HOW PROBABILISTIC CAUSATION CAN ACCOUNT FOR THE USE OF MECHANISTIC EVIDENCE.

Erik Weber

Erik Weber is at the Department of Philosophy and Moral Science, Ghent University.

Correspondence to: Centre for Logic and Philosophy of Science, Universiteit Gent, Blandijnberg 2, B-9000 Gent, Belgium. E-mail: Erik.Weber@UGent.be.

Abstract: In a recent paper in this journal, Federica Russo and Jon Williamson argue that an analysis of causality in terms of probabilistic relationships does not do justice to the use of mechanistic evidence to support causal claims. I will present Ronald Giere's theory of probabilistic causation, and show that it can account for the use of mechanistic evidence (both in the health sciences – on which Russo and Williamson focus – and elsewhere). I also review some other probabilistic theories of causation (of Suppes, Eells and Humphreys) and show that they cannot account for the use of mechanistic evidence. I argue that these theories are also inferior to Giere's theory in other respects.

1. Introduction

In a recent paper in this journal, Federica Russo and Jon Williamson argue that an analysis of causality in terms of probabilistic relationships does not do justice to the use of mechanistic evidence to support causal claims (Russo & Williamson 2007, 164). I will argue that they are

wrong in this respect: probabilistic conceptions of causation can do justice to the use of mechanistic evidence, both in the health sciences (on which Russo & Williamson concentrate) and in other scientific disciplines. Russo & Williamson make an analogous claim about analyses of causation in terms of physical mechanisms: they claim that such an account cannot do justice to the widespread use of probabilistic data (Russo & Williamson 2007, 164). I will not discuss that claim here.

I will proceed as follows. In Section 2 I establish some common ground: I review the most important topics on which Russo & Williamson and I agree. In Section 3 I present Ronald Giere's theory of probabilistic causation. In Section 4 and 5 I show that Giere's theory can account for the use of mechanistic evidence (both in the health sciences and elsewhere). In Section 6 I review some other probabilistic theories of causation (those of Patrick Suppes, Ellery Eells and Paul Humphreys) and show that Russo & Williamson are right about these theories: they cannot account for the use of mechanistic evidence. I also compare them with Giere's theory in other respects, and I argue that Giere's theory is better in those respects too.

2. Evidential pluralism and output monism

Consider the following quotes:

> Evidence is constituted by two complementary elements: probabilities and mechanisms. (Russo & Williamson 2007, 159)

> To establish causal claims, scientists need the mutual support of mechanisms and

> dependencies. (Russo & Williamson 2007, 159)

> Consequently in the health sciences it is now a commonplace that both mechanistic and probabilistic evidence are required to substantiate causal claims. (Russo & Williamson 2007, 161)

I call the position expressed in these quotes *evidential pluralism*, and I agree with it: I think it is correct in the health sciences, but also in several other scientific disciplines (e.g. the social sciences). Russo & Williamson combine evidential pluralism with what I call *output monism*:

> On the contrary, we will argue, only a single notion of cause is used in the health sciences. More specifically, health scientists use two types of evidence *for a single causal claim*, '*C* causes *E*', not for two different types of causal claim, *C*-causes$_1$-*E* and *C*-causes$_2$-*E*. Because only *one* notion of cause is used in the health sciences, pluralism is false. (Russo & Williamson 2007, 165)

Though I agree with the idea expressed here, I think the formulation is sloppy. The second sentence expresses the idea that health scientists want to *produce* only one type of causal claim. They have only one type of *target*, they aim at one kind of *output*. This output monism, which I also endorse, is compatible with claiming that health scientists use mechanistic causal information *as evidence* (so it is compatible with evidential pluralism). Moreover, the conclusion that pluralism in general is false certainly does not follow: it follows that *output* pluralism is false *in the health sciences*. *Evidential* pluralism can still be maintained (that is good, otherwise their position would be incoherent). And *output* pluralism might be true *for other scientific disciplines*

(e.g. the social sciences) or in non scientific contexts (e.g. in legal contexts). I take it that Russo & Williamson mean that *output* pluralism is false in the health sciences. With that I agree, provided that one limits the health sciences to epidemiology and research on the effectiveness of therapies (fundamental clinical research has different aims). For an analysis of different kinds of causation in the biomedical sciences, see De Vreese 2009.

3. Giere's theory of probabilistic causation

3.1 The following definitions constitute the core of Ronald Giere's theory of probabilistic causation in populations:

> C is a *positive causal factor* for E in the population U whenever $\mathbf{P}_X(E)$ *is greater than* $\mathbf{P}_K(E)$.
>
> C is a *negative causal factor* for E in the population U whenever $\mathbf{P}_X(E)$ *is less than* $\mathbf{P}_K(E)$.
>
> **C** is *causally irrelevant* for **E** in the population U whenever $\mathbf{P}_X(E)$ *is equal to* $\mathbf{P}_K(E)$. (Giere 1997, 204)

Though it can be extended to other types of variables, Giere considers only binary variables. So in his definitions, **C** is a variable with two values (C and Not-C); the same for **E** (values E and Not-E). X is the hypothetical population which is obtained by changing, for every member of U that exhibits the value ¬C, the value into C. K is the analogous hypothetical population in which all individuals that exhibit C are changed into ¬C. $\mathbf{P}_X(E)$ and $\mathbf{P}_K(E)$ are the probability of E in respectively X and K. Probabilities are defined as relative frequencies (Giere takes U to be finite, i.e. causal claims are about finite populations).

An example might clarify this. If we claim that smoking (**C**) is a positive causal factor for lung

cancer (**E**) in the Belgian population (U), this amounts to claiming that if every inhabitant of Belgium were forced to smoke there would be more lung cancers in Belgium than if everyone were forbidden to smoke. Conversely for the claim that smoking is a negative causal factor. Causal irrelevance is a relation between variables (represented in bold) rather than a relation between values of a variable (like the first two relations). If we claim that "smoking behaviour" (**C**) is causally irrelevant for "the occurrence or absence of lung cancer" (**E**) this means that we believe that in the two hypothetical populations the incidence of lung cancer is equally high.

