PHILOSOPHICA

2006 Vol. 77

PLURALISM IN THE PHILOSOPHY OF CAUSATION

CONTENTS

ISSN 0379-8402

Philosophica 77 (2006) pp. 5-13

PLURALISM IN THE PHILOSOPHY OF CAUSATION: DESIDERATUM OR NOT?[1]

Leen De Vreese

1. Introduction

For a long time, philosophers working on the topic of causation have been looking for one univocal approach. "Causation" is nonetheless a concept widely used in a variety of reasoning processes. The difficulties experienced in searching for a unique approach able to deal with this diversity have changed the way philosophers think about causation. In the last couple of years, one can notice a shift in attention from the defences and elaborations of opposing univocal approaches towards the development of approaches that leave more room for diversity. As an effect, the topic of causal pluralism has enormously gained interest in the debates on philosophy of causation. This does not mean that pluralistic ideas are generally accepted. Currently, one can discern two camps. On the one hand, some philosophers still swear by causal monism and continue working in this tradition. On the other hand, a different group of philosophers considers a pluralistic view on the matter as a solution to the problems encountered in monistic causal approaches. However, since the debate on causal pluralism is scarcely out of the egg, it is unstructured and confusing, even on what "causal pluralism" itself means.

---

[1] I am greatly indebted to Erik Weber and Steffen Ducheyne for their useful suggestions and insightful comments on earlier versions of this introductory article. The research for this paper was supported by the Research Fund of the Ghent University through research project BOF2001/ GOA/ 008 and by the Fund for Scientific Research - Flanders through research project G.0651.07.

Is it worth the effort to keep searching for a singular approach? Or can a pluralistic view offer us a more accurate picture of causation and our notion of it? And if so, how to develop a pluralistic theory on "causation"? All these questions lie at the basis of this volume of *Philosophica*.

In the following section, I highlight some reasons to become a causal pluralist. I will refer to James Woodward's *Making things happen* (Woodward, 2003) to point to some problems causal monists have to deal with and causal pluralists may be able to solve. I further distinguish between different ways to be a causal pluralist: one can approach causal pluralism from a conceptual, metaphysical or epistemological-methodological point of view. In section 4 of this introduction, I list the questions the contributors to this volume were asked to focus on in their papers. A short overview of the content of the contributions is given in the final section.

2. Reasons for and ways of being a causal pluralist

James Woodward recently made an important contribution to the philosophy of causation with the development of his interventionist theory of causation published in (Woodward, 2003). Woodward gives a univocal conceptual analysis of our notion of "cause" and is, from the beginning of his book, very explicit about his anti-pluralistic stance in this project. In the acknowledgements he consciously remarks:

> Writers in the grip of a single, overarching set of ideas sometimes tend to suppose that these ideas can be used to resolve all of the extant problems in their subject area. I fear that I have not been immune to this impulse. (Woodward, 2003:vi)

The concept "cause" as defined in his interventionist theory is intended to be applicable to causal reasoning in as much disciplines as possible. Woodward clearly experiences the limited applicability of a causal theory as a problem. Only one single concept of causation covering

causal reasoning in all domains of science is acceptable.[2] Clearly, Woodward does not consider thinking in terms of distinct concepts of causation. He presents us a candidate for the one and only theory of causation and causal explanation in which the problems of alternative theories are said to be discarded. He describes the alternative theories as rival and competitive, and proclaims his own theory as "the most promising" over the whole line.

Indeed, Woodward managed wonderfully well to develop a theory applicable to a whole range of cases. However, one can wonder whether this generality is an advantage. As Woodward claims himself with regard to the general concept of explanation: "generality is not always a virtue" (Woodward, 2003:5). The question is whether the same problem does not also occur in the specific case of causation and causal explanation. Woodward's unifying aim leads indeed to some tensions in his approach. On the one hand, he claims that patterns of counterfactual dependence as revealed by his interventionist approach are the "objective core" behind causal judgments, but on the other hand, he is forced to accept the influence of interests and what causal reasoners interpret as "serious possibilities" as contributory in causal reasoning. By explaining causality and "resolving all extant problems" such as causal overdetermination, omission, prevention, etc. along the interventionist lines he seems to redefine the concept "cause" such that it fits the theory rather than the reverse. Consequently, although Woodward claims to lean on causal intuitions and rejects alternative approaches on the basis of their incongruity with these intuitions, some elements intuitively perceived as causes are claimed to be wrongly entitled as "causes" on the basis of his own theory, while other elements intuitively not perceived as such are presented as "real causes" by Woodward. To give just one example as it was criticized by Glymour:

> A woman's sex or race cannot, according to Woodward's constraint, be a cause of her treatment by someone else, since there are no interventions on either feature with suitable invariance.

---

[2] Woodward nonetheless recognizes that one can discern different kinds of causes, e.g. total causes, direct causes, contributing causes, etc. However, these different kinds of causes should all be captured by means of his single notion in terms of interventionist counterfactuals.

> Woodward argues that sex fragments into many different variables, ranging from genotype to employer beliefs about a person's sex. Claims about sex as a cause, he argues, are typically ambiguous. Granted, but philosophers are skilled at disambiguating when they want to and I think the point remains that genotype is not, on his view, even a remote cause of an individual's treatment by others. These last cases are in my view regrettable consequences of trying to found a theory of causal explanations on interventions. (Glymour, 2004:789-790)

The question is then whether Woodward has the right to use these subtle manoeuvres for the benefit of uniformity? Since we can't avoid being confronted with contradictory causal intuitions[3], every monistic causal approach will need to revise our notion of "cause" such that inconsistencies in our intuitions are decided. Hence, a big challenge for a conceptual causal monist is to find an objective and convincing criterion to justify that one should revise one's causal intuitions precisely in the way suggested by the approach defended. In fact, Woodward does not seem to have such a criterion, except for the strong belief in his own theory, to justify his revisions. Such arbitrary revisionism might be overcome by a pluralistic approach, in which contradictory intuitions might be explained by different, but equally acceptable, approaches to causation.

A second aspect of causal reasoning that causal monists do not take into account, is that the context of the reasoning process and/or the context in which the causal event itself arose, can influence our causal judgments. Even if one is convinced that "causation" can be captured by a single concept, what will be selected as "the cause" may vary dependent on these contextual factors. On the other hand, contextual factors may also take part in deciding which approach to causation is appropriate in the situation under consideration, and may hence further justify a non-univocal approach to causation. A second challenge for the causal monist is thus to avoid that his general approach, which is denying contextual elements of possible importance for causal decisions, becomes too uninformative to characterize everyday causal reasoning.

---

[3] Christopher Hitchcock presented in his article *Of Humean Bondage* (Hitchcock, 2003) a whole range of examples on which our causal intuitions can disagree.

However, one can also argue for or against causal pluralism from a totally different point of view, namely from metaphysical convictions. Phil Dowe, for example, clearly defends metaphysical causal monism (Dowe 1992, Dowe 1995, Dowe 2000). He is convinced that, from a metaphysical point of view, causation is a univocal relation describable in terms of conserved quantities at the physical level. Dowe nonetheless admits that we use other concepts of causation, incorporating for example prevention and omission. He labels the latter kinds of "causes" "quasi-causation" and admits that it is not necessary for practical purposes to distinguish "quasi-causation" from real causation (Dowe, 2004).

