Logique & Analyse **216** (2011), 517–536

IS NON-REDUCTIVE CONCEPTUAL ANALYSIS A META-PHILOSOPHICAL PROBLEM FOR THEORIES OF CAUSATION?

ALEXANDER REUTLINGER

1. *Introduction*

In the empiricist tradition in the philosophy of science, causation has been viewed as highly problematic. The concept of causation has usually been conceived as at least unclear in meaning. More extreme views have suspected that the concept of causation plainly contradicts the sciences, because (a) 'causation' has been understood to refer to some kind of metaphysical powers which are not the subject of modern science, and (b) properties of causal relations, such as asymmetry, are not shared by equations in current physics. Nevertheless, empiricist criticisms of the concept of causation are quite different in character and so are their consequences for philosophers who are interested in causation. Let me illustrate this point by introducing two famous critiques by (1) Bertrand Russell and (2) Rudolf Carnap.

(1) According to Russell's challenging classic paper *On the Notion of Cause*, philosophers are plainly wrong to suppose that the concept of causation is used at all in the sciences — to be precise, philosophers are wrong to assume that it is used in *contemporary physics*.[1] Rather the concept of causation should be understood as 'a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm'.[2] In the upshot Russell recommends to eliminate the concept of causation from philosophy of science because nothing in contemporary physics satisfies the description of a cause.[3]

(2) But even if Russell's claim were true with respect to physics, one may still wonder whether it is true for other sciences. In the so-called special sciences (such as biology, psychology and economics) researchers *do* employ

[1] For similar influential criticisms see Mach (1980: 278), (1982: 459), (1900: 435f).

[2] Russell (1912: 1).

[3] See Ladyman & Ross (2007) and Norton (2007: 34) for claims in this Russellian vein.

causal notions in their work.[4] If one accepts that at least in the non-physical sciences causal notions are commonly used, then philosophers of science have to meet the following challenge that Carnap describes in his *An Introduction to the Philosophy of Science*:

> One of the most important tasks of philosophy of science is to analyze the concept of causality and to clarify its meaning. [...] What does a scientist mean when he says that event b was caused by event a? He might mean that event a 'brought about' or 'produced' event b. So you see, when he tries to explain the meaning of 'cause', he falls back to such phrases as 'bringing about', 'bring forth', 'create', and 'produce'. Those are metaphorical phrases, taken from human activity. [...] It is not a very satisfying answer to the question: 'What does it mean that one event caused the other?'[5]

Carnap's view on the problems of causation clearly differs from Russell's claims about causation, because Carnap accepts that 'causation' is a part of scientific language. At the same time Carnap maintains an empiricist suspicion against the concept of causation: The concept of causation is too unclear to be accepted as a primitive concept. Instead 'causation' should be analyzed. In other words, the philosophical project of analyzing the concept of causation arises from (a) the unclear meaning of the concept of causation, (b) the fact that this concept is commonly used in scientific practice and (c) the idea that philosophy has the duty to clarify imprecise scientific notions by the means of conceptual analysis (in the sense of a more precise definition of the meaning of a concept). Note that the analysis of the meaning of causation has to fulfill a certain requirement: the analysis has to be *reductive* in order to be illuminating. Analyzing the concept of causation in a reductive way consists in providing necessary and sufficient truth condition for causation in *acausal* terms, i.e. by using a vocabulary that is free of any causal notions. The metaphors of bringing about and production that Carnap mentions clearly fail to satisfy the requirement of a reductive analysis of the concept of causation. The vocabulary, which is used to analyze the concept of causation reductively, contains terms that are unproblematic (or at least better understood than the *analysandum* itself). In the empiricist tradition such terms are, for instance, fact, event, law of nature, probabilistic

[4] See Williamson (2005) and Hitchcock (2007) for statistics of the use of causal notions in scientific journals.

[5] Carnap (1966: 189).

dependence etc. In this sense, analyzing a concept means to give its truth conditions in terms of *more fundamental* concepts.

In this paper I will argue as follows: In section 2 I will present David Lewis's counterfactual theory of causation. It is paradigmatic in two ways: (1) The counterfactual theory of causation is the current orthodoxy in the debate on causation and the received view of causation in other areas of philosophy (e.g. in philosophy of mind); (2) the counterfactual theory of causation is also a paradigmatic example of a reductive analysis of the concept of causation. In section 3, I will provide two examples of an analysis of causation that is *non-reductive* because both approaches use causal notions in the *analysans*: Nancy Cartwright's and James Woodward's analysis of causation. Further, both non-reductive approaches are widely believed to satisfy two standards of an adequate conceptual analysis of causation successfully. More importantly, these reductive analyses seem to satisfy the standards of adequacy even better than Lewis' reductive theory. But how can that be? Is it not obvious to object to any non-reductive analysis that it is circular? In section 4 I will consider the objection that any non-reductive definition of causation is trivial because it is viciously circular and, thus, it is of no use. I will counter this objection by defending the thesis: We can happily accept non-reductive theories of causation. I defend this thesis in the following way: In section 5 I reconstruct Woodward's argument for the claim that his analysis of causation is not viciously circular although it is non-reductive; on my opinion, the argument also applies to Cartwright's theory of causation. In the final section 6, I will present eight additional reasons to support Woodward's defense of a non-reductive analysis of causation.

