0.5) an explication of the opposing positions in terms of it; and (0.6) the kind of case over which they disagree. Section 0.7 outlines the relevance of these matters to philosophers and others for whom they may not hold intrinsic interest.

## 0.2 Evidence

Evidence is here an (a) subjective and (b) non-causal relation that I'll take (c) to hold between propositions. (a) Whether one thing is evidence for another may vary across persons. (i) You might think that Pete's calling in this hand of poker is evidence that he'll lose; I think it is evidence that he'll win. (ii) I might think that the report of a miracle is evidence of a divine plan whereas you think it is irrelevant.

Such disagreement needn't imply irrationality on either side. What a rational person takes as evidence itself depends upon (i) particular and (ii) theoretical background knowledge or belief, and either may vary interpersonally. (i) Suppose you know that Pete has a worse hand than his opponent's; but I know that Pete has sneaked a look at his opponent's hand. (ii) Suppose my religion says that if God exists then He regularly intervenes in human affairs; but according to yours, the regular course of nature is the only possible expression of His will.

So evidence is subjective. And I won't assume any objective constraints on what can count as evidence for what, beyond the dictates of logic and probability (see sections 1.1, 1.4 and 2.3). On the other hand, there is a limit to what we can learn by considering agents with utterly outlandish beliefs on this point, as the literature on this subject sometimes does (see section 7.1). For this reason I think that the cases worth most attention involve agents whose beliefs you could at least imagine holding. Chapters 4–6 investigate these.

(b) Evidence does not only relate causes to their effects. Of course causes *can* be evidence of effects: my turning the key in the ignition is evidence that the car is going to start. But this is not the only case. (i) This bloodstained knife is evidence that the butler did it. But its being bloodstained is an effect and not a cause of the crime. (ii) The bloodstained knife is also evidence that the butler will confess under questioning. But it may be neither cause nor effect of this. Instead, the deed itself might be a common cause of both. (iii) Certain theories of historical development or of human nature could

have led an educated observer of the French Revolution quite reasonably to expect it to develop along the same lines as the English one that occurred a century before. So the past course of the English Revolution is, for him, evidence of the future course of the French Revolution. But the observer might deny *both* that any causal connection relates these sequences *and* that some particular past state or event caused both of them.<sup>8</sup>

(c) In these examples, and also in ordinary speech, many kinds of thing count as subjects or objects of ‘evidence’, e.g. events (Pete’s calling), objects (a bloodstained knife), states of affairs (the existence of God) or historical sequences (the French Revolution). There may be nothing wrong with that. But here I shall for convenience say that it is *propositions* that evidentially support other *propositions*. Instead of saying that the bloodstained knife was evidence of the butler’s guilt, I’ll say *that the knife was bloodstained* is evidence *that the butler did it*. Similarly I will say that *whether the knife was bloodstained* is evidentially relevant to *whether the butler did it*, meaning by this that the proposition that the knife was bloodstained has some evidential bearing, positive or negative, on the proposition that the butler did it.

### 0.3 Causality

The causal relations that will matter are *causal dependence* and *independence* between possible individual events, facts or states of affairs – or rather, again, between particular propositions. For instance, whether this bottle smashes is causally dependent on whether it is dropped but causally independent of whether that match is struck. Causal relevance is the converse of causal dependence. Whether the match is struck is causally relevant to whether it lights but causally irrelevant to whether the bottle smashes.