Woodward does not offer metaphysical reasons to underpin his monistic conceptual approach. What he does use to underpin his conceptual approach are arguments from a scientific point of view. Witness, for example, Woodward's reaction to Skyrms' pluralistic ideas (see Woodward, 2003:91-93) that all of the criteria for causation have more or less equal weight and that we hence have to think of causation as a cluster concept involving all these approaches:

> whatever the appeal of the cluster concept account as a description of the concept of causation with which we ordinarily operate, it is a problematic account from the point of view of methodology - it is not a concept we should adopt. On the one hand, if we formulate the cluster theory in such a way that satisfaction of all of the above criteria is necessary for the application of the concept "causation," we will exclude a large number of scientifically interesting cases of causation. On the other hand, if we say that "most" or "many" of the criteria must be satisfied or that some criteria are more "important" than other or that "different criteria will be weighted differently in different contexts," then unless we can explain with some precision what the quoted phrases mean, we will end up with a concept of causation that is vague and unclear, and the application of which to specific cases is uncertain and contestable. [...] One of many virtues of a monocriterial view like the manipulability theory is that it forces investigators to be less vague and noncommittal about what they mean when they use this word ["cause"]. (Woodward, 2003:93)

First of all, I think Woodward (just like a lot of other participants in the causal pluralism debate) fails to make a distinction between arguments

underpinning conceptual causal pluralism and arguments underpinning epistemological-methodological causal pluralism. Further, if we read these arguments of Woodward as supporting epistemological-methodological causal monism (along with conceptual causal monism), they can still be questioned. Will all investigators in all domains of science indeed be able to be clearer when they are forced to use a single limited concept of causation? And is it on the other hand really unimaginable to develop some precision with regard to the view that "different criteria will be weighted differently in different [scientific] contexts"? Monistic oriented philosophers of science will probably be easily convinced, but it might not be so for pluralistic oriented ones.

The ultimate question regarding Woodward's recent contribution to the philosophy of causation is then: is the generality with regard to the concept "cause" as purchased by Woodward a real virtue and the right approach to the subject? The contributors to this volume were presented with this and related questions, which will be listed in the following section.

3. The topics of this volume of Philosophica

In the previous sections of this introduction, some questions regarding causal pluralism have already arisen. The following questions served as a guideline for the contributors.

First of all, some questions related to the very general question whether causal pluralism is something we should avoid or rather endorse:

- Is it possible and/or necessary to find one singular and overall concept of causation?
- Do some counterexamples form a thorough reason to entirely reject a theory of causation?
- Is generality a virtue with regard to the characterization of causation?
- Is it necessary and/or possible to develop a pluralistic view on causation?
- Will a pluralistic view offer us a more accurate picture of the causal reality?

Secondly, some questions related to the elaboration of such a pluralistic approach to causation:

- How to develop this pluralistic characterization?

- What kinds of pluralism should one embrace with regard to causation? Conceptual pluralism? Ontological pluralism? Epistemological pluralism? Methodological pluralism?
- What is the importance of the intuitive conceptions of causation? Should a theory of causation try to get grip on their diversity or should it rather redefine the concept to maintain one general criterion? And how to do this?

Lastly, what are the consequences of choosing for a pluralistic approach?

- Will a pluralistic approach to causation necessarily end in a vague and unclear conception which is difficult to apply in practice?
- What will be the concrete consequences for future research when adopting a pluralistic approach?

4. The answers of the contributors

The previous sections already made clear that the dispute on "causal pluralism" is not at all battled out. On the one hand, the idea of "causal pluralism" is not generally accepted and on the other hand, the notion itself can be understood and filled in in many diverse ways. This is also clear from the contributions in the volume at hand. A uniform idea on what "causal pluralism" means, and on whether it should in some way or other be accepted as a fruitful approach at all, certainly does not arise from the aggregate of these contributions. The most important thing these contributions illustrate might precisely be this diversity of ideas on "causal pluralism".

Raffaella Campaner and Francis Longworth both approach causal pluralism from a conceptual point of view. In *Mechanisms and Counterfactuals: A different glimpse of the (secret?) connexion,* Campaner holds a plea for using two concepts of cause at the same time when performing a conceptual analysis, namely a combination of a counterfactual and a mechanical approach. She argues that these approaches complement each other and are of equal value and importance. While the mechanical approach has an important explanatory role to play, counterfactuals are important for their heuristic capacities.

Francis Longworth, in *Causation, Pluralism and Responsibility* begins from Ned Hall's position (2004) that "causation" has to be interpreted as a disjunctive concept in which, contrary to the conceptual pluralism of Campaner, either of the disjuncts (i.e., production or dependence) is sufficient for "causation". Longworth presents some counterexamples to Hall's approach, and then explores whether the introduction of the notion of "responsibility" offers a means of evading these counterexamples.

Jon Williamson argues against causal pluralism in *Causal Pluralism versus Epistemic Causality*. His arguments are mainly based on metaphysical convictions. He maintains that causation is not physically real, but has to be analysed in terms of rational beliefs. Our rational causal beliefs yield one, singular concept of cause, making both metaphysical and conceptual causal pluralism false. However, this singular concept of cause is multifaceted in the sense that there are several different indicators of causal relations. Consequently, Williamson does accept a certain kind of epistemological causal pluralism.

The last two contributions of the volume approach causal pluralism from a scientific point of view. Federica Russo focuses on the social sciences in *The Rationale of Variation in Methodological and Evidential Pluralism* and defends a monistic epistemological account. She argues that, despite methodological and evidential pluralism in the social sciences, one is confronted with a monistic epistemology which is based on the rationale of variation. She further argues that her approach helps in liberating the social sciences from the hallmark of being inferior to the natural sciences.

Lastly, in my own contribution entitled *Causal Pluralism and Scientific Knowledge: an Underexposed Problem*, I argue more generally that an epistemological-methodological approach to causal pluralism should be valued as a line of approach on its own. I further defend epistemological-methodological causal pluralism by demonstrating and arguing that we need different causal concepts to gain scientific knowledge. On the one hand, we need different causal concepts for different scientific domains, and on the other hand, we might even need different causal concepts within singular scientific domains.

5. Conclusion

Although the contributions in this volume will not at all settle the dispute on "causal pluralism", we hope they will help us one step further in the development of arguments pro and contra causal pluralism, and in clarifying what causal pluralism could signify for the philosophy of causation.

Universiteit Gent (UGent)
Leen.DeVreese@UGent.be

REFERENCES

Dowe, P. (1992), "Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory," *Philosophy of Science*, 59, pp. 195-216.

Dowe, P. (1995), "Causality and Conserved Quantities: a Reply to Salmon," *Philosophy of Science*, 62, pp. 321-333.

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

Dowe, P. (2004), "Causation and Misconnections," *Philosophy of Science*, 71, pp. 926-993.

Glymour, C. (2004), "Critical notice. James Woodward, Making Things Happen: A Theory of Causal Explanation," *British Journal for the Philosophy of Science*, 55, pp. 779-790.

Hall, N. (2004) "Two Concepts of Causation," in: Hall, N., Collins, J. and Paul, L.A. (eds.), *Causation and Counterfactuals* (pp. 225-276), Cambridge: MIT Press.

Hitchcock, C.R. (2003), "Of Humean Bondage," *British Journal for the Philosophy of Science*, 54, pp. 1-25.