3.2 Giere's theory is a so-called "average effect theory" of probabilistic causation in populations. To see what this means, we first have to look at how he defines causal relations in individuals, as opposed to populations:

> C is a *positive causal factor* (deterministic) for E in an individual, I, characterized by residual state, S, if in I, C produces E and Not-C produces Not-E.
> C is a *negative causal factor* (deterministic) for E in an individual, I, characterized by residual state, S, if in I, C produces Not-E and Not-C produces E. (Giere 1997, 200)
> If C is neither a positive nor a negative causal factor for E in I, given S, then we say that the variable **C** is causally irrelevant for **E** in I, given S. (Giere 1997, 201)

The residual state S refers to all other characteristics of the individual besides the cause and effect variable. Giere's population claims "always average over individuals and, therefore ignore what might be important differences among individuals" (Giere 1997, 204-205). Giere discusses only one specific case: if in U there are some individuals for which C is a positive causal factor for E, and an equal number for which C is a negative causal factor, **C** is causally irrelevant for **E** in U.

More generally, Giere's definitions leave open the possibility that a population contains individuals for which C is a positive causal factor for E, as well as individuals for which C is a negative causal factor for E. If the first group is larger, C will be a positive causal factor for E in population U. If the second group is larger, C will be a negative causal factor for E in U.

Let us look at an example. Consider a dangerous virus, which threatens a population of humans (H). Some people are immune to the disease (I), but there is no way to find out who is and who is not. It is possible to vaccinate people before they become sick (V). We assume the following probabilities in the hypothetical populations:

$P_V(S|I) = 0.9$ (S stands for survival)

$P_{\neg V}(S|I) = 1$

$P_V(S|\neg I) = 0.8$

$P_{\neg V}(S|\neg I) = 0$

Furthermore, we assume that 50% of the population is immune, so we also have:

$P_V(S) = 0.85 \quad (0.8\times0.5 + 0.9\times0.5)$

$P_{\neg V}(S) = 0.5 \quad (1\times0.5 + 0\times0.5)$

Note that in the subpopulation I of people that are immune, vaccination has negative causal relevance: 10% of this subpopulation would not survive vaccination. In subpopulation ¬I and in H as a whole, vaccination has positive causal relevance. How is this possible? In I there is a group of people (10% of I) whose residual state is such that they die if vaccinated. For the others, vaccination is causally irrelevant at the individual level. Combined, this gives negative causal

relevance at the level of population I. In population ¬I we have a large group (80%) whose residual state is such that vaccination is positively causally relevant at the individual level. For the others, it is irrelevant (their residual state is such that they die anyway). The combination of this gives positive causal relevance at the level of population ¬I. The population H contains a group of 5% (10% of the 50% immune) for whom vaccination has negative causal relevance at the individual level. It also contains a group of 40% (80% of the 50% non-immune) for whom vaccination has positive causal relevance at the individual level. For the others, vaccination is causally irrelevant at the individual level (vaccination makes no difference for them: they survive anyway because they are immune, or they die anyway because the vaccination does not work for them). Because the group with positive relevance is larger (40% as compared to 5%) the result is positive causal relevance at the level of population H.

The vaccination example illustrates that, according to Giere's definitions, causal relevance can be reversed or annihilated in subpopulations: if C is a positive causal factor for E in population U, it can have negative causal relevance or be causally irrelevant in subpopulations of U. The same holds for negative causal relevance and causal irrelevance. In our example, we have positive causal relevance in H, and negative causal relevance in subpopulation I. Theories of probabilistic causation which have this property are called "average effect theories": whether there is a causal relation in a population depends – according to these theories – on the average effect in the populations, no matter what happens in the subpopulations. Other examples of average effect theories can be found in e.g. chapter 9 of Dupré (1993) and chapter 9 of Hausman (1998).

The rivals of average effect theorists are the adherents of so-called "context unanimity theories". The first context unanimity theory can be found in Cartwright (1979). More recent versions can be found in Humphreys (1989) and Eells (1991). I use the latter one here to explain the difference. In chapter 2 of his book, Eells gives the following example:

> To use an example of Cartwright's (1979), ingesting an acid poison (X) is causally positive for death (Y) when no alkali poison has been ingested ($\sim F$), but when an alkali poison has been ingested (F), the ingestion of an acid poison is causally negative for death. I will argue that in a case like this it is best to deny that X is a positive causal factor for Y, even if, overall (for the population is a whole), the probability of death when an acid poison has been ingested (that is, even if $Pr(Y/X) > Pr(Y/\sim X)$). I will argue that it is best in this case to say that X is causally *mixed* for Y, and despite the *overall* or *average* probability increase, X is nevertheless not a positive causal factor for Y in the population as a whole. (Eells 1991, 58)

What he does with this example fits in his conceptual scheme:

> Then we say that X is a *positive causal factor* for Y if and only if, *for each i*, $Pr(Y/K_i\&X) > Pr(Y/K_i\&\sim X)$. *Negative causal factorhood* and *causal neutrality* are defined by changing the "always rises" (>) idea to "always lowers" (<) and "always leaves unchanged" (=), respectively. The idea that the inequality or equality must hold for *each* of the background contexts K_i is sometimes called the condition of *contextual unanimity*, or *context unanimity*. ... Note that these three relations of positive, negative and neutral causal factorhood are not exhaustive of the possible causal significance that a factor X can have for a factor Y: There remains the possibility of various kinds of *mixed* causal relevance, corresponding to various ways in which unanimity can fail. (Eells 1991, 86-87)

The characteristic property of causes in the sense of the context unanimity theory is that the

causal tendency cannot be reversed (from positive to negative) or annihilated (from positive or negative to causally neutral) in a subpopulation.

