Erschienen in: Heuristics, probability and causality : a tribute to Judea Pearl / Dechter, Rina; Geffner, Hector; Halpern, Joseph Y. (Hrsg.). - London : College Publ., 2010. - S. 507-522. - ISBN 978-1-904987-65-9

The Structural Model and the Ranking Theoretic Approach to Causation: A Comparison

WOLFGANG SPOHN

1 Introduction

Large parts of Judea Pearl's very rich work lie outside philosophy; moreover, basically being a computer scientist, his natural interest was in computational efficiency, which, as such, is not a philosophical virtue. Still, the philosophical impact of Judea Pearl's work is tremendous and often immediate; for the philosopher of science and the formal epistemologist few writings are as relevant as his. Fully deservedly, this fact is reflected in some philosophical contributions to this Festschrift; I am glad I can contribute as well.

For decades, Judea Pearl and I were pondering some of the same topics. We both realized the importance of the Bayesian net structure and elaborated on it; his emphasis on the graphical part was crucial, though. We both saw the huge potential of this structure for causal theorizing, in particular for probabilistic causation. We both felt the need for underpinning the probabilistic account by a theory of deterministic causation; this is, after all, the primary notion. And we both came up with relevant proposals. Judea Pearl approached these topics from the Artificial Intelligence side, I from the philosophy side. Given our different proveniences, overlap and congruity are surprisingly large.

Nevertheless, it slowly dawned upon me that the glaring similarities are deceptive, and that we fill the same structure with quite different contents. It is odd how much divergence can hide underneath so much similarity. I have identified no less than fifteen different, though interrelated points of divergence, and, to be clear, I am referring here only to our accounts of deterministic causation, the structural model approach so richly developed by Judea Pearl and my (certainly idiosyncratic) ranking-theoretic approach. In this brief paper I just want to list the points of divergence in a more or less descriptive mood, without much argument. Still, the paper may serve as a succinct reference list of the many crucial points that are at issue when dealing with causation and may thus help future discussion.

At bottom, my comparison refers, on the one hand, to the momentous book of Pearl (2000), the origins of which reach back to the other momentous book of Pearl (1988) and many important papers in the 80's and 90's,s and, on the other hand, to the chapters 14 and 15 of Spohn (forthcoming) on causation, the origins of which reach back to Spohn (1978, sections 3.2 - 3, and 1983) and a bunch of subsequent papers. For ease of access,

Konstanzer Online-Publikations-System (KOPS)
URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-129493

though, I shall substantially refer to Halpern, Pearl (2005) and Spohn (2006) where the relevant accounts are presented in a more compact way. Let me start with reproducing the basic explications in section 2 and then proceed to my list of points of comparison in section 3. Section 4 concludes with a brief moral.

2 The Accounts to be Compared

For all those taking philosophical talk of events not too seriously (the vast majority among causal theorists) the starting point is a *frame*, a (non-empty, finite) set **U** of *variables*; X, Y, Z, W, etc. denote members of **U**, $\vec{X}$, $\vec{Y}$, $\vec{Z}$, $\vec{W}$, etc. subsets of **U**. Each variable $X \in U$ has a *range* Ω_X of values and is a function from some *possibility space* Ω into its range Ω_X. For simplicity, we may assume that Ω is the Cartesian product of all Ω_X and X the projection from Ω to Ω_X. For $x \in \Omega_X$ and $A \subseteq \Omega_X$, $\{X = x\} = \{\omega \in \Omega \mid X(\omega) = x\}$ and $\{X \in A\} = \{\omega \mid X(\omega) \in A\}$ are *propositions* (or events), and all those propositions generate a propositional algebra **A** over Ω. For $\vec{X} = \{X_1, \ldots, X_n\}$ and $\vec{x} = \langle x_1, \ldots, x_n \rangle$ $\{\vec{X} = \vec{x}\}$ is short for $\{X_1 = x_1$ and $\ldots$ and $X_n = x_n\}$. How a variable is to be precisely understood may be exemplified in the usual ways; however, we shall see that it is one of the issues still to be discussed.

The causal theorist may or may not presuppose a temporal order among variables; I shall. So, let $\prec$ be a linear order on the frame **U** representing temporal precedence. Linearity excludes simultaneous variables. The issue of simultaneous causation is pressing, but not one dividing us; therefore I put it to one side. Let, e.g., $\{\prec Y\}$ denote $\{Z \in \mathbf{U} \mid Z \prec Y\}$, that is, the set of variables preceding Y. So much for the algebraic groundwork.

A further ingredient is needed in order to explain causal relations. In the structural-model approach it is a set of structural equations, in the ranking-theoretic approach it is a ranking function.

A set **F** of *structural equations* is just a set of functions F_Y that specifies for each variable Y in some subset $\vec{V}$ of **U** how Y (essentially) functionally depends on some subset $\vec{X}$ of **U**; thus F_Y maps $\Omega_{\vec{X}}$ into Ω_Y. $\vec{V}$ is the set of *endogenous* variables, $\vec{U} = \mathbf{U} - \vec{V}$ the set of *exogenous* variables. The only condition on **F** is that no Y in $\vec{V}$ indirectly functionally depends on itself via the equations in **F**. Thus, **F** induces a DAG on **U** such that, if F_Y maps $\Omega_{\vec{X}}$ into Ω_Y, $\vec{X}$ is the set of parents of Y. (In their appendix A.4 Halpern, Pearl (2005) generalize their account by dropping the assumption of the acyclicity of the structural equations.) The idea is that **F** provides a set of laws that govern the variables in **U**, though, again, the precise interpretation of **F** will have to be discussed below. $\langle \mathbf{U}, \mathbf{F} \rangle$ is then called a *structural model* (*SM*). Note that a SM does not fix the values of any variables. However, once we fix the values $\vec{u}$ of all the exogenous variables in $\vec{U}$, the equations in **F** determine the values $\vec{v}$ of all the endogenous variables in $\vec{V}$. Let us call $\langle \mathbf{U}, \mathbf{F}, \vec{u} \rangle$ a *contextualized structural model* (*CSM*). Thus, each CSM determines a specific world or course of events $\omega = \langle \vec{u}, \vec{v} \rangle$ in Ω. Accordingly, each proposition A in **A** is *true* or *false* in a CSM $\langle \mathbf{U}, \mathbf{F}, \vec{u} \rangle$, depending on whether or not $\omega \in A$ for the ω thereby determined.

