16

Causal Inference

Nancy Cartwright

1. Problems for Causation

Causation has long been a vexed notion across the sciences, so much so that there have been repeated cries that it must be banished forever from both the natural and the social sciences. Yet, whatever is the case with the natural sciences, causation remains at the heart of the social sciences. For example, the London School of Economics and Political Science (LSE), which is one of the foremost social science institutions in the world, has as its motto, *Rerum cognoscere causas*: 'To know the causes of things'.

Why is causality so central? Because, it is widely believed, it is knowledge of the causes of things that allows us to understand the world, to predict the future, to build better social systems, to change the world we live in. Why, then, try to banish causation? Because the notion faces serious problems and there is no widespread agreement how, or even that, the problems can be solved. The problems can be grouped under three headings: meaning, method, and metaphysics.

1.1 *Meaning*

The eighteenth-century Scottish philosopher David Hume argued that our notion of causation is not what we think it is (see e.g. Garrett 2009). He claimed that all of our concepts (he called them 'ideas') come from experience. They are copies of the impressions that make up our experiences. Where, then, in our experience of happenings in the world is the impression of causation? Nowhere, he argued. Look as hard as you will. You only really see shapes, and colours, and motions, never a 'making something happen'. So where does the concept come from?

What we call causes are regularly followed by what we label as their effects, said Hume. Human beings, he believed, are deeply prone to forming habits. So, having observed a regular association between two kinds of events, we come to expect the second when we see the first. Looking inwards at ourselves, we notice this feeling of expectation; we get an impression of it. Our concept of causation, Hume claimed, is a copy of that impression of expectation. All that is happening in the external world that contributes to our coming to have this concept is a regular association of events. The concept itself derives from an impression of our own internal state.

Although no one nowadays subscribes to Hume's theory of concept formation, his conclusion still has a powerful grip, and it is not uncommon to find contemporary thinkers who hold what is called a *regularity theory* of causality, that all there is to causality is association, though many now would allow that the association may be merely probabilistic, not one that holds with full regularity.

Independent of the influence of thinking like Hume's, there is good reason to care about just what the concept of causation is if we are to employ it in social science. For it is one of the central norms of science, whether it be natural science or social science, that its claims be clear and unambiguous and that the concepts employed in them be well defined and well understood. When we talk in science, we are supposed to be clear exactly what we are saying. So philosophers and social scientists alike have invested a great deal of effort trying to get straight the meaning of 'cause'. We shall review some of the central recent attempts in section 2.

1.2 *Method*

Just as good science demands clarity about the meaning of the concepts it employs, it also demands that there be clear, explicit methods for determining just when a concept applies to something in the world and when it does not. Without such methods, claims involving the concept will be untestable. There will be no way to judge whether they are true of the world or not. It will also be difficult to use the concepts to make predictions or to provide instructions for how to bring about change in the world. So much effort is also put into devising sound methods for establishing when relations hold and when they do not. We will look at some of these in section 2.

As Chapter 14 of this book stresses, the two endeavours of characterizing or defining concepts, on the one hand, and of devising methods for determining when they obtain, on the other, must go hand-in-hand. The methods must be geared to showing whether the very concept that has been defined holds, so that we can be sure that our methods are teaching us about what we think we are learning about. You will not be surprised then that some of

the characterizations of causation you will see are almost read directly off from some of our favoured methods for establishing causal relations, and vice versa.

1.3 *Metaphysics*

The problems under this heading concern the nature of the causal relations themselves. Social science studies causal relations at two levels. The first is the *singular*: what are the causes and effects of specific single events? Using an example from Chapter 14, what caused the civil war in Angola in 1975? The second is the *general*: what kind of features are generally connected as cause and effect? For instance, 'Ethnic diversity is among the causes of ethnic civil war', or 'Skill loss during periods of unemployment causes persistence of high levels of unemployment.' Claims in social sciences about general causal relations are often called *causal principles*.

Probably the most central metaphysical problems facing singular causal relations arise from questions about how they behave in time. Are they instantaneous or extended in time? Is the cause temporally contiguous with the effect or is there a gap between them? Bertrand Russell (1912–13) argued that there cannot be a gap. Otherwise what would make the effect pop into existence when the cause was no longer there to produce it? This implies, he argued, that neither the cause nor the effect can be extended in time because the earlier parts of the cause, if they were to matter, would have to have an influence across a time gap; and so too with later parts of the effect. So causes must be instantaneous and contiguous with their effects. But that is impossible because time is continuous. Between any two instants there are infinitely many more, no matter how close together those two instants are. So, said Russell, the notion of one thing causing another does not make sense. We can sensibly talk of what values quantities have at each instant through continuous time, as when we use a line to represent the changing locations or changing velocity of the centre of mass of an object; but not of causality. That's part of why the concept should be banned from both natural and social science.