Except in [Chapter 6](#), it will be clear enough in most cases what causally depends on what. But it’s worth stating that *E* can causally depend on *C* even though it wouldn’t be natural to say that *C* caused *E*, or that *C* was ‘the’ cause of *E*. For instance, the lighting of this match was causally dependent on the presence of oxygen in the atmosphere. But it sounds odd to say that the presence of the oxygen caused the match to light. That is not because

<sup>8</sup> Conversely, something can be a cause without also being a sign. My striking of this particular match might ultimately cause the room to fill with cigarette smoke. But it isn’t evidence that the room will fill with smoke, at least not to anyone who knows that I also have a lighter or can borrow a light from any of a dozen other people, to which person it was already practically certain that the room will fill with smoke whether or not I strike that particular match. But *this* misalignment between the evidential and the causal has little relevance here: what matters is the difference between signs that are causes and signs that are not causes, not the difference between causes that are signs and causes that are not signs.

the oxygen bears *no* causal relation to the lighting but because we select for largely subjective reasons a single event, state or fact, from amongst the many on which the lighting was causally dependent, that we then call '*the*' cause. Causalism doesn't care about the things of which my act may be the cause in this narrower sense, but only about what may causally depend on it in the 'broad and non-discriminatory' sense.<sup>9</sup>

I will take for granted that there *is* such a thing as causal dependence, i.e. that our use in deliberation of causal vocabulary picks out a single metaphysical relation, that that relation is causal dependence, and that sometimes some things really *are* causally dependent on other things. Not because I have any great confidence in that thesis, being myself attracted to the view that not only rational decision but also physical science can manage quite well without it, but because causalism itself presupposes it.

Nor will I raise any concerns, Humean or otherwise, about the epistemology of causal dependence. I take for granted that the normal procedures that we take to establish causal dependence – e.g. controlled trials – do give us reason to think that whether this particular event occurs is causally dependent on whether that other one does. This is not because such doubts are straightforwardly answerable but because they are irrelevant to the assessment of causalism. The question is not whether causation exists or how we can know about it; the question is why we should care.

#### 0.4 Decision theory

Locke wrote:

What we once know, we are certain is so: and we may be secure, that there are no latent proofs undiscovered, which may overturn our knowledge, or bring it in doubt. But, in matters of probability, it is not in every case we can be sure that we have all the particulars before us, that any way concern the question; and that there is no evidence behind, and yet unseen, which may cast the probability on the other side, and outweigh all, that at present seems to preponderate with us . . . And yet we are forced to determine ourselves on the one side or other. The conduct of our lives, and the management of our great concerns, will not bear delay: for those depend, for the most part, on the determination of our judgement in points, wherein we are not capable of certain and demonstrative knowledge, and wherein it is necessary for us to embrace the one side, or the other.<sup>10</sup>

<sup>9</sup> E.g. Lewis (1973a: 162) distinguishes this sense and implicates it in his own version of causalism (1981a: 329–335).

<sup>10</sup> Locke 1975 [1689]: IV.xvi.3.

The second half of this contrast ('And yet . . .') is wrong. The conduct of our lives does *not* force us to embrace one side or the other. We often act, and sometimes act rationally, without any definite opinion on points that matter to the outcome.

This coin lands heads two tosses out of three. I must make a bet that pays my stake if the coin lands heads this time. It is certainly rational for me to put *some* but not *all* of my present wealth on heads. But when I put up the stake I neither have nor feign certainty either that the coin will land heads (otherwise I'd stake *all* of my present wealth) or that it won't (otherwise I'd stake *none* of it). By staking some intermediate amount I am acting rationally without embracing one side or the other.

Normative decision theory tries to say in quite general and abstract terms just *how* to act under such uncertainty. This book conducts the dispute between evidentialism and causalism in the terms of normative decision theory, which from now on I'll just call 'decision theory'. I shall largely set aside *descriptive* decision theory, which tries to make general claims about how agents' *actual* behaviour depends on their values and beliefs.

Decision theory works roughly like this. Suppose an agent has several options in some situation. The situation has many possible outcomes that the agent desires or dreads in varying degrees. For each option, if she takes it, some outcomes are more likely than others. At any rate she *takes* some outcomes to be more likely than others if she takes the option. Suppose we have some idea of *how* good each outcome is for her. And suppose we have some idea of *how* likely she considers each outcome, given each option. Decision theory takes all this as input – the options, the possible outcomes, and how good and how likely she considers them. Its output is a *recommendation* of some option or options. *Contra* Locke, you *can* act rationally in the absence of real or feigned certainty, and decision theory tells you how.