For the structural model approach, causation is essentially related to intervention. Therefore we must first explain the latter notion. An *intervention* always intervenes on a CSM $\langle \mathbf{U}, \mathbf{F}, \vec{u} \rangle$, more specifically, on a certain set $\vec{X} \subseteq \vec{V}$ of endogenous variables, thereby setting the values of $\vec{X}$ to some fixed values $\vec{x}$; that intervention or setting is denoted by $\vec{X} \leftarrow \vec{x}$. What such an intervention $\vec{X} \leftarrow \vec{x}$ does is to turn the CSM $\langle \mathbf{U}, \mathbf{F}, \vec{u} \rangle$ into another CSM. The variables in $\vec{X}$ are turned into exogenous variables; i.e., the set $\mathbf{F}$ of structural equations is reduced to the set $\mathbf{F}^{\vec{X}}$, as I denote it, that consists of all the equations in $\mathbf{F}$ for the variables in $\vec{V} - \vec{X}$. Correspondingly, the context $\vec{u}$ of the original CSM is enriched by the chosen setting $\vec{x}$ for the new exogenous variables in $\vec{X}$. In short, the intervention $\vec{X} \leftarrow \vec{x}$ changes the CSM $\langle \mathbf{U}, \mathbf{F}, \vec{u} \rangle$ into the CSM $\langle \mathbf{U}, \mathbf{F}^{\vec{X}}, \langle \vec{u}, \vec{x} \rangle \rangle$. Again, it will be an issue what this precisely means.

Now, we can proceed to Pearl's explication of actual causation; this is definition 3.1 of Halpern, Pearl (2005, p. 853) slightly adapted to the notation introduced so far (see also Halpern, Hitchcock (2010, Section 3)). Not every detail will be relevant to my further discussion below; I reproduce it here only for reasons of accuracy:

SM DEFINITION: $\{\vec{X} = \vec{x}\}$ is an *actual cause* of $\{Y = y\}$ in the CSM $\langle \mathbf{U}, \mathbf{F}, \vec{u} \rangle$ iff the following three conditions hold:

(1) $\{\vec{X} = \vec{x}\}$ and $\{Y = y\}$ are true in $\langle \mathbf{U}, \mathbf{F}, \vec{u} \rangle$.

(2) There exists a partition $\langle \vec{Z}, \vec{W} \rangle$ of $\vec{V}$ with $\vec{X} \subseteq \vec{Z}$ and some setting $\langle \vec{x}', \vec{w}' \rangle$ of the variables in $\vec{X}$ and $\vec{W}$ such that if $\{\vec{Z} = \vec{z}\}$ is true in $\langle \mathbf{U}, \mathbf{F}, \vec{u} \rangle$, then both of the following conditions hold:

   (a) $\{Y = y\}$ is false in the intervention $\langle \vec{X}, \vec{W} \rangle \leftarrow \langle \vec{x}', \vec{w}' \rangle$ on $\langle \mathbf{U}, \mathbf{F}, \vec{u} \rangle$, i.e., in $\langle \mathbf{U}, \mathbf{F}^{\vec{X},\vec{W}}, \langle \vec{u}, \vec{x}', \vec{w}' \rangle \rangle$. In other words, changing $\langle \vec{X}, \vec{W} \rangle$ from $\langle \vec{x}, \vec{w} \rangle$ to $\langle \vec{x}', \vec{w}' \rangle$ changes $\{Y = y\}$ from true to false.

   (b) $\{Y = y\}$ is true in $\langle \mathbf{U}, \mathbf{F}^{\vec{X},\vec{W}',\vec{Z}'}, \langle \vec{u}, \vec{x}, \vec{w}', \vec{z}' \rangle \rangle$ for all subsets $\vec{W}'$ of $\vec{W}$ and all subsets $\vec{Z}'$ of $\vec{Z}$, where $\vec{z}'$ is the subsequence of $\vec{z}$ pertaining to $\vec{Z}'$.

(3) $\vec{X}$ is minimal; i.e., no subset of $\vec{X}$ satisfies conditions (1) and (2).

This is not as complicated as it may look. Condition (1) says that the cause and the effect actually occur in the relevant CSM $\langle \mathbf{U}, \mathbf{F}, \vec{u} \rangle$ and, indeed, had to occur given the structural equations in $\mathbf{F}$ and the context $\vec{u}$. Condition (2a) says that if the cause variables in $\vec{X}$ had been set differently, the effect $\{Y = y\}$ would not have occurred. It is indeed more liberal in allowing that also the variables in $\vec{W}$ outside $\vec{X}$ are set to different values, the reason being that the effect of $\vec{X}$ on Y may be hidden, as it were, by the actual values of $\vec{W}$, and uncovered only by setting $\vec{W}$ to different values. However, this alone would be too liberal; perhaps the failure of the effect $\{Y = y\}$ to occur is due only to the change of $\vec{W}$ rather than that of $\vec{X}$. Condition (2b) counteracts this permissiveness, and ensures that basically the change in $\vec{X}$ alone brings about the change of Y. Condition (3), finally, is to guarantee that the cause $\{\vec{X} = \vec{x}\}$ does not contain irrelevant parts; for the change described in (2a) all the variables in $\vec{X}$ are required. Note that $\vec{X}$ is a set of

variables so that $\{ \vec{X} = \vec{x} \}$ should be called a *total cause* of $\{Y = y\}$; its parts $\{X_i = x_i\}$ for $X_i \in \vec{X}$ may then be called *contributory causes*.

