Why Think Causally?
Published under the title: "Why Represent Causal Relations?"

Michael Strevens

ABSTRACT

Why do we represent the world around us using causal generalizations, rather than, say, purely statistical generalizations? Do causal representations contain useful additional information, or are they merely more efficient for inferential purposes? This chapter considers the second kind of answer: it investigates some ways in which causal cognition might aid us not because of its expressive power, but because of its organizational power.

Three styles of explanation are considered. The first, building on the work of Reichenbach in *The Direction of Time*, points to causal representation as especially efficient for predictive purposes in a world containing certain pervasive patterns of conditional independence. The second, inspired by work of Woodward and others, finds causal representation to be an excellent vehicle for representing all-important relations of manipulability. The third, based in part on my own work, locates the importance of causal cognition in the special role it reserves for information about underlying mechanisms. All three varieties of explanation show promise, but particular emphasis is placed on the third.

Contents

1.   The Question

Why should the mind represent causal relations? Because they are there, goes the simplest answer. But this is not good enough: there is much that is there in the external world, but that is not represented—though it could be—because it is not important enough to merit a place in the limited space inside the skull.

Thus the next simplest answer: we represent causal information because knowledge of causation is important for getting around the world safely and extracting from it what we need and desire. True enough, but again not entirely satisfactory. Much of the knowledge that enables us to navigate the world for fun and profit can be represented in non-causal form, for example, as information about correlations.

This observation suggests a strategy for answering the question. Compare and contrast information about causal relations with information about correlations, and show that causal information is somehow better, at least sometimes, for getting what we want. What follows is an attempt to implement this strategy.

There are, broadly speaking, two different advantages that causal information might have over statistical information. (In what follows, I use *statistical* more or less interchangeably with *non-causal*.) First, it might be that there are some aspects of the world that can be captured using causal representations but not mere statistical representations, and that knowledge of these properties of our surroundings has some practical use for us. Second, it might be that everything worth knowing for practical purposes can in principle be expressed using statistical representations, but that the causal representation of certain kinds of facts is especially efficient given our particular means and ends.

I call these two explanations of our use of causal representations respectively the external and the internal explanations, since, whereas the external explanation points to aspects of the outside world that can be represented only using causal representations, the internal explanation points to elements of the system of causal representation itself—the system existing inside our heads—that are especially user-friendly.

Which explanation is correct? It is certainly too soon to answer such a categorical question. This chapter will, however, declare an interest, focusing on explanations of the internal variety, for the following reason.

Decades, even centuries of work on the metaphysics of causation by philosophers holds out no great hope for the external approach. It is not that philosophers are agreed that causal claims say nothing that cannot be said by statistical claims—far from it. Debate is as lively as ever as to whether there are sui generis causal relations in the world that it is the privilege of causal language alone to represent, with the realists about causation arguing

for and the empiricists against (Sosa and Tooley 1993). It is rather that, even on realist theories that posit such *sui generis* facts, what is implied by the facts over and above a certain pattern of correlations is not, on the face of it, information that is in itself useful in day-to-day life. In other words, even on realist theories of causation, the practical use of a causal fact lies entirely in the correlations it entails. There is thus a working philosophical consensus that correlations are good enough for everyday life—a consensus that, if correct, implies that an attempt to give an external explanation for our use of causal representations must fail. This consensus has been challenged, recently and vigorously by Woodward (2003) in particular (discussed in section 3.2 below), but as things currently stand, there is good if not conclusive reason to focus on the internal approach to explaining the existence of causal cognition.

A good discipline for an internalist explainer is to assume, for tactical reasons, empiricism about causal language, that is, to assume that there is nothing said by causal claims—more generally, nothing captured by causal representations—that cannot be said by statistical claims. The advantage of the causal way of speaking and thinking must then of necessity be found, not in what is said, but in how it is said. In what follows, I adopt this discipline.

Note that the statistical claims—the claims about correlations and so on—that exhaust the content of a causal claim will be quite complex. Empiricist theories of causation long ago abandoned simple analyses of causal language on which, for example, to say *c is a cause of e* is just to say *c is correlated with e*. It is this complexity that opens the way to an internal explanation of causal representation, since, if a causal scheme organizes the facts about correlation rather differently from, and apparently more simply than, a statistical scheme, then there may be very real advantages to choosing one organization over the other. The challenge is to show that the causal scheme picks out especially important parts of the statistical information and arranges them conveniently for later use.

Imagine, then, a world, perhaps our own, at any rate in many ways not too different from our own, in which every fact can be captured by statistical claims—the kind of world imagined by the metaphysical empiricists. Show that, in such a world, creatures like us would gain some real practical advantage from causal cognition. Show, in other words, that if causation did not exist, it would be necessary, or at least highly desirable, to invent it.

2. Two Features of Causal Representation

The first question to ask is: what are the features that distinguish a causal scheme of representation from a statistical scheme? Perhaps that is too large and unwieldy a way to begin, however. I will ask instead: what are the organizational or logical features that distinguish the causal claim *c is a cause of e* from the statistical claim *c is correlated with e*?

I will organize this chapter around two such features: *asymmetry* and the supposition of an *underlying mechanism*. Each in its own way points to an interesting explanation, or explanations, of the utility of the causal scheme of representation. Let me characterize, loosely, the relation between causation, asymmetry, and underlying mechanism.

First, asymmetry. Correlation is a symmetric relation, in the sense that *c is correlated with e* means exactly the same thing as *e is correlated with c*. Causation is not: *c is a cause of e* means something very different from *e is a cause of c* (though in some cases, both may be true).

As promised at the end of the preceding section, I assume for the sake of the argument that the asymmetric information is represented by the causal claim can also be represented by some set of statistical claims. (Certainly, though correlation itself is a symmetric relation, there is normally no shortage of asymmetry in the complete statistics of the associations between two event types *c* and *e*.) The question I ask in section 3 is what advantage there might be in tracking an asymmetric relation between event types rather than

a symmetric relation such as correlation. The various answers will take the discussion far beyond the advantages of asymmetry itself.

Second, underlying mechanism. Unlike the statistical claim *c is correlated with e*, the causal claim *c is a cause of e* implies the existence, or so I will suppose, of a mechanism connecting *c* and *e*, and in virtue of which *c* causes *e*. You should not think that facts about mechanism are supposed to exist entirely at the metaphysical level, as ineffable necessary connections, hidden strings, or causal *oomph*s. On the contrary, the nature of a causal connection's underlying mechanism is normally amenable to regular empirical investigation: the conditions required for, and the various intermediate steps that constitute, its operation can be inferred or even directly observed.

In any case, for the tactical reasons given above, I assume that the information contained in claims about the workings of a causal connection's underlying mechanism is ultimately statistical information. But a causal claim makes room for this information in its own characteristic way; the question, then, is whether representing the relevant facts as information about an "underlying mechanism" has practical advantages for the user.

Of what follows, the discussion of asymmetry is drawn for the most part from previous work on the philosophy and psychology of causation by Reichenbach (1956), Pearl (2000), Glymour (2001), Woodward (2003), and others; for this reason, I keep the presentation relatively short and simple, directing your attention where appropriate to the primary sources. In the discussion of the inferential role of information about underlying mechanism, I strike out on my own, developing ideas from my earlier work on the psychology of causal reasoning (Strevens 2000).