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

Philosophica 77 (2006) pp. 15-44

MECHANISMS AND COUNTERFACTUALS: A DIFFERENT GLIMPSE OF THE (SECRET?) CONNEXION

Raffaella Campaner

ABSTRACT

Ever since Wesley Salmon's theory, the mechanical approach to causality has found an increasing number of supporters who have developed it in different directions. Mechanical views such as those advanced by Stuart Glennan, Jim Bogen and Peter Machamer, Lindley Darden and Carl Craver have met with broad consensus in recent years. This paper analyses the main features of these mechanical positions and some of the major problems they still face, referring to the latest debate on mechanisms, causal explanation and the relationship between mechanisms and counterfactuals. I shall claim that the mechanical approach can be recognised as having a fundamental explanatory power, whereas the counterfactual approach, recently developed mainly by Jim Woodward and essentially linked to the notion of intervention, has an important heuristic role. Claiming that mechanisms are by no means to be seen as parasitic on counterfactuals or less fundamental than them – as it has been recently suggested –, and that yet counterfactuals can play a part in a conceptual analysis of causation, I shall look for hints in support of the peaceful coexistence of the two.

1. The causal structure of the world: processes and interactions, entities and activities

In the last couple of decades philosophy of science has seen the elaboration of several mechanical accounts. While terminology and emphasis differ, all the theories overlap in believing that mechanisms are complex systems present in nature.

Wesley Salmon's philosophical work is unanimously regarded as the compulsory *locus* for anyone interested in the notion of mechanism since the eighties. As is well-known, Salmon has developed a "process theory" of causation, centred on the notions of causal process, causal

production and causal propagation. In short, causal processes are defined as spatio-temporally continuous processes which exhibit consistency of structure over time, and are capable of transmitting a modification of their structure from the point at which it is performed onwards, without additional interventions. The production of causal influence is accounted for by appealing to causal forks, characterised in statistical terms. Once produced, causal influence is propagated continuously through processes. Interacting processes constitute a mechanical, objective and probabilistic network, underlying phenomena and responsible for their occurrence.

Salmon's conception of causality goes hand in hand with his theory of causal explanation, which comprises two levels: we first need to identify the properties which are statistically relevant with respect to the occurrence of the event to be explained; we then account for them in terms of the net of causal processes underlying the event. A further distinction Salmon makes, not recalled by later authors[1], is that between etiological and constitutive causation. When we aim at explaining a given event *E*,

> we may look at *E* as occupying a finite volume of four-dimensional space-time. If we want to show why *E* occurred, we fill in the causally relevant processes and interactions that occupy the past light cone of *E*. This is the *etiological* aspect of our explanation; it exhibits *E* as embedded in its causal nexus. If we want to show why *E* manifests certain characteristics, we place inside the volume occupied by *E* the internal causal mechanisms that account for *E*'s nature. This is the *constitutive* aspect of our explanation; it lays bare the causal structure of *E*. (Salmon, 1984:275) (emphasis added)

In sum, causal-mechanical explanations of the etiological sort illustrate the causal story leading up to the occurrence of the explanandum, whereas constitutive explanations provide a causal analysis by showing the underlying causal mechanisms that constitute the phenomenon itself. On the whole, Salmon's conception aims at conjugating mechanicism and probability, both indispensable for an adequate picture of the causal structure of the world. Neither statistical relevance relations nor

[1] Apart from Craver (2001:69-70).

connecting causal processes have explanatory import if taken on their own; they do so only together.

In the early nineties, Phil Dowe, among others, has largely criticised Salmon's view, raising objections against its being circular, using vague terms (such as "structure"), characterizing causal production and causal interactions in terms of statistical relations and, last but not least, using counterfactuals in the formulation of the criteria defining causal processes and causal interaction. According to Salmon, a causal process is such that, had a modification of its structure been performed, it would have transmitted it from that point onwards; a causal interaction is an intersection between two causal processes such that, had they intersected, both their structures would have been modified from that point onwards. Dowe has thus advanced a new process theory, called "the conserved-quantity theory", which aims at preserving Salmon's objective and physical idea of causation, while getting rid of counterfactuals. In short, the conserved-quantity theory, embraced with minor modifications by Salmon himself, holds that a causal process is the world line of an object exhibiting a conserved quantity, and a causal interaction is an intersection of processes in which an exchange of a conserved quantity occurs.

The next attempt to elaborate a mechanical position, not intended as a direct further development of Salmon-Dowe's view, has been made by Stuart Glennan roughly in the same years. Glennan wants to substitute Salmon's and Dowe's process causal-mechanical theories with what he calls a "complex-systems account". Its core is the following definition:

> A mechanism for a behaviour is a complex system that produced that behaviour by the interaction of a number of parts, where the interactions between parts can be characterised by direct, invariant, change-relating generalizations. (Glennan, 2002:S344)

The notion of mechanism is here strongly linked to that of behaviour: Glennan believes that a mechanism cannot even be identified without mentioning what it *does*. As in Salmon's view, a central role is played by the notion of interaction, though this is not as precisely defined. According to Glennan, no a priori restrictions are to be put on the sorts of allowable interactions that may take place between the parts of a mechanism. Whereas "interaction" means something very specific and

circumscribed for Salmon and Dowe, Glennan's account takes the relevant modes of interaction between the component parts of mechanisms to always depend upon the behaviour we are interested in explaining. Mechanisms must simply be such that their *"'internal' parts interact to produce a system's 'external' behaviour"* (Glennan, 1996:49), but it is far from clear how we shall make sense of "internal" as opposed to "external", and what can properly count as "parts" of a mechanism.

Glennan's view of mechanical causation is meant to be a theory of causal explanation too. Mechanisms are made up of parts, and events are claimed to be causally related when there is a mechanism that connects them; a good description of a mechanism is believed to provide an adequate causal explanation. As emerges from the key-definition quoted above, the interactions between parts of the mechanism which give rise to its behaviour are characterised by invariant generalizations. Glennan admittedly borrows this notion from Jim Woodward, and takes it to indicate a generalization that would hold were a range of possible interventions to be performed[2]. According to Glennan, a two-way relationship holds between invariant generalizations and mechanisms:

> First, reliable behaviour of mechanisms depends upon the existence of invariant relations between their parts, and change-relating generalizations characterise these relations. Second, many such generalizations are mechanically explicable, in the sense that they are just generalizations about the behaviour of mechanisms. A single generalization can both be explained by a mechanism and characterise the interaction between parts of a larger mechanism. (Glennan, 2005:445-446).

The link between the notions of mechanism and of invariant generalization turns out to be a very strict one.

Being complex, or often very complex, systems, mechanisms can be decomposed into subsystems. Decomposition depends on what is being explained, but, Glennan warns, its context-dependence must not be confused with antirealism or relativism: descriptions of mechanisms are

---

[2] Glennan appeals to such a notion in his (2002). In the definition given in his (1996) mechanisms are claimed to be working "according to direct causal laws" (pp. 50).

adequate descriptions only insofar as they show what is really there. Whereas Salmon recognised a specific part of physics, that of "spooky" actions at-a-distance studied in quantum mechanics, as the only deeply problematic field for his mechanical theory of causation, for Glennan all laws but the fundamental laws of physics can be explained in mechanical terms. Finally, let me just recall that Glennan's most recent work (2005) focuses on the nature and testing of mechanical *models*, where the latter are claimed to consist of both the description of the mechanism's behaviour and the description of the mechanism accounting for that behaviour. A distinction is thus drawn between the mechanism as such, and our conceptual model aimed at representing it.