The details of the SM definition are mainly motivated by an adequate treatment of various troubling examples much discussed in the literature. It would take us too far to go into all of them. I should also mention that the SM definition is only preliminary in Halpern, Pearl (2005); but again, the details of their more refined definition presented on p. 870 will not be relevant for the present discussion.

The basics of the ranking-theoretic account may be explained in an equally brief way: A *negative ranking function* κ for Ω is just a function κ from Ω into $\mathbf{N} \cup \{\infty\}$ such that $\kappa(\omega) = 0$ for at least one $\omega \in \Omega$. It is extended to propositions in $\mathbf{A}$ by defining $\kappa(A) = \min\{\kappa(\omega) \mid \omega \in A\}$ and $\kappa(\varnothing) = \infty$; and it is extended to conditional ranks by defining $\kappa(B \mid A) = \kappa(A \cap B) - \kappa(A)$ for $\kappa(A) \neq \infty$. Negative ranks express degrees of disbelief: $\kappa(A) > 0$ says that A is disbelieved, so that $\kappa(\overline{A}) > 0$ expresses that A is believed in κ; however, we may well have $\kappa(A) = \kappa(\overline{A}) = 0$. It is useful to have both belief and disbelief represented in one function. Hence, we define the *two-sided rank* $\tau(A) = \kappa(\overline{A}) - \kappa(A)$, so that A is believed, disbelieved, or neither according to whether $\tau(A) > 0$, < 0, or $= 0$. Again, we have conditional two-sided ranks: $\tau(B \mid A) = \kappa(\overline{B} \mid A) - \kappa(B \mid A)$. The *positive relevance* of a proposition A to a proposition B is then defined by $\tau(B \mid A) > \tau(B \mid \overline{A})$, i.e., by the fact that B is more firmly believed or less firmly disbelieved given A than given $\overline{A}$; we might also say in this case that A *confirms* or *is a reason for* B. Similarly for negative relevance and irrelevance (= independence).

Like a set of structural equations, a ranking function κ induces a DAG on the frame $\mathbf{U}$ conforming with the given temporal order $\prec$. The procedure is the same as with probabilities: we simply define the set of parents of a variable Y as the unique minimal set $\vec{X} \subseteq \{\prec Y\}$ such that Y is independent of $\{\prec Y\} - \vec{X}$ given $\vec{X}$ relative to κ, i.e., such that Y is independent of all the other preceding variables given $\vec{X}$. If $\vec{X}$ is empty, Y is exogenous; if $\vec{X} \neq \varnothing$, Y is endogenous. The reading that Y directly causally depends on its parents will be justified later on.

Now, for me, being a cause is just being a special kind of conditional reason, i.e., being a reason given the past. In order to express this, for a subset $\vec{X}$ of $\mathbf{U}$ and a course of events $\omega \in \Omega$ let ${}^{\omega}[\vec{X}]$ denote the proposition that the variables in $\vec{X}$ behave as they do in ω. (So far, we could denote such a proposition by $\{\vec{X} = \vec{x}\}$, if $\vec{X}(\omega) = \vec{x}$, but we shall see in a moment that this notation is now impractical.) Then the basic definition of the ranking-theoretic account is this:

RT DEFINITION 1: For $A \subseteq W_X$ and $B \subseteq W_Y$ $\{X \in A\}$ is a *direct cause* of $\{Y \in B\}$ in $\omega \in \Omega$ relative to the ranking function κ (or the associated τ) iff

(a) $X \prec Y$,

(b) $X(\omega) \in A$ and $Y(\omega) \in B$, i.e., $\{X \in A\}$ and $\{Y \in B\}$ are facts in ω,

(c) $\tau(\{Y \in B\} \mid \{X \in A\} \cap {}^{\omega}[\{\prec Y\} - \{X\}]) > \tau(\{Y \in B\} \mid \{X \in \overline{A}\} \cap {}^{\omega}[\{\prec Y\} - \{X\}])$; i.e., $\{X \in A\}$ is a reason for $\{Y \in B\}$ given the rest of the past of Y as it is in ω.

It is obvious that the SM and the RT definition deal more or less with the same explicandum; both are after actual causes, where actuality is represented either by the context $\vec{u}$ of a CSM $\langle \mathbf{U}, \mathbf{F}, \vec{u} \rangle$ in the SM definition or by the course of events ω in the RT definition. A noticeable difference is that in the RT definition the cause $\{X \in A\}$ refers only to a single variable X. Thus, the RT definition grasps what has been called a contributory cause, a total cause of $\{Y \in B\}$ then being something like the conjunction of its contributory causes. As mentioned, the SM definition proceeds the other way around.

Of course, the major differences lie in the explicantia; this will be discussed in the next section. A further noticeable difference in the definienda is that the RT definition 1 explains only direct causation; indeed, if $\{X \in A\}$ would be an indirect cause of $\{Y \in B\}$, we could not expect $\{X \in A\}$ to be positively relevant to $\{Y \in B\}$ conditional on the rest of the past of Y in ω, since that condition would not keep open the causal path from X to Y, but fix it to its actual state in ω. Hence, the RT definition 1 is restricted accordingly. As the required extension, I propose the following

RT DEFINITION 2: $\{X \in A\}$ is a (*direct or indirect*) *cause* of $\{Y \in B\}$ in $\omega \in \Omega$ relative to κ (or τ) iff there are $Z_i \in \mathbf{U}$ and $C_i \in \Omega_{Z_i}$ $(i = 1, \ldots, n \geq 2)$ such that $X = Z_1$, $A = C_1$, $Y = Z_n$, $B = C_n$, and $\{Z_i \in C_i\}$ is a direct cause of $\{Z_{i+1} \in C_{i+1}\}$ in ω relative to κ for all $i = 1, \ldots, n - 1$.

In other words, causation in ω is just the transitive closure of direct causation in ω.

We may complete the ranking-theoretic account by explicating causal dependence between variables:

RT DEFINITION 3: $Y \in \mathbf{U}$ (*directly*) *causally depends* on $X \in \mathbf{U}$ relative κ iff there are $A \subseteq W_X$, $B \subseteq W_Y$, and $\omega \in \Omega$ such that $\{X \in A\}$ is a (direct) cause of $\{Y \in B\}$ in ω relative to κ.