Social scientists have not worried much about Russell's problems. But they do worry about similar ones. There is a general consensus that in the single case there should be some continuous causal process connecting cause and effect. But the causes represented in social science studies are often aggregates, like GDP; they are often characteristics of institutions, like the managerial structure of Enron; and they are often norms or practices, like tax evasion. For causes like these, it can be hard to identify a causal process connecting cause and effect. What, for instance, might connect the high ethnic diversity in Angola in the period before 1975 with

the ethnic civil war there in 1975 that it is supposed to have helped cause? Or suppose we want to argue that the GDP in a country in one quarter caused the rise in consumer spending there next quarter. What temporally continuous causal process connects these aggregate quantities? How even should we think about GDP itself with respect to time? Suppose all the production processes in a country are stopped for a day to mourn a dead leader. Does GDP go to zero then? Although these are live questions from social science, there has not been much headway on them recently and we shall not discuss them further here.

The dominant questions at the general level have to do with, first, the relation between the singular and the general. Must every singular causing fall under a general principle? We may say that the event reported in paragraph three of the lead story in today's *New York Times* caused the event reported in the first paragraph on page five. Surely we do not think there is a general principle that 'Events reported in paragraph three of the lead story of the *New York Times* cause events reported in the first paragraph of page five.' But must there be some feature of the event reported in the lead story—say, that it involves a high level of ethnic diversity—and some feature of the event reported in page five—say that it is an ethnic civil war—which are connected by a general principle? This is an important question for how we approach the understanding of singular events in the social sciences, which is the meat of what is done in anthropology and in policy evaluation (did the policy really produce the improved outcomes we observe?). Philosophers over the last two and a half decades have tended to answer 'yes' to this question. Although there was a lively debate about the question involving both philosophers and social scientists in the 1960s (see e.g. Roberts 1995) concerning historical explanation, social scientists have not had much to say about it lately. The exception is the recent literature on evidence-based policy, to which we will turn.

The second dominant question at the general level is one that confronts all principles in science whether they are causal principles or some other kind of law and whether they are in natural or in social science. What makes a general truth a law? Don't some facts just happen to be true generally? They hold by accident and we would not want to include them as principles in our scientific theories. For instance it is true that Venetian sea levels rise and so do bread prices. But this general association does not seem to be a law. Though this can be an important question, it is widely—and well—addressed in texts on philosophy of science in general, so we shall not pursue it here.

In what follows we shall discuss first general causation and then singular and, because of the close connection between the two, we shall discus meaning and method together.

2. Meaning and Method

What does it mean to say, at the general level, 'Ethnic diversity causes (or is among the causes of) ethnic civil war'? Or 'Rising inflation causes decreases in unemployment'? Or, as Chapter 1 discusses, that wives slimmer than their husbands make for happy marriages? Most recent attempts to answer this question are not *reductive*. They do not attempt to define causation in totally non-causal terms. They suppose, to the contrary, that causation is a basic fact of nature that cannot be properly characterized in totally non-causal terms any more than central physics concepts, such as *electron*, can be characterized without using other concepts from physics theory, like *proton* and *electromagnetic field*. The view that causation cannot be characterized in entirely non-causal language and correlatively that the validity of our methods to test causal claims will always rest on some specific causal assumptions gets summed up in the slogan, 'No causes in, no causes out.' Nevertheless, even if a reductive characterization is not possible, if the concept is to play a role in social science, it is still essential that it have clear unambiguous meaning and that our methods for testing for it be appropriate to the concept as thus understood. Four main ideas are used in recent attempts to characterize causation, which we address in turn: probabilistic association; manipulation or intervention; mechanism; powers.

How do these four different approaches to causation relate? Are they all different theories about the very same thing, perhaps a special kind of relation that holds between events or a special kind of general truth or a special kind of law of nature? If so which, if any, is correct; and might they all, or at least more than one, be correct, perhaps by focusing on different aspects of this one thing? If they are not about the same thing, are each of them—or at least some of them—correct about something? Perhaps they focus on somewhat different kinds of relations in the social world, though are similar enough to go under the same loose title 'causal'? These are important questions requiring both philosophical and empirical input. We will not discuss these here but leave them as topics in a more advanced course since they cannot be tackled till after we have a good idea what the different approaches are, which we do aim to provide here.

We shall also not discuss the relation between causation and explanation, on which there is a large philosophical literature (see e.g. Reiss 2013: chs 5–7).

2.1 *Probabilistic Association*

When a cause is present there should be more of the effect than if it were absent. That is the root idea of the probabilistic theory of causation. If C-type events occurring at some arbitrary time t cause E type of events at a time t'

later, then we should expect $\text{Prob}(E_{t'} \mid C_t) > \text{Prob}(E_{t'} \mid \neg C_t)$. But that's similar to saying $E_{t'}$ and C_t are correlated—they tend to occur together; and it is a well-known mantra that correlation is not causation. High candy consumption in one month is correlated with low divorce rate the next. But eating candy does not prevent divorce. The correlation is due to the fact that candy eating is correlated with being young and young people have generally not been married long enough to get divorced. So, something more needs to be said.