The first advantage of Giere's theory – as compared to context unanimity theories – is that his definitions capture the kind of causal knowledge about population causality that policy makers need. Consider our vaccination example. According to Giere, vaccination has positive causal relevance for survival in H. So the policy advice is: do vaccinate, because that will save lives. According to a context unanimity theorist, the situation is ambiguous (causally mixed). So no policy recommendation follows from that. If we look at Giere's theory from the perspective of a policy maker, it is better than a context unanimity theory. Policy makers only need average effects, not causes in the sense of context unanimity theories.[1]

The issue can be further illustrated by the following example (suggested by one of the referees). Suppose that asbestos is a major problem in the construction of buildings in Belgium, and that in Belgium it is forbidden to smoke inside buildings. If all Belgians were forced to smoke, they would spend much more time outside, thereby reducing their exposure to asbestos. The two effects might cancel each other out, so that in the hypothetical population where everyone smokes, the incidence of lung cancer is not higher than in the hypothetical population where nobody smokes. In this scenario, smoking would turn out to be causally irrelevant in the Belgian population according to Giere's theory (it can be causally relevant in other populations: Gierean causal claims are explicitly linked to a specific population U). This may look strange at first glance (and a problem for Giere's theory) but it isn't. If the scenario were correct, a ban on smoking would *not* be a good policy measure if the Belgian government wants to reduce the incidence of lung cancer: it would simply not help. This is exactly what Giere's theory tells us: causal irrelevance. A clever government then of course would look at other populations (e.g. neighbouring countries). If smoking turns out to have positive causal relevance for lung cancer in

these other populations, it will try to find out why Belgium is an exception to the rule. This shows that a clever policy sometimes requires causal knowledge about other populations than the population of interest. But Giere's account covers the kind of information we need about these different populations.

Another important advantage of average affect theories is that causal relevance as they define it can be discovered by scientific procedures. That is not the case for context unanimity theories. Daniel Steel formulates this as follows:

> A randomized controlled experiment may tell us that the cause is positively relevant in the population overall, but such a result is consistent with that effect being neutralized or even reversed in subpopulations. Indeed, among heterogeneous populations it is quite common that there are unknown factors capable of disrupting the mechanism linking cause and effect. Consequently, if contextual unanimity is part of the meaning of claims concerning positive causal relevance, then it is unclear how one could establish that smoking causes lung cancer, HIV causes AIDS, and so on. (Steel 2008, 28)

A similar point has been made by John Dupré (1993, 200-202): scientific methods cannot provide evidence for causal claims in the sense of unanimity theorists. More generally, it is unclear how such claims can be supported.

3.3 An important feature of Giere's theory is that he defines causation in terms of what would happen in two *hypothetical* populations. In order to see the epistemological consequences of this, it is useful to draw an analogy with the use of observation instruments. Suppose we have a very tiny species A, and claim that individuals belonging to this species have six legs. To support this

claim, we investigate a number of A individuals by means of two types of light microscopes (on the one hand, transmission microscopes in which the light that reaches the eye of the observer is the light that is transmitted by the sample; on the other hand, reflection microscopes in which the light that reaches the eye is reflected by the sample) and by means of an acoustic microscope (which works with sound waves). The results (visual images and patterns of sound waves) have to be interpreted by means of a theory about the functioning of the microscope. We have primary data for which we show, by means of a theoretical argument, that they constitute good evidence for our claim. Furthermore, there is no way to bypass this procedure: the species is too small, so it is unobservable without the aid of instruments.

The situation with Gierean causal claims is the same. All the evidence we obtain (whether it is of a probabilistic nature or of a different type) comes from the *real* world. As a consequence, we need an explicit argument to show the relevance of any kind of evidence for the *hypothetical* worlds which figure in Giere's definitions. There is no type of evidence for which such an argument is superfluous (just like one needs an argument for every type of microscope one wants to use). For each type of evidence one wants to use, the relevance for the hypothetical populations has to be established. There are no exceptions to this rule, so no type of evidence has a special status (in the sense that its relevance is trivial and needs no argument). Moreover, there is no direct access to the hypothetical populations. So there is no bypass: all evidence is indirect, just like in the case of our unobservable species.

3.4 There is one more aspect of Giere's theory that must be discussed here, because it will be important in Sections 4 and 5. Giere rightly stresses that it is often important to know how large the effect of one factor on another is, on top of knowing that there is causal relevance (1997, p. 206). For instance, on top of knowing that smoking causes lung cancer, it is useful – from a

policy perspective – to know how large the effect is. Similarly, on top of knowing that a certain drug has positive causal relevance for recovery in a population of people suffering from a certain disease, it is good to know how many people would benefit from taking the drug. Giere proposes to measure the effectiveness of a causal factor as follows:

$$\mathbf{Ef}(C,E) = \mathbf{P}_X(E) - \mathbf{P}_K(E)$$ (Giere 1997, 206)

So the effectiveness of a causal factor C for E in a population U is the difference between the probabilities in the hypothetical populations. It is clear that, in order to have an idea of the effectiveness of a causal factor, one needs to have rather precise estimates of the value of the $\mathbf{P}_X(E)$ and $\mathbf{P}_K(E)$. However, establishing a causal claim does not require such a precise estimate: given Giere's definitions, all we need is an argument supporting the claim that $\mathbf{P}_X(E)$ is larger than $\mathbf{P}_K(E)$ (in case of a positive cause) or smaller (in case of a negative cause).

4. How probabilistic theories can account for the use of mechanistic evidence, part 1

I claim that Giere's probabilistic theory can account for the use of mechanistic evidence in the health sciences and elsewhere. In Section 4.1 I will make my claim more precise: I clarify what I mean with "to account for", "mechanism" and "mechanistic evidence". Then I proceed to show that Giere's theory can account for the use of mechanistic evidence in contexts in which no *prima facie* relevant probabilistic evidence is available (Sections 4.2 and 4.3). In Section 5 I discuss contexts in which *prima facie* relevant probabilistic evidence is available.

4.1 Consider a mathematician who wants to prove a theorem of the form "Every triangle which has property A also has property B". This *aim* can account for a number of things the mathematician does. Suppose the mathematician decides to try to prove the following lemmas:

> Every triangle which has property A also has property C.
>
> Every triangle which has property C also has property B.