One consequence of RT definition 3 is that the set of parents of Y in the DAG generated by κ and $\prec$ consists precisely of all the variables on which Y directly causally depends.

So much for the two accounts to be compared. There are all the differences that meet the eye. As we shall see, there are even more. Still, let me conclude this section by pointing out that there are also less differences than meet the eye. I have already mentioned that both accounts make use of the DAG structure of causal graphs. And when we supplement the probabilistic versions of the two accounts, they further converge. In the structural-model approach we would then replace the context $\vec{u}$ of a CSM $\langle \mathbf{U}, \mathbf{F}, \vec{u} \rangle$ by a probability distribution over the exogenous variables rendering them independent and extending via the structural equations to a distribution for the whole of $\mathbf{U}$, thus forming a *pseudo-indeterministic system*, as Spirtes et al. (1993, pp. 38f.) call it, and hence a Bayesian net in which the probabilities agree with the causal graph. In the ranking-theoretic approach, we would replace the ranking function by a probability measure for $\mathbf{U}$ (or over $\mathbf{A}$) that, together with the temporal order of the variables, would again induce a DAG or a

causal graph so as to form a Bayesian net. In this way, the basic ingredient of both accounts would become the same: a probability measure;the remaining differences appear to be of a merely technical nature.

Indeed, as I see the recent history of the theory of causation, this large agreement initially dominated the picture of probabilistic causation. However, the need for underpinning the probabilistic by a deterministic account was obvious; after all, the longer history of the notion was an almost entirely deterministic one up to the recent counterfactual accounts following Lewis (1973). And so the surprising ramification sketched above came about, both branches of which well agree with their probabilistic origins. The ramification is revealing since it makes explicit dividing lines that were hard to discern within the probabilistic harmony. Indeed, the points of divergence between the structural-model and the ranking-theoretic approach to be discussed in the next section apply to their probabilistic sisters as well, a claim that is quite suggestive, though I shall not elaborate on it.

3 Fifteen Points of Comparison

All in all, I shall come up with fifteen clearly distinguishable, though multiply connected points of comparison. The theorist of causation must take a stance towards all of them, and even more; my list is pertinent to the present comparison and certainly not exhaustive. Let us go through the list point for point:

(1) The most obvious instances provoking comparison and divergence are provided by *examples*, about preemption and prevention, overdetermination and switches, etc. The literature abounds in cases challenging all theories of causation and examples designed for discriminating among them, a huge bulk still awaiting systematic classification (though I attempted one in my (1983, ch. 3) as far as possible at that time). A theory of causation must do well with these examples in order to be acceptable. No theory, though, will reach a perfect score, all the more as many examples are contested by themselves, and do not provide a clear-cut criterion of adequacy. And what a 'good score' would be cannot be but vague. Therefore, I shall not even open this unending field of comparison regarding the two theories at hand.

(2) The main reason why examples provide only a soft criterion is that it is ultimately left to *intuition* to judge whether an example has been adequately treated. There are strong intuitions and weak ones. They often agree and often diverge. And they are often hard to compromise. Indeed, intuitions play an indispensable and important role in assessing theories of causation; they seem to provide the ultimate unquestionable grounds for that assessment.

Still, I have become cautious about the role of intuitions. Quite often I felt that the intuitions authors claim to have are guided by their theory; their intuitions seem to be what their theory suggests they should be. Indeed, the more I dig into theories of causation and develop my own, the harder it is for me to introspectively discern whether or not I share certain intuitions independently of any theorizing. So, again, the appeal to intui-

tions must be handled with care, and I shall not engage into a comparison of the relevant theories on an intuitive level.

(3) Another large field of comparison is the *proximity to* and the *applicability in scientific practice*. No doubt, the SM account fares much better in this respect than the RT approach. Structural modeling is something many scientists really do, whereas ranking theory is unknown in the sciences and it may be hard to say why it should be known outside epistemology. The point applies to other accounts as well. The regularity theory of causation seems close to the sciences, since they seem to state laws and regularities, whereas counterfactual analyses seem remote, since counterfactual claims are not an official part of scientific theories, even though, unofficially, counterfactual talk is ubiquitous. And probabilistic theories maintain their scientific appearance by ecumenically hiding disputes about the interpretation of probability.

Again, the importance of this criterion is undeniable; the causal theorist is well advised to appreciate the great expertise of the sciences, in general and specifically concerning causation. Still, I tend to downplay this criterion, not only in order to keep the RT account as a running candidate. The point is rather that the issue of causation is of a kind for which the sciences are not so well prepared. The counterfactual analysis is a case in point. If it should be basically correct, then the counterfactual idiom can no longer be treated as a second-rate vernacular (to use Quine's term), as the sciences do, but must be squarely faced in a systematic way, as, e.g., Pearl (2000, ch. 7) does, but qua philosopher, not qua scientist. Probabilities are a similar case. Mathematicians and statisticians by far know best how to deal with them. However, when it comes to say what probabilities mean, they are not in a privileged position.

The point of these three remarks is to claim primacy for theoretical issues about causation as such. External considerations are relevant and helpful, but they cannot release us from the task of taking some stance or other towards these theoretical issues. So, let us turn to them.

(4) Both, the SM and the RT account, are based on a *frame* providing a *framework of variables* and appertaining *facts*. I am not sure, however, whether we interpret it in the same way. A (random) variable is a function from some state space into some range of values, usually the reals; this is mathematical standard. That a variable takes a certain value is a proposition, and if the value is the true one (in some model), the proposition is a fact (in that model); so much is clear. However, the notion of a variable is ambiguous, and it is so since its statistic origins. A variable may vary over a given population as its state space and take on a certain value for each item in the population. E.g., size varies among Germans and takes (presently) the value 6' 0" for me. This is what I call a *generic variable*. Or a variable may vary over a set of possibilities as its state space and take values accordingly. For example, *my* (present) size is a variable in this sense and actually takes the value 6' 0", though it takes other values in other possibilities; I might (presently) have a different size. I call this a *singular variable* representing the possibility range of a

given single case. For each German (and time), size is such a singular variable. The generic variable of size, then, is formed by the actual values of all these singular variables.