Thus suppose that the agent can bet any positive fraction of her current fortune on the next toss of a coin whose chance of landing heads is in her view definitely two-thirds. If the coin lands heads then she is better off by the amount of her stake. If the coin lands tails then she is worse off by the same amount. What fraction of her fortune should she bet?

The point of decision theory is to answer questions like that. In this case the agent has many options (one for every amount that she might stake). And there are many possible outcomes: for each dollar amount  $k$  that she bets out of a dollar fortune  $F$ , there is the good outcome that she ends up with  $\$(F + k)$  (if the coin lands heads) and the bad outcome that she ends up with  $\$(F - k)$  (if the coin lands tails). We know that her confidence in

heads is two-thirds. If we also know the rate at which she values each extra dollar, then decision theory should tell her what to do. And on a simple if not entirely plausible assumption about the rate at which she values money, it turns out that most sensible decision theories tell her to stake one-third of her fortune.<sup>11</sup>

Not every possible decision theory *is* sensible. Consider the theory that tells you always to do what is most costly: that is a stupid theory. So is the one that tells you always to do what is most risky. So is the one that tells you always to do what is *least* risky. In fact infinitely many decision theories make palpably absurd recommendations in all sorts of cases. I'll ignore all of them.

More seriously, I'll largely ignore *objective* decision theories. An objective decision theory specifies the best option independently of the *agent's* uncertainty about the state of the world. Suppose e.g. that at each time each possible future event has an *objective chance*. The objective chance of an event at a time is some quantity that is independent of anyone's confidence that it will occur: for instance, the chance that this radium atom here will decay in the next minute. There are objective decision theories whose outputs depend on chances themselves, and not on the agent's beliefs about chances. One such theory advocates doing *whatever now has the best chance of bringing about the best outcome*, quite independently of what the agent thinks about this. Section 3.3 below states my reasons for setting aside objective theories.

I'll focus instead on theories into which the *facts* of the agent's situation enter *only* in so far as her beliefs reflect them. Such decision theories, unlike objective ones, generally give usable advice to agents that use them. They are *subjective* decision theories. And I'll focus for the most part on the two subjective theories that philosophers have focused upon for the last forty years. These are *Evidential* (sometimes also called *Bayesian*) *Decision Theory* (EDT) and *Causal Decision Theory* (CDT). Neither one is absurd. But they can't both be true because they sometimes disagree.

### 0.5 Evidential Decision Theory and Causal Decision Theory

EDT and CDT are both quantitative theories. They depend upon some numerical measure of the value of each outcome for the agent and also of the uncertainty that she attaches to relevant hypotheses about the state of the world. So I can't explain how they work until I've explained how

<sup>11</sup> The assumption is logarithmic utility.

to quantify both value and uncertainty. Chapters 1 and 2 do that in more detail. Here I explain only the general idea behind each theory.

EDT identifies the value of an option with its *news* value. It recommends what is in the agent's view the most *auspicious* option, the one that good fortune most probably *accompanies*. CDT identifies the value of an option with the value of its believed *effects* (including itself). It recommends the option that the agent considers most *efficacious*, that most probably *produces* good fortune.

It is important to distinguish EDT from 'magical thinking': the false belief that one can *causally* influence the outcome of some process by symbolic gestures or other indirect means.<sup>12</sup> People who suffer from illnesses can't, for the most part, cure themselves by *acting as if* they have recovered. Voting for your candidate doesn't *cause* other like-minded people also to vote for him. EDT can certainly acknowledge that in these cases there is absolutely no *causal* connection between the act and the desired state. But then neither, in these cases, is an agent likely to see much *evidential* connection between them. EDT only insists that when the agent genuinely *does* take some act to be good news, he has reason to perform it.

