JON WILLIAMSON

CAUSALITY

1 INTRODUCTION

Perhaps the key philosophical questions concerning causality are the following:

- what are causal relationships?
- how can one discover causal relationships?
- how should one reason with causal relationships?

This chapter will focus on the first two questions. The last question is equally important — of course we need to know the best way to make predictions, perform diagnoses and make strategic decisions — but in the absence of a well-entrenched mathematical calculus of causality, the answers given to the last question tend to depend on the answers provided to the first two questions.

Standard responses to the first, ontological question are surveyed in §2, while the second, epistemological question is dealt with in §3. I advocate a position I call *epistemic causality* which is sketched in §4, and which is compared to the positions of Judea Pearl in §5 and Huw Price in §6.[1]

2 THE NATURE OF CAUSALITY

There are three varieties of position on causality. One can argue that the concept of causality is of heuristic use only and should be eliminated from scientific discourse: this was the tack pursued by Bertrand Russell, who maintained that science appeals to functional relationships rather than causal laws.[2] Alternatively one can argue that causality is a fundamental feature of the world and should be treated as a scientific primitive — this claim is usually the result of disillusionment with purported philosophical analyses, several of which appeal to the asymmetry of time in order to explain the asymmetry of causation, a strategy that is unattractive to those who want to analyse time in terms of causality. Or one can maintain

[1] Epistemic causality motivates an answer to the last question, how should one reason with causal relationships? The ensuing formalism is presented in detail in [Williamson, 2004].

[2] [Russell, 1913]. Russell later modified his views on causality, becoming more tolerant of the notion.

D.M. Gabbay and F. Guenthner (eds.), Handbook of Philosophical Logic, 2nd Edition, Volume 14, 95–126.

that causal relations can be reduced to other concepts not involving causal notions. This latter position is dominant in the philosophical literature, and there are four main approaches which can be described roughly as follows. The mechanistic theory, discussed in §2.1, reduces causal relations to physical processes. The probabilistic account (§2.2) reduces causal relations to probabilistic relations. The counterfactual account (§2.3) reduces causal relations to counterfactual conditionals. The agent-oriented account (§2.4) reduces causal relations to the ability of agents to achieve goals by manipulating their causes.[3]

2.1 Mechanisms

The mechanistic account of causality aims to understand the physical processes that link cause and effect, interpreting causal statements as saying something about such processes. Proponents of this type of position include Wesley Salmon[4] and Phil Dowe.[5] They argue that a causal process is one that transmits[6] or possesses[7] a conserved physical quantity, such as energy-mass, linear momentum or charge, from start (cause) to finish (effect).

The mechanistic account is clearly a physical interpretation of causality, since it identifies causal relationships with physical processes. Such a notion of cause relates single cases, since only they are linked by physical processes, although causal regularities or laws may be induced from single-case causal connections.

The main limitation of this approach is its rather narrow applicability: most of our causal assertions are apparently unrelated to the physics of conserved quantities. While it may be possible that physical processes such as those along which quantities are conserved could suggest causal links to physicists, such processes are altogether too low-level to suggest causal relationships in economics, for instance. One could maintain that the economists' concept of causality is the same as that of physics and is reducible to physical processes,[8] but one would be forced to accept that the epistemology of such a concept is totally unrelated to its metaphysics. This is undesirable: if the grounds for knowledge of a causal connection have little to do with the nature of the causal connection as it is analysed then one can argue that it cannot be the causal connection that we have knowledge

[3] See the introduction to [Sosa and Tooley, 1993] for more discussion on the variety of interpretations of causality.

[4] [Salmon, 1980], [Salmon, 1984], [Salmon, 1997], [Salmon, 1998].

[5] [Dowe, 1993], [Dowe, 1996], [Dowe, 1999], [Dowe, 2000], [Dowe, 2000b].

[6] [Salmon, 1997] §2.

[7] [Dowe, 2000b] §V.1.

[8] This was the tack Salmon took in connection with his earlier theory that conceived of causal processes as involving the transmission of marks rather than conserved quantities. See [Salmon, 1998], page 206.

of, but something else.[9] On the other hand one could keep the physical account and accept that the economists' causality differs from the physicists' causality. But this position faces the further questions of what economists' causality is, and why we think that cause is a single concept when in fact it isn't. These problems clearly motivate a more unified account of causality.

2.2 Probabilistic Causality

Probabilistic causality has a wider scope than the mechanistic approach: here the idea is to understand causal connections in terms of probabilistic relationships between variables, be they variables in physics, economics or wherever. There is no firm consensus among proponents of probabilistic causality as to what probabilistic relationships among variables constitute causal relationships, but typically they appeal to the intuitions behind the *Principle of the Common Cause*: if two variables are probabilistically dependent then one causes the other or they are effects of common causes which screen off the dependence (i.e. the two variables are probabilistically independent conditional on the common causes). Indeed Hans Reichenbach applied the Principle of the Common Cause to an analysis of causality, as a step on the way to a probabilistic analysis of the direction of time.[10] Similarly Patrick Suppes argued that causal relations induce probabilistic dependencies and that screening off can be used to differentiate between variables that are common effects and variables that are cause and effect.[11] However, both these analyses fell foul of a number of criticisms,[12] and more recent probabilistic approaches adopt *Causal Dependence* (cause and direct effect are probabilistically dependent conditional on the effect's other direct causes) and the *Causal Markov Condition* (each variable is probabilistically independent of its non-effects, conditional on its direct causes) as necessary conditions for causality, together with other less central conditions which are sketched in §3.[13] Sometimes Causal Dependence is only implicitly adopted: the causal relation may be defined as the smallest relation that satisfies the Causal Markov Condition, in which case Causal Dependence must hold.

Probabilistic causality is normally applied to repeatably-instantiatable rather than single-case variables — in principle either is possible, as long as the chosen interpretation of probability handles the same kind of variables. Invariably causality is interpreted as a physical, mind-independent concept.

[9] See [Benacerraf, 1973] for a parallel argument in mathematics.

[10] [Reichenbach, 1956].

[11] [Suppes, 1970].

[12] See [Salmon, 1980b], §§2-3

[13] See [Pearl, 1988], [Pearl, 2000], [Spirtes *et al.*, 1993], [McKim and Turner, 1997] and [Korb, 1999]. Note that the concept of *direct* cause does not require that causal chains be discrete. It is merely presumed that Causal Dependence or the Causal Markov Condition will hold where the direct causes are taken to be a set of causes that are sufficiently close to the effect, with one direct cause per causal chain that leads to the effect.

The chief problem that besets probabilistic causality is the dubious status of the probabilistic conditions to which the account appeals. While the conditions seem intuitive and might be expected to hold much of the time there are clear cases where they fail. The Principle of the Common Cause and the Causal Markov Condition are widely acknowledged to fail in certain cases that crop up in quantum mechanics, but they also fail more generally wherever probabilistic dependencies are induced by non-causal relationships: where variables are semantically, logically or mathematically related, or they are related by non-causal physical laws (as in the quantum mechanics case) or boundary conditions.[14] Causal Dependence fails for instance where an event must be caused by one of two equally efficacious physical processes: if a machine can be activated by precisely one of two fully reliable power supplies, then the choice of power supply will not change the probability of its direct effect, the machine being activated.[15] Of course it is not good enough for a probabilistic analysis of causality if the defining connection between probability and causality admits exceptions — we are left with the question as to how causality is to be analysed in the exceptional cases.

2.3 Counterfactuals

The counterfactual account, developed in detail by David Lewis,[16] reduces causal relations to subjunctive conditionals: C is a direct cause of E if and only if (i) if C were to occur then E would occur (or its chance of occurring would be significantly raised) and (ii) if C were not to occur then E would not occur (or its chance of occurring would be significantly lowered). The subjunctive conditionals (called *counterfactual* conditionals if the antecedent is false) are in turn given a semantics in terms of possible worlds: 'if C were to occur then E would occur' is true if and only if (i) there are no possible worlds in which C is true or (ii) E holds at all the possible worlds in which C holds that our closest to our own world. So causal claims are claims about what goes on in possible worlds that are close to our own.[17]

Lewis's counterfactual theory was developed to account for causal relationships between single-case events (which can be thought of as single-case variables which take the values 'occurs' or 'does not occur'), and the causal relation is intended to be mind-independent and objective.

Many of the difficulties with this view stem from Lewis' reliance on possible worlds. Possible worlds are not just a dispensable façon de parler for

[14] These counterexamples are explained in detail in [Williamson, 2004], §4.2.

[15] [Williamson, 2004] §7.3.

[16] [Lewis, 1973].

[17] Lewis modified his account in [Lewis, 2000], but the changes made have little bearing on our discussion. See [Lewis, 1986b] for Lewis' account of causal explanation.

Lewis, they are assumed to exist in just the way our world exists. But we have no physical contact with these other worlds, which makes it hard to see how their goings-on can be the object of our causal claims and hard to see how we discover causal relationships. Moreover it is doubtful whether there is an objective way to determine which worlds are closest to our own if we follow Lewis' suggestion of measuring closeness by similarity — two worlds are similar in some respects and different in others and choice or weighting of these respects is a subjective matter. Causal relations, on the other hand, do not seem to be subjective. Instead of analysing causal relations, of which we have at least an intuitive grasp, in terms of subjunctive conditionals and ultimately possible worlds, which many find mysterious, it would be more natural to proceed in the opposite direction. Thus we might be better-off appealing to causality to decide whether E would (be more likely to) occur were C to occur,[18] and depending on the answer we could then say whether a world in which C and E occurs is closer to our own than one in which C occurs but E does not.