The above RT account exclusively speaks about singular variables and their realizations; generic variables simply are out of the picture. By contrast, the ambiguity seems to afflict the SM account. I am sure everybody is fully clear about the ambiguity, but this clarity seems insufficiently reflected in the terminology. For instance, the equations of a SM represent laws or ceteris paribus laws or invariances in Woodward's (2003) terms or statistical laws, if supplemented by statistical 'error' terms, and thus state relations between generic variables. It is contextualization by which the model gets applied to a given single case; then, the variables should rather be taken as singular ones; their taking certain values then are specific facts. There is, however, no terminological distinction of the two interpretations; somehow, the notion of a variable seems to be intended to play both roles. In probabilistic extensions we find the same ambiguity, since probabilities may be interpreted as statistical distributions over populations or as realization propensities of the single case.

(5) I would not belabor the point if it did not extend to the causal relations we try to capture. We have causation among facts, as analyzed in the SM definition and the RT definitions 1 - 2; they are bound to apply to the single case. And we have causal relations among variables, i.e., causal dependence (though often and in my view confusingly the term "cause" is used here as well), and we find here the same ambiguity. Causal dependence between generic variables is a matter of causal laws or of *general causation*. However, there is also causal dependence between singular variables, something rarely made explicit, and it is a matter of *singular causation* applying to the single case just as much as causation between facts. Since its inception the discussion of probabilistic causality was caught in this ambiguity between singular and general causation; and I am wondering whether we can still observe the aftermath of that situation.

In any case, structural equations are intended to capture causal order, and the order among generic variables thus given pertains to general causation. Derivatively these equations may be interpreted as stating causal dependencies also between singular variables. In the SM account, though, singular causation is explicitly treated only as pertaining to facts. By contrast, the RT definition 3 explicates only causal dependence between singular variables. The RT account is so far silent about general causation and can grasp it only by generalizing over the causal relations in the single case. These remarks are not just pedantry; I think it is important to observe these differences for an adequate comparison of the accounts.

(6) I see these differences related to the issue of the role of *time* in an analysis of causation. The point is simply that generic variables as such are not temporally ordered, since their arguments, the items to which they apply, may have varying temporal positions; usually, statistical data do not come temporally ordered. By contrast, singular variables are temporally ordered, since their variable realizability across possibilities is tied to a fixed time. As a consequence, the SM definition makes no explicit reference to time,

whereas the RT definitions make free use of that reference. While I think that this point has indeed disposed Judea Pearl and me to our diverging perspectives on the relation between time and causation, it must be granted that the issue takes on much larger dimensions that open enough room for indecisive defenses of both perspectives.

Many points are involved: (i) Issues of analytic adequacy: while Pearl (2000, pp. 249ff.) argues that reference to time does not sufficiently further the analytic project and proposes ingenious alternatives (sections 2.3 - 4 + 8 - 9), I am much more optimistic about the analytic prospects of referring to time (see my 1990, section 3, and forthcoming, section 14.4). (ii) Issues of analytic policy (see also point 10 below): Is it legitimate to refer to time in an analysis of causation? I was never convinced by the objections. Or should the two notions be analytically decoupled? Or should the analytic order be even reversed by constructing a causal theory of time? Pearl (2000, section 2.8) shows sympathies for the latter project, although he suggests an evolutionary explanation, rather than Reichenbach's (1956) physical explanation for relating temporal direction with causal directionality. (iii) The issue of causal asymmetry: Is the explanation of causal asymmetry by temporal asymmetry illegitimate? Or incomplete? Or too uninformative, as far as it goes? If any of these, what is the alternative?

(7) Causation always is causation within given *circumstances*. What do the accounts say what the circumstances are? The RT definition 1 explicitly takes the entire past of the effect except the cause as the circumstances of a direct causal relationship, something apparently much too large and hence inadequate, but free of conceptual circularity, as I have continuously emphasized. In contrast, Pearl (2000, pp. 250ff.) endorses the circular explanation of Cartwright (1979) that those circumstances consist of the other causes of the effect and hence, in the case of direct causation, of the realizations of the other parents of the effect variable in the causal graph. Pearl thus accepts also Cartwright's conclusion that the reference to the obtaining circumstances does not help explicating causation; he thinks that this reference at best provides a kind of consistency test. I argue that the explicatory project is not doomed thereby, since Cartwright's circular explanation may be derived from my apparently inadequate definition (cf. Spohn 1990, section 4). As for the circumstances of indirect causation, the RT definition 2 is entirely silent, since it relies on transitivity; however, in Spohn (1990, Theorems 14 and 16) I explored how much I can say about them. In contrast, the SM definition contains an implicit account of the circumstances that applies to indirect causal relationships as well; it is hidden in the partition $\langle \vec{Z}, \vec{W} \rangle$ of the set $\vec{V}$ of endogenous variables. However, it still accepts Cartwright's circular explanation, since it presupposes the causal graph generated by the structural equations. So, this is a further respect in which our accounts are diametrically opposed.

(8) The preceding point contains two further issues. One concerns the distinction of *direct and indirect causation*. The SM approach explicates causation without attending to this distinction. Of course, it could account for it, but it does not acquire a basic importance. By contrast, the distinction receives analytic significance within the RT approach

that first defines direct causation and then, only on that basis, indirect causation. The reason is that, in this way, the RT approach hopes to reach a non-circular explication of causation, whereas the SM approach has given up on this hope (see also point 10 below) and thus sees no analytic rewards in this distinction.

(9) The other issue already alluded to in (7) is the issue of transitivity. This is a most vexed topic, and the community seems unable to find a stable attitude. Transitivity had to be given up, it seemed, within probabilistic causation (cf. Suppes 1970, p. 58), while it was derivable from a regularity account and was still defended by Lewis (1973) for deterministic causation. In the meantime the situation has reversed; transitivity has become more respectable within the probabilistic camp; e.g., Spirtes et al. (1993, p. 44) simply assume it in their definition of "indirect cause". By contrast, more and more tend to reject it for deterministic causation (cf., e.g., McDermott 1995 and Hitchcock 2001).