That *what* mechanisms do is closely linked to *how* they do it, namely to how they are organised, is very strongly highlighted by Peter Machamer, Lindley Darden and Carl Craver, whose theory is probably nowadays the most famous and debated mechanical account. The articles that, part jointly and part separately, Machamer, Darden and Craver have written in the last six years present mechanisms as

> entities and activities organised such that they are productive of regular changes from start-up to finish or termination conditions [...]. Mechanisms are composed of both *entities* (with their properties) and *activities*. Activities are the producers of change [...] Entities are the things that engage in activities. (Machamer et. al., 2000:3)

In describing mechanisms it is crucial to specify how parts are related to wholes, and how the activities of the parts concur in the performance of the activities of the whole. The internal organization of mechanisms is hence given special attention: the building of mechanical accounts proceeds through accumulation of constraints on the space of possible mechanisms. A constraint is a finding that either shapes the boundaries of the space of such mechanisms, or changes the probability distribution over the space. Constraints can be, for example, spatial (compartmentalization, location, structural orientation, …), temporal (order, rate, duration, frequency, …) and hierarchical (integration of levels). Such constraints are meant to specify precisely how the relevant entities and activities are organized, and to exclude the possibility, given what we know about given entities and activities, that some kinds of mechanisms hold in certain portions of space. Information on spatio-

temporal conditions and background knowledge on entities and their features are then required for a mechanical account to be adequate.

Instead of a process ontology like that put forward by Salmon or Dowe, Machamer, Darden and Craver advance a so-labelled "dualistic" theory, in which *both* entities and activities are claimed to be essential. The components of the mechanical system are understood in virtue of their membership in the whole, their location in space and time, the order and rhythm with which they operate, their duration, and so on. Moreover, while Salmon's explanatory representation of mechanisms can include the etiological dimension and/or the constitutive one, Machamer, Darden and Craver insist on mechanisms being organised along a number of hierarchically ordered levels, all holding at the same time. Different stages can also be identified in the development in time of a mechanism. Continuity is regarded as a crucial feature of mechanisms, and a gap in the description of the sequence of the mechanisms' stages is seen as a sign of the fact that the mechanism is not fully understood. If we knew more precisely how the mechanism actually worked, we would be able to draw an account of it in which each stage follows another continuously.

An important aspect that has been often highlighted in the recent literature concerns the relationship between the level of fine-graininess that the description of a mechanism reaches and the contexts in which such a description is sought. It is now widely acknowledged that no inventory of *all* causal factors is ever produced. Machamer, Darden and Craver also stress how mechanisms' descriptions are usually approximated: they present a certain level of abstraction and can be more correctly labelled "mechanism schemata". According to the purposes for which the mechanism has to be identified, or, more generally, the context in which the enquiry is carried out, a truncated, more or less abstract, description is provided which can later be filled in with more specific details. While explanatory accounts of mechanisms are context, interest or purpose relative, mechanisms as such are strictly objective. In Jim Bogen's words,

> complete enumerations of a mechanism's components, their activities, and the exogenous factors that influence their operations are never required. Which questions a causal explanation must answer, and in how much and what kind of detail varies from context to context with people's interests and background knowledge, cultural factors, and social settings. By contrast, what

> parts belong to a mechanism, what they do, how they do it, and how their activities contribute to the production of an effect are matters of fact that neither depend upon nor vary with the contextual factors which determine what should be included in an explanation. (Bogen, 2005:398, footnote 2)

In general, Machamer, Darden and Craver put forward an apparently less ambitious view than Glennan's or Salmon's. Although they maintain that much of the history of science can be seen as written in mechanical terms, they do not hold that *all* sciences and/or *all* scientific explanations are of a mechanical sort. Their theory aims at fitting mechanical accounts of phenomena studied in fields such as Mendelian genetics, molecular biology, cell biology, neuroscience and cognitive science. At the same time, they express the hope that their theory will be applied to disciplines different from the biological and biomedical sciences, possibly to social disciplines such as psychology and economics.

Although they are largely indebted to Salmon's remarkable contribution to the revival of mechanical causation in contemporary philosophy of science, Glennan and Machamer, Darden and Craver barely mention it. An important difference between them seems, in any case, worth stressing. While for Salmon probability and its link with causality are among the main themes to be analysed and his overall position is developed precisely as a *probabilistic* mechanical theory, the relation between causality and probability is not explicitly addressed and clearly dealt with in the last two theories illustrated. Neither Machamer, Darden and Craver nor Glennan discuss probabilistic causality as such. Nevertheless, it seems that Glennan's and Machamer, Darden and Craver's discourses can work in a probabilistic context: they do so once we think of complex systems and of organised entities as exhibiting a probabilistic behaviour like that of probabilistic causal processes described by Salmon.

Glennan's and Machamer, Darden and Craver's views, which currently play an important part in the debate, do not need to be understood as opposed to each other. James Tabery, for example, has recently shown how the two positions could be combined. Although Glennan, on the one hand, and Machamer, Darden and Craver, on the other, conceptualize mechanisms in different ways, their aims are not divergent, and their views could be "synthesized" in a single account of mechanisms. Glennan takes an intervention to be a change in a property

of one part which produces a change in a property of another part; Machamer, Darden and Craver regard activities as bringing about change, and as best rendering the concept of productivity, which is taken to lie at the heart of their account. According to Tabery,

> the dualists' requirement of productivity, rather than demanding an ontological switch from Glennan's interactions to activities, only reveals the need for interactions as Glennan conceives them *alongside* activities. (Tabery, 2004:9) (emphasis added)

The idea of activity stresses the dynamicity, focusing on the role of bringing about that mechanisms perform, whilst the concept of interaction helps us understand in which respect activities are productive in a mechanism, namely what property changes are involved. In this sense, each view emphasizes an important element lacking in the other, and both could adopt the "synthesizing" concept of "interactivity", with interactions being defined as occasions

> on which a change in a property of one part *dynamically produces* a change in a property of another part (Tabery, 2004:12) (emphasis added)

These two – possibly complementary – mechanical theories are not without difficulties. Troubles seem to arise if we ask what the notions they employ to define mechanisms exactly stand for. What we are presented with are not very strict definitions, but rather general ideas of "interaction" and "behaviour", of "entity" and "activity", which do not seem to be satisfactorily spelt out and show somehow loose borders. Glennan's "mechanisms' parts" are vague and very comprehensive: as he says himself, "*it is important that a very wide variety of entities may be parts of mechanisms. [...] Parts must be objects*" (Glennan, 1996:53). Unlike what Machamer, Darden and Craver claim, mechanisms' parts are not required by Glennan to be spatially localizable; unlike what Salmon and Dowe maintain, they need not be describable in a purely physical vocabulary. Glennan seems to be leaving us with a very wide set indeed.