Suppose you were able to control for all the other factors causing or preventing $E_{t'}$.[1] Imagine you could look in a population where all of these took some fixed value. So everyone is the same age, the same religion, is undergoing the same stress at work, has the same number of children, etc. In this population it seems reasonable to expect that if candy consumption does cause low divorce rate, it will increase the probability; and the probability will increase only if it does so since in this population there's no other way to account for an increase in probability. Somewhat more formally, let K_i represent a population where all the causes of E_t other than C_t take some fixed value, where i ranges over all the sets of values these causes can take. Then C_t causes $E_{t'}$ in K_i if and only if $\text{Prob}(E_{t'} \mid C_t + K_i) > \text{Prob}(E_{t'} \mid \neg C_t + K_i)$.

This is the gist of the probabilistic theory of causation, though more details need to be ironed out. For instance, note that this characterization is relative to the population satisfying K_i. What then about a larger population that contains K_i? That is one of the points of dispute. Some philosophers and social scientists insist on keeping the relativization to the K_i. Others allow that C_t causes $E_{t'}$ in any population that contains K_i, though this can lead to it being true both that C_t causes $E_{t'}$ in a population and also prevents $E_{t'}$ there in cases where in one subpopulation C_t increases the probability of $E_{t'}$ and in another it decreases it. Under any circumstances this kind of characterization will always be relativized to a population since different populations will have different probabilities over the same variables and also they may well have other general causal relations holding, hence different factors for determining the K_is. Chapter 3 worries about taking studies that show that a policy worked—i.e. caused the targeted outcomes—in one situation as evidence that it will have the same effects elsewhere. It is just this kind of relativity of causal claims to other causal factors (those that pick out the K_is) that creates this worry: whether C_t causes $E_{t'}$ or prevents it or does nothing to $E_{t'}$ at all in a given situation can depend very much on what other causal factors are there in the situation along with C_t.

[1] Except those on the causal pathway between C_t and $E_{t'}$ if it exists.

Note also that seen as a characterization of general level causality—as a way of providing clear unambiguous sense for the concept—the probabilistic theory of causation is not reductive since we need to refer to general causal relations in specifying what K_i is and thus we refer to other general causal relations in explaining what is required for any given general causal relation to hold. It is, however, very constraining. Each of the factors in K_i must satisfy a similar formula with respect to C_t and the remaining factors in K_i since each of these is itself meant to be a cause of $E_{t'}$. Although this may not narrow the choice of causes for $E_{t'}$ to a single choice, it will rule out a huge number of alternatives. And adding some information about a few factors, that they are indeed a cause of $E_{t'}$ or that they are not, can sometimes fix the set entirely given the probabilities.

This way of characterizing causality has the advantage that it connects immediately with standard statistical methods used in the social sciences to test for causal relations. These methods are used in what are called 'observational studies', which means that the data come from populations in their natural settings and not from specially selected populations enrolled in experiments. In the populations under study, social scientists measure *correlations* or *regressions* between factors to begin to test whether there is a causal connection between them in that population. These are weaker notions than conditional probabilities, like $\text{Prob}(E_{t'} \mid C_t)$ and $\text{Prob}(E_{t'} \mid \neg C_t)$, but are closely related. In a better test, they measure *partial correlations* or *partial regressions*, holding fixed other variables that they hope represent the other causes. This is akin to the partial conditional probabilities $\text{Prob}(E_{t'} \mid C_t + K_i)$ and $\text{Prob}(E_{t'} \mid \neg C_t + K_i)$. Similar kinds of ideas are used in econometrics, in estimating the coefficients that appear in economic equations that will hopefully represent causal relations.

The two assumptions of the probabilistic theory—that an effect is always probabilistically dependent on its cause and that this kind of dependency disappears when other causes are held fixed—are also at the core of what are called 'Bayes nets methods' for causal inference, for which computer programmes can generate all sets of causal relations among a given set of variables that are possible, given information about probabilistic dependencies among the variables, supposing the fundamental assumptions linking causes and probabilities are satisfied.

These two assumptions are also at the core of the reasoning behind randomized controlled trials (RCTs), which are highly touted as the gold standard for causal testing in medicine and, as we see in Chapter 3, in evidence-based policy, and are being pushed throughout the social sciences now, especially in development economics. The neat thing about RCTs is that they help with one of the central problems that the statistical methods discussed so far face: that we generally don't know what the other causal factors are and so don't know what to hold fixed.

In an RCT, the individuals in a population enrolled in the experiment—which could be individuals, schools, countries, etc.—are randomly assigned either to the treatment group, which will be subject to the cause (C_t), or to the control group, which will not have the cause ($\neg C_t$), but may perhaps receive a placebo. There should be as much masking as possible: the individuals in the experiment should not know which group they are in, nor should anyone involved in further monitoring or treatment, or in reading out the results to see if $E_{t'}$ obtains or not, or in carrying out the statistical analysis. This is to guard against conscious or unconscious bias that may influence the results that are finally recorded. The aim is to ensure that both groups have the same distribution for all the other causal factors influencing the outcome, so that every arrangement of them—every K_i—has the same frequency in both groups. This aim can very rarely be achieved and moreover, we won't know when it has been: statistics can tell us how often in the long run random assignment will produce any particular imbalance but we do not have even that kind of assessment when it comes to how well the masking has succeeded. So, as always in scientific work, we must not place too much confidence in the results of a single study, even one that has been very well conducted.