The mathematician's practice of working on these two lemmas is rational given his aim, because the proof of the theorem is easy once the two lemmas are proven. In the same way the fact that a scientist (e.g. an epidemiologist or social scientist) wants to make a causal claim about populations in the sense of Giere can account for the fact that this scientist takes certain steps to gather evidence for this claim. What I will argue is that it is perfectly rational for such a scientist to gather *mechanistic* evidence.

Stuart Glennan has defined mechanisms as follows:

> A mechanism underlying a behavior is a complex system which produces that behavior by the interaction of a number of parts according to direct causal laws. (Glennan 1996, 52)

A mechanism in this sense is different from what Wesley Salmon used to call a causal mechanism (Salmon 1984), because of the idea of levels of reality (a complex system which has parts) that it incorporates. Therefore, people that adopt a definition like Glennan's are often called "complex system mechanicists".

As an example, let us look at Dan Steel's characterization of social mechanisms:

> Social mechanisms in particular are usually thought of as complexes of interactions among individuals that underlie and account for aggregate social regularities. ... But there is more to social mechanisms than just individual interactions: typically, the individuals are categorized into relevantly similar groups defined by a salient position their members occupy vis-à-vis other members of the society (Steel 2004, 57-58)

Or briefly:

> Social mechanisms are complexes of interacting individuals, usually classified into specific social categories, that generate causal relationships between aggregate-level variables. (Steel 2004, 59)

This is a definition of a specific type of mechanisms which fits the general idea of Glennan.

Now that I have clarified what a mechanism is here in this paper, I can also clarify what mechanistic evidence is. Mechanistic evidence is bottom-up evidence: it consists in using information about the behaviour of the parts in order to support causal claims about the system as a whole. For instance, in the social sciences, mechanistic evidence consists in using knowledge about the behaviour of individuals to support causal claims about societies as a whole. In the biomedical sciences, it consists in using knowledge about the functioning of parts of the human body to make claims about populations of individuals (note that this assumes at least three levels: the population, individuals and parts of individuals).

4.2 Consider the following claim:

> The reliable water supply in Taiwan (due to Japanese irrigation projects in the 1930s) caused a breakdown of the joint-family system in rural areas of the island.

This is a claim about a causal relation in a population in Giere's sense. There is a population (Taiwanese peasants in the 1930s), a binary cause variable (reliable water supply or not) and an effect variable (being part of a nuclear family or being part of a joint family). This example is taken from Little 1991 (p. 141), who, in turn, relies on Pasternak 1978. Till 1930, a joint-family system was dominant in rural areas of Taiwan: parents and married sons continued to live together and farm their holdings together, rather than dividing into two or more nuclear families. From 1930 on, there is a continuing trend toward divided families. Pasternak explains this change in family structure as a rational adaptation to a change in circumstances of the rural economy: the availability of reliable irrigation water. Indeed, Taiwan was invaded by Japan, and the Japanese established large-scale irrigation projects in the 1930s. An adherent of this explanation must accept the causal claim above.

Whether we accept the causal claim or not depends on the evidence we have. If the Japanese had irrigated half of Taiwan (each half consisting of randomly chosen part, so that other possible relevant factors are equally distributed), and would have randomly distributed the farmers over the irrigated and non-irrigated part, that would have been a very useful randomised experiment to decide about the causal relation. Performing the experiment would be unethical. But that has become irrelevant nowadays: the appropriate experiment (involving actual division in irrigated and non-irrigated parts, and random division of all Taiwanese peasants over the two parts) was in principle possible 70 years ago, but now it is too late. The only type of evidence we can deliver now is a thought experiment along the following lines:

Premisses

(1) In both parts (irrigated and non-irrigated) an equally great substantial part of the population makes rational decisions about family structure.

(2) In both parts, normal frictions in social life (e.g., between sisters-in-law) occur often and with equal frequency.

(3) In the irrigated part, farmers are convinced that the irrigation system protects them against crop failure due to draught.

(4) In the non-irrigated part, farmers know that, in a nuclear family system, their rice crop will fail (due to insufficient labour supply) if there is less than *15* days of consecutive rainfall.

(5) In the non-irrigated part, farmers know, in a joint family system, their rice crop will fail (due to insufficient labour supply) if there is less than *10* days of consecutive rainfall.

(6) Periods with more than 10 but less than 15 days of consecutive rainfall occur often.

Conclusion

(I) The irrigated part evolves towards nuclear families more rapidly than the non-irrigated part.

In this argument, we reason with imaginary cases: both the experimental and control group are virtual (that is why I call it a thought experiment) , and we argue what would happen if they were real. But there is a telling difference: the characteristic property of the experimental group (irrigation) has been present in the real world. I will come back to this immediately.

The thought experiment uses a “bottom-up” approach: we start from assumptions about how rational decision making and/or nonrational psychological processes determine the behaviour of

individuals in a social group. Then these assumptions are used to infer a causal relation on the higher level, the level of the population. More precisely, the way the thought experiment works is this:

(1) We know what would happen in the experimental group, because its characteristic property has been present in the real world: the result in the experimental group would be the same as in the real world (e.g. evolution towards nuclear family).

(2) We do not know *why* this result occurred in the real world, only *that* it occurred.

(3) We try to find out which mechanism produced the result in the real world (and thus would produce the result in the experimental group).

(4) We use this mechanism to argue that the result would be different in the control group.

4.3 We can draw three conclusions from this example. The first is that, in history and more broadly in the social sciences, it happens that we want to make causal claims about populations while no probabilistic evidence is available. The second conclusion is that, in such cases, mechanistic evidence can help: a bottom-up argument as specified above can lead to conclusions about the hypothetical populations X and K. It cannot give us a precise estimate of $\mathbf{P}_X(E)$ and $\mathbf{P}_K(E)$. However, as we have seen in Section 3.4, this is not a problem: all we need is an argument that $\mathbf{P}_X(E)$ is different from $\mathbf{P}_K(E)$. The third conclusion follows from the second: for a historian or social scientist that wants to make a Gierean causal claim about a population, it is perfectly rational to gather mechanistic evidence, since that evidence can help him to establish the claim.