2. *A Reductive Analysis of Causation: Lewis' Counterfactual Theory*

The currently most influential analysis of causation is David Lewis' counterfactual theory of causation.[6] Lewis analyzes causation in terms of counterfactual dependence between propositions that state the occurrence of distinct events (let A be the proposition that event c occurs, and let B be the proposition that event e occurs). According to Lewis, event c causes event e iff

> (1) Either B counterfactually depends on A, i.e. the following two counterfactual conditionals are true:
>
> (a) If A were the case, then B would be the case as well.
>
> (b) If A were not the case, then neither would B be the case.

[6] See Lewis (1973b) and (2004).

> (2) Or there is a chain of counterfactual dependencies leading from A to event B via propositions that express the occurrence of further intermediary events which are spacio-temporally located between the events c and e.

Lewis' counterfactual theory of causation is reductive because it makes no use of causal terms. Merely the acausal notion of counterfactual dependence is used. But suppose one might object:

> Well, the definition of causation itself does not use causal notions. But what about the truth conditions for counterfactuals? Is it not the case that one needs causal terms or, at least, causal intuitions in order to evaluate counterfactuals?

According to Lewis, this is not the case. Lewis' account of truth conditions for counterfactuals meets the conditions of a reductive analysis as well as his theory of causation.

> Lewisian Semantics: The counterfactual *if ϕ were the case, then ψ would be the case*[7] is (non-vacuously) true at a possible world w if and only if there is some possible world u, where ϕ and ψ are both true, and u is closer to w than any possible world v, where ϕ is true but ψ is false.[8]

The relation of closeness is spelled out in terms of similarity between possible worlds. The criteria of similarity consist in shared laws of nature and the amount of shared facts (or property instantiations in space-time regions, as Lewis puts it) in each world. 'Facts' and 'laws of nature' are part of the vocabulary that is approved by proponents of reductive analysis.[9] Therefore Lewis' analysis of causation including his semantics for counterfactual conditionals obeys the rules of a reductive methodology.

[7] Italics indicate expressions mentioned in the meta-language.

[8] See Lewis (1973b: 164) and (1973a: 16). According to Lewis (1973a: 24–26), a counterfactual is *vacuously true* iff its antecedent expresses an impossible proposition. This might be a controversial claim. But it need not concern us here, because causal claims have to express contingent propositions.

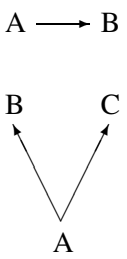
[9] See Lewis (1979: 47f).

3. *Non-Reductive Approaches to Analyzing 'Causation'*

In the debate on causation, the adequacy of an analysis is judged with respect to two standards:

(1) Does the analysis apply to paradigmatic causal claims in the sciences (and also to causal claims in everyday contexts)?

(2) Does the analysis distinguish correctly between intuitively different causal structures? For instance, does the analysis distinguish correctly between (i) causal chains and (ii) conjunctive forks as illustrated by the following causal graphs?[10]



(i) structure of a causal chain

(ii) structure of a conjunctive fork

Both standards have caused severe problems for Lewis's counterfactual theory of causation.[11] Of course, the failure to meet these standards in some cases is no knock-down argument against the counterfactual theory. The counterfactual theory might be improved by modifications in order to deal with counterexamples.[12]

However, independently of how the prospects of Lewis's counterfactual theory are, there is a problem: Some theories of causation in the current debate appear to be more successful in meeting the standards of adequacy of an analysis than Lewis' reductive analysis. It is a characteristic of these theories that they are *non-reductive*, because they make use of causal notions in the *analysans*. Let me illustrate the non-reductive analysis of causation by two widely acknowledged examples: The accounts of Nancy Cartwright and James Woodward.

[10] This example has been decisive in the debate: The failure of distinguishing causal chains and conjunctive forks has been one reason against regularity theories of causation. See Spohn (2006).

[11] See Collins, Hall & Paul (2004b) for a detailed list of counterexamples to Lewis' counterfactual theory of causation.

[12] Most papers in Collins, Hall & Paul (2004a) are attempts to do precisely this.

Example 1: Nancy Cartwright's Theory of Causation

In *Causal Laws and Effective Strategies* Cartwright analyzes causation (on the type level):[13]

> 'C causes E' if and only if C increases the probability of E in every situation which is otherwise *causally homogeneous* with respect to E.[14]

A situation is 'causally homogeneous with respect to E' if other causes $\{U_i\}$ of E are (experimentally or hypothetically) held fixed such that the other causes $\{U_i\}$ of E are not statistically relevant for C. Cartwright motivates the need for information about the other causes of E by the following kind of example.