In order to see the logic of the causal inference, let us suppose, though, that the other causal factors have the same frequency in the treatment and control groups. The $\text{Prob}(E_{t'})$ in the treatment group will then be an average across the probability of $E_{t'}$ given C_t in each of the subpopulations in it—i.e. each of the K_i. So it will be an average over $\text{Prob}(E_{t'} \mid C_t + K_i)$. Similarly, the $\text{Prob}(E_{t'})$ in the control is the average of the probability of $E_{t'}$ given $\neg C_t$ in each of the subpopulations represented there—each of the K_i. So it is an average over $\text{Prob}(E_{t'} \mid \neg C_t + K_i)$. By our hypothesis that the other causal factors have the same distribution in the two groups, the frequency of each K_i subpopulation is the same in both. So if the probability of $E_{t'}$ is greater in the treatment group than in the control group, that implies that for one of the K_i subpopulations, $\text{Prob}(E_{t'} \mid C_t + K_i) > \text{Prob}(E_{t'} \mid \neg C_t + K_i)$. It follows that there is at least one subpopulation—one K_i—of the experimental population in which 'C_t causes $E_{t'}$ in K_i' is true, or at least is true under the probabilistic theory's way of characterizing causality.

Can we conclude from the experiment 'C_t causes $E_{t'}$ in the experimental population'? That depends on the decision referred to earlier about what to say about 'C_t causes $E_{t'}$' in a population given that C_t causes $E_{t'}$ in one of its subpopulations. What matters is that, however this decision is made, those using the claim should understand it and not read more out of it than the study supports. In particular, the higher probability in the treatment over the control group shows only that C_t causes $E_{t'}$ in some subpopulation of the population enrolled in the experiment. It may hold in specific other populations

or even across most. But C_t may have exactly the opposite effect on $E_{t'}$ in some subpopulations than it does in others. Finding out whether that is true requires a great deal more social science work.

2.2 *Intervention and Manipulation*

The manipulation view of causation revolves around the idea that causes give us effective strategies for producing effects we want, or preventing those we do not; by manipulating the cause we can manipulate the effect in a predictable way. So manipulation theories characterize general-level causation roughly this way, where again there is a dispute about the exact details of the formulation: for any two times t and t', 'C_t causes $E_{t'}$ in situations of type S' just in case manipulating C at t—making it bigger or smaller, or bringing it in or taking it away—is regularly followed by appropriate changes in E at t' '. For this to be true, we must be careful how these manipulations are carried out. 'Reducing class sizes causes improved reading scores' is true in many populations. But not if you lower teacher quality at the same time as reducing class size. Alternatively the manipulation of one factor may be followed by improvement in another without any causal connection. For instance new programmes for teaching reading may be followed by better reading scores not because the programmes cause better scores but because the teachers adopting them are regularly the better teachers or they become more enthusiastic when trying out new programmes.

This is where the concept of *intervention* comes in. 'Intervention' has been given a variety of formal definitions, but the basic idea is that an intervention is a manipulation that is done in the 'right way' to make the causal relation, or lack of it, apparent. The right way will be one in which neither other causes nor other preventatives of the effect change, an idea we are familiar with from the probabilistic theory. Nor can any causal relations involving the effect change during the intervention. That's because it is not a fair test of a causal relation if suddenly there are a lot of new causal relations produced between the effect and other factors that were not causes before. This includes changes in the very relation under test. One familiar way in which the requirement that causal relations involving the effect not change during intervention is violated is when the manipulation of the cause is so ham-fisted that it busts up the causal relation we are trying to test. One familiar example is when we wind up the toy soldier to see if that will make it march, but we wind too tightly and break the mechanism that makes it work.

Nobel-prize-winning Chicago School economist Robert Lucas claims this is frequently true with attempts to manipulate economic variables to bring about desired change. For instance, inflation, when it occurs naturally, can

in the short run cause reductions in unemployment. That's because, so the story goes, entrepreneurs mistakenly see the universal rise in prices as a real rise in prices in their sector and so hire more workers to produce more goods to meet increased demand. If, though, the government manipulates inflation to improve unemployment, entrepreneurs will recognize the rise in prices for what it is—just inflation—and will not open new jobs. The government's very attempt to use the causal relation between inflation and unemployment will break it. Or at least so says Lucas's model.

As with probabilistic theory, the characterization of general causation using the concept of manipulation is not reductive since the definition of an intervention, however the details are worked out, will have to refer to other general causal relations.