Given this result, Russo & Williamson can still hold on to the following claim:

> [T]he proponent of the probabilistic theory can't account for the fact that mechanisms are required even when appropriate probabilistic associations are well established. (Russo &

Williamson 2007, 164)

Indeed, my example only shows that the use of mechanistic evidence is rational – from the point of view of Giere's theory – in cases where no probabilistic evidence is present. In Section 5 I will argue that it can also account for the use of mechanistic evidence when relevant probabilistic associations have been established.

5. How probabilistic theories can account for the use of mechanistic evidence, part 2

Biomedical scientists investigating the causes of diseases face a fundamental ethical problem. Randomised experiments with the target population (i.e. humans) are the best experimental method for establishing causal relations in the biomedical sciences:

> A decisive test of whether smoking causes heart disease, then, would be to take a large sample of human infants randomly selected from the human population, divide them into two equal groups, and force one group to smoke for the rest of their - no doubt abbreviated - lives. (Dupré 1993, 202-203)

However, these randomised experiments are usually impossible for ethical reasons: they may cause physical harm to the experimental subjects, as in Dupré's example. Biomedical scientists can avoid the unethical experiments by doing non-random experiments with humans (prospective or retrospective designs) and by doing randomised experiments with animals. I discuss these solutions in Sections 5.1 and 5.2. The reason why I discuss them is that they constitute two

typical cases in which scientists combine probabilistic evidence with mechanistic evidence.

5.1 If we perform a non-random experiment, we are confronted with the well known problem of confounders, which originates in the fact that in a prospective or retrospective design the individuals either put themselves into the experimental or control group by the way they act or belong to one of the groups because of the properties of the environment they live in. For instance, people that decided to smoke end up in the experimental group, non-smokers in the control group. Because of this non-random selection, there may be disturbing factors. For instance, if there are more heart diseases among the smokers, this may be due to the fact that both smoking and heart disease are positively influenced by coffee drinking (i.e., coffee drinking is a common cause of smoking and heart disease). Randomised experiments avoid this problem by the random division into experimental and control group.

The standard solution to this problem is “conditioning on potential confounders”. But this solution has its limitations. Dan Steel formulates these limitations as follows:

> I agree that there are cases in which one can draw reasonable conclusions about what causes what without the aid of experiment or substantial knowledge of underlying mechanisms. However, the usefulness of conditioning on potential causes does not undermine the proposal that mechanisms significantly aid causal inferences in the social science, since social scientists are rarely able to measure all potential common causes. Indeed, the inability to exhaustively consider all potential common causes is a basic element of the problem of confounders, to which mechanisms are being considered as a partial solution. (Steel 2004, 63)

This limitation is not specific for the social sciences: potential disturbing factors (confounders) can be eliminated by means of statistical methods on a one-by-one basis. However, we can never be sure that untested variables will not turn out to be confounders, and we cannot test all possible variables. For instance, we can exclude the possibility that coffee drinking is a common cause in the above example, but we cannot be sure that there is no other variable which is correlated with smoking and heart disease and which is responsible for the correlation (i.e. we cannot exclude the possibility that smoking and heart disease have a common cause; we can only test variables individually and exclude them as common causes).

How can mechanisms help here? Steel 2004 distinguishes two possible roles of mechanisms relating to the problem of confounders. The first possible role is a negative one: if we don't find a plausible mechanism linking the two variables, we can conclude that the correlation between them is spurious (i.e. there is a common cause). Steel thinks that this negative role does not work, because we can always find plausible mechanisms. The second possible role is positive: if we find a mechanism for which we have good evidence we can conclude that there is a causal relation between the two correlated variables. Because finding good evidence that a mechanism is present is much more difficult than coming up with a plausible mechanism (without showing that it is present) Steel claims that the positive role can really work.

In order to understand the positive role which Steel has in mind properly, it is important to see that it is part of a larger procedure, which also includes conditioning on potential confounders. Consider a scientist who starts from evidence gathered in a prospective or retrospective study, and wants to arrive at a conclusion about a Gierean causal claim. This scientist should first check some potential confounders. If the causal hypothesis survives all these tests (each test can falsify the hypothesis) the scientist can start to look for a mechanism which connects the alleged cause variable with the alleged effect variable. If such a mechanism is found, this strengthens the case

for the causal claim. So the best overall procedure to process data from prospective and retrospective designs includes mechanistic evidence, at least if the hypothesis survives the statistical confounding tests (otherwise the procedure results in the rejection of the causal hypothesis before the stage at which mechanistic evidence comes in).

What is the upshot of all this? Giere's definitions are certainly compatible with the use of results of prospective and retrospective studies as evidence. This means that Giere's theory can account for the use of mechanisms, even in cases where appropriate probabilistic associations are known from prospective or retrospective studies: mechanistic evidence plays a role in the process by which data from such studies are used to build an argument for causal claims. So for a scientist who wants to make Gierean causal claims about populations, it is perfectly rational to look for mechanistic evidence: it will be useful for processing the results of prospective and retrospective studies.

5.2 As already indicated, experiments with laboratory animals are very common in the biomedical sciences for ethical reasons. Extrapolating the results of such studies to humans is not trivial. If we can show that a substance (e.g., saccharin) is dangerous for rats (e.g., causes bladder cancer), this does not logically entail that this substance causes the same disease in humans (this is one of Giere's examples). So we need a warrant to justify the extrapolation. The need for such a warrant becomes very clear if we realise that different animal species may suggest different causal relations. For instance, aflotoxin B_1 causes liver cancer in rats but not in mice (Steel 2008, 82). So when we extrapolate without limitation in this case, we arrive at the conclusion that aflotoxin B_1 both causes and does not cause liver cancer in humans. One of these must be false. A sophisticated procedure is necessary to decide which model species we will use for extrapolation, and which one we will disregard. According to Steel, this procedure must include the use of

mechanistic evidence. More precisely, it must contain what he calls *comparative process tracing* (Steel 2008, 88-92).