A final clarification, a terminological note, and a hint to the reader: Some claims of the form *c is a cause of e* are what philosophers call *singular* causal claims, concerning particular events or occurrences, and some are causal generalizations, concerning types of event. For example, *my eating eggplant last night prevented me from sleeping* is a singular claim, whereas

eating vegetables causes insomnia is a causal generalization. Both kinds of causal claims play an essential role in causal cognition, singular claims because the end product of causal reasoning is so often the prediction and control of individual facts and happenings, generalizations because reasoning about singular events is invariably guided by information about the causal tendencies of the event types to which they belong.

Some of the literature on causal generalizations, including much work on Bayesian networks, talks about *variables* rather than *event types*. The differences between these two notions are not important for the purposes of the discussion in this chapter.

Each of sections 3.1, 3.2, and 4 is self-contained, and so may be read independently of the others. If you have time for nothing else, read section 4.

3. ASYMMETRY

What asymmetric aspects of the world's statistical web might be especially usefully represented by a causal schema? Usefully, you might ask, with respect to what end? The three cardinal aims of science are said to be explanation, prediction, and control. I will organize the discussion of asymmetry around the more practical goals of prediction and control.

3.1 Asymmetry and Prediction

Hans Reichenbach suggested in *The Direction of Time* (Reichenbach 1956) that the roots of causal thinking could be found in certain pervasive asymmetrical statistical patterns in our world. I will focus on one such pattern, which I will call the *Reichenbach asymmetry* (invidiously, since Reichenbach investigated several such patterns, and synecdochically, since there is more to

the pattern than its lack of symmetry).[1] Reichenbach's view that causal representations always go along with certain statistical asymmetries has been put to work in various ways by philosophers and other students of causality. Preeminent for my purposes is the asymmetries' use to give an internal account of the purpose of causal cognition that invokes the utility of causal representation as either an explanatory or a predictive tool.

Reichenbach himself emphasized the explanatory importance of causal representation (see, in particular, p. 152 of *The Direction of Time*). I will focus rather on prediction—not much of a departure, since for logical empiricists such as Reichenbach, explanation and prediction have much in common. What follows does not capture anything like the full range and subtlety of Reichenbach's attempt to understand causality and causal thinking in terms of statistical relations, and indeed, the particular position I lay out cannot be ascribed to Reichenbach at all; it is a deliberately simplified version of section 22 of *The Direction of Time*.[2]

Let me begin by characterizing the statistical pattern that I am calling the Reichenbach asymmetry. The fundamental notion employed in the characterization is a statistical relation captured by a construction that I call a *Reichenbach dyad*. A Reichenbach dyad consists of two entities: a single event that I call the *focal event*, and a set of events that I call the *parent events*. The term *parent* hints at the causal interpretation that is to come, but you should bear in mind that the definition of a Reichenbach dyad is purely statistical— it makes no reference to causal relations.

---

1. Reichenbach connected the asymmetry of his statistical patterns to the statistical mechanical roots of the second law of thermodynamics and ultimately to the direction of time, but this aspect of his work will be passed over here.

2. I might add, though, that the ideas about causality presented in *The Direction of Time*, a work left unfinished at the time of Reichenbach's death, are not easily understood as a unified whole. Like another well-known philosopher of causality, he could be accused of having given several incompatible definitions of cause. I suppose that Reichenbach would have replied, in the logical empiricist spirit, that each of his definitions has its advantages and disadvantages, and that it would be a philosophical error to insist that any one definition must be uniquely correct.

A focal event and a set of parent events form a Reichenbach dyad just in case

1. The parent events occur before the focal event,[3] and

2. Conditional on the parent events, the occurrence of the focal event is statistically independent of the occurrence of any previous event, that is, any event occurring before the focal event. In other words, once the parent events are taken into account, the occurrence of any other event preceding the focal event can be ignored in determining the focal event's probability.

This is only a rough version of condition (2); the precise condition will be specified shortly.

Three remarks. First, the definition of a Reichenbach dyad should remind you of the *causal Markov condition* from the Bayesian networks literature discussed elsewhere in this volume; however, unlike that condition, it makes no reference to causal relations. (In this respect, it is closer to the original acausal Markov condition used to represent purely statistical information in acausal Bayes nets.)

Second, on Reichenbach's view, the probabilities attaching to events are simply the probabilities attaching to the corresponding event types. The statistical clause of the definition of a Reichenbach dyad, then, is satisfied in virtue of probabilistic facts about event types, whereas the temporal clause is satisfied in virtue of facts about the temporal ordering of the particular events in the dyad.[4] For a philosopher with a frequency-based notion of probability—such as Reichenbach—the probability distributions attached

---

3. It might be better to say *before or at the same time as* the focal event, but it will be simpler to leave things as they are in the main text.

4. The probabilistic facts about event types, note, will themselves often refer to the temporal order of the probabilified events. For example, the probability of hearing a loud bang *after* the trigger is pulled is much higher than the probability of hearing a bang before the trigger is pulled, and so on.

to event types will be determined in turn by the statistics of singular events, so the asymmetry as a whole will be a matter of the patterning of singular events.

Third, the definition needs the following refinement: the independence relation must hold not only conditional on the occurrence of all of the parent events, but also conditional on any combination of the parent events' occurrence and non-occurrence. If there are two parent events, for example, then the relation must hold conditional on the occurrence of both, on the non-occurrence of both, on the occurrence of the first and the non-occurrence of the second, and on the non-occurrence of the first and the occurrence of the second.

Reichenbach notes that the world we live in is statistically patterned in a certain way: it is full of Reichenbach dyads. More exactly, for almost every event, there is a Reichenbach dyad for which it is the focal event, and—this is what gives the claim bite—these dyads have relatively small sets of parent events.[5]

This pattern is the Reichenbach asymmetry. That it is indeed an asymmetry is due to the temporal asymmetry in the definition of a dyad. (It was in virtue of this asymmetry, or something close to it, that Reichenbach suggested that the time order of events can be determined from information about conditional independences; unlike Reichenbach, I am of course taking a time ordering of events as given.)

In a Reichenbach-asymmetric world, Reichenbach held, there is a very close relationship between the dyads and the causal structure of the world. To simplify somewhat,[6] there is a Reichenbach dyad in the actual world just

---

5. Without the additional claim, nothing has been said, since there is a trivial dyad for any focal event in which the set of parent events contains every other event.

6. What follows posits a relationship that is closer than anything Reichenbach would have endorsed; as I remarked earlier, I am simplifying his view considerably. Specifically, what I identify as a necessary and sufficient condition for one event to be the cause of another is for Reichenbach only a necessary condition. See *The Direction of Time*, §22 for the details.

in case there is a causal structure in which the parent events of the dyad are the direct causes of the focal event, that is, just in case there is a causal structure of the form shown in figure 1.
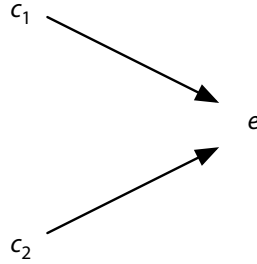

Figure 1: The causal structure for a Reichenbach dyad with c_1 and c_2 as parent events and e as the focal event. Arrows run from cause to effect.