Suppose that smoking (C) causes lung cancer (E). Suppose further that exercising (U) prevents lung cancer. If one does not hold U constant (e.g. by examining only populations of smokers who exercise), then it could be the case that smoking does not increase the probability of lung cancer *although* smoking does cause lung cancer. This situation might occur if Prob(no lung cancer | exercising) > Prob(lung cancer | smoking). Consequently, a situation with such a probability distribution is a counterexample to a naïve probabilistic theory of causation which analysis causes as mere probability raisers for the occurrence of the effect. Cartwright concludes that one cannot analyze the concept of causation without referring to information about other causes of the effect E.[15] For the reason that Cartwright's theory of causation essentially refers to other causes of the effect E, her analysis is clearly non-reductive.

[13] Notice that Lewis deals with actual causation, i.e. the relata of a causal relation are events. Cartwright and Woodward focus on type level causation, i.e. the causal relata are random variables. Nevertheless their theories can also account for actual causation (see Woodward 2003: 74–85; Halpern & Pearl 2005: 583). But this difference is a minor point here, because the question whether one is tied to a reductive methodology applies to the analysis of actual as well as of type level causation.

[14] Cartwright (1983: 25).

[15] Cartwright (1983: 23f). The general problem behind this example is Simpson's Paradox, i.e. 'any association [...] between two variables which holds in a given population can be reversed in the sub-populations by finding a third variable which is correlated with both' (Cartwright 1983: 24).

Example 2: Woodward's Interventionist Theory of Causation

Interventionists like Woodward claim to advocate a (modified) counterfactual theory of causation, because their definitions of causation depend essentially on counterfactual conditionals.[16] In this respect the interventionist project is close to Lewis' counterfactual theory of causation. For instance, take Woodward's definition of a direct type-level cause that is formulated in terms of random variables[17] X and Y (being elements of a set V of random variables) and their possible values $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_k\}$:[18]

> 'A necessary and sufficient condition for X to be a direct cause of Y with respect to some variable set V is that there be a possible *intervention* on X that will change Y (or the probability distribution of Y) when *all other variables* are held fixed at some value by intervention.'[19]

What is the counterfactual involved in this definition of a direct cause? It is what one might call an interventionist counterfactual.[20] Such interventionist counterfactuals, which Woodward also calls 'active'[21] counterfactuals, are of the following form:

[16] See Hitchcock (2001), Woodward (2003), Halpern & Pearl (2005).

[17] The use of 'variable' in the sense of a random variable in the literature on causation and probability theory is not to be confused with its use in first order predicate logic, i.e. in the sense of an individual constant. In the first case variables refer to properties. In the second case they refer to individuals. See, for instance, Sinai (1992: 5) for the algebra that defines random variables.

[18] Usually the number of values contained in the set is considered to be finite. These values are required to be *mutually exclusive and exhaustive*. The values associated with a variable X are exclusive, if X can only have one value (at a time). They are exhaustive, if the variable has to take one of its possible values. Let me illustrate this by an example. Suppose that Mary's income of September 2008 is either high or low, i.e. the set of values associated with the income variable I is {high; low}. The set of values is exclusive, because Mary cannot *both* have high income in September 2008 and a low one as well. The set is exhaustive if — as we suppose in the example — Mary's income has to be either high or low, and there is no further value, that it might take, e.g. "medium" income.

[19] Woodward (2003: 55), my italics highlight causal notions in the *analysans*.

[20] Woodward (2003: 15).

[21] Woodward (2000: 199); see also his (2003: 122) for the same point.

> If the value of X were changed to be x_i by an intervention, then the value of Y would change to y_i.

Interventionists think that the antecedent proposition is made true or is the outcome of an intervention into the variable(s) referred to in the antecedent (I will say more about the notion of intervention in the next paragraph). However, there is a striking difference between Lewisian counterfactual theories and interventionist counterfactual theories of causation. Although the interventionist theory of causation is a counterfactual approach, the interventionist definition differs from the orthodox Lewisian counterfactual theory of causation (and various regularity theories of causation) in the methodological constraints on conceptual analysis. It is not *reductive* because it involves two kinds of causal concepts: (1) Interventions and (2), in the same way as Cartwright, information about other causes of the effect Y. Any definition with such features is clearly at odds with the reductive methodology. Because the presented interventionist definition of causation makes use of causal notions in the *analysans* it is clearly not reductive. Let me explicate the first kind of causal notion — i.e. interventions — in a more detailed manner, because it is an genuine feature of Woodward's approach. As already mentioned Woodward's idea of information about other causes ('*all other variables* are held fixed at some value by intervention'[22]) is close, if not equivalent, to Cartwright's 'causally homogeneous situation'.

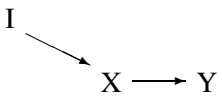
Let me elucidate the definition of a direct type-level cause by introducing the notion of intervention. I will explain (a) why one needs interventions and (b) what is precisely meant by an intervention.

(a) The motivation to talk about interventions is strongly tied to counterfactuals. Interventions are (experimental or merely hypothetical) means to bring about or to imagine the situation that is expressed in the antecedent of a counterfactual. For instance, suppose that the antecedent of a counterfactual expresses the proposition that the inflation rate in the Netherlands in 2008 is 2.5%. The idea of an intervention is to set the variable representing the inflation rate in the Netherlands on the value 2.5%. An intuitive way to understand an intervention is to think of it as a command like 'suppose that the inflation rate is 2.5%!'.