2.4 Agency

The agency account, whose chief recent proponents are perhaps Huw Price and Peter Menzies,[19] analyses causal relations in terms of the ability of agents to achieve goals by manipulating their causes. According to this account, C causes E if and only if bringing about C would be an effective way for an agent to bring about E. Here the strategy of bringing about C is deemed effective if a rational decision theory would prescribe it as a way of bringing about E. Menzies and Price argue that the strategy would be prescribed if and only if it raises the 'agent probability' of the occurrence of E.[20] (The events they consider are single-case.)

Menzies and Price do not agree as to the interpretation of these probabilities: Menzies maintains that they are chances, while Price seems to have a Bayesian conception.[21] Consequently it is not entirely clear whether they view causality as a physical or mental notion. On the one hand they claim that there would be causal relations without agents,[22] while on the other they say, 'we would argue that when an agent can bring about one event as a means to bringing about another, this is true in virtue of certain basic intrinsic features of the situation involved, these features being essentially

[18] See [Pearl, 2000], chapter 7, for an analysis of counterfactuals in terms of causal relations. [Dawid, 2001] argues that counterfactuals are irrelevant and misleading for an analysis of causality.

[19] [Price, 1991], [Price, 1992], [Price, 1992b], [Menzies and Price, 1993].

[20] [Menzies and Price, 1993].

[21] [Menzies and Price, 1993] pg. 190.

[22] [Menzies and Price, 1993] §6.

non-causal though not necessarily physical in character',[23] and maintain that the concept of cause is a 'secondary quality', relative to human responses or capacities.[24] From this relativity one might expect cause to be subjective, but they say that causation is significantly more objective than other secondary quantities like colour or taste.[25] We shall examine Price's views on these matters in more detail in §6.

The main problems that beset the agency approach are inherited from those faced by the probabilistic and counterfactual approaches. First, the agency approach assumes a version of Causal Dependence for agent probabilities — we saw in §2.2 that this condition does not always hold.[26] Of course, where a causal connection is not accompanied by probabilistic dependence, such as in the power supply example of §2.2, bringing about a cause is not a good strategy for bringing about its effects. Second, the agency account appeals to subjunctive conditionals[27] (C causes E if and only if, *were* an agent to bring about C, that *would* be a good strategy for bringing about E) and so qualms about the utility of a counterfactual account can equally be applied to the agency approach.

3 DISCOVERING CAUSAL RELATIONSHIPS

Different views on the nature of causality lead to different suggestions for discovering causal relationships. The mechanistic view of causality, for example, leads naturally to a quest for physical processes, while proponents of probabilistic causality prescribe searching for probabilistic dependencies and independencies.

However there are two very general strategies for causal discovery which cut across the ontological positions. Whatever view one holds on the nature of causality, one can advocate either *hypothetico-deductive* or *inductive* discovery of causal relationships. Under a hypothetico-deductive account (§3.1) one hypothesises causal relationships, deduces predictions from the hypothesis, and then tests the hypothesis by seeing how well the predictions accord with what actually happens. Under an inductive account (§3.2), one makes a large number of observations and induces causal relationships directly from this mass of data. We shall discuss each of these approaches

[23] [Menzies and Price, 1993] pg. 197.
[24] [Menzies and Price, 1993] pp. 188,199.
[25] [Menzies and Price, 1993] pg. 200.
[26] In fact the version assumed by the agency approach does not restrict attention to direct causes and does not demand that dependence be conditional on the effect's other causes. This type of dependence condition is rarely advocated since it faces a wider range of counterexamples than Causal Dependence in the form used here — see the references given in §2.2.
[27] [Menzies and Price, 1993] §5.

in turn in this chapter, and give an overview of some recent proposals for discovering causal relationships.

3.1 Hypothetico-Deductive Discovery

According to the hypothetico-deductive account, a scientist first hypothesises causal relationships and then tests this hypothesis by seeing whether predictions drawn from it are borne out. The testing phase may be influenced by views on the nature of causality: a causal hypothesis can be supported or refuted according to whether physical processes are found that underlie the hypothesised causal relationships, whether probabilistic consequences of the hypothesis are verified, and whether experiments show that by manipulating the hypothesised causes one can achieve their effects.

Karl Popper was an exponent of the hypothetico-deductive approach. For Popper a causal explanation of an event consists of natural laws (which are universal statements) together with initial conditions (which are single-case statements) from which one can predict by deduction the event to be explained. The initial conditions are called the 'cause' of the event to be explained, which is in turn called the 'effect'.[28] Causal laws, then, are just universal laws, and are to be discovered via Popper's general scheme for scientific discovery: (i) hypothesise the laws; (ii) deduce their consequences, rejecting the laws and returning to step (i) if these consequences are falsified by evidence. Popper thus combines what is known as the *covering-law* account of causal explanation with a hypothetico-deductive account of learning causal relationships.

The covering-law model of explanation was developed by Hempel and Oppenheim[29] and also Railton,[30] and criticised by Lewis.[31] While such a model fits well with Popper's general account of scientific discovery, neither the details nor the viability of the covering-law model are relevant to the issue at stake: a Popperian hypothetico-deductive account of causal discovery can be combined with practically any account of causality and causal explanation.[32] Neither does one have to be a strict falsificationist to adopt a hypothetico-deductive account. Popper argued that the testing of a law only proceeds by falsification: a law should be rejected if contradicted by observed evidence (i.e. if falsified), but should never be accepted or regarded as confirmed in the absence of a falsification. This second claim of Popper's

[28] [Popper, 1934] §12.
[29] [Hempel and Oppenheim, 1948].
[30] [Railton, 1978].
[31] [Lewis, 1986b] §VII.
[32] Even Russell's eliminativist position of [Russell, 1913], in which he argued that talk of causal laws should be eradicated in favour of talk of functional relationships, ties in well with Popper's logic of scientific discovery. Both Popper and Russell, after all, drew no sharp distinction between causal laws and the other universal laws that feature in science.

has often been disputed, and many argue that a hypothesis is confirmed by evidence in proportion to the probability of the hypothesis conditional on the evidence.[33] Given this probabilistic measure of confirmation — or indeed any other measure — one can accept the hypothesised causal relationships according to the extent to which evidence confirms the hypothesis. Thus the hypothetico-deductive strategy for learning causal relationships is very general: it does not require any particular metaphysics of causality, nor a covering-law model of causal explanation, nor a strict falsificationist account of testing.

Besides providing some criterion for accepting or rejecting hypothesised causal relationships, the proponent of a hypothetico-deductive account must do two things: (i) say how causal relationships are to be hypothesised; (ii) say how predictions are to be deduced from the causal relationships.

Popper fulfilled the latter task straightforwardly: effects are predicted as logical consequences of laws given causes (initial conditions). The viability of this response hinges very closely on Popper's account of causal explanation, and the response is ultimately inadequate for the simple reason that no one accepts the covering-law model as Popper formulated it: more recent covering-law models are significantly more complex, coping with chance explanations.[34]

Popper's response to the former task was equally straightforward, but perhaps even less satisfying:

> my view of the matter, for what it is worth, is that there is no such thing as a logical method of having new ideas, or a logical reconstruction of this process. My view may be expressed by saying that every discovery contains 'an irrational element', or 'a creative intuition'[35]

Popper accordingly placed the question of discovery firmly in the hands of psychologists, and concentrated solely on the question of the justification of a hypothesis.

The difficulty here is that while hypothesising may contain an irrational element, Popper has failed to shed any light on the rational element which must surely play a significant role in discovery. Popper's scepticism about the existence of a logic need not have precluded any discussion of the act of hypothesising from a normative point of view: both Popper in science and Pólya in mathematics remained pessimistic about the existence of a precise logic for hypothesising, yet Pólya managed to identify several imprecise but important heuristics.[36] One particular problem is this: a theory may be refuted by one experiment but perform well in many others; in such

[33] See [Howson and Urbach, 1989], [Earman, 1992].
[34] [Railton, 1978] for example.
[35] [Popper, 1934] pg. 32.
[36] [Polya, 1945], [Polya, 1954], [Polya, 1954b].

a case it may need only some local revision, to deal with the domain of application on which it is refuted, rather than wholesale rehypothesising. Popper's account says nothing of this, giving the impression that with each refutation one must return to a blank sheet and hypothesise afresh. The hypothetico-deductive method as stated neither gives an account of the progress of scientific theories in general, nor of causal theories in particular.

Any hypothetico-deductive account of causal discovery which fails to probe either the hypothetico or the deductive aspects of the process is clearly lacking. These are, in my view, the key shortcomings of Popper's position. I shall try to shed some light on these aspects when I present a new type of hypothetico-deductive account in §4.5. For now, we shall turn to a competing account of causal discovery, inductivism.