The International Agency for Research on Cancer (IARC) uses mechanistic evidence to extrapolate the results of animal experiments to human beings (see IARC 2006). The basic idea in their procedure is that, if one has good evidence of carcinogenicity of a substance in experimental animals and also good evidence that the carcinogenesis (in the animals) is mediated by a mechanism that also operates in humans, then it is safe to extrapolate. For a detailed analysis of the IARC procedures (and their shortcomings) see Leuridan & Weber (2010).

Let us take stock. Extrapolation from animal experiments is a context in which mechanistic and probabilistic evidence are combined in practice (cfr. the IARC procedures). There is a sound justification for this practice (conflicting information from different animal species). Giere's theory of probabilistic causation has no problem to account for this practice. Data about animals do not automatically give us reliable estimates of the value of the $\mathbf{P}_X(E)$ and $\mathbf{P}_K(E)$ (where X and K are hypothetical populations of humans). So a scientist who wants to make a Gierean causal claim, needs a warrant to link the animal data to the hypothetical human populations. Mechanistic evidence can provide this link. So it is rational for such a scientist to look for mechanistic evidence.

A final note must be made about qualitative versus quantitative extrapolation. Suppose we discover that for a certain substance S we find liver cancer in 70% of the treated rats, while in the controls the incidence of liver cancer is only 10%. We can attempt either a quantitative or a qualitative extrapolation from this. The qualitative extrapolation consists in the following claim:

Substance S is a positive causal factor for liver cancer in humans.

The quantitative extrapolation consists in a stronger claim:

> Substance S is a positive causal factor for liver cancer in humans, and its effectiveness is 0.6.

Effectiveness is understood here as defined in Section 3.4. Mechanistic evidence, if used properly (cfr. Steel's comparative process tracing) warrants qualitative extrapolation: combined with the data about laboratory animals, it provides evidence for the claim that there is *a difference between* $\mathbf{P}_X(E)$ and $\mathbf{P}_K(E)$ (where X and K are hypothetical populations of humans). That is, according to Giere's theory, sufficient to claim that there is positive or negative causal relevance. We don't get precise estimates of the values of $\mathbf{P}_X(E)$ and $\mathbf{P}_K(E)$. So animal experiments allow us to make causal claims, but do not warrant claims about the effectiveness of a causal factor.

5.3 In this section and the previous one, I have argued that Giere's theory can account for the use of mechanistic evidence: the way mechanistic evidence is used by scientists in various contexts, is rational from a Gierean point of view. Let us now imagine that we live in the ideal world for experimenters. In this world, time travel is possible (so we can do randomised experiments in the past; cfr. the absence of data in the example in Section 4) and there are no ethical restrictions (so there is no need to do animal experiments or prospective or retrospective studies; the Ideal World for Experimenters is obviously not ideal for experimental subjects). In such a world, one could say, a scientist who wants to make a Gierean population claim, does not have any use for mechanistic evidence. Even if that would be true, that would not count as an argument against Giere's theory: Giere wants to deal with real science and scientists in the real world. Moreover, it can be argued that even in the ideal world for experimenters, mechanistic evidence would play a

role from a Gierean point of view (see Weber 2007 for this; the argument is connected to the stability of causal generalisations over time).

6. Giere's theory compared to other probabilistic theories of causation

While Giere defines causation in terms of what would happen in hypothetical populations, other probabilistic theories, such as the ones of Patrick Suppes, Ellery Eells and Paul Humphreys, define causation in terms of probabilistic relations in the real world. In this section I discuss these theories and argue that this difference makes Giere's theory better than these rivals.

6.1 In her recent book *Hunting Causes and Using Them*, Nancy Cartwright says that in the philosophical study of causation "[m]etaphysics, methods and use must march hand in hand" (Cartwright 2007, 1). By this she means that philosophers of causation should try to develop an integrated account, which answers three interrelated questions about causation: What do causal claims mean? How do we confirm them? What use can we make of them? I think that Giere's theory scores better from this point of view than the others.

Let us start with the theory of Patrick Suppes. In his view, genuine (probabilistic) causes are prima facie causes that are not spurious (Suppes 1970, 24). The definition of a prima facie cause is:

The event $B_{t'}$ is a prima facie cause of the event A_t if and only if:

(i) $t' < t$,

(ii) $P(B_{t'}) > 0$,

(iii) $P(A_t|B_{t'}) > P(A_t)$. (Suppes 1970, 12)

Spurious causes are defined as follows[2]:

An event $B_{t'}$ is a spurious cause in sense one of A_t if and only if $B_{t'}$ is a prima facie cause of A_t and there is a $t'' < t'$ and an event $C_{t''}$ such that:

(i) $P(B_{t'}C_{t''}) > 0$,

(ii) $P(A_t|B_{t'}C_{t''}) = P(A_t|C_{t''})$,

(iii) $P(A_t|B_{t'}C_{t''}) \geq P(A_t|B_{t'})$. (Suppes 1970, 23)

These definitions are Suppes' answer to the question "What do probabilistic causal claims mean?". As already mentioned, the crucial difference between Giere's definitions and those of Suppes, is that the latter refers to the presence and absence of positive statistical relevance relations in the real world. If we adopt Suppes' definition, the policy relevance of probabilistic causal claims is not clear. Why should policy makers want causal knowledge? If we adopt Giere's theory, the answer is clear: the hypothetical populations X and K correspond to populations a policy maker may create by means of some direct intervention (e.g. a ban on smoking, a mandatory inoculation...). There is no such link between causation and policy if we adopt Suppes's definition, because it refers to the real world, not to a hypothetical world that policy makers can create.