(b) The basic idea of an intervention consists in a local change, which *sets* a variable to a certain value. This local change is called an intervention or, more precisely, an intervention into the value of a variable. Roughly, an intervention is an exogenous causal influence on some endogenous variable

[22] Woodward (2003: 55), my italics highlight causal notions in the *analysans*.

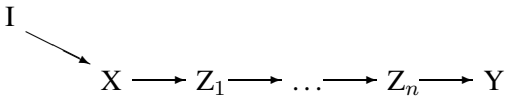
X.[23] In a causal graph this exogenous causal influence on X can be represented by an additional exogenous random variable I having the possible values $\{i_1, \ldots, i_n\}$. Suppose the case that X causes Y, because $\mathrm{Y} = y_1$ counterfactually depends on $\mathrm{X} = x_1$. That X takes the value x_1 — the antecedent of an interventionist counterfactual — is itself the effect of another cause: it is the effect of the intervention variable I.



Since notion of an intervention has to meet a number of formal conditions, it is a technical term.[24] Some possible value i_i of I, i.e. the event statement $I = i_i$, is an intervention into the variable X relative to Y iff the following condition are fulfilled:

(1) $I = i_i$ directly causes $\mathrm{X} = x_i$. I having one of its possible values i_i is defined as a direct cause of X iff $I = i_i$ changes the value of X while all other variables are held constant.[25]

(2) $I = i_i$ is an indirect cause of $\mathrm{Y} = y_1$ such that there is a directed path leading from I through X to Y. This does imply the possibility of a set of variables $\mathsf{Z} = \{\mathrm{Z}_1, \ldots, \mathrm{Z}_n\}$, which is intermediate between X and Y. This situation can be presented graphically as follows.

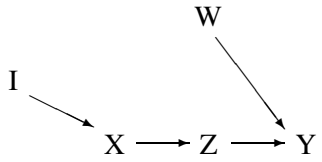


[23] It is common practice to partition the set of variables V into a set of *endogenous* variables E and a set of *exogenous* variables U, i.e. the set V is the union set of all endogenous and exogenous variables. Exogenous variables are taken as given in the model, because their values are not described as being caused by other events. They have no causal predecessors in the graph. Rather, the role they play is restricted to being the causes of endogenous variables. In contrast, each endogenous variable does have a cause in the graph. Note that there is a clear distinction between exogenous and endogenous variables *within* a given graph. But there is no principled distinction between endogenous and exogenous variables relative to different graphs: One variable might be considered as endogenous in graph G and as exogenous in another graph G*. Whether a variable is endogenous or not depends mostly on the pragmatic decisions of modeling scientists.

[24] See Woodward (2003: 98), Hitchcock & Woodward (2003a: 12f).

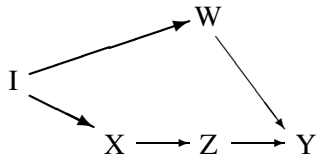
[25] See Woodward (2003: 42 and 55) for a more detailed definition of a direct cause.

(3) I is causally independent of other causal paths along a set W of variables to Y, which do not go through X. The following graph illustrates a situation where X is causally independent of W.

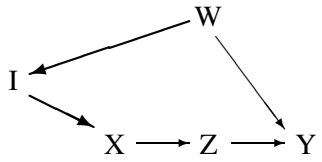


In other words, condition (3) is supposed *to rule out* the following causal relations (the relevant causal arrows are highlighted as thick arrows):[26]

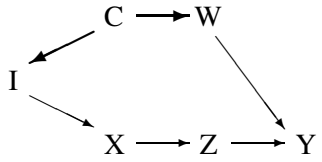
(i) I is a common cause of W and X.



(ii) I is an effect of W, i.e. W is an indirect cause of X.



(iii) I is an effect of common cause C of W and I.



Any value i_i of I, which satisfies the above stated three conditions, deserves to be called an intervention on X relative to Y.

[26] See Woodward (2003: 99–102).

Let me sum up the result of this section so far. The interventionist definition of a direct type-level cause involves counterfactuals whose antecedent is conceived as the outcome of an intervention. Interventions are formal recipes for imagining or bringing about a situation that is described by the antecedent of a counterfactual.

Obviously, Woodward's definition is non-reductive, i.e. causation is not defined by using acausal terms. Rather, it is something that one might call a *non-reductive definition* of causation. Woodward describes his own project as clarifying the meaning of causal notions:

> 'My enterprise is, roughly, to provide an account of the meaning or content of just those qualitative causal notions that Pearl (and perhaps Spirtes et al.) [whose projects are to discover causal relations] take as primitive.'[27]

According to the theory of meaning presupposed by Woodward, the meaning of a term is determined by its truth conditions. Lewis and Woodward agree on truth conditional semantics in order to determine the meaning of a concept. But they disagree on the question whether these necessary and sufficient truth conditions for a causal notion can be provided in causal terms.