3.2 Inductive Learning

Francis Bacon developed a rather different account of scientific learning. First one makes a large amount of careful observations of the phenomenon to be explained, by performing experiments if need be. One compiles a table of positive instances (cases in which the phenomenon occurs),[37] a table of negative instances (cases in which the phenomenon does not occur)[38] and a table of partial instances (cases in which the phenomenon occurs to a certain degree).[39]

> We have chosen to call the task and function of these three tables the *Presentation of instances to the intellect.* After the *presentation* has been made, *induction* itself has to be put to work. For in addition to the *presentation* of each and every instance, we have to discover which nature appears constantly with a given nature or not, which grows with it or decreases with it; and which is a limitation (as we said above) of a more general nature. If the mind attempts to do this affirmatively from the beginning (as it always does if left to itself), fancies will arise and conjectures and poorly defined notions and axioms needing daily correction, unless one chooses (in the manner of the Schoolmen) to defend the indefensible.[40]

Thus Bacon's method consists of presentation followed by induction of a theory from the observations. It is to be preferred over a hypothetico-deductive approach because it avoids the construction of poor hypotheses in the absence of observations, and it avoids the tendency to defend the indefensible:

[37] [Bacon, 1620] §II.XI.
[38] [Bacon, 1620] §II.XII.
[39] [Bacon, 1620] §II.XIII.
[40] [Bacon, 1620] §II.XV.

> Once a man's understanding has settled on something (either because it is an accepted belief or because it pleases him), it draws everything else also to support and agree with it. And if it encounters a larger number of more powerful countervailing examples, it either fails to notice them, or disregards them, or makes fine distinctions to dismiss and reject them, and all this with much dangerous prejudice, to preserve the authority of its first conceptions.[41]

Note that while Bacon's position is antithetical to Popper's hypothetico-deductive approach, it is compatible with Popper's falsificationism — indeed Bacon claims that 'every *contradictory instance* destroys a conjecture'.[42] The first step of the inductive process, *exclusion*, involves ruling out a selection of simple and often rather vaguely formulated conjectures by means of providing contradictory instances.[43] The next step is a *first harvest*, which is a preliminary interpretation of the phenomenon of interest.[44] Bacon then produces a seven-stage process of elucidating, refining and testing this interpretation — only the first stage of which was worked out in any detail.[45]

Present-day inductivists claim that causal relationships can be inferred algorithmically from experimental and observational data, and that suitable data would yield the correct causal relationships. Usually, but not necessarily, the data takes the form of a database of past cases: a set V of repeatably instantiatable variables are measured, each entry of the database $D = (u_1, \ldots, u_k)$ consists of an observed assignment of values to some subset U_i of V. Such an account of learning is occasionally alluded to in connection with probabilistic analyses of causality and has been systematically investigated by researchers in the field of artificial intelligence, including groups in Pittsburgh,[46] Los Angeles[47] and Monash,[48] proponents of a Bayesian learning approach,[49] and computationally-minded psychologists.[50]

These approaches seek to learn various types of causal model. The simplest type of causal model is just a *causal graph* (i.e. a directed acyclic graph in which nodes correspond to variables and there is an arrow from one node

[41] [Bacon, 1620] §I.XLVI.
[42] [Bacon, 1620] §II.XVIII.
[43] [Bacon, 1620] §§II.XVIII-XIX.
[44] [Bacon, 1620] §II.XX.
[45] [Bacon, 1620] §§II.XXI-LII.
[46] [Spirtes *et al.*, 1993], [Scheines, 1997], [Glymour, 1997], [Mani and Cooper, 1999], [Mani and Cooper, 2000], [Mani and Cooper, 2001].
[47] [Pearl, 2000], [Pearl, 1999].
[48] [Dai *et al.*, 1997], [Wallace and Korb, 1999], [Korb and Nicholson, 2003].
[49] [Heckerman *et al.*, 1999], [Cooper, 1999], [Cooper, 2000], [Tong and Koller, 2001], [Yoo *et al.*, 2002].
[50] [Waldmann and Martignon, 1998], [Waldmann, 2001], [Tenenbaum and Griffiths, 2001], [Glymour, 2001], [Hagmayer and Waldmann, 2002].

to another if the former directly causes the latter) which shows only qualitative causal relationships. A *causal net* is slightly more complex, containing not only a qualitative causal graph but also quantitative information, the probability distribution $p(a_i|par_i)$ of each variable A_i conditional on its parents Par_i, the direct causes of A_i in the graph. A *structural equation model* is a third type of causal model — this can be thought of as a causal graph together with an equation for each variable in terms of its direct cause variables, $A_i = f_i(Par_i, E_i)$, where f_i is some function and E_i is an error variable.

The mainstream of these inductivist AI approaches have the following feature in common. In order that causal relationships can be gleaned from statistical relationships, the approaches assume the Causal Markov Condition.[51] A causal net contains the Causal Markov Condition as an inbuilt assumption; in the case of structural equation models the Causal Markov Condition is a consequence of the representation of each variable as a function just of its direct causes and an error variable, given the further assumption that all error variables are probabilistically independent.

The inductive procedure then consists in finding the class of causal models — or under some approaches a single 'best' causal model — whose probabilistic independencies implied via the Causal Markov Condition are consistent with independencies inferred from the data. Other assumptions are often also made, such as minimality (no submodel of the causal model also satisfies the Causal Markov Condition), faithfulness (all independencies in the data are implied via the Causal Markov Condition), linearity (all variables are linear functions of their direct causes and uncorrelated error variables), causal sufficiency (all common causes of measured variables are measured), context generality (every individual possesses the causal relations of the population), no side effects (one can intervene to fix the value of a variable without changing the value of any non-effects of the variable) and determinism. However these extra assumptions are less central than the Causal Markov Condition: approaches differ as to which of these extra assumptions they adopt and the assumptions tend to be used just to facilitate the inductive procedure based on the Causal Markov Condition, either by helping to provide some justification of the inductive procedure or by increasing the purported efficiency or efficacy of algorithms for causal induction.[52]

The brunt of criticism of the inductive approach tends to focus on the Causal Markov Condition and the ancillary assumptions outlined above. I

[51] There are inductive AI methods that take a totally different approach to causal learning, such as that in [Karimi and Hamilton, 2000] and [Karimi and Hamilton, 2001], and [Wendelken and Shastri, 2000]. However, non-Causal-Markov approaches are well in the minority.

[52] See Chapter 8 of [Williamson, 2004] for a more detailed overview of inductive algorithms for causal discovery.

have already mentioned the difficulties that beset the Causal Markov Condition; in cases where this condition fails the inductive approach will simply posit the wrong causal relationships. It is plain to see that the ancillary conditions are also very strong and these face numerous counterexamples themselves. The proof, inductivists claim, will be in the pudding. However, the reported successes of inductive methods have been questioned,[53] and these criticisms lend further doubt to the inductive approach as a whole and the Causal Markov Condition in particular as its central assumption.[54]

Unfortunately neither Popper's hypothetico-deductive approach nor the recent inductivist proposals from AI offer a viable account of the discovery of causal relationships. Popper's hypothetico-deductive approach suffers from underspecification: the hypothesis of causal relationships remains a mystery and Popper's proposals for deducing predictions from hypotheses were woefully simplistic. On the other hand, the key shortcoming of the inductive approach is this: given the counterexamples to the Causal Markov Condition the inductive approach cannot guarantee that the induced causal model or class of causal models will tally with causality as we understand it — the causal models that result from the inductive approach will satisfy the Causal Markov Condition, but the true causal picture may not. While this objection may put paid to the dream of using Causal Markov formalisms for learning causal relationships, an alternative formalism may yet ground the inductive approach. In §4.5 we shall see that the inductive and hypothetico-deductive approaches can be reconciled by using new inductive methods as a way of hypothesising a causal model, then deducing its consequences and restructuring the model if these are not borne out.

4 EPISTEMIC CAUSALITY

In this section I shall sketch my own view of causality, *epistemic causality*. A more detailed exposition can be found in [Williamson, 2004].

As I see it, current theories of causality suffer from over-compartmentalisation. Current theories analyse causality in terms of just one of the indicators of causal relationships — mechanisms, probabilistic dependencies or independencies, counterfactuals or agency considerations — to the expense of the others. While one indicator may be more closely connected with causality than the others, our causal beliefs are clearly based on several indicators, not exclusively on one. It seems that if we are to understand

[53] [Humphreys and Freedman, 1996], [Humphreys, 1997], [Freedman and Humphreys, 1999], [Woodward, 1997].

[54] See [Dash and Druzdzel, 1999], [Hausman, 1999], [Hausman and Woodward, 1999], Part Three of [Glymour and Cooper, 1999], [Lemmer, 1996], [Lad, 1999], [Cartwright, 1997], [Cartwright, 1999] and [Cartwright, 2001] for further discussion of the inductive approach.

the complexity of causality we must focus on our causal beliefs and the role these indicators have in forming them.

Epistemic causality focusses on causal beliefs. It provides an account of causal beliefs in informal causal reasoning (§4.1), as well as a more formal account of how we ought to determine causal beliefs (§4.2). It takes causality to be an objective notion (§4.3) yet primarily a mental construct (§4.4). And it provides an account of the discovery of causal relationships (§4.5).

4.1 Informal Causal Reasoning

Why do we have causal beliefs? The answer to this fundamental question, according to the epistemic view, is based on the following doctrines:

Convenience It is convenient to represent the world in terms of cause and effect.

Explanation Humans think in terms of cause and effect because of this convenience, not because there is something physical corresponding to cause which humans experience.