This uncertainty is also reflected in the present comparison. Pearl (2000, p. 237) rejects transitivity of causal dependence among variables, but, as the argument shows, only in the sense of what Woodward (2003, p. 51) calls "total cause". Still, Woodward (2003, p. 59), in his concluding explication **M**, accepts the transitivity of causal dependence among variables in the sense of "contributory cause", and I have not found any indication in Pearl (2000) or Halpern, Pearl (2005) that they would reject Woodward's account of contributory causation. However, all of them deny the transitivity of actual causation between facts.

I see it just the other way around. The RT definition 2 stipulates the transitivity of causation (with arguments, though; cf. Spohn 1990, p. 138, and forthcoming, section 14.12), whereas the RT definition 3 entails the transitivity of causal dependence among variables in the contributory sense only under (mild) additional assumptions. Another diametrical opposition.

(10) A much grander issue is looming behind the previous points, the issue of *analytic policy*. The RT approach starts defining direct causation between singular facts, proceeds to indirect causation and then to causal dependence between singular variables, and finally only hopes to thereby grasp general causation as well. It thus claims to give a non-circular explication or a reductive analysis of causation. The SM approach proceeds in the opposite direction. It presupposes an account of general causation that is contained in the structural equations, transfers this to causal dependence between singular variables (I mentioned in points 4 and 5 that this step is not fully explicit), and finally arrives at actual causation between facts. The claim is thereby to give an illuminating analysis of causation, but not a reductive one.

Now, one may have an argument about conceptual order: which causal notions to explicate on the basis of which? I admit I am bewildered by the SM order. The deeper issue, though, or perhaps the deepest, is the feasibility of reductive analysis. Nobody doubts that it would be most welcome to have one; therefore the history of the topic is full of attempts at such an analysis. Perhaps, though, they are motivated by wishful thinking. How to decide? One way of assessing the issue is by inspecting the proposals. The proponents

are certainly confident of their analyses, but their inspection revealed so many problems that doubts preponderate. However, this does not prove their failure. Also, one may advance principled arguments such as Cartwright's (1979) that one cannot avoid being entangled in conceptual circles. For such reasons, the majority, it seems, has acquiesced in non-reductive analysis; cf., e.g., Woodward (2003, pp. 104ff.) for an apology of non-reductivity or Glymour (2004) for a eulogy of the, as he calls it, Euclidean as opposed to the Socratic ideal.

Another way of assessing the issue is more philosophical. Are there any more basic features of reality to which causation may reduce? One may well say no, and thereby justify the rejection of reductive analysis. Or one may say yes. Laws may be such a more basic feature; this, however, threatens to result either in an inadequate regularity theory of causation or in an inability to say what laws are beyond regularities. Objective probabilities may be such a feature – if we only knew what they are. What else is there on offer? On the other hand, it is not so easy to simply accept causation as a basic phenomenon; after all, the point has deeply worried philosophers for centuries after Hume.

In any case, all these issue are involved in settling for a certain analytic policy. It will become clearer in the subsequent points why I nevertheless maintain the possibility of reductive analysis.

(11) The most conspicuous difference of the SM and the RT approach is a direct consequence of their different policies. The SM account bases its analysis on *structural models* or *equations*, whereas the RT account explicates causation in terms of *ranking functions*. These are entirely different things!

Prima facie, structural equations are easier to grasp. Despite its non-reductive procedure the SM approach incurs the obligation, though, to somehow explain how the structural equations can establish causal order among generic variables. They can do this, because Pearl (2000, pp. 157ff.) explicitly gives them an interventionistic interpretation that, in turn, is basically a counterfactual one, as is entirely clear to Pearl; most interventions are only counterfactual. Woodward (2003) repeatedly emphasizes the point that the interventionistic account clarifies the counterfactual approach by forcing a specific interpretation of the multiply ambiguous counterfactual idiom. Still, despite Woodward's (2003, pp. 121f.) claim to use counterfactuals only when they are clearly true of false, and despite Pearl's (2000, section 7.1) attempt to account for counterfactuals within structural models, the issue how counterfactuals acquire truth conditions remains a mystery in my view.

By contrast, it is quite bewildering to base an analysis of causation on ranking functions that are avowedly to be understood only as doxastic states, i.e., in a purely epistemological way. One of my reasons for doing so is that the closer inspection envisaged in (10) comes out, on the whole, more satisfactorily than for other accounts, that is, the overall score in dealing with examples is better. The other reason why I find ranking functions not so implausible a starting point lies in my profoundly Humean strategy in dealing with causation. There is no more basic feature of reality to which causation might reduce. The issue rather is how modal facts come into the world – where modal facts

pertain to lawhood, causation, counterfactuals, probabilities, etc. We do not find 'musts' and 'cans' in the world as we find apples and pears; this was Hume's crucial challenge. And his answer was what is now called Hume's projectivism (cf. Blackburn 1993, in particular the essays in part I). Ranking functions are well suited for laying out this projectivist answer in detail. This fundamental difference between the SM and the RT approach further unfolds in the final four points.

(12) A basic idea in our notion of causation between facts is, very roughly, that the cause does something for its effect, contributes to it, makes it possible or necessary or more likely, in short: that the cause is somehow positively relevant to its effect. One fact could also be negatively relevant to another, in which case the second obtains despite the first. As for causal dependence between variables, it is only required that the one is relevant for the other. What are the notions of *relevance* and *positive relevance* provided by the SM and the RT approach?

Ranking theory has a rich notion of positive and negative relevance, analogous and equivalent in formal behavior to the probabilistic notions. Its relevance notion is much richer and, I find, more adequate to the needs of causal theorizing than those provided by the key terms of other approaches to deterministic causation: laws, counterfactuals, interventions, structural equations, or whatever. This fact grounds my optimism that the RT approach is, on the whole, better able to cope with all the examples and problem cases.