Besides the tradition of providing necessary and sufficient truth conditions in a reductive way, there is also a non-reductive tradition. Along the lines of this tradition an informative conceptual analysis relates important concepts, such as 'intervention' and 'direct cause' without pointing out a class of more fundamental concepts. In the literature one finds many philosophers sympathetic to such a methodology. For instance, Strawson distinguishes a reductive way to analyze concepts from a 'connective', i.e. non-reductive, way.[28] A contemporary ancestor of Strawson's connective analysis is the idea of the theoretical role played by a concept as advocated by functionalists in the philosophy of mind and philosopher's engaged in the *Canberra Plan*, most importantly by Jackson (1998). To determine the theoretical role of a concept allows loosening the reductive methodology. Let me briefly sketch Jackson idea of the theoretical role of concepts. According to Jackson,[29] it is the goal of conceptual analysis to explicate the role of a concept

[27] Woodward (2003: 38).

[28] See Strawson (1992: Chapter 2); and also Glock (2003: 115, 244).

[29] See Jackson (1998: chapter 2). Hitchcock (2006: 450) also indicates that Jackson's theoretical role view is a proper methodological reflection of non-reductive analysis.

in a (folk or scientific) theory. For instance, Jackson takes 'fish' to be analyzed in the following way: 'x is a fish iff x has the important properties out of or descended from or explanatory of F_1, F_2, F_3, ..., according to the best true theory.'[30] Describing the meaning of a concept by its theoretical role is not necessarily reductive. A reason to actually accept a non-reductive analysis of a concept can be due to the nature of the theory which the concept belongs to: If 'direct cause' is part of a theory that also employs other causal concepts, such as 'intervention', then determining the role of 'direct cause' also involves other causal notion. This view seems to fit Woodward's description of his own project: The analysis of a causal concept has 'to provide an account of the meaning [...] of just those qualitative causal notions that [methodologists of causal inference] take as primitive.'[31] One could read Woodward from a Jacksonian point of view in the following way: The 'best true theories' (Jackson) involved in the case of causal concepts are methods of causal modeling in the sciences.[32]

4. *The Objection Against Non-reductive Analysis*

The objection that might arise against a non-reductive definition is obvious:

> A non-reductive definition of causation is trivial because it is viciously circular and, thus, it is of no use.[33]

This is a serious objection that has to be refuted in order to maintain theories of causation à la Cartwright and Woodward. But is this objection justified? Is Cartwright's theory and the interventionist notion of cause really worthless, because it uses causal notions to detect whether there is a causal relationship between two distinct events? Couldn't we accept a non-reductive definition, because it tells us something interesting?

Let me explore an answer that simply accepts that a good definition may be non-reductive: I call it the *Easy Answer*. In order to defend the Easy Answer, I will, in a first step, explore Wodward's argument against the objection that

[30] Jackson (1998: 35).

[31] Woodward (2003: 38).

[32] For a detailed survey of the methods of causal modelling the social sciences see Russo (2009).

[33] See, e.g., Psillos (2002: 104f., 182f) and Psillos (2007: 99).

non-reductive definitions are always viciously circular (in section 5). Secondly, I will provide additional eight reasons to be happy with accepting the Easy Answer (in section 6).

5. *Woodward's Defense of Non-Reductive Analysis*

Woodward defends the non-reductive analysis of causation by the following argument: the analysis is *non-reductive*, but it is — contrary to the objection raised in section 4 — not viciously circular or a *petitio principii*. Let me explain what is meant by the distinction between being non-reductive and being viciously circular. We could, as Woodward[34] does himself, insist that circularity is nothing bad at all, if it only leads to admitting the fact that we are not capable of providing a completely reductive analysis of causation. A reductive analysis deletes all causal notions from the *analysans*. But also a non-reductive analysis may reveal (or help to reveal) interesting conceptual connections between the notions of causation, intervention, law, explanation, stability of counterfactuals etc.[35] *The crucial point is that we can be non-reductive without being viciously circular.* For instance, Woodward's definition of direct causation says roughly that X causes Y iff Y is counterfactually dependent on X when we change the value of X by an intervention and all other variables are held fixed. We may willingly admit that 'to intervene' is itself a causal notion. But it is by no means true that we have *to presuppose* that X causes Y in order to spell out what is meant by 'X causes Y'. Doing so would indeed be viciously circular. I think that Woodward has a good point here.

In addition to Woodward's defense, many more arguments for the acceptability of non-reductive definitions can be presented. Let me present eight reasons for the easy answer in the following section.

6. *Eight Reasons for Embracing Non-Reductive Conceptual Analysis*

Here are some reasons to be happy with a non-reductive analysis of causation and, thus, to accept the Easy Answer without philosophical grief.

Reason 1: Although the definition by Woodward and Cartwright are non-reductive, we *intuitively* think they are informative and that we learn something about causation. Intuitively they are not trivial.

[34] See Hitchcock & Woodward (2003b: 197).

[35] See Woodward (2003: 103f).