It is convenient to represent the world in terms of cause and effect because a causal representation, if correct, enables us to make successful causal inferences: it allows us to make correct predictions, correct diagnoses and successful strategic decisions. Correct predictions and diagnoses are possible since, typically, cause and direct effect are probabilistically dependent. Successful strategic decisions are possible since, typically, manipulating a cause is a good way of changing its direct effects. (Note that here it is enough that these associations are *typical*; on the other hand an analysis of causality in terms of these associations would be flummoxed by the existence of counterexamples.)

It is clear why the convenience of causality explains our having causal beliefs: successful causal reasoning has survival value. It doesn't take us long as babies to learn that crying brings us food. The value of correctly predicting the effect of a fault in a power plant, correctly diagnosing an ulcer, or successfully manipulating the economy is equally apparent.

The Explanation thesis divorces causal beliefs from any physical, mind-independent notion of causality. While one might remain agnostic as to whether there are physical causal relationships, one might instead adopt an *anti-physical* position, claiming that in the interests of ontological parsimony one should reject physical causality. I leave the selection of an appropriate stance here entirely open.

4.2 Formal Causal Reasoning

The starting-point of a more formal account of causal beliefs is to ask how one might determine a directed acyclic causal graph $\mathcal{C}_\beta$ that depicts the

causal beliefs that an agent ought to adopt on the basis of her background knowledge β.

Arguably $\mathcal{C}_\beta$ should be compatible with background knowledge β, but should otherwise be as non-committal as possible. The agent's causal beliefs should include those causal claims warranted by her background knowledge but no unwarranted causal claims. Since each arrow in a causal graph makes a causal claim, $\mathcal{C}_\beta$ should be a graph that contains fewest arrows, from all those graphs that are compatible with β.

Thus we need to determine which graphs are compatible with background knowledge β. Given the above discussion of informal causal reasoning it seems natural to suppose that a causal graph that is compatible with β should be a good causal representation of β, in the sense that its causal claims should represent any predictive, diagnostic and strategic relationships that can be gleaned from β. We can explicate this thought by insisting that the causal graph include an arrow from A to B if:

- A and B represent non-overlapping physical events (so A and B are the kinds of things that might be causally related, rather than semantically, logically or mathematically related),[55]
- B is *strategically dependent* on A: intervening to change A can change the probability of B, when B's other direct causes are controlled for,
- this dependence is not otherwise accounted for by the agent's background knowledge or other beliefs, and
- the inclusion of this arrow is not inconsistent with other background knowledge. It is here that the other various indicators of causality get taken into account: for instance if it is known that there is no physical mechanism linking A with B, or if it is known that A only occurs after B, then the agent should not deem A to be a direct cause of B.

In sum then, the agent's causal belief graph $\mathcal{C}_\beta$ should be a graph, from all those that are compatible with β in the sense outlined above, that has fewest arrows.

Given this concept of a causal causal belief graph, it is not hard to see that the Causal Markov Condition and the Principle of the Common Cause will hold when $\mathcal{C}_\beta$ contains an arrow for each strategic dependency, and that Causal Dependence will hold if furthermore each arrow in $\mathcal{C}_\beta$ corresponds to a strategic dependency. In this latter case $\mathcal{C}_\beta$ will be a minimal graph satisfying the Causal Markov Condition. We thus have a *qualified* justification of the three controversial principles that connect causality and probability, and a *qualified* justification of inductive methods for causal learning that infer a minimal graph satisfying the Causal Markov Condition.

[55] In fact this is too strict. A causal graph can also feature as a cause or effect — see [Williamson, 2004], Chapter 10.

4.3 The Objectivity of Causality

Clearly a primary desideratum of any theory of causality is that it account for the apparent objectivity of causal notions: causal claims do not appear to be arbitrary, a matter of personal opinion. It might be thought that epistemic causality, focussing as it does on causal beliefs, suffers in this respect. On the contrary, epistemic causality leads to an objective concept of cause as we shall see now.

The word 'objectivity' is routinely used to mean many different things, but the meaning most relevant to discussions of causality is *lack of arbitrariness*. It is important that causal claims are not arbitrary in a pathological way. Note that objectivity in this sense is a matter of degree: if any set of causal claims is correct then causality is *fully subjective*; at the other end of the scale if only one set of causal claims is correct then causality is *fully objective*; degree of objectivity increases as arbitrariness, i.e. the proportion of causal claims that are correct, decreases. We shall be interested in two points on this scale:

Epistemic Objectivity If two agents with the same background knowledge disagree as to causal relationships then at least one of them must be wrong.

Full Objectivity If two agents disagree as to causal relationships then at least one of them must be wrong.

The causal belief graph $\mathcal{C}_\beta$ that an agent ought to adopt on the basis of background knowledge β is epistemically objective (rather, close to epistemically objective: there may be more than one minimal graph compatible with β, but there tends to be little room for subjectivity).

Note that epistemic objectivity is enough for the requirements of science. Sciences demand that disagreements should be resolvable on the basis of current background knowledge in the scientific literature: if there is a disagreement as to whether or not the claim that smoking causes cancer is warranted by current evidence, at least one party should be wrong, for otherwise arbitrariness would render such debates pointless.

Philosophical preconceptions require more though — something close to full objectivity. Intuitively there is a fact of the matter as to what causes what, and if indeed causality is fully objective, a theory of causality should be able to capture this characteristic. The standard way of explaining full objectivity of a scientific concept is to suppose that the concept refers to something physical and mind-independent. Then if there is disagreement as to claims about the concept, the correctness of these claims are decided on the basis of their truth when taken as claims about physical reality.

But projecting a concept onto the physical world is not the only way to account for its full objectivity. Full objectivity can also be generated

from epistemic objectivity. A (close to) fully objective causal graph $\mathcal{C}^*$ can be interpreted as $\mathcal{C}_{\beta^*}$, the causal belief graph one ought to adopt on the basis of some *ultimate background knowledge* β^*. This is the *ultimate belief* interpretation of causality.

What constitutes ultimate background knowledge? There are two possible approaches here.

One might choose β^* to be *limiting* background knowledge, to which an agent's background knowledge tends as time progresses. Now different agents' knowledge might be expected to tend to different limits, so one needs to distinguish a special agent. When C.S. Peirce wanted to analyse truth as the limit of belief, he chose science as the agent whose beliefs are privileged.[56] In our context we might take β^* to be the limit of scientific inquiry. The difficulties with this suggestion are (i) that science is not unanimous: different scientific parties and different scientific theories contradict each other, making it difficult to extract a consistent body of knowledge from science at any particular time, and (ii) that scientific knowledge is no longer considered to be accumulative: science undergoes revolutions, radical changes in scientific knowledge, and thus it is by no means clear that scientific knowledge will tend to a fixed limit. A further problem with this general strategy stems from the way it ties causality very closely to a *particular* agent (science or whomsoever): if the agent had been different, her background knowledge may have been very different, in which case her limiting beliefs and thus causality itself would be very different. This seems counter-intuitive. Under the epistemic account, a causal model is a convenient way of representing the world. While causal relations might be expected to depend on the contingencies of the world, they should not be expected to depend on non-epistemic contingencies of a particular agent.

A natural alternative strategy is to consider the characterising feature of causality, its convenience, and choose β^* that optimises the convenience of $\mathcal{C}^* = \mathcal{C}_{\beta^*}$. (This approach corresponds to William James' analysis of truth: '*The true is the name of whatever proves itself to be good in the way of belief.*'[57]) Now causal beliefs will provide the most convenient representation of the world if they are based on the fullest knowledge of the world, i.e. if β^* contains knowledge of all the indicators of causality. Thus we can take β^* to consist of knowledge of all probabilities, physical mechanisms, temporal relations, non-causal inducers of probabilistic dependencies (semantic, logical and mathematical relationships, non-causal physical laws and boundary conditions) and so on. This strategy has the advantages that β^* is well defined (as long as the indicators of causality can be delimited) and that causality is not tied to a particular agent — indeed causality is not tied even to there *being any agents*.

[56] [Peirce, 1877].
[57] [James, 1907] 30.

We see then how epistemic causality can provide the (close to epistemic) objectivity required for science and the (close to full) objectivity required to satisfy our intuitions about causality.

4.4 What Causality Is

To summarise, epistemic causality provides both an account of causal beliefs and of a fully objective notion of causality. It deals with the causal beliefs an agent ought to adopt on the basis of her background knowledge, and considers causality itself to be the causal beliefs that an agent ought to adopt on the basis of full knowledge of the indicators of causality.

In that sense causality is a mental notion, not a physical notion. This mental metaphysics for causality stands shoulder to shoulder with causal epistemology: the causal relation is just an ultimate set of causal beliefs. Moreover the anti-physical version of epistemic causality makes the further claim that this is the only notion of cause — there is no such thing as physical causality.

But causality is not mental in any degenerate psychologistic sense. Causality does not depend on the mind of any particular agent — it is a normative notion and causal relations are as mind-independent as the laws of logic. Causality is not subject to the whim of an agent: a rational agent can exercise little or no choice when she forms her causal beliefs; there is little or no arbitrariness as what the correct causal relationships are. Causality is objective.

Note that although epistemic causality can be construed as a subjunctive theory, claiming that *were* an agent to know β and *were* she rational then she *would* believe $\mathcal{C}_\beta$, it does not suffer from the problems that beset a counterfactual analysis of causality. This is because its subjunctive conditional claims are not given a semantics in terms of possible worlds — instead a theory of rational causal belief is developed to explicate their meaning. Thus worries about possible worlds do not translate into worries about the claims of epistemic causality.