I just said that the relevance notion provided by the SM approach is poorer. What is it? Clause (2b) of the SM definition says, in a way, that the effect $\{Y = y\}$ had to occur given the cause $\{\vec{X} = \vec{x}\}$ occurs, and clause (2a) says that the effect might not have occurred if the cause does not occur and, indeed, would not have occurred if the cause variable(s) $\vec{X}$ would have been realized in a suitable alternative way. In traditional terms, we could say that the cause is a necessary and sufficient condition of the effect provided the circumstances – where the subtleties of the SM approach lie in the proviso; that's the SM positive relevance notion. So, roughly, in SM terms, the only 'action' a cause can do is making its effect necessary, whereas ranking theory allows many more 'actions'. This is what I mean by the SM approach being poorer. For instance, it is not clear how a fact could be negatively relevant to another fact in the SM approach, or how one fact could be positively and another negatively relevant to a third one. And so forth.

(13) Let's take a closer look at what "action" could mean in the previous paragraph. In the RT approach it means comparing ranks conditional on the cause $\{X \in A\}$ and on its negation $\{X \in \overline{A}\}$; the rank raising showing up in that comparison is what the cause 'does'. In the SM approach we do not conditionalize on the cause $\{\vec{X} = \vec{x}\}$ and some alternative $\{\vec{X} = \vec{x}'\}$; rather, in clauses (2a-b) of the SM definition we look at the consequences of the interventions $\vec{X} \leftarrow \vec{x}$ and $\vec{X} \leftarrow \vec{x}'$, i.e., by replacing the structural equation(s) for $\vec{X}$ by the stipulation $\vec{X} = \vec{x}$ or, respectively, $= \vec{x}'$. The received view by now is that *intervention* is quite different from *conditionalization* (cf., e.g., Goldszmidt, Pearl 1992, and Meek, Glymour 1994), the suggestion being that interven-

tion is what causal theorizing requires, and that all approaches relying on conditionalization such as the RT approach therefore are misguided (cf. also Pearl 2000, section 3.2).

The difference looks compelling: intervention is a real activity, whereas conditionalization is only a mental, suppositional activity. But once we grant that intervention is mostly counterfactual (i.e., also suppositional), the difference shrinks. Indeed, I tend to say that there never is a real intervention in a given single case; after a real intervention we deal with a different single case than before. Hence, I think the difference the received view assumes is spurious; rather, interventions may be construed in terms of conditionalization:

Of course, the intervention $\vec{X} \leftarrow \vec{x}$ differs from conditioning on $\{ \vec{X} = \vec{x} \}$; in this, the received view is correct. However, the RT and other conditioning approaches do not simply conditionalize on the cause, but on much more. What the intervention $X_1 \leftarrow x_1$ on the single variable X_1 does is change the value of X_1 to x_1 while at the same time keeping fixed the values of all temporally preceding variables as they are in the given context, or, if only a causal graph and not temporal order is available, either of all ancestors of X_1 or of all non-descendants of X_1 (which comes to the same thing in structural models, and also in probabilistic terms given the common cause principle). Thus, the intervention is equivalent to conditioning on $\{X_1 = x_1\}$ *and* on the fixed values of those other variables.

Similarly for a double intervention $\langle X_1, X_2 \rangle \leftarrow \langle x_1, x_2 \rangle$. For assessing the behavior of the variables temporally between X_1 and X_2 (or being descendants of X_1, but not of X_2) under the double intervention, we have to look at the same conditionalization as in the single intervention $X_1 \leftarrow x_1$, whereas for the variables later than X_2 (or descending from both X_1 and X_2) we have to condition on $\{X_1 = x_1\}$, $\{X_2 = x_2\}$, the past of X_1 as it is in the given context, and on those intermediate variables taking the values as they are after the intervention $X_1 \leftarrow x_1$. And so forth for multiple interventions (that are so crucial for the SM approach).

Given this translation, this kind of difference between the SM and the RT approach vanishes, I think. Consider, e.g., the definition of direct causal dependence of Woodward (2003, p. 55): Y directly causally depends on X iff an intervention on X can make a difference to Y, provided the values of all other variables in the given frame **U** are somehow fixed by intervention. Translate this as proposed, and you arrive at the conditionalization I use in the above RT definitions to characterize direct causation.

(14) The preceding argument has a gap that emerges when we attend to another topic that I find crucial, but nowhere thoroughly discussed: the *frame-relativity* of causation. Everybody agrees that the distinction between direct and indirect causation is frame-relative; of course, a direct causal relationship relative to a coarse-grained frame may turn indirect under refinements. What about causation itself, though? One may try some moderate antirealism, e.g., general thoughts to the effect that science only produces models of reality and never truly represents reality as it really is; then causation would be model-relative, too.

However, this is not what I have in mind. The point is quite specific: The RT definition 1 refers, in a way I had explained in point 7, to the obtaining circumstances, however

only insofar as they are represented in the given frame **U**. This entails a genuine frame-relativity of causation as such; $\{X = x\}$ may be a (direct) cause of $\{Y = y\}$ within one frame, but not within another or more refined frame. As Halpern, Hitchock (2010, Section 4.1) argue, this phenomenon may also show up within the SM approach.

I do not think that this agrees with Pearl's intention in pursuing the SM account; an actual cause should not cease to be an actual cause simply by refining the frame. Perhaps, the intention was to arrive at a frame-independent notion of causation by assuming a frame-independent notion of intervention. My translation of the intervention $X_1 \leftarrow x_1$ into conditionalization referred to the past (or the ancestors or the non-descendants) of X_1 as far as they are represented in the given frame **U**, and thus reproduced only a frame-relative notion of intervention. However, the intention presumably is to refer to the entire past of X_1 absolutely, not leaving any hole for the supposition of $\{X_1 = x_1\}$ to backtrack. If so, there is another sharp difference between the SM and the RT approach with repercussions on the previous point.