Reason 2: The charge of being viciously circular usually means that one commits a *petitio principii*, i.e. one already presupposes the concepts and statements that have to be the shown in the analysis of a concept. But although our interventionist definition is non-reductive, it is no *petitio principii*. The definition of *X (directly) causes Y* does not presuppose that X (directly) causes Y. If so, the non-reductive analysis by Cartwright and Woodward can counter the objection that it is trivial because it is viciously circular. It is not viciously circular because it does not presuppose the concept to be analyzed. What is really the case is: The non-reductive analysis does presuppose *other* causal notions than the analysans. The notion of a *petitio principii* is supposed to clarify Woodward's argument as presented in section 5. Woodward's rather vague claim that a non-reductive analysis provides interesting connections among different causal concepts can now be replaced by a more precise formulation: A non-reductive analysis of a causal concept provides necessary and sufficient truth conditions for this concept by (a) relying of other causal concepts and (b) avoiding a *petitio principii* with respect to the *analysandum*. Note that Cartwright and Woodward agree with philosophers who accept the reductive methodology that any alleged analysis which is viciously circular is automatically ruled out. A viciously circular analysis is not tested whether and how it meets the standards of adequacy for an analysis. It is no analysis at all.

Reason 3: Although Woodward's 'direct cause' (the *analysandum*) and 'intervention' (part of the *analysans*) are both causal concepts, they do not have the same intension. Being a direct cause is defined as 'that there be a possible intervention on X that will change Y (or the probability distribution of Y) when all other variables are held fixed at some value by intervention.'[36] An intervention is defined as (1) a direct cause of X and (2) as an indirect cause of Y that is (3) causally independent of other causes of Y (that are not a path between X and Y). These definitions are obviously not equivalent. Hence, 'direct cause' and 'intervention' have different meanings.

Based on these considerations we can develop an argument against the thesis that the non-reductive analysis is trivial. If both sides of a biconditional use (causal) terms with different intension, then the biconditional cannot be trivially true. Our examples for the non-reductive analysis of causation use different (causal) terms on the RHS and the LHS of the biconditional. Thus, the non-reductive analyses in question are not trivially true.

Reason 4: Suppose we reject a non-reductive analysis in the case of causation for purely methodological reasons. In other words, we accept a reductive constraint on conceptual analyses. Suppose further that we do not want to restrict this reductive methodology to the analysis of causation: If

[36] Woodward (2003: 55), my italics.

this methodology is justified for the analysis of causation, then it is also justified in other areas of philosophy. But there are well-established non-reductive analyses in other areas of philosophy (e.g. for semantic concepts such as meaning and truth, for epistemic concepts such as knowledge and belief, for moral concepts such as good and virtue, etc.). For instance, take the following examples of conceptual analysis from philosophy of mind and philosophy of probability:

- *Mental States and their causal role*: Something is, e.g., a belief iff it is typically caused by perceptions and itself typically causes actions.
- *Chance*: Prob(A) = 0.8 is the chance of a proposition A to occur iff an ideally rational agent's degree of belief that A will occur (conditional on the history of agent's world) equals 0.8.
- *Hoping that p*: If there is something such that (i) someone desires that p, (ii) someone does not know that p and (iii) believes that it is possible that p, then someone hopes that p.

These examples of analysis use notions of the same family of mental and probabilistic notions in the *analysans* and in the *analysandum*. Thus, they are examples of a non-reductive analysis. It is obvious that there is a tension between holding the reductive methodology true and accepting well-established examples of non-reductive analyses. We face two alternatives in order to resolve this tension:

(i) One may maintain the reductive methodology and reject the non-reductive analyses altogether (whether not they deal with causation);
(ii) One might reject the reductive methodology and maintain various well-established non-reductive definitions.

Alternative (i) appears to have higher costs than (ii). If we opted for (i), then a lot of philosophers in other areas of research besides causation would have to pay a high price. Maybe it's not worth paying it — maybe we should be less restrictive in our methodology by accepting alternative (ii).

Reason 5: Suppose one accepts non-reductive conceptual analyses. Is it true that one has to give up the *preference* for reductive analysis? Certainly not. In defense of non-reductive analysis, one merely has to claim that there are standards of adequacy that a non-reductive analysis might meet in a better way than a reductive analysis. Those standards consist in — as introduced above — (1) reconstructing paradigmatic examples of causal claims in the sciences and (2) distinguishing causal structures correctly. Nevertheless a proponent of non-reductive analysis can allow for the preference for a reductive analysis in the following way: Were these standards of adequacy

equally met by a reductive as well as by a non-reductive analysis, one would choose the reductive analysis. One reason not to buy Lewis' counterfactual theory is the fact that it struggles with the second standard (especially with respect to various scenarios of preemption).[37]

Reason 6: Having Reason 5 in mind one could ask: How do we test theories of causation? Again, we do so by asking whether and how the standards of adequacy are fulfilled. Being reductive or non-reductive does not seem to matter for passing these tests. Various non-reductive approaches can distinguish causal structures more successfully than, e.g., the Lewisian counterfactual theory of causation.[38] Thus, they are to be preferred on these grounds.