Even this vague exposition of their differences reveals EDT and CDT as versions of evidentialism and causalism respectively. The news value of an act is a function of its evidential bearing upon outcomes of interest, irrespective of whatever causal relations lie beneath. So EDT cashes out the central commitment of evidentialism: that only its diagnostic or evidential import need be relevant to the assessment of an act. Similarly, CDT crystallizes the central idea of causalism, that the practical value of an act is sensitive to its causal bearing upon outcomes of interest. There *are* some causalists who think that CDT does not correctly spell out the nature of the sensitivity, and I will discuss three arguments to this effect in [Chapter 3](#). But CDT is the most popular version of causalism.<sup>13</sup> In what follows I shall for the most part identify the issue between evidentialism and causalism with the issue between EDT and CDT.

Putting them in the way that I have, both decision theories sound reasonable. At least neither is as absurd as the theory that always recommends the most costly option, or the one that always recommends the least risky

<sup>12</sup> Shafir and Tversky 1992: 463.

<sup>13</sup> Causalist defenders of CDT include Nozick (1969: 222 ff.); Gibbard and Harper (1978: 355–7); Lewis (1981b: 308–12); Joyce (1999: Chapter 5); Pearl (2000: 108–10); Sloman (2005: Chapter 7); and many others. Causalists who reject CDT, or at least don't plainly endorse it, include Cartwright (1979; see Lewis 1981b: 325 n. 15); Egan (2007; see section 3.1 below); and Mellor (1983, 2005; see Section 3.3 below).

option. The way to settle the issue between them is to look at situations where they conflict. There a contemplated act typically has an *evidential* bearing on an outcome that it *does nothing to bring about*. We have already seen (at [section 0.2](#)) possible situations where *A* is evidentially but not causally relevant to *B*. What we seek now is a case of this sort in which *A* is an option and *B* an outcome to whose occurrence the agent is not indifferent.

It is a surprising fact that such cases are fairly difficult to find. In fact the most widely discussed case, which I cover at [section 2.5](#) and also in [Chapter 7](#), is explicitly science-fictional. Variants on this example involve stories about God, Satan, angels, fantastically powerful computers and unusually able psychologists.

Whilst its being thus hypothetical makes it attractively simple, it also raises a concern about relevance. If *in practice* EDT and CDT never disagree, then the disagreement between them might seem relatively trivial. If so, we have reason to prefer evidentialism, which says that an agent *needn't* care about the causal relevance of his options once their evidential bearing is firmly in view. I discuss this point briefly at the start of [Chapter 4](#) and in more detail at [section 7.1](#). In any case, and as [Chapters 4–6](#) argue, disagreement between the theories is feasible and, in at least some cases, reasonably realistic. But given the present purposes of introduction and illustration, it is worth citing a clear and straightforward example of disagreement between EDT and CDT, even if it is as fanciful as I expect most readers to find the following.<sup>14</sup>

## 0.6 Predestination

Predestination was historically the most important feature of Calvinism.<sup>15</sup> Here it is, in the Westminster Confession of 1647:

By the decree of God . . . some men and angels are predestinated unto everlasting life, and others foreordained to everlasting death. Those of mankind that are predestinated unto life, God before the foundation of the world was laid, according to his eternal and immutable purpose, and the secret counsel and good pleasure of his will, hath chosen . . . out of his free grace and love, without any foresight of faith and good works, or perseverance in either of them, or any other thing in the creature as conditions or causes moving Him thereunto. The rest of mankind God was pleased . . . to pass by. All those

<sup>14</sup> Resnik (1987: 112) calls it a real-life case, presumably because many real people genuinely believed themselves to be in it.

<sup>15</sup> Weber 1992 [1920]: 57.

Table 0.1 *Calvinist problem*

	Salvation	Damnation
Virtue	2	0
Sin	3	1

whom God hath predestined unto life, and those only, He is pleased . . . to call by His word and spirit . . . renewing their wills, and by His almighty power determining them to that which is good. As for those wicked and ungodly men . . . He not only with-holdeth His grace . . . but sometimes also withdraweth the gifts which they had and . . . gives them over to their own lusts, the temptations of the world and the power of Satan.