Another disadvantage of Suppes' theory is that it has problems with mechanistic evidence. In Section 3.3 I used microscopes to clarify the relation between different types of evidence and Giere's definitions. Here it is useful to draw an analogy with the use of telescopes. Suppose we first observe an object with a telescope, and then manage to come close enough to it to observe it

without the aid of instruments. The evidence we gather in the second stage (without the telescope) is more direct than the one we gather in the first stage (with the aid of the telescope). As soon as we have observed the object with the naked eye, the observations through the telescope become superfluous. In a similar way, mechanistic evidence becomes superfluous according to Suppes's account, as soon as we have probabilistic evidence. The reason for this is that a specific type of information (probabilistic dependencies) is used to define what causation means. This way of defining establishes a direct evidential link between this type of information and causal claims.

6.2 As already mentioned in Section 3, Ellery Eells (1991) has developed a context unanimity theory of probabilistic causation. Let me repeat the crucial passage:

> Then we say that X is a *positive causal factor* for Y if and only if, *for each i*, $Pr(Y/K_i\&X) > Pr(Y/K_i\&\sim X)$. *Negative causal factorhood* and *causal neutrality* are defined by changing the "always rises" (>) idea to "always lowers" (<) and "always leaves unchanged" (=), respectively. The idea that the inequality or equality must hold for *each* of the background contexts K_i is sometimes called the condition of *contextual unanimity*, or *context unanimity*. ... Note that these three relations of positive, negative and neutral causal factorhood are not exhaustive of the possible causal significance that a factor X can have for a factor Y: There remains the possibility of various kinds of *mixed* causal relevance, corresponding to various ways in which unanimity can fail. (Eells 1991, 86-87)

Like Suppes, Eells defines causation in terms of positive statistical relevance relations in real populations. This means that his theory faces the same problems as Suppes' theory: the policy

relevance of causal claims is not clear, and there is a direct evidential link between probabilistic data and causation.

As has been argued in Section 3.2, the context unanimity requirement causes additional problems. A counterfactual version of the context unanimity theory (which would not suffer from the two problems just mentioned) would still be worse than Giere's theory with respect to the link between policy and causation, because policy makers need average effects. Furthermore, the context unanimity theory entails that scientific inquiry cannot support probabilistic causal claims.

Finally, let us look at the definition of Paul Humphreys (1989):

> *B* is a *direct contributing cause* of A just in case
>
> (i) *A* occurs;
>
> (ii) *B* occurs;
>
> (iii) *B* increases the chance of *A* in all circumstances *Z* that are physically compatible with *A* and *B*, and with *A* and B_0, where B_0 is the neutral state of B, i.e., $P(A/BZ) > P(A/B_0Z)$ for all such *Z*; and
>
> (iv) *BZ* and *A* are logically independent.
>
> Similarly, *B* is a *direct counteracting cause* of *A* just in case (i), (ii), (iii), and (iv) hold, with 'increases' replaced by 'decreases' and with the inequality reversed. (Huphreys 1989, 74)

This definition also includes a context unanimity requirement. And like the definitions of Suppes and Eells, it refers to real populations. So this definition faces the same problems as the theory of Eells. And a counterfactual version of it would still be worse than Giere's theory.

6.3 Elliot Sober has argued that so-called "frequency-dependent causation" is a serious threat to

Giere's theory (the version Sober is criticising is Giere 1980). Here is one of his examples:

> The first example is from the study of mimicry. The monarch butterfly *Danaus plexippus* tastes terrible to blue jays. Another butterfly species, *Limenitis archippus*, has evolved the characteristic appearance of the monarch, but without the bad flavor. The selective advantage of this form of mimicry depends on the frequency of the mimics relative to the models. If the unpalatable Monarchs predominate, mimicry will be advantageous, since the blue jays will be fooled. But if the tasty mimics predominate, the blue jays will learn how nice they are to eat. So the fitness of mimicry increases with its rarity. (Sober 1982, 249)

According to Sober, this poses a problem for Giere's theory: if all the butterflies in the second species would be mimics, there would be no selective advantage anymore. So if we apply Giere's theory, mimicry is judged not to be a cause of survival. Giere himself (1984) and Deborah Mayo (1986) have replied to this criticism by constructing refinements of Giere's account which have to be applied in cases of frequency-dependent causation. This type of reply fits into Giere's general view on what philosophy of science should be. In his view, philosophers of science have to develop concepts, and then apply them to the empirical world, in casu scientific practice. He sees his definitions as part of such philosophy of science (Giere 1984, 385). In Giere's view, his simple model is applicable to a lot of scientific research (especially in the biomedical sciences). And in cases where it is not applicable, we need a more sophisticated model. James Fetzer has offered a different kind of reply to Sober's criticism. In his view, frequency-dependent causation shows that Gierean causal claims should be taken as accidental generalisations which hold for a population at a certain time. Extrapolation in time requires specific arguments (Fetzer 1986, 122). Humphreys (1989, pp. 87-88) agrees with this. Taken together these reactions and proposal allow

us to safely conclude that frequency-dependent causation is not a serious threat for Giere's theory.

6.4 Given the strengths of Giere's theory, it is surprising that it is not popular among philosophers. The main reason for this, I think, is that probabilistic accounts in general are not very popular. In the philosophy of causation, the scene is dominated by adherents of a causal mechanical view which finds its roots in the work of Wesley Salmon (e.g. Dowe 2000 and complex system mechanicists like Glennan 1996)), adherents of a counterfactual view that originates in the work of David Lewis (Lewis 1973, expanded in Lewis 1986) and dualists that combine both views (e.g. Hall 2004). This dominance results from a focus – in the philosophy of causation – on the conceptual analysis of causal claims in everyday language and on metaphysical issues about causation. However, if we want to analyse how causation functions in science (this is what Giere wants to do; see also the quote from Nancy Cartwright in Section 6.1), we need a probabilistic conception. In my view, Giere's theory is the best candidate for fulfilling this role. I have given part of my reasons for this above: Giere's account is better than that of Suppes, Eells and Humphreys.[3]

7. Conclusion

I have presented Ronald Giere's theory of probabilistic causation, and argued that it can account for the use of mechanistic evidence in various contexts in which scientists use such evidence. This undermines the thesis of Russo & Williamson that probabilistic theories of causality cannot account for the use of mechanistic evidence. If we adopt Giere's theory of probabilistic causation, we can be output monists and evidential pluralists, just like Russo & Williamson. Other probabilistic theories of causation (Suppes, Eells and Humphreys) cannot account for the use of mechanistic evidence. I have also highlighted some other advantages of Giere's theory, especially with respect to the policy relevance of causal claims.[4]

Acknowledgements

I thank Leen De Vreese, Bert Leuridan and three anonymous referees of this journal for their comments on earlier drafts of this paper. The research for this paper was supported by the Research Fund - Flanders (FWO) through project nr. G.0651.07.