Also, as with the probabilistic theory, the characterization of causality in terms of manipulation is closely linked with familiar methods for testing, in this case with real non-statistical experiments where other causes are held fixed and only the cause under test is varied. Here too we are often stymied by our lack of knowledge of what these other causes are. This is the problem that plagues standard *before-and-after studies*. In before-and-after studies the cause is administered and we look to see if the effect changes. But, are we sure that nothing else changed at the same time as the cause? One of the tricks of social science is to find situations where we can have reasonable confidence that nothing else that could influence the effect has changed even if we do not have a catalogue of what all the other influences might be to check on them. This is the strategy that is used in what are called 'instrumental variables' models in economics. For instance Joshua Angrist in a classic paper uses the Vietnam era draft lottery to measure the effect of veteran status on earnings (Angrist 1990). Since whether an individual was drafted was determined on the basis of a randomly assigned number, we can be fairly confident that factors that affect earnings, like education, were equally distributed between the group of individuals who were drafted and the group of individuals who were not, as in an RCT. And so we can be fairly confident that the comparison between the average earnings in one group and the average earnings in the other tells us something about the effect of veteran status on earnings.

The manipulation theory of general-level causation is closely akin to the *counterfactual theory* of singular causation, which is based on the idea that one specific event, c, is a cause of a later event, e, just in case e would not have occurred unless c had. You can see immediately that this kind of claim will be hard to test since c and e are specific events that either occur or do not; so we can never observe what would have happened had things been different. There are also huge problems of formulation. There is a vast literature available in philosophy on counterfactuals and causes in which you can read more about these (see e.g. Paul 2009 and references therein).

Similar counterfactuals have come to the fore in social science lately because of the rise of the evaluation industry for social policy. As Chapter 3 explains, there is a great demand to know whether the policies we have implemented have 'worked': Has the policy genuinely produced the change intended? To answer that, it helps to know what would have happened otherwise. RCTs are often employed to this end. But of course they cannot tell us of a specific case whether the policy brought about the effect there, since they only look at groups. If an RCT shows more positive results in the treatment group, the group in which the policy was implemented, than in the group where it was not, we can conclude that in some of the cases in the treatment group the policy produced the effect. But we do not know which; we have no way to sort cases where the policy brought about the effect from those where something else was responsible. So for reliable evaluation of the single case, some different way of reasoning is required. This often involves process tracing which we shall discuss in the next section.

Not only do we not know from the RCT for which individuals the policy worked. We do not know what sets of characteristics matter. We know that there is at least one arrangement of characteristics that individuals might have—one K_i—for which the policy promotes the desired outcome. We call these 'support factors'. But we do not know which it is. This matters when it comes to putting the claims established in RCTs to use in predicting what will happen if the policy were to be implemented elsewhere. If the population elsewhere has no individuals with just the right arrangement of support factors (no individuals described by the successful K_is), or not enough to make it pay, the policy will not produce the desired outcomes in the new setting. Given good reason to think that the new setting does have enough individuals with the right arrangements, we have good reason to suppose the policy will work for some individuals there. But without good reason to suppose this, the RCT results will not be much use in predicting results in the new setting.

Consider for instance Oportunidades, a Mexican programme for poverty reduction. Because Oportunidades was itself designed as an RCT, there is very good evidence that it did contribute to reducing poverty where it was implemented. Does the success of Oportunidades provide evidence supporting predictions regarding what would happen were a similar programme implemented in some other population? Only if enough individuals in this target population have the right arrangement of support factors. And one cannot assume without evidence that this will be so. This point is illustrated by the case of Opportunity NYC, a poverty reduction programme modelled—and named—after Oportunidades which was implemented in New York City in 2007 but discontinued in 2010 because it failed to produce the expected effects. The success of Oportunidades in Mexico did not, by itself, provide sufficient evidence to support the prediction that Opportunity NYC would

produce similar results, and the mistaken assumption that it did led to a poor policy decision.

2.3 *Mechanisms*

There are three senses of the term 'mechanism' in play currently in work on causation in philosophy and in the social and in the biomedical sciences. These include notions of *causal process, invariance,* and *underlying structure.*

CAUSAL PROCESSES

For singular causation it is widely held that a cause must be connected with its effect by some spatio-temporally continuous process between them. So tracing the causal process between them can be a good way to tell whether two events are related as cause and effect, a good way IF we know how to tell a causal process when we see one. So what distinguishes a continuous sequence of events that constitutes a causal process from one that does not? One popular philosophical account answers that in causal processes energy is transferred at each step in between. But this is not of much help in social science (see e.g. Dowe 2009).

A standard procedure in social science and in policy evaluation is to break the sequence into small steps where it is easier to determine for each step whether it is a genuine cause of the next. For instance, the Global Environment Facility evaluated the effectiveness of one of its programmes, the end-goal of which was to 'establish a long-term conservation finance mechanism to support biodiversity conservation' in the Bwindi national park (Uganda), by determining whether this programme had an effect on four intermediate outcomes (e.g. the establishment of a Bwindi Trust) which in turn were supposed to contribute to achieving the end-goal of the program (GEF 2007: 6).