Nothing that anyone does can *bring about* his salvation. For instance, nothing that I do now can somehow *cause* God to have foreseen that I would do it and to decide my salvation on that basis: for God has already settled that ‘*without any foresight . . . or any other thing in the creature as conditions or causes moving Him thereunto*’. But although the doing of good works is not a *cause* of salvation, it is a *sign* of salvation, because if God has chosen to save me then he has already *determined* me to do good works. (Note that only incompatibilists about free will would think that this derogates from my freedom to choose in this matter.) Similarly, a sure sign of damnation, though again not a cause of it, is indulgence in one’s own lusts etc. In Weber’s summary: ‘however useless good works might be as a means of attaining salvation . . . nevertheless, they are indispensable as a sign of salvation’.<sup>16</sup>

Let some Calvinist agent believe this. Imagine him facing what he considers the decisive temptation. If he yields then his life (on Earth) will be pleasurable. But yielding is a sure sign that he was damned for all eternity and so will suffer everlasting death after this life. If he declines the temptation then his life on Earth will certainly be dull. But declining is a sure sign of everlasting superlunary life.

Table 0.1 summarizes the possible outcomes and their relative value to the agent. The entries in the top row represent God’s possible decrees: that the agent is saved or damned. The entries in the left-hand column represent the agent’s options: to decline the temptation and to yield to it. The body of the table has four cells. Each corresponds to one of four possible outcomes.

<sup>16</sup> Weber 1992 [1920]: 69.

For instance, the top left-hand cell corresponds to the outcome that the agent is virtuous and is saved. And the bottom right-hand cell corresponds to the outcome that the agent is sinful and is damned. The number inside each cell represents the agent's evaluation of the outcome that it represents. At this point you should interpret the numbers as follows: for any two outcomes, the agent *prefers* the one with the *bigger* number.

So what [Table 0.1](#) tells us about the agent's preferences over the four outcomes is just what you would expect. He prefers each outcome involving salvation to both outcomes involving damnation. And he prefers sin under any dispensation to virtue under that same dispensation. The best outcome (which scores 3) is the bottom left-hand box: the agent gets to sin *and* to be saved. The next-best outcome is the top left-hand box: salvation more than compensates for his forgoing of worldly pleasures. The third-best outcome is the bottom right-hand box. Here, sin is insufficient compensation for damnation: he would rather have the salvation and forgo the sin. Finally, the worst outcome is associated with the top right-hand box: the agent is virtuous but still damned. And the problem for the Calvinist agent is: given all these preferences, should he take the path of virtue or the path of sin? (I'm not yet concerned to justify an answer, but only to illustrate the EDT/CDT conflict.)

Virtue is a sign of salvation. Sin is a sign of damnation. So the agent's ultimate fate is *evidentially dependent* on his present choice. The agent prefers virtue and salvation (the top left-hand cell) to sin and damnation (the bottom right-hand cell). So, according to EDT, the agent should choose virtue.

But CDT advises sin. God's decree for the agent took place many years ago, or perhaps is outside time altogether. Either way, no good works can now *make* the agent's next life go any better. And sin cannot make it go any *worse*. But sin would make *this* life better for the agent than virtue would, whatever God has decreed. So the agent should be sinful.

This is a clear if unrealistic case where EDT and CDT disagree. It is a version of *Newcomb's problem*, which has attracted plenty of attention from philosophers since about 1970 and which in various realizations gets plenty here too, at sections [2.5](#), [4.2–6](#) and [5.1](#), and throughout [Chapter 7](#). The strange thing about Newcomb's problem is that there is widespread agreement that it is obvious what the agent should do, but not so much over what that is.