The co-occurrence— not necessarily an equivalence, note— of causal and statistical structure is the linchpin of everything that follows. It can be exploited in various ways, all suggested by Reichenbach in some form.

First, it can be used to construct a metaphysics of cause, or as Reichenbach would say, a semantics for causal language: define *cause* so that for one event to be a cause of another simply is for certain Reichenbach dyads to exist. On the most straightforward definition, one event is a (direct) cause of another just in case the one belongs to the parents in a dyad for which the other is focal.

Second, an epistemologist may use the statistical structure to discover causal structure, inferring the existence of the sort of structure shown in figure 1 wherever the right Reichenbach dyads exist. Spirtes et al. (2000)'s system for inferring causal structure is, in effect, a very sophisticated version of this proposal (though it employs statistical subtleties that I have not even hinted at here).

Third, the coincidence of causal and statistical structure underwrites an alternative representational scheme for statistical facts, a scheme that repre-

sents the fact that a set of events constitute a Reichenbach dyad using the sort of graph structure shown in figure 1 rather a set of probability statements. It is of course this third use that interests me, suggesting as it does the beginnings of an internal explanation for causal cognition.

The sort of internal explanation I have in mind hinges on three posits:

1. That the world is Reichenbach asymmetric,

2. That a causal network diagram—a *directed acyclic graph*, or DAG, in the Bayesian networks parlance—is an especially compact way of representing Reichenbach dyads in a Reichenbach-asymmetric world, and

3. That representing the Reichenbach dyads has great practical utility in a Reichenbach-asymmetric world, given the asymmetries in our own epistemic situation.

These three premises imply, if not that a causal schema is indispensable for certain kinds of reasoning about the world, then at least that it is highly advantageous. (A caveat: I will not have much to say about the relationship between DAGs and causal representation; I simply assume that they tend to go together, leaving the hard work to the neo-Reichenbachians.)

In what follows I assume the existence of the Reichenbach asymmetry for the sake of the argument. What of the other two premises? The advantage of a causal network diagram, or DAG, as a representation of the facts about Reichenbach dyads and other more complex conditional facts about independence of the same sort is as follows. A many-noded DAG— a graph that represents the causal structure in which a number of events are embedded—identifies a Reichenbach dyad for every event whose parents are represented. To record this information as a set of statements about probability would require every such event to be mentioned a number of times, once as a focal event and then many times as parent. By contrast, the DAG "mentions" every event just once. The DAG thus provides a very compact (and, I should add, computationally convenient) representation of

the dyadic relations. The compactness of the DAG is due in considerable part, note, to that aspect of the world's Reichenbach asymmetry which guarantees that almost every parent in a dyad is the focal event of some other dyad.

To the last premise, then: why is tracking the Reichenbach dyads so useful, and useful in particular in a world pervaded by Reichenbach asymmetry? As I said above, I will focus on predictive utility, that is, the usefulness of dyadic information in your trying to ascertain whether some event will occur in the future, given what you know now.

Consider how predictions are made using purely statistical information. You wish to predict whether or not some event e will occur. What you want to calculate is the probability of e; if it is high, you predict that e will occur, and act accordingly, if low, not. In calculating the probability, you want to take into account all of your background knowledge, as this will result in the most accurate predictions (the *principle of total evidence*). Thus you want to calculate the probability of e conditional on everything you know. An onerous task! You must not only keep track of every event that has occurred, but you must make use of a probability distribution so fantastically detailed that it is well defined over all of these events.

Your task is much simpler, however, if you know that a very small subset of your background knowledge screens off the rest from the event e that you are trying to predict—if you know that, once you conditionalize on the small subset, conditionalizing on the rest will make no difference to the probability of e. Then you may, with a serene heart and a clear epistemological conscience, track only the events in this subset, and invoke a probability distribution defined over only these events (and of course over e).

It is precisely this predictive advantage that a Reichenbach dyad supplies, to organisms like us who have direct knowledge of past but not of future events. In the simplest case, the event e that you want to predict is the focal event of a Reichenbach dyad in which all of the parent events are known to have occurred and you have no knowledge of events occurring at the same

time or after the focal event. The probability of *e* conditional on the parent events is then the best predictor of *e* available to you. Thus only *e*'s parents need to be taken into account, enormously simplifying your predictive task.

The more complex case where you have knowledge about only some of the parent events cannot be handled so easily, but again, information about the conditional independences is extremely helpful.

In either case, note, the *asymmetry* of the Reichenbach asymmetry comes into its own: it is because the temporal asymmetry of the dyad reflects the temporal asymmetry in our knowledge of the world—we have far more knowledge of the past than of the future—that the dyads are so useful, despite the temporally qualified nature of the independence relations they represent.[7]

Let me summarize the internal explanation of the role of causal cognition sketched above. A would-be predictor has great need of information about conditional independences. In a Reichenbach-asymmetric world, the class of such information most useful to asymmetric knowers like us can be stored very compactly using a causal network representation. The role of causal representations in our psychology is to exploit this efficiency. (For a fuller discussion of the psychological utility of DAGs, see Glymour (2001).)

Two non-psychological remarks. Observe that a world containing the Reichenbach asymmetry will be symmetric overall if it also contains the pattern you might call the reverse Reichenbach asymmetry: for (almost) every event *e*, there is a set of "child" events that (a) occur after *e*, and (b) create a conditional independence between *e* and any other event occurring after *e*. In a deterministic world, you ought to expect precisely such a pattern. As writers on causation and statistical mechanics beginning with

---

7. An exception to this rule, the kind of case in which you have some direct knowledge of events caused by the focal event—a case of retrodiction, presumably—can contribute to the internal explanation upon the introduction of statistical relations more complex than those represented by Reichenbach dyads. These relations are also very efficiently represented by a DAG.

Reichenbach have noted, however, though a deterministic world may have the reverse asymmetry at the microlevel, if it conforms to the second law of thermodynamics it will not have the asymmetry nearly so extensively at the macrolevel, that is, relative to a coarse-graining of events (since any candidate set of children will have to specified very finely for the independence relation to hold, so will be lost in a coarse-graining). Such a world will have both the forward and backward-looking asymmetries, then, but the forward-looking asymmetry may be far more apparent to macroscopic beings like us. (Some writers would argue that the second law also explains the asymmetry of our epistemic situation in virtue of which, as argued above, the forward-looking asymmetry is far more *useful* to macroscopic beings.)

As noted at the beginning of this section, I have discussed just one asymmetric pattern used in Reichenbach's investigation of the causal and statistical order of the world. There are other such patterns. (*The Direction of Time* itself discusses two more, namely, the *conjunctive fork* and the *mark asymmetry* that inspires Salmon (1984)'s account of causation.) The work described in this section should, then, be regarded as part of a wider research program, which is perhaps still in its early stages.