The list of verbs which are recognised as indicating "activities" by Machamer, Darden and Craver is also extremely long: entities and activities

> are organised such that they *do* something, *carry out* some task or process, *exercise* some faculty, *perform* some function or *produce* some end product. (Craver, 2000:S84)

Mentioned activities are: attracting, repelling, binding, breaking, diminishing, retarding, eliminating, disabling, destroying, augmenting, intensifying, multiplying, and many others. To help us find out what characterises activities, Darden suggests we should bear in mind that

> for a given scientific field, there are typically entities and activities that are accepted as relatively fundamental or taken to be unproblematic for the purpose of a given scientist, research group or field. That is, descriptions of mechanisms in that field typically bottom out somewhere. Bottoming out is relative: different types of entities and activities are where a given field stops when constructing its descriptions of mechanisms (Darden, 2002:S356).

But do entities involved in mechanisms share something inter-fields? Do they have some common feature that makes them all relevant for a causal account? Is it just their being involved in one activity or another? Or must there be something which the various activities share, such that it makes them all "activities"? Machamer too holds that activities often have identification criteria specific to a given enquiry or discipline:

> One might try to do something more general by giving the conditions for all productive changes. […] It is not clear that they all have one thing in common or are similar in any significant way, but neither commonality nor similarity are necessary conditions for an adequate category. (Machamer, 2004:29)

What other conditions should an adequate category of "activity" (and of "entity") meet?

The obvious risk to be avoided is that of any observable behaviour or output of activity whatsoever counting as causal. Shall *anything* which affects or influences something else in any scientific context be acknowledged causal power in a mechanical sense? Machamer himself recognises the difficulties in getting close to a definition:

> We could say that activities are the happenings that, singularly or in concert with other activities, produce changes in or bring into

> existence other entities and /or activities. [...] We might say that activities are ways of acting, processes, or behaviours; they are active rather than passive; dynamic rather than static. However, even this way of talking, while maybe helpful, seems a far distance from providing necessary or sufficient conditions or from definitionally characterizing activities in terms of things even more generically ontological. (Machamer, 2004:29)[3].

Things do not get any better with respect to entities, which are not given a precise definition either and are simply claimed to be things that act. If entities are what engage in activities and activities are what entities do, we also run the risk of running in a narrow circle.

> People learn to pick out and categorise activities as well as they do entities, and independently. [...] People, including children, categorize the world into running, breaking and boozing just as they do into flowers, bears and bootstraps. (Machamer, 2004:32)

Even if we all gain an idea of causation from common, everyday, ostensive knowledge, what happens when we are asked to deal with causal relations in complex systems and/or within advanced sciences? It seems that further features are called for to circumscribe "activities" and "entities", "parts" and "interactions" in a causal sense as things that make a fundamental contribution to the correct functioning of a mechanism as a whole.

Both Glennan and Machamer, Darden and Craver attempt to give a more articulated idea of mechanisms, one which aims at being more applicable to concrete cases and that substantiates Salmon's and Dowe's notions of "process" with a closer look at scientific uses of the idea of "mechanism" in a range of scientific disciplines much larger than physical or natural sciences. As a drawback, though, we are left with notions of entity, interaction, activity and mechanism which prove much looser and, as a consequence, imprecise. Glennan actually criticises Salmon for not adequately distinguishing between causal processes and

---

[3] Machamer insists that *"activities are better off ontologically than some people ontic commitments to capacities, dispositions, tendencies, propensities, powers or endeavours. All these concepts are derivative from activities"* (Machamer, 2004:30).

pseudo-processes. Salmon's requirement – he states – does not explain why the former transmit marks and the latter do not: *"the true difference [...] can only be explained by considering the differences in the mechanisms underlying them"* (Glennan, 1996:70, footnote 14). According to Glennan, a genuinely causal relation differs from a merely coincidental one in that only the former derives from an underlying mechanism, revealed by empirical investigation. But how shall such empirical investigation work? Won't it have to presuppose the presence of the very same things (i.e. mechanisms) that it is supposed to reveal? Neither of the two mechanical theories provides us with criteria for distinguishing between what counts as causal and what does not.

2. Mechanisms *and* Counterfactuals?

The previous section raised some issues regarding the identification of criteria for interactions and behaviours, entities and activities. Machamer states:

> The problem of causes is not to find a general and adequate ontological or stipulative definition, but a *problem of finding out*, in any given case, *what are the possible, plausible, and actual causes* at work in any given mechanism [...] The problem of causes, in our terms, is to discover the entities and activities that make up the mechanism. (Machamer, 2004:27-28)

Glennan, on his part, says:

> analysis of causal connections in terms of mechanisms is only meaningful when there are ways (even if indirect) of acquiring knowledge of their parts and the interactions between them. (Glennan, 1996:51)

How is such knowledge acquired? How do we find out what the entities and activities at work are? To look for some suggestions, let us turn to one of the currently most successful and widely debated conceptions of causation, that elaborated by Jim Woodward in roughly the last decade. This theory takes a combination of manipulation and counterfactuals as

crucial for an adequate understanding of causation and causal explanation[4].

According to Woodward, what makes a variable *A* causally relevant to a variable *B* is the invariance of the relation between the values of *A* and of *B*: a relationship between two variables is said to be causal if, were an intervention to change *A* appropriately, then the value of *B* would change too, while the relationship between *A* and *B* would still hold. If the relation holding between *A* and *B* in *actual* cases had been only coincidental, then it would not have remained the same under some range of interventions. Invariance is meant as invariance under either actual or just possible interventions, and hence presented as a modal or counterfactual notion, having to do with *"whether a relationship would remain stable if, perhaps contrary to actual fact, certain changes or interventions were to occur"* (Woodward, 2000:235). In sum, if *A* is a cause of *B* then the values of *B* must vary counterfactually with the values *A* assumes under interventions. Woodward links counterfactuals that are relevant to grasping causation with experimental interventions:

> rather than being understood in terms of similarity relations among possible worlds, counterfactuals are understood as claims about what would happen if a certain sort of experiment were to be performed (Woodward, 2004:44)

---

[4] I shall here confine my attention to the counterfactual theory of causal explanation developed by Woodward (recently in collaboration also with Christopher Hitchcock; see (Hitchcock, 2003a) and (Hitchcock, 2003b)), since this is the theory which has been most widely and directly confronted with the mechanical approaches developed by Glennan and Machamer, Darden and Craver. A different theory of causal explanation that appeals to counterfactuals has recently also been put forward by Joseph Halpern and Judea Pearl, who suggest using structural equations to model counterfactuals. In this approach, the causal influence that random variables in the world can have on others is modelled by a set of structural equations, with each equation representing *"a distinct mechanism (or law) in the world, which may be modified (by external actions) without altering the others"* (Halpern and Pearl, 2005b:891). Structural equations are taken to represent causal mechanisms and to support a counterfactual interpretation. See (Halpern and Pearl, 2005a) and (Halpern and Pearl, 2005b).

and are hence labelled "interventionist counterfactuals" or "active counterfactuals". In this account causal claims are connected to counterfactual claims concerning what would happen under interventions even when such interventions are merely hypothetical and not physically possible.