Many people think it obvious that the agent should sin, since whether he sins has no effect on whether he is saved. Obviously, if virtue were to *cause* his salvation, e.g. by causing God to have foreseen his virtue, then

the pay-offs in [Table 0.1](#) would make virtue rational. But it doesn't, so they don't. He may as well have fun before meeting a future that would have been his whatever he had done.

Others find it completely opaque just what *about* the causal relation could make this difference. After all, the situation is practically indistinguishable from one in which virtue really does cause salvation. Everyone who is virtuous is saved, including the agent. Everyone who sins is damned, including the agent. That connection falls short of *being* causal, but how could genuine causation add anything of practical relevance?

It couldn't – not here and not anywhere else. At least not according to EDT, not according to the evidentialism that it represents, and not according to Ramsey when he wrote that 'all the practical man wants to know is that all people who take arsenic die, not that this is a causal implication, for a universal of fact is *within its scope* just as good a guide to conduct as a universal of law'.<sup>17</sup> Evidential Decision Theory is essentially a vast generalization of Ramsey's point. This book does not defend that theory against all objections but only its superiority to CDT, since it is on that point that the dispute between evidentialism and causalism seems to me, as it has seemed to many, to turn.

[Chapter 2](#) presents EDT and CDT in slightly greater detail as rival developments of Savage's original decision theory, as outlined in [Chapter 1](#). My reasons for proceeding this way are, first, that Savage's is still the most widely known and applied decision theory and so represents a convenient point of entry to this philosophical debate for decision theorists who are not familiar with it; and second, that Savage's theory represents an agreed background against which the formal difference between EDT and CDT stands out more clearly.

The main argument of the book is, then, for these two premises:

- (1) If causalism is true then the causal bearing of an act is relevant to its assessment in the way that CDT claims.
- (2) The causal bearing of an act is not relevant to its assessment in the way that CDT claims.

It follows from (1) and (2) that causalism is false. [Chapter 3](#) argues for (1) against those who assert that the causal bearing of one's acts *is* relevant to their assessment, only not in the way that CDT describes. [Chapters 4–7](#) argue for the second premise by examining cases where EDT and CDT

<sup>17</sup> Ramsey 1990 [1928]: 144. His later paper on this subject (1990 [1929]) takes a more nuanced though still evidentialist line, some elements of which I outline and criticize in [Chapter 8](#).

seem to disagree. I argue that in every such situation, *either* they really give the same advice *or* EDT gives better advice. Either way, the examples all support premise (2) and also the superiority of EDT over CDT.

Chapter 8 considers an alternative ‘dualistic’ version of evidentialism that seeks to defuse the debate. That view assigns a traditionally anti-Humean significance to a contrast between the *agent’s* perspective and the *observer’s* perspective. And when EDT accommodates this point, it turns out not to disagree with CDT anywhere. I reject both this conclusion and also its philosophical starting point. There *is* a genuine dispute between EDT and CDT. And for the purpose of evaluating his options, an agent need *not* see any deeper significance in the distinction between those bodily movements and other events that occur because he wants them to, and those on the other hand that take place willy-nilly.

## 0.7 Why it matters

The relevance of the dispute between Evidential Decision Theory and Causal Decision Theory outruns its bearing on causalism. It should interest (i) non-philosophers who care about decision theory, (ii) non-decision theorists who care about philosophy, and (iii) everyone else.

(i) First, decision theory standardly plays a more *descriptive* than normative role in those sciences, like economics and political science, that make widespread use of it; whereas this book is largely concerned with the foundations of the normative theory. Nevertheless, normative questions should matter to those sciences in so far as they are concerned to map human conformity to, or deviation from, ideals of practical *rationality*. For instance: when people vote in large elections, are they acting rationally or irrationally? The answer depends in part on what counts as acting rationally. So if EDT and CDT disagree about that then they disagree about which areas of human conduct are deviations from that ideal.<sup>18</sup>