4.5 Discovery of Causal Relationships

Epistemic causality breaks the barriers between the hypothetico-deductive and inductive accounts of discovering causal relationships.

On the one hand epistemic causality advocates an inductive approach to causal discovery. Given observations β, epistemic causality prescribes an algorithmic way of generating a causal theory $\mathcal{C}_\beta$. This is a different inductive approach to the causal-Markov methods most widely advocated today, but as I have argued in §3.2, those methods are based on questionable assumptions, and (§4.2) the epistemic causality approach explains the special cases where causal-Markov methods work.

On the other hand epistemic causality is hypothetico-deductive: a causal theory $\mathcal{C}_\beta$ is at best a tentative hypothesis, a set of beliefs, and needs testing before it can become entrenched as causal knowledge. Moreover epistemic causality provides a way of filling in the gaps of a hypothetico-deductive approach. The hypothetico phase is no mystery — we have an account of how a hypothesis $\mathcal{C}_\beta$ can be determined by knowledge of the indicators of causality.[58] The deductive phase is no mystery either: we test a causal hypothesis by the inverse mapping from causality to indicators. ¿From a causal relation we can predict a strategic dependency, the existence of a physical mechanism, a temporal relation, and so on, and the causal hypothesis is confirmed to the extent that those predictions are borne out.

5 PEARL'S DETERMINISM

In this section I shall compare epistemic causality with the position recently advocated by Judea Pearl, a pioneer of one of the inductive approaches for discovering causal relationships discussed in §3.2.

It is important to note that Pearl's recent views (as of 2000) differ significantly from his original conception of causality (of 1988).

Pearl's original position stressed the convenience of causality and had much in common with epistemic causality:[59]

> We take the position that human obsession with causation, like many other psychological compulsions, is computationally motivated. Causal models are attractive mainly because they provide effective data structures for representing empirical knowledge — they can be queried and updated at high speed with minimal external supervision.[60]

However, Pearl then changed his mind about causality altogether:

> Ten years ago, when I began working on *Probabilistic Reasoning in Intelligent Systems* (1988), I was working within the empiricist tradition. In this tradition, probabilistic relationships constitute the foundations of human knowledge, whereas causality simply provides useful ways of abbreviating and organizing intricate patterns of probabilistic relationships. Today, my view is quite different. I now take causal relationships to be the fundamental building blocks both of physical reality and of human

[58] Machine learning techniques can be used here to automate the generation of a hypothesis from a database of observations in conjunction with other background knowledge. See [Stankovski *et al.*, 2001] for an analogous proposal.

[59] Epistemic causality is compared to Pearl's early views in [Williamson, 2004], §9.4.

[60] [Pearl, 1988] 383.

> understanding of that reality, and I regard probabilistic relationships as but the surface phenomena of the causal machinery that underlies and propels our understanding of the world.[61]

Thus Pearl's new view is that causality is mind-independent and physical, not to be understood in terms of convenience of belief after all:

> . . . causal relationships are more "stable" than probabilistic relationships. We expect such difference in stability because causal relationships are *ontological*, describing objective physical constraints in our world, whereas probabilistic relationships are *epistemic*, reflecting what we know or believe about the world. Therefore, causal relationships should remain unaltered as long as no change has taken place in the environment, even when our knowledge about the environment undergoes changes.[62]

Interestingly, here Pearl appears to be invoking a physical notion of cause in order to account for the objectivity of causality. As I have pointed out in §4.3, this move is by no means necessary — equally one can account for objectivity by taking an epistemic approach. While for epistemic causality causal beliefs may change as knowledge changes, the induced fully objective notion of cause is independent of any particular agent's knowledge.

Pearl's recent view is that causal models are structural equation models (introduced in §3.2). Pearl's new account thus not only embraces physical causality, but also universal determinism:

> causal relationships are expressed in the form of deterministic, *functional* equations, and probabilities are introduced through the assumption that certain variables in the equations are unobserved. This reflects Laplace's (1814) conception of natural phenomena, according to which nature's laws are deterministic and randomness surfaces owing merely to our ignorance of the underlying boundary conditions.[63]

Pearl subsequently describes his reasons for preferring a deterministic approach to his more stochastic 1988 approach which took causal models to be causal nets rather than structural equation models:[64]

> First, the Laplacian conception is more general. Every stochastic model can be emulated by many functional relationships (with stochastic inputs), but not the other way around; functional relationships can only be approximated, as a limiting case,

[61] [Pearl, 2000] xiii-xiv.
[62] [Pearl, 2000] 25.
[63] [Pearl, 2000] 26.
[64] See also [Pearl, 2000], 31.

> using stochastic models. Second, the Laplacian conception is more in tune with human intuition. The few esoteric quantum mechanical experiments that conflict with the predictions of the Laplacian conception evoke surprise and disbelief, and they demand that physicists give up deeply entrenched intuitions about locality and causality. Our objective is to preserve, explicate, and satisfy — not destroy — those intuitions.
>
> Finally, certain concepts that are ubiquitous in human discourse can be defined only in the Laplacian framework. We shall see, for example, that such simple concepts as "the probability that event B occurred *because* of event A" and "the probability that event B would have been *different* if it were not for event A" cannot be defined in terms of purely stochastic models. These so-called *counterfactual* concepts will require a synthesis of the deterministic and probabilistic components embodied in the Laplacian model.[65]

While functional models may be desirable and appropriate in many circumstances, it seems perverse to develop a theory of causality that is inconsistent with indeterminism when indeterminism is advocated by our best scientific theories. Far better, in my view, to develop an account of causality that is consistent with indeterminism but to use deterministic functional models where possible. This is one of the advantages of epistemic causality over Pearl's later position: it leaves open the choice of model. According to epistemic causality, an agent's causal belief graph is purely qualitative, involving neither probabilistic relationships nor deterministic functional relationships. But this does not stop one from quantifying the causally connections using either type of relationship if it is appropriate to do so. Clearly an account that does not restrict one to appealing to just probabilistic relationships or to just deterministic relationships (i) is more general than either the purely stochastic or the purely deterministic approach, (ii) satisfies the demands of science as well as intuition, and (iii) can support Pearl's semantics for counterfactuals wherever deterministic models are appropriate.

Pearl's advocacy of the Causal Markov Condition is another point that sets it apart from epistemic causality. Because Pearl uses only structural equation models and assumes that the error variables are probabilistically independent, the Causal Markov Condition follows.[66] There are three difficulties with this justification. First it depends on the acceptance of universal determinism which, as we have seen, is problematic. Second, no independent argument is given for the assumption that error variables are independent. Pearl merely points out the utility of this assumption: it yields the Causal

[65] [Pearl, 2000] 26-27.
[66] [Pearl, 2000] Theorem 1.4.1.

Markov Condition and thereby agrees with the Principle of the Common Cause and the properties that ensue.[67]

Third, there are the counterexamples to the Causal Markov Condition referred to in §2.2. Pearl attempts to salvage the condition by arguing that counterexamples either belong to quantum mechanics (in which case they are ignorable for practical purposes) or they can be explained away by invoking latent variables (dummy variables that act as common causes).[68] However, the first response is undesirable both because the quantum domain is becoming increasingly important for technology (there is already considerable interest in applications of quantum computation and quantum cryptography), and because as yet it is just a matter of conjecture that quantum indeterminacy fails to infect the macroscopic world. The second response fails because while introducing latent variables can salvage the independencies posited by the Causal Markov Condition, the condition itself often still fails since it is often the case that a *causal interpretation* of a latent variable remains implausible (analogously if A and B are probabilistically dependent but neither causes the other, then the Principle of the Common Cause requires both that there be variables that render A and B independent, *and that these variables are interpretable as common causes* of A and B, not just dummy variables).[69] Pearl also claims that the continuing interest in probabilistic analyses of causality, which often invoke the Causal Markov Condition or an equivalent, lends weight to the condition: 'The intellectual survival of probabilistic causality as an active philosophical program for the past 30 years attests to the fact that counterexamples to the Markov condition are relatively rare and can be explained away through latent variables.'[70] This is rather flimsy evidence though: the history of philosophy is littered with failed attempts (lasting longer than 30 years) to produce a viable version of an initially attractive analysis.

Epistemic causality takes a different view. It accepts that counterexamples to the Causal Markov Condition do arise, but as we saw in §4.2, the condition demonstrably holds in certain special cases. This justifies a qualified use of Pearl's methods for causal reasoning and causal discovery (but not his ontology).

I have argued, then, that Pearl need not have changed his mind about the nature of causality in order to produce an objective notion of cause: epistemic causality, which does yield objectivity, can be viewed as close to Pearl's early approach. Moreover the unqualified adoption of deterministic causal models and the Causal Markov Condition leads to a formalism that is at best a first approximation to the complexity of causality. Epistemic causality aims to capture that complexity.

[67] [Pearl, 2000] 61.
[68] [Pearl, 2000] 62.
[69] [Williamson, 2004] §4.2.
[70] [Pearl, 2000] 62-63.

6 PRICE'S PRAGMATISM

Huw Price, a proponent of the agency theory discussed in §2.4, has developed an interesting 'perspectival' conception of causality that is based on pragmatism.