3.2   Asymmetry and Control

To have control over an event is to be able to manipulate nature in such a way that the event occurs. Woodward (2003) has argued that the most important function of causal representation is to encode manipulability relations. Because the relation between causality and manipulability is discussed at length elsewhere in this volume, in the Campbell and Woodward chapters, I will confine myself in this section to the briefest sketch of the way in which manipulability might provide an internal explanation of causal cognition.

A simple example demonstrates the difference between manipulability and prediction. Consider the relationship between a switch and light. There is a very strong correlation between the switch's being in the *on* position and

the light's shining. As a result, you can use your knowledge of the switch's position to predict whether the light is shining, or just as surely, you can use your knowledge as to whether the light is shining to predict the switch's position. But this predictive symmetry breaks down when it comes to control: you can change whether not the light is shining by changing the state of the switch, but you cannot change the position of the switch by changing the state of the light, for example, by removing the bulb or breaking the mains circuit. When we say that toggling the switch causes the light to go on, but deny that changing the state of the light causes the switch to be toggled, we are asserting something like this asymmetry, according to Woodward.

The purpose of the asymmetric language of causation, then, is to capture the asymmetric facts about manipulability. (An exercise for the reader: how are the facts about manipulability related to Reichenbach's concerns in *The Direction of Time*: the statistical asymmetries, the second law of thermodynamics, and the direction of time itself? Horwich (1987) is a useful resource on such questions, and also contains a brief discussion of the manipulationist approach to causation.)

Woodward suggests that causal language can capture facts about manipulability that are beyond the reach of any mere statistical claim, in the sense that facts about manipulability are not reducible to statistical facts (Woodward 2003, 28). Perhaps, then, he would endorse an external explanation of causal cognition: we think causally because causal representations capture facts about manipulability that are both of great practical utility and beyond the expressive range of statistical language.

I think, however, that this cannot be correct. Even if Woodward is right in holding that the facts about manipulability are not purely statistical facts, a simple and familiar argument shows that the practical aspect of manipulability facts—that is, the consequences of manipulability facts that are relevant to our practical decision-making—can be captured statistically.

Very briefly, the argument is as follows:

16

1. Anything that makes a practical difference makes a publicly observable difference. You and I can both see, for example, that my flicking the switch changes the state of the light, but that my removing the bulb does not change the state of the switch.

2. Any observable pattern can be represented by a statistical claim.

3. Therefore, everything contained in a manipulability claim that is of practical use to us could be stated using purely statistical language.

In short, all we need to know about manipulability in our world can be represented statistically. The fact that flicking a switch turns on a light is stated as a correlation not between the *on* state of the switch and the light's shining, but between the fact of flicking and the subsequent change of the light's state. The fact that breaking the light bulb does not toggle the switch is stated as a lack of correlation between the breaking and a subsequent change of the switch's state. (For a way of making the same point from within the Bayes nets framework, see Spirtes et al. (2000), §3.72.)

There is no external explanation of causal cognition to be found in the insights of Woodward and others about the relation between causality and manipulability, then. There may, however, very well be an internal explanation.

Consider what I am calling the practical content of our knowledge about manipulability. This content can, I have argued, be given a statistical representation; it could be encoded in a list of correlations. But might it not be more economical to encode it in a causal DAG?

A good case can be made that the causal representation is more efficient. But it is a complicated issue, since even the causal representation is rather more complex than you might think: in order to provide useful information about manipulability, a DAG must represent not only the various features of the world that are to be manipulated, but also the manipulating actions themselves, in particular, the various actions that can be taken by the

manipulator—pushings, pullings, switchings, and so on. There are a great number of these, and there is no short cut to the causal representation of this information, or at least, no short cut that is not also a short cut for the probabilistic representation of the same information.

Thus I will, having given the barest sketch of the explanation, leave the hard work to the modern day proponents of the view that the impetus for our causal thinking is the need to represent relations of manipulability.

4. Underlying Mechanism

Assume that for every type of causal relation, that is, every fact of the form *c causes e*, we causal cognizers suppose that there is an underlying mechanism in virtue of which occurrences of *c* bring about occurrences of *e*. Perhaps this is not always true—perhaps causal relations considered fundamental are not thought to have an underlying mechanism—but ignore these exceptions for the sake of the discussion.

On the face of it, imagining the existence of underlying mechanisms might seem to be a species of metaphysical daydreaming. You may think that there is a mechanism or you may not, but it will not make a difference to the serious business of everyday causal inference, and in particular, it will not make a difference to your use of causal knowledge to predict and control the aspects of the world that matter to you.

My first goal in this section is to refute such a view. Everyday causal inference of the most mundane and utilitarian sort, I will show, makes use of information about underlying mechanisms on a regular basis. Any account of practical causal reasoning must be in part an account of the inferential role of information about mechanisms.

In the course of the demonstration, I employ a notion of underlying mechanism on which there is more to a mechanism than "hidden strings" attaching cause to effect. In my sense, the operation of a mechanism may

be in large part entirely observable, in the form of various intermediate steps in a causal process. In accordance with my overall strategy in this chapter, I go even further and assume that information about underlying mechanisms—certainly, the useful part of information about underlying mechanisms—can be captured completely by non-causal, probabilistic representations. But although it can be organized as a body of statistical claims, conceiving of and organizing this information instead as though it concerns an underlying mechanism is, I will argue, more productive in a number of ways.

4.1 The Inferential Role of Information about Mechanisms

What is the inferential role of our knowledge of a causal relation's underlying mechanism? I will answer this question by discussing an example that I have written about elsewhere, the causal relations that underwrite the characteristic appearances of a biological species member,[8] for example, a lemon's yellow color or a tiger's ferocity.

An emerging consensus in the psychological literature on natural kind concepts, and on species concepts in particular, holds that there is a natural human tendency to conceive of the relation between a species and its characteristic properties as causal. This is an assumption common to the *psychological essentialists* (Medin and Ortony 1989; Gelman 2003) and to my own work on this topic (Strevens 2000). If it is allowed that wherever we humans posit a causal relation, we also tend to posit a mechanism, then both the essentialists and I hold that, for every species and known characteristic property, humans typically believe in the existence of a mechanism, common to all members of the species, that causes the property. For example, all humans who know that tigers are ferocious posit the existence of a single

---

8. By *species* here I mean what might be better called a folk genus or a generic-specieme (Medin and Atran 1999).

mechanism, common to all tigers, that causes ferocity. In what follows, I assume this without argument.

When children first learn that, say, lemons are yellow, they normally know little or nothing of the mechanism by which lemons acquire their characteristic color. Their commitment to the existence of a mechanism, then, is solely that: a commitment but nothing more. As they develop, they learn something about the workings of the mechanism. Immature lemons are green, but they develop their characteristic color over time, their skin gradually becoming colored by some internal process as they mature. This knowledge is slight and superficial, but it can have a considerable influence on ordinary causal reasoning, as I will now show.