Of course, I admit that our intuitive notion of causation is not frame-relative; we aim at an absolute notion. However, this aim bars us from having a reductive analysis of causation, since the analysis would have to refer then to the rest of the world, as it were, to many things outside the frame that are thus prevented from entering the analysis. In fact, any rigorous causal theorizing is thereby frustrated in my view. For, how can you theoretically deal with all those don't-know-what's? For this reason I always preferred to work with a fixed frame, to pretend that this frame is all there is, and then to say everything about causation that can be said within this frame. This procedure at least allows a reductive analysis of a frame-relative notion.

How, then, can we get rid of the frame-relativity? I propose, by ever more fine-graining and extending the frame, studying the frame-relative causal relations within all these well-defined frames, and finding out what remains stable across all these refinements; we may hope, then, that these stable features are preserved even in the maximally refined, universal frame (cf. Spohn forthcoming, section 14.9; for Halpern, Hitchcock (2010, Section 4.1) this stability is also crucial). I would not know how else to deal with the challenge posed by frame-relativity, and I suspect that considerable problems in causal theorizing result from not explicitly facing this challenge.

(15) The various points may be summarized in the final opposition: whether causation is to be *subjectivistically* or *objectivistically* conceived. Common sense, Judea Pearl, and many others are on the objectivistic side: "I now take causal relationships to be the fundamental building blocks both of physical reality and of human understanding of that reality" (Pearl 2000, pp. xiiif.). And insofar as structural equations are objective, the SM approach shares this objectivism. By contrast, frame-relativity is an element of subject-relativity; frames are chosen by us. And the use of only epistemically interpretable ranking functions involves a much deeper subjectivization of the topic of causation. (The issue of relevance, point 12, is related, by the way, since in my view only epistemic relevance is rich enough a concept.)

The motive of the subjectivistic RT approach was, I said, Hume's challenge. And the gain, I claimed, is the feasibility of a reductive analysis. Any objectivistic approach has to tell how else to cope with that challenge and how to make peace with non-reductivity. Still, we cannot simply acquiesce in subjectivism, since it flies in the face of everyone keeping some sense of reality. The general philosophical strategy to escape pure subjectivism has been aptly described by Blackburn (1993, part I) as Humean projectivism leading to so-called quasi-realism that is indistinguishable from 'real' realism.

This general strategy may be precisely explicated in the case of causation: I had indicated in the previous point how I propose to get rid of frame-relativity. And in Spohn (forthcoming, ch. 15) I develop an objectification theory for ranking functions, according to which some ranking functions, the objectifiable ones, may be said, to truly (or falsely) represent causal relations. No doubt, this objectification theory is disputable, but it shows that the subjectivistic starting point need not preclude us from objectivistic aims. Maybe, though, these aims are more convincingly served by approaching them in a more direct and realistic way, as the SM account does.

4 Conclusion

On none of the fifteen differences above could I seriously start discussion; obviously nothing below book length would do. Indeed, discussing these points was not my aim at all, let alone treating anyone conclusively (though, of course, I could not hide where my sympathies are). My first intention was simply to display the differences, not all of which are clearly seen in the literature; already the sheer number is surprising. And I expressed my second intention between point 3 and point 4: namely to show that there are many internal theoretical issues in the theory of causation. On all of them one must take and argue a stance, a most demanding requirement. My hunch is that those theoretical considerations will eventually override issues of exemplification and application. All the more important it is to take some stance; no less will do for reaching a considered judgment. Judea Pearl has paradigmatically shown how to do this. His brilliant theoretical developments have not closed, but tremendously advanced our understanding of all these issues pertaining to causation.

Acknowledgment: I am indebted to Joe Halpern for providing most useful comments and correcting my English.

References

Blackburn, S. (1993). *Essays in Quasi-Realism*, Oxford: Oxford University Press.

Cartwright, N. (1979). Causal laws and effective strategies. *Noûs 13*, 419-437.

Glymour, C. (2004). Critical notice on: James Woodward, Making Things Happen, *British Journal for the Philosophy of Science 55*, 779-790.

Goldszmidt, M., and J. Pearl (1992). Rank-based systems: A simple approach to belief revision, belief update, and reasoning about evidence and actions. In B. Nebel,

C. Rich, and W. Swartout (Eds.), *Proceedings of the Third International Conference on Knowledge Representation and Reasoning*, San Mateo, CA: Morgan Kaufmann, pp. 661-672.

Halpern, J. Y., and C. Hitchcock (2010). Actual causation and the art of modeling. This volume, chapter 22.

Halpern, J. Y., and J. Pearl (2005). Causes and explanations: A structural-model approach. Part I: Causes. *British Journal for the Philosophy of Science 56*, 843-887.

Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy 98*, 273-299.

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

McDermott, M. (1995). Redundant causation. *British Journal for the Philosophy of Science 46*, 523-544.

Meek, C., and C. Glymour (1994). Conditioning and intervening. *British Journal for the Philosophy of Science 45*, 1001-1021.

Reichenbach, H. (1956). *The Direction of Time*. Los Angeles: The University of California Press.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann.

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

Spirtes, P., C. Glymour, and R. Scheines (1993). *Causation, Prediction, and Search*. Berlin: Springer, 2nd ed. 2000.

Spohn, W. (1978). *Grundlagen der Entscheidungstheorie*, Kronberg/Ts.: Scriptor. Out of print, pdf-version at: http://www.uni-konstanz.de/FuF/Philo/Philosophie/philosophie/files/ge.buch.gesamt.pdf.

Spohn, W. (1983). *Eine Theorie der Kausalität*. Unpublished Habilitationsschrift, University of München, pdf-version at: http://www.uni-konstanz.de/FuF/Philo/Philosophie/philosophie/files/habilitation.pdf.

Spohn, W. (1990). Direct and indirect causes. *Topoi 9*, 125-145.

Spohn, W. (2006). Causation: An alternative. *British Journal for the Philosophy of Science 57*, 93-119.

Spohn, W. (forthcoming). *Ranking Theory. A Tool for Epistemology*.

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

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