Second, the argument of this book covers various misunderstandings of what EDT in fact recommends. For instance, the theory does *not* recommend that you shouldn’t visit the doctor for fear that this will confirm that you are ill, or that you shouldn’t read the newspaper for fear that you will get bad news; more generally, it does not always advise against actions that are statistically associated with bad outcomes (see section 4.1). Chapter 4 attempts to identify where and why it *does* proscribe options that bring bad news without causing bad outcomes. And

<sup>18</sup> For further discussion of the voting case see section 4.6.3 below.

whilst that point is here put in service of a normative doctrine, it also has implications for the descriptive version of EDT, which predicts the behaviour that its normative counterpart recommends. Once we understand the true consequences of EDT we are better placed to assess its descriptive bearing on the choices that people do in fact make; hence also to appreciate its potential relevance to predictive or explanatory projects in social science.<sup>19</sup>

Decision theory as applied in the social sciences does not normally recognize the EDT/CDT distinction. What gets much more usually applied is Savage's original theory, which is a common ancestor of EDT and CDT in a sense that [section 2.8](#) spells out. But this is not because the difference between EDT and CDT is irrelevant to those sciences but because the general application of the Savage theory *already presupposes* a stand on the EDT/CDT dispute. For instance (and putting it in terms that [Chapter 1](#) explains): the application of the Savage theory to the Calvinist's problem demands a decision on whether the appropriate partition of the state space is (a) into cases where you are damned and cases where you are saved, or (b) into cases where you *now act* as God decreed and cases where you now act against God's decree concerning your action. Thinking of the problem in terms of the (a)-possibilities is *already* to settle the issue in favour of CDT; thinking of it in (b)-like terms is to settle it in favour of EDT.<sup>20</sup>

(ii) Turning to its strictly philosophical interest: first, at least some of the cases I cover demand a stand on matters of wider philosophical interest. For instance, what you think evidentialism prescribes in a variety of cases may depend on your views on self-knowledge, in particular on your sympathy for a broadly Cartesian auto-epistemology: see [sections 4.3.2](#) and [4.5](#) for further discussion.

More generally, subjective decision theory should take account of the sense, if any, in which a deliberating agent's beliefs, desires and intentions are transparent to her. So the foundations of decision theory constitute a novel field in which to apply work in this live area of philosophy of mind. Here those questions arise most prominently at [sections 3.2](#) and [8.2](#).

Second, some of the cases that I discuss are themselves relevant to metaphysics. At least, they are for pragmatists: in particular those who want to know what ways of thinking about time, or about causality, best serve our practical interests. Time: [section 5.4](#) argues that a rational decision-maker thinks of the past as open to deliberation much as everyone thinks of

<sup>19</sup> For an extensive working out of the applications of EDT to questions in political science, see Grafstein 1999.

<sup>20</sup> For a more realistic example that makes the same point, see Joyce 2002: 71–5.

the future as being thus open. For only this view of openness is consistent with a rational approach to the decision problems that sections 5.1 and 5.2 discuss. Causality: section 6.5.1 argues that in some cases what the *causal* theory recommends turns on what causality is, making the Chapter 6 cases of interest even to those whom they don't convince to abandon CDT.

(iii) Normative decision theory distinguishes rational from irrational behaviour against a fixed background of beliefs and desires. Although the argument in this book presupposes some idealization as outlined at section 1.1, still in many cases EDT's advice holds good even for agents that do not meet these standards. For instance, situations of the 'Prisoners' dilemma' type arise frequently, and whilst EDT largely agrees that it is more rational to defect than to co-operate in these cases, section 4.6.3 identifies marginal cases where co-operation may be rational. Marginal they may be, but it is easy enough to see how they might arise in real life. And when they do, the argument here gives practical advice about what to do in them: for instance, and continuing the example in (i) above, when it would be rational to vote in a large election.

So the subject is not only of interest to philosophers who study decision theory. But whatever your reason for approaching this subject, the proper starting point is the first systematic and still the most influential treatment of decision under subjective uncertainty, viz Savage's theory.