According to Woodward, invariance is not an all-or-nothing matter, but admits of degrees. A generalization can be more or less invariant according to whether it is invariant under a larger or narrower range of interventions. Such degrees are taken as a symptom of different degrees of explanatoriness: the more invariant a generalization, the better the explanation within which it is included. The range of interventions taken into account admittedly depends on the disciplinary field and the subject matter under consideration. Contextual factors play an important role. As we have seen, mechanical views recognise that context plays a part in the description of mechanisms and their behaviour, by affecting which portions of mechanisms are examined and at which level of detail they are described. Instead, in Woodward's theory the context enters into the identification of the degree of invariance a given generalization exhibits and, hence, determines to what extent the generalization is explanatory. In this theory the manipulation "core" and the use of counterfactuals are closely intertwined. Only relations that are invariant under interventions are stated to be causal, and this is meant as a proper criterion, something I noted is missing in the latest mechanical theories.

Could reflections of this sort make sense within a mechanical perspective? Can "active counterfactuals" and "mechanical activities" be combined? I shall try to show how some aspects of Glennan's and Machamer, Darden and Craver's theories, on the one hand, and Woodward's, on the other, can coexist more peacefully than has been claimed by many.

As I mentioned, Glennan directly appeals to the notion of invariant generalizations used by Woodward and expressed by counterfactuals to characterise mechanisms. He presents counterfactuals as easily admissible and utterly unproblematic: generalizations governing mechanisms sustain counterfactuals, and no serious issue needs to be raised about them. Glennan points out that we are usually justified in asserting, for example, "if we were to turn the key, the car would start" because we know a mechanism exists which connects key-turning with

car-starting. Likewise, we know that a given sort of circumstances exists in which

> the counterfactual would turn out to be false, namely breakdown conditions for the mechanism which explains it. […] Counterfactual generalizations can be understood in this way without appealing to unanalysed notions of cause, propensity, possible worlds, and the like. (Glennan, 1996:63)

No such thing as a counterfactual analysis of causation separated from the mechanical one is put forward, and Glennan admits of counterfactuals only insofar as they stand in a close relation with mechanisms.

The matter is a little more complicated if one turns to Machamer's, Darden's and Craver's position, although I believe that some – rather indirect – link can be found here too. Machamer, Darden and Craver speak of mechanisms and, in a few cases, of interventions, but not *directly* of counterfactuals. Counterfactuals can be inserted, though, by a careful analysis of their discourse on strategies devised to understand the hierarchical multilevel organization of complex mechanisms. To provide a description of mechanisms accounting for the various levels in which they articulate, Craver, for instance, presents a *"taxonomy of interlevel experimental strategies"*, which includes *"activation strategies, interference strategies and additive strategies"* (Craver, 2002:S6)[5] and can interact and reinforce each other. Experiments for testing mechanisms are taken to have three fundamental aspects: (1) an experimental model; (2) an intervention technique and (3) a detection technique. Elements (2) and (3) are applied to different levels of the hierarchical structure. Cases in which an intervention is performed which perturbs a component at a lower level to detect consequences at a higher level is labelled "bottom-up"; an intervention which perturbs a component of a higher level to detect variations in activities or entities at a lower level is labelled a "top-down" experiment. This seems to be near to saying that we hypothesise what the relations and interactions between different levels could be and then test them: we perform an experiment in

[5] Specifically, Craver wants to use such a taxonomy to explore how levels integrate in neural mechanisms.

a given way, according to a given model and following a given strategy because we believe that, were we to intervene at a certain level, something would occur at the previous or next one. The productive continuity in time between the various stages characterising a mechanism's functioning can also be reconstructed by means of specific strategies, which Machamer, Darden and Craver call "forward chaining" or "backward chaining". The former appeals to the earlier stages of a mechanism to find out something about the entities and activities that could be present later on; backward chaining, on the contrary, starts from what we know about entities and activities in later stages to understand what could be present earlier. Such strategies can be adopted when

> anything is known, or can be *conjectured*, about entities and activities, in the *hypothesized* mechanism" (Darden, 2002:S363)(emphasis added).

We could reasonably think that these strategies are employed on the basis of conjectures on what entities and activities could have been at a certain stage *n*, if they had been such and such at a stage *n+1* or *n-1*.

Summing up, it seems that experiments Machamer, Darden and Craver appeal to could also be described in terms of the interventionist counterfactuals James Woodward proposes. A mechanism's sketchy description can be filled in by means of both actual *and hypothetical* interventions. Recently, Machamer himself has claimed:

> intervention is a good strategy for uncovering mechanisms or for finding causal connections", meaning by "intervention" something "whereby one stops or changes a putative activity to find out what happens. (Machamer, 2004:28)[6]

---

[6] Machamer emphasizes – as does Woodward – that not only researchers, but also nature itself can bring about interventions, and thus avoid any form of anthropocentrism. Interventions are also taken by the later Machamer as crucial, in general, for any kind of knowledge gaining, from perception to children learning, to scientific experiments. *"We have too long been misled by passive pictures regarding the fundamental epistemic processes of perception and, one might add, cognition. […] knowledge representations are not static traces deposited by incoming signals, but active representations that must include activities on the part of the knower. […] That is, acting is a major part of*

He claims interventions are, in any case, just epistemic or methodological tools, devoid of any ontological import for the definition of what causality really *is*, but they still play a part:

> epistemologically or methodologically, by experimentation or other means, one rules out possibilities that are first promising, or could be thought to be the cause, in order to find out what cause or causes are more probable. One may hope, then, that after enough work one may discover what the actual cause is. (Machamer, 2004:31)

Although Glennan appeals to invariant generalizations, his position remains a genuinely mechanical one. Machamer's, Darden's and Craver's account also remains utterly mechanical: it is entities and activities which are responsible for what goes on among the various levels, for what options are to be ruled out and what, instead, the hierarchical structure of the mechanism is. At the same time, the fact that scientists consider which changes might have occurred as a consequence of interventions performed at a given level or temporal stage seems to leave room for interventionist counterfactuals. In other words, it is activities which are regarded as responsible for changes that experiments bring about, and experimental interventions are interesting only insofar as they help uncover activities and hierarchically structured systems of entities involved in them. Yet, counterfactuals can be used to infer the presence of causal links and they play an important heuristic role, as has been recognised in various contexts. For example, in diagnostic and engineering reasoning evaluation of counterfactuals has been held to be

> a means to find the relevant causal factors and justify that they rather than others are consistent with data. [...] [Counterfactuals conditionals] are cues to the real causal relations. (Pederson et. al., 1984:241-242.)

---

knowing." (Machamer, 2004:33) *"It is important to learn how to intervene and manipulate in experimental settings. Much scientific training and subsequent practice involves pursuing that goal."* (Machamer, 2004:36)

According to Stathis Psillos, Machamer, Darden and Craver cannot avoid counterfactuals, which *"may enter at two places"* (Psillos, 2004:314) in their theory, that is with respect to activities themselves and in the characterisation of interactions within the mechanism. Counterfactuals not only *may,* but *do enter* in their view: they do so in forward and backward chaining and in bottom-up and top-down experimental strategies.

Experimental manipulations presented by means of counterfactuals can hence be acknowledged also within mechanical perspectives as useful for assessing causal relations. Invariance under intervention does not seem to shed enough light, though, on causal relations as employed for explanatory purposes. Woodward and Christopher Hitchcock have recently claimed that

> successful explanation has to do with the exhibition of patterns of counterfactual dependence describing how the system whose behaviour we wish to explain would change under various conditions. (Woodward and Hitchcock, 2003a:2)

As a matter of fact, though, when having a look at science books, textbooks and even scientific reports on novelties in many fields, we do not find a series of counterfactual claims, but the description of entities engaged in activities and the ways these activities are carried out. It cannot be denied that in a number of scientific textbooks, for example in biology, medicine and neuroscience, it is mechanical terms that actually abound. Although the existence of causal relationships is often assessed in the absence of knowledge about mechanisms, it is the latter which is foremost sought for explanatory purposes: when explanations of systems must be provided, it is the behaviour of such systems which is presented first of all, not how such behaviour varies, or, even less, how it would have varied if certain conditions, which did not obtain, had done so.