While pragmatism is normally associated with Peirce's and James' attempts to analyse truth in terms of belief (alluded to in §4.3), Price delineates his pragmatism as follows:[71]

> A third form of pragmatism, and the one that interests me here, is the view that a philosophical account of a problematic notion — that of causation itself, for example — needs to begin by playing close attention to the role of the concept concerned in the *practice* of the creatures who use it. Indeed, the need to explain the use of a notion in the lives of ordinary speakers is often the original motivation for an account of this kind. Causal notions and their kin are ubiquitous in the everyday talk of ordinary people. Pragmatists argue that we cannot hope to explain this anthropological fact if we begin where metaphysics traditionally begins, at the level of the objects themselves — if we ask what causation *is*, if we begin by looking for something for causation to *be*, which will explain all these uses. Instead, pragmatists think, we need to start with the practise of *using* such notions, and to ask what role such notions play in the lives of the creatures concerned — why creatures like us should have come to describe the world in these causal terms.[72]

The last sentence portrays pragmatism as the rather uncontroversial methodological claim that philosophical investigation of a problematic notion should start with an investigation of its use. Indeed epistemic causality takes practice (the convenience of causal representations) as a starting point and only then develops a more formal account of causality and of what causality is. However, there is more to Price's pragmatic account of causality than this advice as to where to begin. Price maintains that not only should one not *start* by asking what causality is, one should not ask what causality is at all — this is the wrong question and one should instead focus on how causal notions are *used*. (Epistemic causality, in contrast, makes no such claim; indeed it provides an account of what causality is.) On the other hand Price does narrow down what causality is. For Price causality is *perspectival*: causal models are viewed from an agent's standpoint,[73] but are projected onto the world,[74] and like fictions the perspectival aspect may not

[71] [Price, 2003] describes the relationship between his form of pragmatism and truth.
[72] [Price, 2001] 105.
[73] [Price, 2004] §3.1.
[74] [Price, 2004] §3.2.

be obvious to the agent:[75]

> Perhaps causal asymmetry isn't really in the world at all, but the *appearance that it is* is a product of our own standpoint. Perhaps it is like the warmth that we see when we look at the world through rose-tinted spectacles.[76]

Yet Price's notion of causality is not mental:

> let me emphasise that pragmatism about causation is not the view that when we talk of causation we are talking *about* ourselves, in whole or in part.[77]

> I simply want to emphasise that the view is not ... that talk of causation is talk *about* agents or agency, but rather the ... doctrine that we don't understand the notion of causation — as philosophers, as it were — until we understand its origins in the lives and practice of agents such as ourselves.[78]

This is another point of difference between Price's pragmatism and epistemic causality. Epistemic causality *is* a mental notion, in the sense that talk about causality is talk about what agents ought to believe. Since Price's conception of causality is not mental, his view is not analogous to the Bayesian view that probability is rational degree of belief.[79] In contrast, epistemic causality *is* analogous to this view: just as an agent ought to adopt a certain probability function as a representation of her degrees of belief, she ought to adopt a certain directed acyclic graph as a representation of her causal beliefs.[80] Moreover just as David Lewis viewed fully objective probabilities as those degrees of belief an agent ought to adopt were she to know everything relevant,[81] so too epistemic causality views a fully objective notion of cause as those causal beliefs an agent ought to adopt were she to know everything relevant.[82]

Note though that epistemic causality does not imply that if there were no agents there would be no causation — for epistemic causality causal beliefs are idealised, the beliefs that an agent ought to adopt, which remain well-defined in the absence of agents. Price concurs on this point:

> If the concept of causation is essentially tied to our experience as agents, as my kind of ... pragmatism maintains, then of course

[75] [Price, 2004] §3.3.
[76] [Price, 1996] 153.
[77] [Price, 2001] 107.
[78] [Price, 2001] 107.
[79] [Price, 2001] 107.
[80] [Williamson, 2004] §9.10.
[81] [Lewis, 1980].
[82] [Williamson, 2004] §9.9.

> the concept would not arise in a world without agents. But this does not make it appropriate to say that if there had been no agents there would have been no causation. Pragmatism does not conflict with realism in that sense.[83]

On the other hand Price goes on to argue that only an extremely weak form of realism remains tenable:

> This view simply takes the existence claims of science at face value, and rejects any 'additional' metaphysical or philosophical viewpoint from which it would really make sense to ask 'Do these things (electrons, for example) *really* exist?' The key to weak realism is a rejection of a standpoint for ontology beyond that of science.[84]

As Price acknowledges this is not much of a realist position:

> I am following convention in calling this view a species of realism. However, it is also instructive to see the view as rejecting the traditional realist-antirealist debate altogether, at least as that debate arises within the empiricist tradition.[85]

Epistemic causality is less radical. For epistemic causality the question of whether causal relations exist in the physical world does make sense; different varieties of epistemic causality (agnosticism and anti-physicalism) give different answers to this question.

Price advocates his 'weak realism' on the basis of the following problem with the more usual 'strong realism':

> the main argument for strong realism about theoretical entities goes in terms of inference to explanatory *causes*. But this reason simply takes the notion of causation for granted, and therefore can't be applied in *support* of realism about causation. In this context, the supposed role of inference to the best explanation is epistemological — it is supposed to *justify* a belief in the reality of entities of a certain kind. My point is that such an attempt at justification would be viciously circular in the case of causation itself, in virtue of the fact by the realist's own lights, the inference presupposes realism about explanatory causes.[86]

Note though that while Price does identify a potential problem for the view that causality is a physical relation, a dismissal of strong realism leaves

[83] [Price, 2001] 108.
[84] [Price, 2001] 112.
[85] [Price, 2001] 112.
[86] [Price, 2001] 113-114.

several tenable views — Price's own weak realism (a rejection of the realism-antirealism question) but also the anti-physical and agnostic varieties of epistemic causality — none of which appeal to inference to the best causal explanation. So Price's argument does not on its own decide between weak realism and epistemic causality.

A rather counter-intuitive relativity of the agency notion of causality might provide one deciding factor:

> Suppose that the world had developed in such a way that we had fewer manipulative abilities and skills than we actually possess but that we still applied our concept of causation roughly in conformity with the agency approach. In this case the reference of the expression 'relation between events such that an actual agent could manipulate one event as a means to bringing about the other' would have been fixed on different relations, even though our way of fixing the reference would have been the same.[87]

Thus the agency theory possesses a form of subjectivity: agents with different capacities may rationally disagree about causal relationships. This looks to be a problem not just across possible worlds but across agents in this world. Just as the capacities of a human, a robot and a Venus fly trap differ, so too would causality-for-a-human, causality-for-a-robot and causality-for-a-Venus-fly-trap. Such subjectivity is attributable to Price's view of causality as a secondary quality, like colour:

> we shall take as our reference point a simple version of the orthodox dispositional theory, namely the view that to be red is to be disposed to look red to a normal observer under standard conditions. This embodies the insight that colour is a secondary quality, defining the colour concept in terms of human capacities and responses. ... Our claim is simply that the agency theory correctly portrays causation as something analogous to a secondary quality — *as* a secondary quality, in fact, on a suitably extended understanding of that notion.[88]

However, while the subjectivity of colour does not clash strongly with intuition, causality does intuitively seem to be objective. Menzies and Price reply to this objection as follows:

> Our response is to accept that this kind of relativity is a consequence of the theories concerned, but to deny that it is untoward. We make two main points in support of this conclusion.

[87] [Menzies and Price, 1993] 199.
[88] [Menzies and Price, 1993] 188-189.

> The first, as usual, is that the characteristic of causation thus identified is already a non-problematic feature of colour and the other classical secondary qualities. It is something we live with in those cases, and may be expected to accommodate ourselves to in the case of causation. Secondly, however, we want to point out that there is an important difference of degree between the two cases. As we shall explain in a moment, it turns out that causality is very much less sensitive than colour, say, to the accidents of the human situation. In this we find a basis for the intuition that causation is significantly more 'objective' than the usual secondary qualities — an intuition with which we thus concur.[89]

Although the subjectivity of the agency theory of causality may be more limited than that of the dispositional theory of colour, and although some philosophers may be able to bite the bullet and live with the subjectivity, one can avoid the subjectivity altogether. Epistemic causality does not define causality in terms of agents' capacities and is not subjective in this problematic respect. Thus the objectivity of causality provides a reason to prefer epistemic causality over the agency account.

In sum, Price's objection to strong realism about causality need not force one to adopt his rather radical rejection of the realism-antirealism debate. Epistemic causality, which views causality as mental rather than physical, remains a contender. Moreover epistemic causality might be preferred over Price's agency theory, since the latter notion of causality suffers from relativity to the capacity of agents.

7 CONCLUDING REMARKS

We have seen that contemporary theories tend to explain causality in terms of just one of its indicators, in particular physical mechanisms, probabilistic relationships, functional relationships, counterfactual relationships or agency considerations. These approaches then find it hard to explain how all the other indicators can have a bearing on our causal judgements. However, by looking first at causal beliefs and the ways in which they are constrained by knowledge of these indicators, one can account for the complexity of causality. Moreover the ensuing approach, epistemic causality, provides an account of the objectivity of causality and an answer to fundamental questions about what causality is and how we can discover causal relationships.

There are a couple of philosophical concerns one might have with epistemic causality, to do with circularity.

[89] [Menzies and Price, 1993] 199-200.