The causal relation between membership of a species k and a given characteristic observable property p will be represented by a causal generalization roughly of the form *An organism's being a member of species k causes it to have property p*.[9] For example, the connection between lemonhood and yellowness is represented as the mental equivalent of the sentence *A fruit's being a lemon causes it to be yellow*, and the connection between tigerhood and ferocity as *An animal's being a tiger causes it to be ferocious.*

A causal claim of this form is good for two things: inferring from an organism's species that it has certain observable properties, and inferring from an organism's observable properties that it is a member of a certain species.

You perform the first sort of inference, which may be called *projection*, when you stay well away from tigers on the assumption that they are ferocious, or avoid eating lemons on the assumption that they are sour. You perform the second sort of inference, which is always called *categorization*, when you classify something with all the characteristic observable properties

---

9. Psychological essentialists maintain in addition that the causation is represented as going by way of an essence, but the truth or otherwise of this posit will make no difference to what I have to say here.

of a lemon—something that is yellow, football-shaped, sour, and so on—as a lemon.

A projection takes you from the antecedent to the consequent of the represented causal relation, inferring from the presence of a cause, namely, species membership, the presence of a characteristic effect. A categorization, by contrast, takes you from the consequent to the antecedent of the causal representation, inferring from the presence of a cluster of characteristic effects, the presence of something that typically causes those effects. (For some other ways to integrate causal thinking into the process of categorization, see the chapters by Danks and Rehder, this volume.)

The significance of knowledge of an underlying mechanism lies in its ability to modify both kinds of inference, most often, though not always, in its ability to defeat them—to give you reason not to infer the presence of the effect from the presence of the cause or vice-versa. In illustrating this inference-mediating power, I will discuss projection and categorization separately.

First, projection. The basic form of a projective inference is as follows:

1. Organism x belongs to species k,

2. An organism's being a member of species k causes it to have observable property p, therefore

3. x will have p.

By following this inferential pattern whenever you can—by inferring, of any member of k, that it has p—you will not do too badly in this world. But you could do better. The reason is that, even when the premises of the inference hold true, the conclusion may not. Some members of a species will lack the characteristic properties of that species.

For example, lemons are characteristically yellow, but immature lemons are green. Skunks are characteristically four-legged, but injured skunks may

be three-legged. Tigers are characteristically ferocious, but sedated tigers are not. Uncharacteristic specimens of a species are, you will see from these examples, far from rare.

What makes an uncharacteristic specimen possible is that the existence of an underlying mechanism connecting species membership with the appearance of a characteristic property does not guarantee the presence of the property. Very broadly, there are three reasons that the property might not be present. First, the conditions required for the mechanism to operate properly may not have been present. This is true for immature lemons, for example, where the yellowing mechanism has not had the time it needs to do its work. Second, something may have interfered with the mechanism, preventing it from operating properly. This is true for the sedated tiger, in which the sedatives temporarily disable the behavioral or other mechanisms responsible for ferocity. Third, the mechanism may have operated properly, so that the characteristic property was present at one time, but some outside force may have since undone the mechanism's work. This is true for the injured skunk: it originally had four legs, but one has been lost.

The more you know about underlying mechanisms, the better you are able to predict the breakdown of the relation between species membership and observable properties, and so to know when not to make a projection on the grounds of species membership: you avoid the error of projecting the yellowness of immature lemons, or the ferocity of tigers that are sedated, or ill, or which have been tamed.

The utility of mechanism knowledge is quite general: the mechanism underlying any causal generalization *c is a cause of e* can break down or have its effects undone, so that some instances of *c* are not accompanied by instances of *e*. The more you know about the workings of the mechanism, the better you will be able to recognize the circumstances in which a breakdown or a reversal is likely, and so the better you will be at recognizing cases in which the presence of an *e* should not be inferred from the presence of a *c*.

Your projective prowess as a mechanistic reasoner about a given causal connection, then, will increase in proportion to, first, your knowledge of the connection's underlying mechanism, and second, the frequency and systematicity of the connection's exceptions.

Let me now discuss categorization. The basic form of a categorical inference is as follows:

1. Organism x has observable properties p.

2. An organism's being a member of species k causes it to have observable properties p,

3. There is no other likely cause of x's having p, therefore

4. Organism x belongs to species k,

(Note that in most categorizations, p is a complex of observable properties—though categorization is sometimes possible on the basis of a single characteristic property, as when you recognize a fruit by its taste.)

In virtue of premise (3), a categorical inference is more complex than a projective inference. The need for this premise is an entirely general feature of inferring from effects to causes as opposed to inferring from causes to effects. To see this, suppose that c causes e. If you know that a c has occurred, then the presence of other potential causes of e makes your inference that an e occurs no less secure. But if you know that an e has occurred, the presence of other potential causes of e should most definitely deter you from inferring the presence of a c, unless you have further information.

The further information is often, if not always, information about or pertinent to underlying mechanisms. If you have some knowledge of how this particular e was caused, you may be able to rule out potential causes of e other than c, or alternatively, you may be able to rule out c itself as a cause of the e.

Let me illustrate this claim by returning to causal connections between species and their observable properties. Suppose you observe some green, lemon-shaped fruit on a tree. Are they lemons or limes? You cannot taste them, so you have only the usual visual information to go on, namely, their color and shape. They have the characteristic color and shape of limes, so a diagnosis of limehood would seem apt. But suppose that you know that it is early in the growing season. Then, if you know something about the mechanism underlying color in lemons, that color takes time to develop, you know that you cannot use the inference schema above to categorize the fruit as limes, because premise (3) does not hold: there is a possible cause of the fruit's greenness other than limehood. If, by contrast, you know nothing about the underlying color mechanism, but simply think of lemons as yellow, you will not gasp the precariousness of the inference to limehood.

A natural reaction to this description of the lemon/lime inference is to question whether an elaborate causal framework is necessary to encode the relevant information. Why not simply record a correlation between lemon-hood, immaturity, and greenness? Then the fruit in question will resemble two different prototypes (using the term *prototype* loosely), the prototypes for *lime* and *immature lemon*, and the inferrer will exhibit the appropriate level of uncertainty as to the fruit's category.

As I have said several times now, if you want an internalist explanation of causal cognition, it is unhelpful to contrast causal with statistical information as though each were sui generis; better to contrast a causal representation of statistical information with other schemas for representing the same sort of information. The question then becomes one of organization rather than extent: what is the most flexible and efficient way of storing statistical information about, say, lemons, their color, and their maturity?

This question is properly the subject of the next section, but I will lay some of the groundwork here. Let me begin by making a case for the great flexibility of the causal schema, and in particular the flexibility of the part of

the schema inhabited by underlying mechanisms, as a means for representing inferentially relevant information, by considering some other ways that information about mechanism affects categorization.

In the example above, information about mechanism rightly inhibited categorization, but in other cases, it opens the way to categorizations that would otherwise fail to be made. I have in mind categorization tasks involving uncharacteristic specimens: members of a species that lack some characteristic property of the species.

Consider, for example, the three-legged skunk. Knowing that the mechanism underlying skunks' four-leggedness, though it causes skunks to grow four legs, does not maintain the legs once grown—so that a severed leg will not grow back—provides the foundation for understanding that a skunk or other quadruped may easily lose a leg despite being characteristically four-legged, and so allows us to categorize an otherwise skunk-like three-legged animal as a skunk. (Somewhat deeper knowledge of the mechanism shows that there are other ways, such as congenital defects, in which the mechanism may fail to ensure four-leggedness.)

