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.[1] 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

[1] 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.[2] 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.[3] 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.[4] In connection with Russellian eliminativism, one might deny both (a) that only physics can settle what causes what[5] and (b) that physics makes no room for causality.[6] 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.[7]

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.

[2] Russell 1913.
[3] 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).
[4] Michotte 1963. [5] Cartwright 1979. [6] Hitchcock 2007.
[7] 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.[8]

(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

[8] 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.[9]

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.[10]

[9] E.g. Lewis (1973a: 162) distinguishes this sense and implicates it in his own version of causalism (1981a: 329–35).
[10] 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.[11]

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

[11] 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.[12] 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.[13] 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

[12] Shafir and Tversky 1992: 463.

[13] 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.[14]

0.6 Predestination

Predestination was historically the most important feature of Calvinism.[15] 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

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

[15] 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'.[16]

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.

[16] Weber 1992 [1920]: 69.