The first concern is that the characterisation of epistemic causality might be circular. Epistemic causality provides an ultimate belief interpretation of a fully objective notion of cause. Thus causality is characterised in terms of causal beliefs. But if causal beliefs are beliefs about causality then the relationship between causality and causal beliefs is circular.

While this argument is valid, it does not tell against epistemic causality, for two reasons. First and foremost, epistemic causality provides an independent route to causal beliefs: in §4.1 causal beliefs are characterised independently of ultimate belief causality, in terms of knowledge of strategic dependencies, mechanisms, temporal relations, and so on.[90] Second, epistemic causality does not claim that causal beliefs are beliefs about causality. For epistemic causality, causal beliefs are a *type* of belief, not necessarily beliefs *about* anything in particular: 'causal' modifies 'beliefs' and does not specify an object of the beliefs. The claim that causal beliefs are beliefs about ultimate belief causality is in any case implausible: it is simply implausible to suggest that when Audrey believes that smoking causes cancer, she believes that were she to know about all the relevant indicators she ought to believe that smoking causes cancer. This latter point is perhaps more obvious when made regarding the Bayesian view of probability that is analogous to epistemic causality. Here the terminology 'degrees of belief' is used for 'probabilistic beliefs' while 'chance' is used for 'probability': degrees of belief are a type of belief and are not beliefs about chances. If they were beliefs about chances, then an ultimate belief characterisation of chance in terms of degrees of belief (such as that of Lewis) would be circular. But in any case it is implausible to suggest that when Bill believes that England will win the cricket to degree 0.8, he believes that were he to know the entire history of the world and all history-to-chance conditionals he would believe that England will win the cricket to degree 0.8.

The second worry is that the relationship between epistemic causality and its indicators might be circular. According to epistemic causality, causal beliefs depend on knowledge of the multifarious indicators of causality. If these indicators are themselves reducible to causal notions then it is natural to suspect circularity. For example, we might want to understand temporal direction in terms of causality — but how can this be possible if temporal knowledge helps delimit the causal relation? In contrast, if we simply reduce causality to counterfactuals then an account of temporal direction in terms of causality is more obviously non-circular.

In fact though, epistemic causality leaves open the question of which

[90] [Williamson, 2004] §9.8 deals with the case in which positive causal knowledge can constrain causal beliefs. In that case causal beliefs can depend upon ultimate belief causality. But there is no circularity there either, because ultimate belief causality is characterised in terms of causal beliefs relative to background knowledge that includes all knowledge of strategic dependencies, mechanisms and so on, but that does not include knowledge of ultimate causal relations.

reductive relationships obtain amongst its indicators. Epistemic causality offers a functional explanation of causality in terms of its convenience, and a characterisation of the causal relation in terms of rational beliefs, but not a reductive analysis of causality in terms of its indicators. Consider an analogy in medicine. When a condition is poorly understood, one may posit a *syndrome* and characterise it in terms of its indicators. For example, Tourette's syndrome is characterised (implicitly defined) in terms of involuntary tics and uncontrollable verbalisation, in particular the use of obscene language and the tendency to repeat uttered words. No commitment is made as to what actually causes what — indeed the causal picture regarding Tourette's syndrome is still unclear. As long as the characterisation of the syndrome latches onto something objective, it will suffice for diagnosis and treatment. Similarly, a characterisation of causality that latches onto something objective can offer a way of handling causality without presupposing relationships amongst its indicators: temporal direction can be a good indicator of causal direction whether or not the former is reducible to the latter.[91]

Thus epistemic causality offers a powerful alternative to the standard accounts of causality, yet one that is compatible with a range of philosophical agendas.[92]

Jon Williamson
University of Kent

BIBLIOGRAPHY

[Bacon, 1620] Francis Bacon: 'The New Organon', Lisa Jardine and Michael Silverthorne (eds.), Cambridge: Cambridge University Press 2000.

[Benacerraf, 1973] Paul Benacerraf: 'Mathematical truth', in [Benacerraf and Putnam, 1983], pages 403-420.

[Benacerraf and Putnam, 1983] Paul Benacerraf and Hilary Putnam (eds.): 'Philosophy of mathematics: selected readings', Cambridge: Cambridge University Press, second edition.

[Cartwright, 1997] Nancy Cartwright: 'What is a causal structure?', in [McKim and Turner, 1997], pages 343-357.

[Cartwright, 1999] Nancy Cartwright: 'Causality: independence and determinism', in [Gammerman, 1999], pages 51-63.

[Cartwright, 2001] Nancy Cartwright: 'What is wrong with Bayes nets?', The Monist 84(2), pages 242-264.

[91] Of course for the the anti-physicalist form of epistemic causality, any reduction of temporal direction to causal direction will yield a *mental* notion of temporal direction. In contrast, the agnostic form of epistemic causality leaves this issue open too: the mental epistemic account is one interpretation of causality, but there may be other viable interpretations of causality to which time (or other indicators) can be reduced.

[92] I am very grateful to Donald Gillies, Huw Price and Federica Russo for comments, and to Oxford University Press for permission to reprint passages from [Williamson, 2004] in §2 and §3.

[Cooper, 1999] Gregory F. Cooper: 'An overview of the representation and discovery of causal relationships using Bayesian networks', in [Glymour and Cooper, 1999], pages 3-62.

[Cooper, 2000] Gregory F. Cooper: 'A Bayesian method for causal modeling and discovery under selection', in Proceedings of the Conference on Uncertainty in Artificial Intelligence 2000, pages 98-106.

[Corfield and Williamson, 2001] David Corfield and Jon Williamson (eds.): 'Foundations of Bayesianism', Kluwer Applied Logic Series, Dordrecht: Kluwer Academic Publishers.

[Dai *et al.*, 1997] Honghua Dai, Kevin Korb, Chris Wallace and Xindong Wu: 'A study of causal discovery with weak links and small samples', in Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI-97), Nagoya, Japan, August 23-29 1997.

[Dash and Druzdzel, 1999] Denver Dash and Marek Druzdzel: 'A Fundamental Inconsistency Between Causal Discovery and Causal Reasoning', in Proceedings of the Joint Workshop on Conditional Independence Structures and the Workshop on Causal Interpretation of Graphical Models, The Fields Institute for Research in Mathematical Sciences, Toronto, Canada.

[Dawid, 2001] A.P. Dawid: 'Causal inference without counterfactuals', in [Corfield and Williamson, 2001], pages 37-74.

[Dowe, 1993] Phil Dowe: 'On the reduction of process causality to statistical relations', British Journal for the Philosophy of Science 44, pages 325-327.

[Dowe, 1996] Phil Dowe: 'Backwards causation and the direction of causal processes', Mind 105, pages 227-248.

[Dowe, 1999] Phil Dowe: 'The conserved quantity theory of causation and chance raising', Philosophy of Science 66 (Proceedings), pages S486-S501.

[Dowe, 2000] Phil Dowe: 'Causality and explanation: Review of Salmon', British Journal for the Philosophy of Science 51, pages 165-174.

[Dowe, 2000b] Phil Dowe: 'Physical causation', Cambridge: Cambridge University Press.

[Earman, 1992] John Earman: 'Bayes or bust?', Cambridge, Massachusetts: M.I.T. Press.

[Freedman and Humphreys, 1999] David Freedman and Paul Humphreys: 'Are there algorithms that discover causal structure?', Synthese 121, pages 29-54.

[Gammerman, 1999] Alex Gammerman (ed.): 'Causal models and intelligent data management', Berlin: Springer.

[Glymour, 1997] Clark Glymour: 'A review of recent work on the foundations of causal inference', [McKim and Turner, 1997], pages 201-248.

[Glymour, 2001] Clark Glymour: 'The Mind's Arrows: Bayes nets and graphical causal models in psychology', Cambridge, Massachusetts: The M.I.T. Press.

[Glymour and Cooper, 1999] Clark Glymour and Gregory F. Cooper (eds.): 'Computation, causation, and discovery', Cambridge, Massachusetts: The M.I.T. Press.

[Hagmayer and Waldmann, 2002] York Hagmayer and Michael R. Waldmann: 'A constraint satisfaction model of causal learning and reasoning', in Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society, Mahwah, NJ: Erlbaum.

[Hausman, 1999] Daniel M. Hausman: 'The mathematical theory of causation', review of [McKim and Turner, 1997], British Journal for the Philosophy of Science 50, pages 151-162.

[Hausman and Woodward, 1999] Daniel M. Hausman and James Woodward: 'Independence, invariance and the causal Markov condition', British Journal for the Philosophy of Science 50, pages 521-583.

[Heckerman *et al.*, 1999] David Heckerman, Christopher Meek and Gregory Cooper: 'A Bayesian approach to causal discovery', in [Glymour and Cooper, 1999], pages 141-165.