There are a thousand ways that a specimen might turn out to be uncharacteristic. Although it is possible, in principle, to store a statistical profile for every one of these possibilities, so that examples of each will be correctly classified (or at least, classified as intelligently as background information allows) should they be encountered, it is far more efficient—especially given that most varieties of uncharacteristic specimen will be rather rare—to store a great deal of information about characteristic specimens. This information gives you your best chance of recognizing that a particular observable but uncharacteristic property was not produced naturally, that is, not produced in accordance with the kind of causal law specified in premise (2) of the categorization schema above. Once you know that a property is not naturally produced, you know that it is not a clue to species membership. You are free—in fact, you are obliged—to ignore it, and to use whatever other in-

formation you have to make a categorization. In this way, information about the normal operation of mechanisms is used to classify abnormal specimens correctly. I will return to the question of the efficiency of the causal schema in section 4.2 below.

An extreme case of an uncharacteristic specimen is the sort of organism described in Frank Keil's "transformation" experiments (Keil 1989). In these experiments, a raccoon is supposedly subjected to a cosmetic makeover so comprehensive that it is visually (and olfactorily) indistinguishable from a skunk. Such an animal has all the characteristic properties of a skunk, but knowing that it came to have these properties unnaturally—that is, not in accordance with the mechanism by which real skunks come to have them— you resist the inference from skunk appearance to skunkhood. Such a case is too bizarre to play a part in explaining the practical function of causal cognition, but it does a very good job of exposing the causal underpinnings of our reasoning about uncharacteristic properties. The case that this reasoning is driven by knowledge about underlying mechanism is made in Strevens (2000).

Let me summarize some of the ways described above that knowledge of underlying mechanism can impact everyday projection and categorization. Take as a paradigm the causal connection between lemonhood and yellowness. The more I know about the mechanism underlying this connection, the better I am able:

1. Given a lemon, to see when the mechanism will fail to operate, and so to see that I should resist the projection from lemonhood to yellowness, as when I know that the lemon has not had time to develop its characteristic color,

2. Given a lemon, to see when the successful operation of the mechanism may have been later permanently undone, and so to see that I should resist the projection from lemonhood to yellowness, as when I

know that the lemon has been daubed with a non-yellow dye (because I know that a lemon cannot "grow its color back"),

3. Given a yellow fruit, to see when the lemonhood/yellowness mechanism was not responsible for the yellowness, and so to see that I should resist the categorization from yellowness to lemonhood, as when I know that the fruit has been daubed with a yellow dye.

4. Given a non-yellow fruit, to see that the color of the fruit is not produced in accordance with a fruit/color mechanism, and so to proceed (on other grounds) to categorize the fruit as a lemon despite its color, as when I know that an otherwise lemony fruit owes its non-yellow color to its being daubed with a dye.

In the first three cases, knowledge of mechanism gives you reason to refrain from making inferences that would otherwise seem reasonable; in the fourth case, knowledge of mechanism gives you reason to make an inference that would otherwise seem questionable.

Note that the last, inference-enabling function requires general knowledge of the mechanisms by which fruits acquire their colors; such knowledge may in fact play a role in any of the cases described. Note also that knowledge of mechanism is even more useful in categorization than in projection; this reflects the greater in-principle complexity of categorical inferences.

4.2   Virtues of the Mechanism Schema: Efficiency

The information that we conceive of as concerning a causal connection's underlying mechanism is relevant, I have shown, to the task of everyday inference. By my working assumption, this information could in principle be represented in statistical form. Why, then, represent it causally?

I will give two quite different, though complementary, answers to this question. The first answer, presented in this section, continues the strategy

I have pursued so far: it makes a case that the causal scheme does an excellent job of representing practically significant information in an efficient way. The next section presents the second answer: the causal character of a representation might encourage certain especially good search strategies for the information that is to be represented.

To the explanation from efficiency, then. Because information about underlying mechanism can be used in so many ways, a case for the overall efficiency of the causal representation of such information would be quite complex. Let me focus instead on a single application, already discussed above: the use of mechanism information to recognize that an organism is an uncharacteristic, though genuine, member of a species, and more particularly, to recognize that the uncharacteristic appearance of the specimen is no barrier to the categorization. When you conclude that an immature, green lemon is a lemon, that a three-legged skunk is a skunk, or that a raccoon transformed to look exactly like a skunk is a raccoon, you use mechanism information in this way.

Consider three ways of dealing with uncharacteristic specimens:

1. Ignore them. Simply represent the characteristic observable properties of each species, and use these as the basis for categorization. This is how uncharacteristic specimens are dealt with by, for example, the prototype theory of concepts (Rosch 1978). Using this strategy, an uncharacteristic specimen may still be classified correctly if it resembles the characteristic specimens of the correct category more than those of any other category.

2. Catalogue the exceptions. Store a separate prototype, for example, for immature lemons and three-legged skunks.

3. Think causally: represent the mechanisms by which a species' characteristic properties are caused, and ignore, for the purposes of categorization, properties that are not caused in these ways.

Of the three, I suggest that the third, causal strategy gives you the best ratio of accurate categorizations to cognitive effort. The second strategy, maintaining a complete list of exceptions, is extremely resource-intensive. The first strategy involves a rather modest commitment of cognitive resources—though not that much more modest than the causal strategy—but offers a much less sophisticated handling of uncharacteristic specimens. The sources of the uncharacteristic properties are entirely ignored, and a crude resemblance heuristic is used to classify non-paradigmatic specimens. Such a strategy may be good enough much of the time, but the causal strategy offers far more inferential control for little additional investment.

The causal, or mechanism-based strategy, then, offers a promising middle point between the prototype and the exception list strategies. Like the prototype strategy, it requires relatively few resources, because it stores information only about normal specimens, not about abnormal specimens. Unlike the prototype strategy, it stores information not only about the observable properties of normal specimens, but also about the process that leads to the appearance of the normal properties. Since the appearance of uncharacteristic properties is necessarily due either to some kind of irregularity or abnormality in this process, or to an overriding or reversal of the process, the causal strategy is able to reliably distinguish unusual category members from category non-members.

To appreciate the efficiency of the causal strategy, it is important to understand that a very small amount of knowledge about mechanisms can go a long way. (This is just as well, since humans tend not to know much about mechanisms (Wilson and Keil 2000).) No deep insight into developmental biology is needed to see that a normally four-legged creature can lose a leg, or that a dyed lemon does not have its color naturally. Much sophistication can be added to your causal inferences, then, at a very low price. Observe also that there is never an obligation to learn about mechanisms. If a certain causal connection is unimportant to you, you need not seek out and retain

any information about its underlying mechanism.