It is undeniable that an interest in elucidating causal relevance, which was a major task in Salmon's theory, is missing in Glennan's and Machamer's, Darden's and Craver's, and counterfactuals can prove extremely helpful in this respect[7]. Although it can be readily agreed that

[7] Woodward goes as far as to say that causal relevance cannot be handled in any other way. See (Woodward, 2004:48).

invariant generalizations can play a role in explaining, and that in presenting an explanation one is committed to a set of counterfactual claims concerning what would have happened to the effect if the cause had been different, this is not to say that explanations consist just in exhibiting patterns of counterfactual dependence. It is one thing to say that its being invariant under intervention is what distinguishes a causal generalization from a non-causal claim, and another to say that only this very feature on its own is that which explains. Woodward maintains that Machamer's, Darden's and Craver's account cannot *"capture the idea that there is an overall productive relationship [...] without explicitly invoking the idea of counterfactual dependence"* (Woodward, 2000:35). It seems they *can* capture productivity, although they might need counterfactual dependence in order to identify *what* performs the production.

> On our account the aim of explanation is to provide the resources for answering what-if-things-had-been-different questions by making explicit *what* the value of the explanandum variable depends upon. (Woodward and Hitchcock, 2003b:190) (emphasis added)

Adequate explanations are able to provide answers to what-if-things-had-been-different-questions, and hypothetical, idealised experiments allow us to gain insights into properties we would like to control, to deal with matters such as control groups, to respond to why we chose one experimental strategy rather than another, and so on. When providing an explanation, though, we do not want to make explicit only *what* the explanandum depends upon, but also as much as possible how, when and where it depends on it. Counterfactuals can exhibit explanatory relevant information, which is then usually organised within a mechanical framework and filled in with mechanical details. Saying that "*had* C *not occurred* E *would not have occurred either*" (or would have had a much lower chance of occurring) informs us about the existence of a relationship between *C* and *E*. Everything going on between *C* and *E* is not normally expressed by means of a sequence of counterfactual claims, but rather by a series of – unfortunately often vague or confused – claims involving notions such as "process", "interaction", "entity", "activity", and the like.

Finally, counterfactuals could have the resources to deal with cases of causation-per-absence, which puzzle mechanical views.

> Mechanisms are sometimes described by things that are absent, are not done, or fail to occur. […] These all would seem to be cases where causality is attributed not via an activity, but by virtue of a non-activity. (Machamer, 2004:35)

Counterfactuals can turn out to be particularly useful in cases where the exact functioning of the alleged mechanism cannot be displayed: they help reveal that, had something been in place, some effect would have occurred.

> Non-existent activities cannot cause anything – but – […] failures and absences can be used to explain why another mechanism, if it had been in operation, would have disrupted the mechanism that actually was operating. (Machamer, 2004:35-36)[8]

By the counterfactual clause, Machamer himself admits that failures and absences can be used to indicate which mechanism, *had it been* in operation, *would have performed* a productive activity. This might be a hint indicating that the *"anti-counterfactual Pittsburgh tradition"* (Woodward, 2004:43) may not be as monolithic as Woodward suggests.

3. More on interventionist counterfactuals and mechanisms

As is well-known, counterfactuals have traditionally given rise to a number of philosophical puzzles which also lie heavy on any attempt to reconcile them with mechanical views of causation. According to Woodward, what we shall be looking for is a *"basis for assessing the truth of counterfactual claims concerning what would happen if various interventions were to occur"* (Woodward, 2003:130), where such interventions can be either performed or merely possible.

---

[8] That counterfactuals can solve cases of omissions is also maintained by Ned Hall (2004:248-249, 256). Counterfactuals are not acknowledged any role in these cases by Humphreys (see Humphreys, 2006:42).

In a recent paper, entitled *"A Glimpse of the Secret Connexion: Harmonizing Mechanisms with Counterfactuals"*, Stathis Psillos has criticised Woodward's use of counterfactuals, claiming that he does not give a clear account of what he takes their evidence-conditions and their truth-conditions to be. Psillos states:

> there seems to be a conceptual distinction between causation and invariance-under-intervention: there is an intrinsic feature of a relationship in virtue of which it is causal, an extrinsic symptom of which is its invariance under interventions. (Psillos, 2004:302).

More precisely, he accuses Woodward of keeping evidence-conditions and truth-conditions apart: evidence-conditions of Woodward's active counterfactuals are fully specified in terms of experiments, whereas truth-conditions are not. What aspects of Woodward's theory is Psillos referring to? The problem arises from statements like the following:

> doing the experiment corresponding to the antecedents of [counterfactual claims] doesn't *make* [them] have the truth-values they do. Instead the experiments look like ways of *finding out* what the truth values of [the counterfactual claims] were all along. On this view of the matter, [counterfactual claims] have non-trivial values [...] even if we don't do the experiments of realizing their antecedents. Of course, we may not *know* which of [two counterfactual claims] is true and which false if we don't do these experiments and don't have evidence from some other source, but this does not mean that [they] both have the same truth values. (Woodward, 2004:46).

Psillos concludes that, while he gives us a relatively detailed account of the evidence-conditions of counterfactuals, Woodward does not provide anything remotely like that for their truth-conditions.

Among other critics of Woodward's use of counterfactuals, I shall briefly recall Paul Humphreys and Jim Bogen. Humphreys (2006) suggests that a distinction could be drawn between understanding and explaining, and that Woowdard could be construed as giving an account of the former but not of the latter: providing answers to what-if-things-had-been-different questions increases our understanding of phenomena, but belongs to a realm of no relevance to explanations of why

phenomena occur. According to Bogen (2005), on the other hand, what actually goes on when a mechanism operates is sufficient for its effects, and for explaining them: it is simply *a fact* that some things exert causal influence and others do not, that some parts of mechanisms contribute to the production of their outcome and others do not, and what could have resulted if other things had occurred cannot make any difference.

Does it mean that we should confine ourselves to describing causally productive activities and get rid of any use of counterfactuals whatsoever? This – as I have tried to show in section two – seems too radical. Even opponents such as Bogen acknowledge that counterfactuals as related to experimental interventions can play an important role: in this sense,

> it is certainly plausible that counterfactual reasoning is important to the design and execution of experiments, and to the interpreting of data, modifying old hypotheses, developing new ones, and so on. (Bogen, 2005:416)[9]

Among others, counterfactuals allow a comparison between results obtained by means of actual experiments and results derived from ideal manipulations, i.e. interventions that for some reason cannot be performed in a particular time and place, or from mental experiments, which will never be carried out in practice. Interventionist counterfactuals can also improve our causal knowledge when we are faced with plausible competing mechanical accounts of the same observable behaviour, insofar as they can shed light on how a mechanism varies between different conditions, and set the limits between conditions in which the mechanism will continue to hold and function properly, and those in which the mechanism will break.