[Hempel and Oppenheim, 1948] Carl G. Hempel and Paul Oppenheim: 'Studies in the logic of explanation', with Postscript in [Pitt, 1988], pages 9-50.
[Howson and Urbach, 1989] Colin Howson and Peter Urbach: 'Scientific reasoning: the Bayesian approach', Chicago: Open Court, Second edition, 1993.
[Humphreys, 1997] Paul Humphreys: 'A critical appraisal of causal discovery algorithms', in [McKim and Turner, 1997], pages 249-263.
[Humphreys and Freedman, 1996] Paul Humphreys and David Freedman: 'The grand leap', British Journal for the Philosophy of Science 47, pages 113-123.
[James, 1907] William James: 'What pragmatism means', in 'Pragmatism: A new name for some old ways of thinking', New York: Longman Green and Co, pages 17-32.
[Karimi and Hamilton, 2000] Kamran Karimi and Howard J. Hamilton: 'Finding Temporal Relations: Causal Bayesian Networks versus C4.5', Proceedings of the Twelfth International Symposium on Methodologies for Intelligent System (ISMIS'2000), Charlotte, NC, USA, October 2000.
[Karimi and Hamilton, 2001] Kamran Karimi and Howard J. Hamilton: 'Learning causal rules', Technical Report CS-2001-03, Department of Computer Science, University of Regina, Saskatchewan, Canada.
[Korb, 1999] Kevin B. Korb: 'Probabilistic causal structure', in H. Sankey (ed.): 'Causation and laws of nature', Dordrecht: Kluwer, pages 265-311.
[Korb and Nicholson, 2003] Kevin B. Korb and Ann E. Nicholson: 'Bayesian artificial intelligence', London: Chapman and Hall / CRC Press UK.
[Lad, 1999] Frank Lad: 'Assessing the foundation for Bayesian networks: a challenge to the principles and the practice', Soft Computing 3(3), pages 174-180.
[Lemmer, 1996] John F. Lemmer: 'The causal Markov condition, fact or artifact?', SIGART Bulletin 7(3), pages 3-16.
[Lewis, 1973] David K. Lewis: 'Causation', with postscripts in [Lewis, 1986], pages 159-213.
[Lewis, 1980] David K. Lewis: 'A subjectivist's guide to objective chance', in [Lewis, 1986], pages 83-132.
[Lewis, 1986] David K. Lewis: 'Philosophical papers volume II', Oxford: Oxford University Press.
[Lewis, 1986b] David K. Lewis: 'Causal explanation', in [Lewis, 1986], pages 214-240.
[Lewis, 2000] David K. Lewis: 'Causation as influence', The Journal of Philosophy 97(4), pages 182-197.
[Mani and Cooper, 1999] Subramani Mani and Gregory F. Cooper: 'A study in causal discovery from population-based infant birth and death records', in Proceedings of the AMIA Annual Fall Symposium 1999, Philadelphia: Hanley and Belfus Publishers, pages 315-319.
[Mani and Cooper, 2000] Subramani Mani and Gregory F. Cooper: 'Causal discovery from medical textual data', in Proceedings of the AMIA annual fall symposium 2000, Philadelphia: Hanley and Belfus Publishers, pages 542-546.
[Mani and Cooper, 2001] Subramani Mani and Gregory F. Cooper: 'Simulation study of three related causal data mining algorithms', in Proceedings of the International Workshop on Artificial Intelligence and Statistics 2001, San Francisco: Morgan Kaufmann, pages 73-80.
[McKim and Turner, 1997] Vaughn R. McKim and Stephen Turner: 'Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences', Notre Dame: University of Notre Dame Press.
[Menzies and Price, 1993] Peter Menzies and Huw Price: 'Causation as a secondary quality', British Journal for the Philosophy of Science 44, pages 187-203.
[Pearl, 1988] Judea Pearl: 'Probabilistic reasoning in intelligent systems: networks of plausible inference', San Mateo, California: Morgan Kaufmann.
[Pearl, 1999] Judea Pearl: 'Graphs, structural models, and causality', in [Glymour and Cooper, 1999], pages 95-138.

[Pearl, 2000] Judea Pearl: 'Causality: models, reasoning, and inference', Cambridge: Cambridge University Press.

[Peirce, 1877] Charles Sanders Peirce: 'The fixation of belief', Popular Science Monthly 12, pages 1-15.

[Pitt, 1988] Joseph C. Pitt (ed.): 'Theories of explanation', Oxford: Oxford University Press.

[Polya, 1945] George Polya: 'How to solve it', second edition, Penguin 1990.

[Polya, 1954] George Polya: 'Induction and analogy in mathematics', volume 1 of 'Mathematics and plausible reasoning', Princeton: Princeton University Press.

[Polya, 1954b] George Polya: 'Patterns of plausible inference', volume 2 of 'Mathematics and plausible reasoning', Princeton: Princeton University Press.

[Popper, 1934] Karl R. Popper: 'The Logic of Scientific Discovery', with new appendices of 1959, London: Routledge 1999.

[Price, 1991] Huw Price: 'Agency and probabilistic causality', British Journal for the Philosophy of Science 42, pages 157-176.

[Price, 1992] Huw Price: 'Agency and causal asymmetry', Mind 101, pages 501-520.

[Price, 1992b] Huw Price: 'The direction of causation: Ramsey's ultimate contingency', Philosophy of Science Association 1992(2), pages 253-267.

[Price, 1996] Huw Price: 'Time's arrow and Archimedes' point: new directions for the physics of time', New York: Oxford University Press.

[Price, 2001] Huw Price: 'Causation in the special sciences: the case for pragmatism', in Domenico Costantini, Maria Carla Galavotti and Patrick Suppes (eds.): 'Stochastic Causality', Stanford, California: CSLI Publications, pages 103-120.

[Price, 2003] Huw Price: 'Truth as convenient friction', Journal of Philosophy 100, pages 167-190.

[Price, 2004] Huw Price: 'Models and modals', in Donald Gillies (ed.): 'Laws and models in science', London: King's College Publications.

[Railton, 1978] Peter Railton: 'A deductive-nomological model of probabilistic explanation', in [Pitt, 1988], pages 119-135.

[Reichenbach, 1956] Hans Reichenbach: 'The direction of time', Berkeley and Los Angeles: University of California Press 1971.

[Russell, 1913] Bertrand Russell: 'On the notion of cause', Proceedings of the Aristotelian Society 13, pages 1-26.

[Salmon, 1980] Wesley C. Salmon: 'Causality: production and propagation', in [Sosa and Tooley, 1993], chapter 9.

[Salmon, 1980b] Wesley C. Salmon: 'Probabilistic causality', in [Salmon, 1998], pages 208-232.

[Salmon, 1984] Wesley C. Salmon: 'Scientific explanation and the causal structure of the world', Princeton: Princeton University Press.

[Salmon, 1997] Wesley C. Salmon: 'Causality and explanation: a reply to two critiques', Philosophy of Science 64(3), pages 461-477.

[Salmon, 1998] Wesley C. Salmon: 'Causality and explanation', Oxford: Oxford University Press.

[Scheines, 1997] Richard Scheines: 'An introduction to causal inference', in [McKim and Turner, 1997], pages 185-199.

[Sosa and Tooley, 1993] Ernest Sosa and Michael Tooley (eds.): 'Causation', Oxford: Oxford University Press.

[Spirtes *et al.*, 1993] Peter Spirtes, Clark Glymour and Richard Scheines: 'Causation, Prediction, and Search', Cambridge, Massachusetts: The M.I.T. Press, second edition 2000.

[Stankovski *et al.*, 2001] V. Stankovski, I. Bratko, J. Demsar and D. Smrke: 'Induction of hypotheses concerning hip arthoplasty: a modified methodology for medical research', Methods of Information in Medicine 40, pages 392-396.

[Suppes, 1970] Patrick Suppes: 'A probabilistic theory of causality', Amsterdam: North-Holland.

[Tenenbaum and Griffiths, 2001] Joshua B. Tenenbaum and Thomas L. Griffiths: 'Structure learning in human causal induction', in T. Leen, T. Dietterich, and V. Tresp (eds.): Advances in Neural Information Processing Systems 13, Cambridge, Massachusetts: The M.I.T. Press, pages 59-65.

[Tong and Koller, 2001] Simon Tong and Daphne Koller: 'Active Learning for Structure in Bayesian Networks', in B. Nebel (ed.): Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, San Francisco: Morgan Kaufmann, pages 863-869.

[Waldmann, 2001] Michael R. Waldmann: 'Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm', Psychonomic Bulletin and Review 8, pages 600-608.

[Waldmann and Martignon, 1998] Michael R. Waldmann and Laura Martignon: 'A Bayesian network model of causal learning', in M.A. Gernsbacher and S.J. Derry (eds.): 'Proceedings of the Twentieth Annual Conference of the Cognitive Science Society', Mahwah, New Jersey: Erlbaum, pages 1102-1107.

[Wallace and Korb, 1999] Chris S. Wallace and Kevin B. Korb: 'Learning linear causal models by MML sampling', in [Gammerman, 1999], pages 88-111.

[Wendelken and Shastri, 2000] Carter Wendelken and Lokendra Shastri: 'Probabilistic Inference and Learning in a Connectionist Causal Network', In Proceedings of the Second International Symposium on Neural Computation, Berlin, May 2000.

[Williamson, 2004] Jon Williamson: 'Bayesian nets and causality: philosophical and computational foundations', Oxford: Clarendon Press.

[Woodward, 1997] James Woodward: 'Causal models, probabilities, and invariance', in [McKim and Turner, 1997], pages 265-315.

[Yoo *et al.*, 2002] Changwon Yoo, V. Thorsson and G. Cooper: 'Discovery of Causal Relationships in a Gene-regulation Pathway from a Mixture of Experimental and Observational DNA Microarray Data', in Proceedings of the Pacific Symposium on Biocomputing, New Jersey: World Scientific, pages 498-509.