Reason 7: The reason for accepting a non-reductive analysis of causation is due to the standard of reconstruction paradigmatic causal claims in scientific practice. Cartwright and Woodward define causation relative to models that represent the relevant causal factors for a phenomenon (e.g., lung cancer). In the sciences one finds very restricted (i.e. abstract and/or idealized) models of phenomena (for instance, the Mundell-Flemming model or the Solow-Model in economics).[39] It is a characteristic of these models to include information about other relevant factors besides the causal factors X and Y in question. I think it is an advantage of non-reductive analyses to capture these properties of scientific representation.

Notice that Lewis' reductive approach seems to have to pay a price at this point: Lewis cannot refer straightforwardly to causal information about other relevant factors. Instead Lewis substitutes information about other causal factors by a rather complicated and vague similarity ranging over *acausally described* possible worlds. (See Lewisian Semantics in section 2.)

Reason 8: Some proponents[40] of Bayesian Nets correctly observe that there are (at least) two kinds of theories of causation: (1) the methodology of causal inference and (2) the conceptual analysis of causation.

These philosophers and scientists claim that these two theories of causation do not have equal rights: They think that algorithms for causal inference

[37] See Collins, Hall & Paul (2004b) for a detailed list of counterexamples (most importantly various scenarios of preemption and probabilistic causation). See Cartwright (1983), Hitchcock (2001), Woodward (2003) and Halpern & Pearl (2005) for non-reductive account that propose convincing treatments for those counterexamples to Lewisian counterfactual theories of causation.

[38] See Fn. 35.

[39] See, e.g., Frigg & Hartmann (2006) and Frigg (forthcoming).

[40] See Scheines (1997), Spirtes, Glymour & Scheines (2000) and Reiss (2006).

should *replace* the time-honored philosophical enterprise of conceptual analysis. Following this idea would rule out reductive as well as non-reductive analysis.

But we may disagree that these projects could replace one another: Discovering causes and clarifying concepts (that might be used in the methodology of causal inference) seem to be entirely different projects. Techniques of causal inference are part of methodology to discover causal relations and — reductive as well as non-reductive — conceptual analysis belong to philosophy. Moreover, methodology of discovery and conceptual analysis do not seem to be substitutes. Their relationship is better understood as complementary. On the one hand, Woodward describes the relation of analysis and causal inference adequately as: It is the task of analysis 'to provide an account of the meaning or content of just those qualitative causal notions that Pearl (and perhaps Spirtes et al.) take as primitive.'[41] The project of researchers as Pearl and Spirtes, Glymour and Scheines is to develop methods in order to discover causal relations. Coherence with methods of causal inference is required for any analysis of causation by the second standard of adequacy (i.e. reconstructing paradigmatic causal claims from scientific practice). On the other hand, the clarification of the central concepts of causal modeling might be of use for the methodologists. Thus, it seems that both theories of causation can peacefully and productively coexist.

There is also some evidence for the peaceful and productive coexistence from the debate on both theories of causation: Some philosophers and scientists engage in methodology of causal inference *as well as* in conceptual analysis.[42]

However, distinguishing two theories of causation and believing that they can peacefully and productively coexist does neither commit us to the reductive nor to the non-reductive brand of conceptual analysis. On the other hand, if one gives up analysis and replaces it by methods of causal inference, then the issue of a reductive or non-reductive methodology for analysis does naturally not arise. Either way methods of causal inference cannot directly decide whether one should accept non-reductive analysis.

To sum up this section, I have provided eight reasons for the easy answer to the objection presented in section 4. The most important result is: Non-reductive analysis can be defended against being viciously circular. Whether one is satisfied with a non-reductive analysis à la Woodward and Cartwright

[41] Woodward (2003: 38).

[42] See Pearl (2000), Halpern & Pearl (2005) and Spohn (2006).

depends on the same standards that are applied to a reductive analysis of causation.

7. *Conclusion*

I started from Russell's and Carnap's concerns about the unclear concept of causation. Russell's claim that nothing in scientific practice satisfies the description of being a cause seems too strong, because causal vocabulary is used at least in the special sciences. Carnap's strategy to analyse 'causation' as an accepted part of scientific language seems more adequate. Along with the task of analyzing causation comes a reductive constraint on the analysis. One of the orthodox theories of causation is David Lewis' counterfactual theory. It is also a paradigm case of a reductive analysis. But there are also theories of causation on the market that violate the reductive methodology. Woodward's and Cartwright's theories of causation are paradigm cases of non-reductive analyses. The objection against a non-reductive methodology is obvious: The non-reductive analysis of causation is trivial because it is viciously circular and, thus, it is of no use. I doubt that this objection is justified. I have argued for the thesis that one can accept a non-reductive analysis of causation quite happily. Most importantly non-reductive approaches can be defended against the objection that they are viciously circular. A viciously circular analysis is no analysis at all. But since we can distinguish viciously circular analyses from non-reductive ones, the latter qualify as serious attempts to clarify the meaning of causal notions. I have provided eight reasons to belief that these non-reductive theories are acceptable. The result of these eight reasons is basically that the non-reductive character of an analysis does *not* prevent that an analysis satisfies two standards of adequacy for conceptual analysis: (1) The analysis may apply to paradigmatic causal claims in the sciences (and also in everyday contexts). (2) The analysis may distinguish correctly between intuitively different causal structures. Reductive as well as non-reductive analyses are tested by the same standards. I dare to conclude that one can accept a non-reductive analysis of causation as long as there is no reductive analysis that meets both standards of adequacy of conceptual analysis equally well.