In a somewhat different vein, note that there is nothing mysterious about mechanism information, by which I mean that it is not hard to see how the same information, or at least its practically useful component, might be stated in statistical form. That lost legs are not regrown, or that the natural color of a lemon is achieved without outside influence, are observable phenomena, though less easily observed, of course, than the number of legs or the color themselves. What is recorded, in a representation of an underlying mechanism, is not in the first instance something essentially metaphysical, but rather the stages and symptoms of the sequence of events that leads to the appearance of the relevant characteristic observable property. These are the clues—the observable clues, since if they were unobservable they would be useless—that distinguish the normal from the abnormal production of the property.

What distinguishes the representation of the mechanism, then, is first, an attention to the details of a process, and second, a concern with representing what is normal rather than what is exceptional about the process. My corresponding claims are, first, that attending to some of the details of production can have a real practical payoff, and second, that by recording the details of normal or paradigmatic production processes only, these benefits come at relatively little cost in cognitive resources.

Let me bolster this discussion of the uses of the mechanism-based strategy in biological reasoning with a few words on physical reasoning, drawing on the investigations of Shultz (1982). (See also Ahn et al. (1995).)

Shultz showed children of various ages between 3 and 10 scenarios in which three events occurred, two of which were candidate causes for the third. The subjects had to decide which of the events was the actual cause, the aim of the experiment being to pit against one another two different rules for causal attribution. What Shultz called the *Humean rule* picks out as the cause of an effect another event that is spatiotemporally contiguous

with the effect and that is of a type that covaries with events of the same type as the effect. The *generative transmission* rule, by contrast, picks out as the cause the event that is connected to the effect by way of the appropriate mechanism. The signs of mechanical connection in Shultz's experiments are all observable, and take the form of certain conditions' holding. For example, for a tuning fork to cause a hollow container to resonate, the fork had to be vibrating and situated in front of the open end of the container, and the space between them had to be unobstructed.

Shultz's older subjects tended overwhelmingly to favor the generative transmission rule: when the conditions for the operation of the mechanism obtained for one putative cause and not the other, the event for which they obtained was named the actual cause, even though, thanks to the clever design of the experimental scenarios, the Humean rule pointed to the other event.[10] The children were using information regarding mechanism to reason causally about Shultz's physical scenarios, then, in much the same way that they are using information regarding mechanism in the biological scenario considered above.

Shultz's experiments, I must point out, take information about causes as an inferential end in itself. Thus these are not cases where thinking causally is a means to a non-causal inferential end; they do not directly support my conclusion that thinking in terms of mechanisms would be valuable, given the event patterns in our world, even if everything worth knowing could be cast as a statistical fact. What they do show is that mechanisms play a similar role in reasoning about physical correlations as they do in reasoning about biological correlations; what remains to be demonstrated is how great an advantage in efficiency such reasoning might, on the whole, provide by comparison with non-causal styles of statistical reasoning. What I have laid out here is the beginning, not the end, of a mechanism-based explanation of

---

10.   Only the three-year-olds failed to show a definite preference for the generative transmission rule, and then only in some scenarios.

the efficiency of causal cognition.

4.3 Virtues of the Mechanism Schema: Search Strategies

Humans believe that for every causal connection, there is an underlying mechanism, or so I have assumed, and will continue to assume, in this discussion. The belief in an underlying mechanism can manifest itself in three ways. First, as information pertinent to the nature of the mechanism arrives, you retain it and file it in the appropriate place. Second, when making inferences to which mechanism information is relevant, you retrieve the information and put it to work. Third, and more proactively, you may go looking for further information to apply in this way, that is, you may search for more information about mechanisms.

So far, I have focused on the first two aspects of the commitment to mechanism, arguing that the information we regard as concerning an underlying mechanism is particularly useful in our practical causal inferences. This postulate about the practical utility of the information can equally well be used to explain the third aspect of causal thinking—if the information is good to have, there is every reason to seek it out.

In what follows, I want to focus on the way that we search for information about mechanism, describing a further element of causal cognition that influences not only the way that the information is put to use in later inference, but also the way in which we go about acquiring the information in the first place. As you will expect, I want to suggest that there is something about the causal way of thinking that makes for an especially efficient search strategy, that is, a search strategy that turns up a great deal of information for relatively little investment.

I propose that our conceiving of mechanism information as causal motivates us, when looking for such information, to adopt what I call the *constraint from below*. The content of this constraint is roughly that any postulated mechanism ought to be "implementable", and indeed implemented, by

mechanisms at the appropriate *basic level*. For example, any physical mechanism must be at root constructed from basic physical mechanisms, any biological mechanism from basic biological mechanisms, any psychological mechanism from basic psychological mechanisms, and so on.

Clearly, the content of the constraint very much depends on what is meant by *appropriate basic level*. Three important remarks on this notion. First, as suggested by my examples, there may be different basic levels for different kinds of phenomena. The basic level for mental phenomena, for example, may be quite distinct from the basic level for physical phenomena.

Second, that a level is basic does not entail that it is metaphysically fundamental. That the basic level for mental phenomena, for example, is different from the basic level for physical phenomena allows—though it certainly does not require—that basic level mental processes are themselves physically implemented. There may be a hierarchy of implementation, then, among the basic levels themselves. I will not explore the inferential role of such a hierarchy here, although belief in a hierarchy would clearly affect the cognitive significance of the constraint from below.

Third, the identity of the basic levels is not built into the constraint from below, and indeed, is never known apodictically. Your beliefs as to the number, nature, and extent of the basic levels are always being revised, sometimes radically.[11]

What is useful about the constraint from below? There are two separate questions to address. The first concerns the validity of the constraint. The second asks how, even if correct, a doctrine that sounds more like a metaphysical thesis than a piece of helpful advice could play a role in improving our everyday inference.

---

11. A philosopher would say that everything that is known about the basic levels is known a posteriori rather than a priori, but in a psychological context, the use of the term *a priori* tends to run together the question of innateness and the question of immunity to empirical refutation. My claim is that all beliefs about the basic levels are considered subject to revision in the light of the empirical evidence.

I want to focus on the second question, so let me assume without any argument that, with the basic levels properly identified, the assumption about universal implementability explicit in the constraint from below is more or less true. (The more radical anti-reductionists among philosophers of science will demur.) How is this fact practically relevant?

The wrong way to apply the constraint is to insist that no mechanism be postulated—perhaps even no causal connection posited—until an underlying implementation is found, that is, until the operation of the mechanism is completely understood in terms of the appropriate basic level. Pursuers of this policy will indeed be lost in thought (insofar as teaching responsibilities and administrative duties allow).

The aim of the constraint is not to place impediments on the path to finding an underlying mechanism, but on the contrary, to clear the path and to speed the search. Given a few beliefs about the applicable basic level, the constraint from below will point to certain areas, and away from certain others, as sources of information about underlying mechanism. To learn the truth about the mechanism underlying the yellowness of lemons, look inside, not outside, the lemon. Look for characteristically biological processes, not mental processes. Ignore the possibility that lemons get their yellowness the same way that gold gets its yellowness, but take seriously the possibility that lemons get their yellowness in the same way as grapefruit. And so on.