Process tracing necessarily focuses on singular causation. But it can be a tool for general causation as well. If in situations of kind S, an event of kind C at t is regularly followed by an event of kind E at t' and it regularly happens that the same kind of causal process connects the individual events, then we can conclude at the general level that 'In S-type situations, C_t causes $E_{t'}$'. This kind of reasoning has played a significant role in the biomedical sciences. For instance, tracing the process by which the chemicals in cigarette smoke lead to lung cancer provided a central piece of the evidence that showed that smoking causes lung cancer.

INVARIANCE

Suppose we observe that in a particular kind of situation S one kind of feature F2 regularly changes after another, F1, changes. That is mere correlation and we know that correlation does not equal causation. But suppose that F1

changes by intervention. If F1 causes F2 we expect that F2 will change following the intervention in the way that it always has; but not, if F1 does not. Changing levels of candy consumption may be regularly followed by changing probabilities of divorce. But this relationship is not invariant under intervention. If we intervene to change candy consumption—'intervene', so the only change is in the amount of candy consumed—the divorce rate will not change in tandem. This observation has given rise to the invariance account of causality now in fashion in philosophy of science. The basic idea is that, supposing C_t and $E_{t'}$ are regularly associated in S, then 'C_t causes $E_{t'}$ in S' just in case the association between C_t and $E_{t'}$ is invariant under interventions on C_t.

This account is clearly closely connected with the manipulation account, and it shares one major drawback with that account that we have not yet discussed. The manipulation account is grounded in the idea that causes provide strategies for changing the world. Suppose we do characterize causation as invariance under intervention. Then changing C_t will be a good way to change $E_{t'}$—but that follows only if C_t is changed by intervention. But interventions are hard to come by. We are in a far more powerful position to predict what will happen and to engineer what we want if we know about a regular association that is invariant under the method of implementation that we will in fact employ. This gives rise to another, different characterization of general-level causation, one that we can find in contemporary economics (see e.g. Hoover 2001: ch. 2). Roughly, if C_t and $E_{t'}$ are regularly associated in S, then C_t causes $E_{t'}$ in S relative to ϕ just in case the association between C_t and $E_{t'}$ is invariant when C_t is changed by ϕ, where ϕ can then represent the method by which we will in fact change C_t. Knowing that C_t causes $E_{t'}$ under this characterization of causation immediately provides an effective strategy for changing E at t'.

Notice that these different versions of the invariance account can give different verdicts about whether C_t causes $E_{t'}$ in any given situation. This underlines the importance of being clear exactly what we mean when we make causal claims in social science.

UNDERLYING STRUCTURES

Different causal relations hold in different social, cultural, and economic settings. What is a required politeness in one culture can cause insult in another. Why? What causal relations hold in a social setting depends on the underlying system, the underlying social, economic, and cultural structure. It is now standard in philosophy of biology to call the structures that underlie causal relations there 'mechanisms', and sometimes this usage is adopted in the social sciences as well. Though there is no widely used mechanistic characterization of causation, there is widespread sense that, if we are to follow the injunction of the LSE motto, to know the causes of things, we should come

to understand the underlying mechanism that make the particular causal principles that hold for a given population possible.

This is clearly of central importance in social planning and policy. Consider recent problems in child welfare in the UK, where attention tended to focus on causal processes. When a tragic child death occurs, 'The standard response is to hold an inquiry... trying to get a picture of the causal sequence of events that ended in the child's death... We are tracing a chain of events back in time to understand how it happened.' So says Eileen Munro (2005: 377), author of Chapter 3 and of the 2012 UK government report on child protection. Munro recommends instead a 'systems' approach. That means focusing on the underlying mechanism or social structure that makes these kinds of processes possible, or even probable. As a US National Academy of Science report on building safer health systems explains, 'The focus must shift from blaming individuals for past errors to a focus on preventing future errors by designing safety into the system' (Kohn et al. 2000: 5).

So it is clear that it is important to study social mechanisms in social science. What is not clear is what kinds of concepts should be used to describe these mechanisms and their operation, nor what tool kit of methods will help us to understand them. This is one of the central tasks right now on which philosophy and social science intersect.

2.4 *Powers*

John Stuart Mill argued that women have the same powers of leadership and imagination as men, and we have these powers even should they be seldom displayed. Give women the same education, upbringing, situation, and opportunities as men and they will display these qualities equally with men. And he used this claim to argue for dramatic changes in the social policies and practices in play at the time regarding the role of women in society.

As with all concepts, if the concept *power* is to play an important role like this in social science or in social policy deliberation, we should be clear what is meant by it. This is a big topic in metaphysics nowadays and, to a lesser extent, in philosophy of science (see e.g. Mumford 2009). It is usual to make at least four assumptions about powers. First, a system—an object, a person, an institution—can have a power without displaying it. Second, the power will be displayed if the conditions are right. Third, what actually happens when a power is displayed depends on the setting. So what happens when a woman displays her imagination will depend on whether she is at a laboratory work bench or addressing Parliament or tending her children. Fourth, when a power operates in a situation, what happens there will depend in some intelligible way on the display of the power (as when a heavy metal object does not fall to the ground because the power of a magnet is displayed).