References

Cartwright, N. 1979. Causal Laws and Effective Strategies. *Noûs* 13: 419-437.

Cartwright, N. 2007. *Hunting Causes and Using Them. Approaches in Philosophy and Economics*. Cambridge: Cambridge University Press.

De Vreese, L. 2009. Epidemiology and Causation. *Medicine, Health Care and Philosophy.* DOI: 10.1007/s11019-009-9184-0 (published online first 15 February 2009).

Dowe, P. 2000. *Physical Causation*. Cambridge: Cambridge University Press.

Dupré, J. 1993. *The Disorder of Things*. Cambridge & London: Harvard University Press.

Eells, E. 1991. *Probabilistic Causality.* Cambridge: Cambridge University Press.

Fetzer, J. 1986. Methodological Individualism: Singular Causal Systems and their Population Manifestations. *Synthese* 68: 99-128.

Giere, R. 1979. *Understanding Scientific Reasoning* (1st edition). New York: Holt, Rinehart & Winston.

Giere, R. 1980. Causal Systems and Statistical Hypotheses. In *Applications of Inductive Logic,* edited by L. Cohen and M. Hesse, 251-270. New York: Oxford University Press.

Giere, R. 1984. Causal Models with Frequency Dependence. In *Journal of* Philosophy 81: 384-391.

Giere, R. 1997. *Understanding Scientific Reasoning* (4th edition). Fort Worth: Harcourt Brace College Publishers.

Glennan, S. 1996. Mechanisms and the Nature of Causation. *Erkentniss* 44: 49-71.

Hall, N. 2004. Two Concepts of Causation. In *Causation and Counterfactuals,* edited by N. Hall, J. Collins and L.A. Paul, 225-276. Cambridge: MIT Press.

Hausman, D. 1998. *Causal Asymmetries*. Cambridge: Cambridge University Press.

Humphreys, P. 1989. *The Chances of Explanation*. Princeton: Princeton University Press.

IARC (2006), *Preamble to the IARC Monographs on the Evaluation of Carcinogenic Risks to Humans.* Http://monographs.iarc.fr/ENG/Preamble/CurrentPreamble.pdf. Last accessed on 4 December 2008.

Leuridan, B., E. Weber, and M. Van Dyck. 2008. The Practical Value of Spurious Correlations: Selective versus Manipulative Policy. *Analysis* 68: 298-303.

Leuridan, B., and E. Weber. 2010. IARC, Mechanistic Evidence and the Precautionary Principle. In *Causality in the Sciences,* edited by P. McKay Illari, F. Russo and J. Williamson (eds.). Oxford: Oxford University Press (in print).

Lewis, D. 1973. Causation. *Journal of Philosophy* 70: 556-567.

Lewis, D. 1986. Causation. In *Philosophical Papers II,* 159-213. New York & Oxford: Oxford University Press.

Little, D. 1991. *Varieties of Social Explanation*. Boulder: Westview Press.

Mayo, D. 1986. Understanding Frequency-dependent Causation. *Philosophical Studies* 49: 109-124.

Morgan, S., and C. Winship. 2007. *Counterfactuals and Causal Inference. Methods and Principles for Social Research*. Cambridge: Cambridge University Press.

Pasternak, B. 1978. The Sociology of Irrigation: Two Taiwanese Villages. In *Studies in Chinese Society*, edited by A. Wolf. Stanford: Stanford University Press.

Russo, F., and J. Williamson Jon. 2007. Interpreting Causality in the Health Sciences. *International Studies in the Philosophy of Science* 21: 157-170.

Salmon, W. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton, New Jersey: Princeton University Press.

Sober, E. 1982. Frequency-dependent Causation. *Journal of Philosophy* 79: 247-253.

Steel, D. 2004. Social Mechanisms and Causal Inference. *Philosophy of the Social Sciences* 34: 55-78.

Steel, D. 2008. *Across the Boundaries. Extrapolation in Biology and Social Science*. New York: Oxford University Press.

Suppes, P. 1970. *A Probabilistic Theory of Causality*. Amsterdam: North-Holland Publishing Company.

Weber, E. 2007. Social Mechanisms, Causal Inference, and the Policy Relevance of Social Science. *Philosophy of the Social Sciences* 37: 348-359.

Notes

1. In Leuridan, Weber & Van Dyck (2008) a distinction is made between manipulative policy and selective policy. Manipulative policy requires causation (in the average effect sense), selective policy uses a specific type of non-causal information.

2. Suppes distinguishes a second sense of spurious cause, but that type is not important for my purposes.

3. Giere's model is not popular in the traditional philosophy of causation, but the situation is clearly different when we look at scientists dealing with methodological issues in their own discipline. For instance, Stephen Morgan and Christopher Winship (2007, chapter 2) use a "potential outcome model" of causation which is very similar to Giere's theory; this model is very common in the social methodological literature on causation.

4. I do not discuss interventionist accounts (such as Pearl 2000 and Woodward 2003) here because that would bring us too far from the original aim of this paper. Pearl and Woodward do not require context unanimity, and their definitions have a counterfactual nature. So they share the important advantages of Giere's account discussed here. Giere avoids the concept of intervention which Pearl and Woodward use in their definition of causation. This seems to be an advantage, though Pearl and Woodward insist on a non-anthropocentric interpretation of the concept of intervention. A detailed comparison may reveal other relative advantages and disadvantages.