University of Münster
University of Cologne
Germany
E-mail:
Alexander.Reutlinger@uni-muenster.de
areutlin@smail.uni-koeln.de

REFERENCES

Carnap, Rudolf (1966): *An Introduction to the Philosophy of Science*, New York: Dover Publications.

Cartwright, Nancy (1983): 'Causal Laws and Effective Strategies', in: Cartwright (1983): *How the Laws of Physics Lie*, Oxford: Clarendon Press, 21–43.

Collins, John, Ned Hall & Laurie A. Paul (eds) (2004a): *Causation and Counterfactuals*, Cambridge (MA): MIT Press.

Collins, John, Hall, Ned & Paul, Laurie A. (2004b): 'Counterfactuals and Causation. History, Problems and Prospects', in: Collins, Hall & Paul (2004a), 1–58.

Frigg, Roman (2010): 'Models and Fiction', *Synthese* 172, 251–268.

Frigg, Roman & Hartmann, Stephan (2006): 'Models in Science', *The Stanford Encyclopedia of Philosophy*.

Glock, Hans-Johann (2003): *Quine and Davidson on Language, Thought and Reality*, Cambridge, Cambridge University Press.

Halpern, Joseph & Pearl, Judea (2005): "Causes and Explanations: A Structural-Model Approach. Part I: Causes", *British Journal for the Philosophy of Science* 56, 843–887.

Hitchcock, Christopher (2001): "The Intransitivity of Causation Revealed in Equations and Graphs", *The Journal of Philosophy* 98, 273–299.

Hitchcock, Christopher (2006): 'Conceptual Analysis Naturalized: A Methodological Case Study', *The Journal of Philosophy*, Vol. CIII, No. 9, 427–451.

Hitchcock, Christopher & Woodward, James (2003): 'Explanatory Generalizations, Part I: A Counterfactual Account', *Nous* 37.1, 1–24.

Jackson, Frank (1998): *From Metaphysics To Ethics. A Defense of Conceptual Analysis*, Oxford: Clarendon Press.

Ladyman, James & Ross, Don (2007): Everything Must Go: Metaphysics Naturalized, Oxford: Oxford University Press.

Lewis, David (1973a): *Counterfactuals*, Oxford: Blackwell.

Lewis, David (1973b): 'Causation', in: Lewis (1986), 159–172.

Lewis, David (1979): "Counterfactual Dependence and Time's Arrow", in: Lewis (1986), 32–51.

Lewis, David (1986): *Philosophical Papers II*, Oxford: Oxford University Press.

Lewis, David (2004): "Causation as influence", in: Collins, John et al. (ed.) (2004).

Mach, Ernst (1980): *Erkenntnis und Irrtum*, Darmstadt: Wissenschaftliche Buchgesellschaft.

Mach, Ernst (1982), *Die Mechanik*, Darmstadt: Wissenschaftliche Buchgesellschaft.

Mach, Ernst (1900), *Principien der Wärmelehre*, Leipzig: Barth.

Norton, John, 'Causation as Folk Science', in: Huw Price & Richard Corry (eds) (2007): *Causation, Physics and the Constitution of Reality*, Oxford: Oxford University Press, 11–44.

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

Psillos, Stathis (2002): *Causation and Explanation*, Chesham: Acumen.

Psillos, Stathis (2007): 'Causal Explanation and Manipulation', in: Johannes Person & Petri Ylikoski (eds) (2007): *Rethinking Explanation*, in: *Boston Studies in the Philosophy of Science*, vol. 252. Springer 2007, 97–112.

Reiss, Julian (2007): *Causation. An Opinionated Introduction*, URL: http://jreiss.org/Causality%20Manuscript.pdf.

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

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

Scheines, Richard (1997): "An introduction to causal inference", in: McKim, Vaughan (Hrsg.): *Causality in Crisis*, Notre Dame: University of Notre Dame Press, 185–199.

Sinai, Yakov G. (1992): *Probability Theory. An Introductory Course*, Berlin/New York: Springer.

Spirtes, P., Glymour, Clark & Scheines, Richard (2000): *Causation, Prediction and Search*, New York: Springer.

Spohn, Wolfgang (2006): 'Causation. An Alternative', *British Journal for the Philosophy of Science* 57, 93–119.

Strawson, Peter F. (1992): *Analysis and Metaphysics*, Oxford: Oxford University Press.

Williamson, Jon (2005): *Bayesian Nets and Causality. Philosophical and Computational Foundations*, Oxford: Oxford University Press.

Woodward, James (2000): 'Explanation and Invariance in the Special Sciences', *The British Journal for the Philosophy of Science* 51, 197–254.

Woodward, James (2003): *Making Things Happen*, Oxford: Oxford University Press.