The most typical characterizations of powers 'back' define them from their displays. Roughly, A has the power to α just in case when properly enabled and failing something stopping it doing so (which we sometimes express as 'nothing interferes with the display'), A does α. This characterization is dramatically non-reductive because of the concepts of *enablement* and *interference*. Contrast this with an alternative: 'A has the power to α just in case there are some circumstances Cα such that it is a law of nature that in Cα, A does α', where Cα is to be filled in by some possibly very long list of features that can be characterized without any power-related notions, like 'is schooled at Eton', 'learns Latin', 'does not play with dolls', etc. Many powers advocates do not adopt this alternative because, they argue, there is never a list that does the job. In each actual situation there is a fact of the matter about whether the circumstances that obtain there constitute an interference and whether they are enabling, but there is nothing these facts share in common that we could fill in for Cα other than that they are enabling or they interfere. This is taken by many to be particularly problematic since philosophers have not succeeded in having much enlightening to say about interferences and enablers at the general level.

Why then take powers seriously? Philosophers offer a panoply of metaphysical arguments. But for social science the answer seems to lie in the usefulness of the concept, as with Mill, who argued for the centrality of the related notion of *tendency* in political economy. Another word used for much the same idea is 'capacity'. In order to link it with our contemporary discourse, I shall use the word 'power' throughout for all these related notions. Mill's model for powers comes from Newtonian mechanics. The sun has the power to attract the planets. Nothing ever interferes with it so the sun constantly displays this power (called 'gravity'). What happens when it displays it? The planets move around it in elliptical orbits. By contrast, when the earth displays its power to attract a cannonball shot at an enemy ship, the cannonball moves along a parabola. What actually happens when the power of gravitational attraction is displayed (i.e. when the sun or earth attract other objects) depends on the setting. To make a link with our immediately previous discussion, we may think of the earth–cannonball pair as a mechanism, in the sense of an underlying structure, giving rise to the causal principle, 'The earth displaying its power of gravitational attraction causes cannonballs to move along parabolas'; and the planetary system as a mechanism giving rise to the causal principle 'The sun displaying its power of gravitational attraction causes the planets to move along ellipses around it'.

How do we know what happens when a power is displayed in a given setting? That's easy, at least in principle, in Newtonian mechanics. All that's relevant to how a power affects motions is what other powers are displayed in the situation. The display of the power is represented by what we call a force.

represented by a vector. When several forces are displayed at once, we add these vectors, call the result 'the total force, F_T', and calculate the acceleration of any object in the situation by the familiar formula $F_T = ma$, where m is the mass of the object. That is what we do, for instance, when we want to know how a pin will behave when a magnet attracts it upward and the earth attracts it downward, or how an electron in an atom behaves when it is attracted by both the mass of the nucleus and the charge on the protons there.

That's all physics. Life is not so easy in social science, which may have some knowledge of powers but little knowledge of methods for calculating just what happens in different situations when the powers are displayed. Consider an economics example already mentioned, the one involving skill loss during unemployment. Earlier I suggested that economics entertains a causal principle that skill loss perpetuates high levels of unemployment. That is probably a mistaken way to look at it. It is seldom the case that skill loss is the only thing going on that affects unemployment. There are, for instance, sometimes vigorous successful government efforts to combat unemployment. In these cases skill loss may be followed by improving levels of employment, contrary to what is suggested by this causal principle. A better way to put it might be 'Skill loss has the power to cause continuing high unemployment'. This will not lead us to expect that skill loss actually produces unemployment whenever it occurs, but only that, to understand the levels of employment that actually occur, we shall have to take the downward push of skill loss into account.

So, we return to the question, 'How do we take it into account?' How do we predict what happens when a socio-economic power is displayed? There is no systematic answer. This is one of the problems social scientists constantly confront in trying to put their isolated parcels of hard-won knowledge—like the knowledge of specific powers such as the power of skill loss to perpetuate unemployment—to use.

3. Putting Causal Knowledge to Use

Chapter 14 in this book teaches us that the meaning we assign a scientific concept must be closely paired with our methods for finding out whether it obtains. We should have good arguments to show that the methods we employ provide a good way to find out about just what we claim we are finding out about. That message has been echoed throughout this chapter with respect to the concept of causation when it is used in the social sciences. It is not just meaning and method that must match, however. So too must use. The use to which we put our social scientific claims—the inferences we draw from them and the practices we adopt on the basis of them—should be

supported by what those claims actually say; and what they say that supports our inference should in turn be supported by our methods for deciding that what they say is true. It is no good making a very broad claim whose first half is well supported by the evidence but then draw our inferences from the second half, which is unsupported.