Even if you are quite wrong in many of your beliefs about the basic levels, the influence of the constraint from below will tend to be largely or wholly positive. On almost any view that anyone has ever had about the workings of biological mechanisms, the previous paragraph's sage counsel as to where to look for the lemon/yellowness mechanism holds. The constraint from below, for all its appearance of heavy-duty metaphysicking, is a fount of good, practical advice in the search for causal information. Ironically, the constraint is liable to cause trouble only when you know—or think you know—much more about the workings of the world than a modest, amateur

causal inquirer. But that is a story for another time.

In short, then, the assumption, which I take to be built into our system of causal reasoning, that every causal connection has an underlying mechanism, together with the meat added to the notion of *underlying* by the constraint from below and some rudimentary knowledge about the basic level, prompts a search strategy for mechanistic information that is better than most.

4.4   Virtues of the Mechanism Schema: Overview

Let me now step back to summarize the various ways in which thinking about causal connections as having underlying mechanisms improves day-to-day causal inference, even if the mechanistic information is nothing but a certain kind of statistical information in another guise.

1. The mechanism schema provides a compact representation of certain information about normal or characteristic specimens that can be used to decide whether to proceed with various projections and categorizations in an intelligent way.

2. The mechanism schema invites us to make mechanism-based inferences using the information that it encodes, by representing it in a form that makes its relevance to simple causal inference—inference from cause to effect or effect to cause—immediate.

3. This relevance made clear, the mechanism schema not only invites us to make mechanism-based inferences but also to search out the information we need to do so—to find the mechanical reality underlying a causal connection.

4. Because the mechanical information is subject to the constraint from below, the search for information about mechanism is guided in part by wider beliefs about the workings of the appropriate basic level.

It has been my working assumption in this chapter that all of the information mentioned above—the information contained in a simple causal claim, the information about the mechanism that underlies the causal connection asserted by the claim, and the information contained in beliefs about the appropriate basic level brought to bear by way of the constraint from below—can be represented in statistical form.[12]

What, then, is the origin of the peculiar phenomenological character of causal beliefs? It cannot be in their content (at least, not in their content extensionally conceived—for example, not in their truth conditions). This leads naturally to the suggestion that the causal phenomenology is due to the particular inferential role played by causal information in the human mind, that is, the role characterized immediately above.

I am not sure that this suggestion can account for the sense of causal *oomph* we experience when one billiard ball strikes another. Perhaps the oomph ought to be explained in some completely different way, for example, as an aspect of our sensory phenomenology, as suggested by Leslie (1994). I do think that the inferential role investigated in this section is at least a part of the explanation for the sense of hidden connectedness in the world as we experience it causally, the sense that behind the scenes—or below the scenes—something is going on, something on which everything we see depends.

---

12. How might the information contained in the constraint from below be reduced to statistical language? Relative to a set of basic levels and some information about their workings, the answer might go, roughly, as follows: the constraint contains information about where certain kinds of correlations—the correlations that make up information about mechanism—are found. The constraint identifies clues, then, and says: look for your correlations where you find these clues. You might say that it asserts a kind of meta-correlation, a correlation between the clues and other correlations.

5. Conclusion

I have surveyed a number of quite different approaches to explaining the prevalence of causal cognition. They are all internal explanations: they focus on the organizing power, rather than the expressive power, of causal representation schemas.

Each of these explanations assumes that, even if the world is Humean, in the sense that every fact can be captured by some purely statistical claim or other, it is a very particular Humean world: its pattern of correlations is a very particular pattern, with very particular properties. It has the Reichenbach asymmetry, it exhibits certain asymmetric patterns of manipulability, or it is the sort of pattern that is amenable to characterization in terms of the language of underlying mechanism, consistent with the constraint from below.

It is this special, perhaps very unusual property of the worlds' correlations that makes causal cognition so useful. The practical value of causal thinking lies, then, not in its ability to capture facts in principle out of the reach of statistical vocabulary, but in its ability to organize statistical facts—given certain unusual patterns in those facts—in an especially effective way, and perhaps also to organize the search for those facts just as efficiently.

It is striking how much can be said in favor of each of the proposed internal explanations of causal cognition's practical value, and how many opportunities there are for still more proposals. The complete explanation of the form of causal cognition looks to be rich and complex indeed.

That the existence and structure of causal cognition is best explained internally allows, but does not imply, that the world is Humean. A modern causal realist, or opponent of Humeanism, will aim to explain the particular pattern of correlations we see around us—the pattern that is so amenable to causal encoding—as a consequence of something bigger than statistics at work in the external world. It is only because of a metaphysics that posits more than can be said in Humean or statistical terms that the world con-

tains these special kinds of correlations, or so the causal realist argues; they are our best clue that there is something else out there. Our system of causal inference has developed to exploit the unusual correlations for practical purposes only; yet, despite its purely instrumental rationale, it points past mere patterns of fact and gestures, however vaguely, at a causal world beyond.

References

Ahn, W., C. W. Kalish, D. L. Medin, and S. A. Gelman. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition* 54:299–352.

Gelman, S. A. (2003). *The Essential Child: Origins of Essentialism in Everyday Thought*. Oxford University Press, Oxford.

Glymour, C. (2001). *The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology*. MIT Press, Cambridge, MA.

Hirschfeld, L. and S. A. Gelman (eds.). (1994). *Mapping the Mind*. Cambridge University Press, Cambridge.

Horwich, P. (1987). *Asymmetries in Time*. MIT Press, Cambridge, MA.

Keil, F. C. (1989). *Concepts, Kinds and Conceptual Development*. MIT Press, Cambridge, MA.

Leslie, A. M. (1994). ToMM, ToBy, and agency: Core architecture and domain specificity. In Hirschfeld and Gelman (1994), pp. 119–148.

Medin, D. and A. Ortony. (1989). Psychological essentialism. In S. Vosniadou and A. Ortony (eds.), *Similarity and Analogical Reasoning*, pp. 179–195. Cambridge University Press, Cambridge.

Medin, D. L. and S. Atran (eds.). (1999). *Folkbiology*. MIT Press, Cambridge, MA.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge.

Reichenbach, H. (1956). *The Direction of Time*. University of California Press, Berkeley, CA.

Rosch, E. (1978). Principles of categorization. In E. Rosch and B. Lloyd (eds.), *Cognition and Categorization*, pp. 27–48. Lawrence Erlbaum, Hillsdale, NJ.

Salmon, W. (1984). *Explanation and the Causal Structure of the World.* Princeton University Press, Princeton, NJ.

Shultz, T. R. (1982). *Rules of Causal Attribution*, volume 47:1 of *Monographs of the Society for Research in Child Development.* Chicago University Press, Chicago.

Sosa, E. and M. Tooley. (1993). *Causation.* Oxford University Press, Oxford.

Spirtes, P., C. Glymour, and R. Scheines. (2000). *Causation, Prediction, and Search.* Second edition. MIT Press, Cambridge, MA.

Strevens, M. (2000). The essentialist aspect of naive theories. *Cognition* 74:149–175.

Wilson, R. A. and F. C. Keil. (2000). The shadows and shallows of explanation. In F. C. Keil and R. A. Wilson (eds.), *Explanation and Cognition.* MIT Press, Cambridge, MA.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation.* Oxford University Press, Oxford.