We have looked at a number of accounts of what causal claims say. The claims say different things because they employ different concepts of causation. These concepts may all go under the same title, 'general causal relation', but they require different conditions to be met in order to obtain. So different methods must be used to find out about them and different uses made of them. So, of what use is knowledge of the causes of things? Knowing causes helps us to understand how things work. But it is also supposed to be of practical value. Knowing the causes of things should help us to change the world.

Here we must be cautious though. There is a tendency to too much haste, to try to read off directly from our causal claims just what we must do to bring about the change we desire. That will work with some kinds of causal knowledge but not most. Consider the manipulation concept of causation found in economics that I mentioned in discussing invariance: C_t causes $E_{t'}$ relative to ϕ just in case manipulating C_t by ϕ is followed by $E_{t'}$, where ϕ is one of the methods available to us to change C at t. Suppose we know that. Then we are in a powerful position to change E at t'.

But suppose we have established instead a different causal claim, say a manipulationist claim of the kind favoured in philosophy—C_t causes $E_{t'}$ just in case if C_t changes by intervention $E_{t'}$ changes—or in invariance-type claim—C_t causes $E_{t'}$ just in case the association between them is invariant when C_t changes by intervention. That too can be of immediate use—if we have an intervention to hand, which we usually do not. We have seen the same kind of problem with knowledge of powers. Knowing about what powers a system has will not by itself tell us what to expect to happen in particular circumstances nor how to build circumstances to get what we want. Nor does knowing the causal relations, on the probabilistic theory, that hold in one population, perhaps the population enrolled in an RCT, give us immediate guidance about how to bring about change in a different population.

This does not, however, in any way make causal knowledge in social science useless. It just puts it more on a par with similar knowledge in the natural sciences, where strategies are built by splinting together a very great many different pieces of knowledge and often undertaking designated new researches and developing new theory to fill in gaps. The great nineteenth-century physicist Lord Kelvin laid the first Atlantic cable in 1866. His knowledge of causes in physics was essential to his success but it definitely did not tell him just what to do. Similarly, the large team of physicists, engineers, mathematicians, and technicians who developed the US Second World War radar took

many months to do so, despite the vast repository of causal knowledge the team members brought to the project. The conclusion I propose is that the causal knowledge we work so hard to gain in the social sciences can be of immense use. But there is no reason to think it should be easier to put our social science knowledge to use than it is to use our knowledge in the natural sciences.

References

Angrist, J. (1990). 'Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records', *American Economic Review*, 80(3): 313–36.

Dowe, P. (2009). 'Causal Process Theories' in H. Beebee et al. (eds), *Oxford Handbook of Causation*. Oxford: Oxford University Press, 213–33.

Garrett, D. (2009). 'Hume', in H. Beebee et al. (eds), *Oxford Handbook of Causation*. Oxford: Oxford University Press, 73–91.

GEF (2007). *Case Study: Bwindi Impenetrable National Park and Mgahinga Gorilla National Park Conservation Project*. Impact Evaluation Information Document, 7. Washington, DC: Global Environment Facility, Evaluation Office.

Hoover, K. (2001). *Causality in Macroeconomics*. Cambridge: Cambridge University Press.

Kohn, L., Corrigan, J., and Donaldson, M., eds (2000). *To Err is Human: Building a Safer Health System*. Washington, DC: Committee on Quality of Health Care in America, Institute of Medicine.

Mumford, S. (2009). 'Causal Powers and Capacities', in H. Beebee et al. (eds), *Oxford Handbook of Causation*. Oxford: Oxford University Press, 265–78.

Munro, E. (2005). 'Improving Practice: Child Protection as a Systems Problem', *Child and Youth Services Review*, 27: 375–91.

Paul, L. A. (2009). 'Counterfactual Theories', in H. Beebee et al. (eds), *Oxford Handbook of Causation*. Oxford: Oxford University Press, 158–84.

Reiss, J. (2013). *Philosophy of Economics: A Contemporary Introduction*. London: Routledge.

Roberts, C. (1995). *The Logic of Historical Explanation*. University Park, PA: Pennsylvania State University Press.

Russell, B. (1912–13). 'On the Notion of Cause', *Proceedings of the Aristotelian Society*, 13: 1–26.

Further Readings

Beebee, H., Hitchcock, C., and Menzies, P., eds. (2009). *Oxford Handbook of Causation*. Oxford: Oxford University Press.

Campaner, R., and Galavotti, M. C. (2007). 'Plurality in Causality', in P. Machamer and G. Wolters (eds), *Thinking about Causes: From Greek Philosophy to Modern Physics*. Pittsburgh: University of Pittsburgh Press. 178–99.

Cartwright, N. (2007). *Hunting Causes and Using Them*. Cambridge: Cambridge University Press.

McKim, V., and Turner, S., eds (1997). *Causality in Crisis: Statistical Methods and the Search for Causal Knowledge in the Social Sciences*. South Bend, IN: University of Notre Dame Press.

Russo, F. (2009). *Causality and Causal Modelling in the Social Sciences*. New York: Springer.

Index