> Counterfactual reasoning can be *epistemically* important to the discovery of causal structures. But Counterfactualism is not an epistemological idea. It is an ontological idea, one piece of

[9] Bogen is still very critical of Woodward's requirement of modularity in interventions, which erroneously assumes the possibility of what Bogen labels "immaculate manipulation" (Bogen, 2004:19).

> conceptual analysis to the effect that there is no causality without counterfactual regularities. (Bogen, 2005:415).

What everything considered so far seems to suggest, however, is that counterfactuals can be employed in Woodward's sense, without necessarily becoming "counterfactualists".

Let me now return to Psillos' criticisms of Woodward's position. He points out that Woodward suggests some sort of relationship with intrinsic features exists that we exploit when aiming at bringing about *Y* by bringing about *X*. Woodward states:

> what matters for whether X causes […] Y is the 'intrinsic' character of the X-Y relationship, but the attractiveness of an intervention is precisely that it provides an extrinsic way of picking out or specifying this intrinsic feature. (Woodward, 2000:204)

Psillos takes Woodward to be drawing an explicit conceptual distinction between causation and invariance-under-intervention: an intrinsic feature is thought to exist by virtue of which a relationship is causal; invariance under intervention is considered its extrinsic symptom.

> So there is something more to causation – *qua* an intrinsic relation – than just invariance under intervention" (Psillos, 2004:302).

Although this is not maintained by Woodward, who remains ambiguous on the matter, couldn't we take the working of mechanisms as just this intrinsic feature? We have seen how Machamer's, Darden's and Craver's mechanical theory can admit of interventionist counterfactuals as a means to uncover mechanisms. What if in Woodward's counterfactual account, on the other hand, evidence-conditions were experimental tests, and truth-conditions were underlying mechanisms, which evidence-conditions reveal? Such an interpretation does not seem to clash with Woodward's general position. His view about the use of counterfactuals in connection with understanding causation is grounded in pragmatic and experimental considerations:

> when I say […] that good explanations should provide counterfactual information about what would happen to their explananda under interventions […], I mean information about

> what would really in fact happen, as an empirical matter (where this information might be provided by physics or some other relevant science or by experimental manipulation) under such interventions. (Woodward, 2006:58)

Among the goals he thinks counterfactuals should have, we can recall their being useful in solving problems, clarifying concepts and facilitating inference, and interventionist counterfactuals are deemed to have non-trivial truth values as long as we can describe how to test them. This way of conceiving of counterfactuals is entirely compatible with a mechanical approach: according to Woodward himself, there is *"no reason to believe that we can dispense with counterfactuals in understanding causation and explanatory claims"*(Woodward, 2004:48), but there is no reason to dispense with mechanisms either. He admits of them simply maintaining that, to explore the operation of a mechanism, the key-idea we will appeal to is that of invariance. Woodward thus suggests that a mechanism shall be an organised set of components, where the behaviour of a component must be described by an invariant under interventions generalization and each generalization must be changeable independently of the others[10].

Section 2 showed how recent mechanical theories can be seen as leaving room for experimental counterfactuals. Let me come full circle and conclude with some remarks on Salmon's view. As mentioned, counterfactuals were a major threat to his theory. This was why he embraced Dowe's conserved-quantity theory, where there is no trace of counterfactuals. Salmon, though *"with great philosophical regret"* (Salmon, 1997:18), appealed to counterfactuals to formulate his principle of mark transmission – providing a criterion to distinguish processes which qualify as causal from processes which do not – and his principle of causal interaction – proving a criterion to distinguish causal interactions from mere intersections. Salmon emphasizes that such distinctions are fully objective and warns against interpreting the use of counterfactuals in the opposite direction. Counterfactuals do not present serious difficulties, he believes, once they are linked with experiments:

---

[10] See (Woodward, 2002).

> science has a direct way of dealing with the kinds of counterfactual assertions we require, namely the experimental approach. In a well-designed experiment, the experimenter determines which conditions are to be fixed for purposes of the experiment and which are allowed to vary. The result of the experiment establishes some counterfactual statements as true and others as false under well-specified conditions. [In] the kinds of cases that concern us […] counterfactuals can readily be tested experimentally. (Salmon, 1984:149-150)

Thus, Salmon too addresses concerns over counterfactuals by interpreting them in an experimental sense. Like the other authors we have mentioned, Salmon stresses that causal processes and causal interactions are objectively present in the world long before any experiment is performed. Importantly, Salmon also invokes counterfactuals when formulating criteria for drawing a clear distinction between genuine causal processes and pseudo-processes and between genuinely causal interactions and mere intersections, that is, precisely when providing those strict criteria for identifying mechanisms' components and behaviours missing in the more recent mechanical accounts. A possible link between counterfactuals and causation as manipulation, however, is totally ignored by both Salmon and Dowe, who are not interested in such an aspect of causation. Counterfactuals for Salmon are to be used only insofar as they serve to identify causal mechanisms, which are to appear in causal explanations, his major concern.

> One of the most important insights of Salmon's extensive work on explanation is that it recognised that notions like difference-making and relevance are central to the ideas of cause and explanation, and that some elucidation of them was necessary if we were ever to construct an adequate treatment of causal explanation. It is regrettable that recent accounts of causal explanation in the mechanist tradition seem not to engage with this point. (Woodward, 2004:49)

In examining Glennan's and Machamer's, Darden's and Craver's approaches, I have searched for possible traces of such an engagement.

Darden and Craver have recently argued for the existence of a very strict relation between mechanisms and experimentation, a relation

which, through experimentation, we can extend to involve counterfactuals as well. Not only do they claim that mechanisms are often explored through experimentation; they also claim that

> the rise of the mechanical philosophy was closely associated with the rise of experimental science. The observable phenomena of the natural world are to be explained in terms of hidden mechanisms, and these mechanisms are to be inferred using well controlled experiments to sort *how-actually* from *how-possibly* descriptions of mechanisms. (Craver and Darden, 2005:236) (emphasis added)

Hence, the issue of experimentation, which was not a priority in Salmon's and Dowe's mechanical accounts, is gaining increasing importance within more recent mechanical theories, and could constitute a tentative bridge with a counterfactual approach to causation like Woodward's.

4. Concluding remarks. A different glimpse of the causal connexion

While Woodward believes that there is

> a fundamental split between, on the one hand, those (e.g. Salmon 1984; Dowe 2000) who think that explanation (and perhaps causation as well) has to do just with what actually happens, and those, like [himself] who think that causal and explanatory claims must be understood (at least in part) in terms of the counterfactual commitments that they carry. (Woodward, 2006:54)

Perhaps the split is narrower than one would be inclined to think, given that traces of the (allegedly) opposite attitude can be found on each side.

Facing possible objections, Woodward says:

> it might be claimed that the account I've offered captures aspects of how we test causal claims, it has nothing to do with the content of those claims. (Woodward, 2004:63)

I have tried to show how counterfactuals, interpreted in Woodward's experimentalist sense, have to do *both* with the tests *and* content of

causal claims, finding some role within mechanical theories as well. Tests tell us something about causal connexions: saying, for example, that, had the patient not been given an amount x of a given drug y, she would not have recovered, is to say something about the fact that a causal relation holds between the drug intake and the recovery. To explain adequately *why* the drug intake caused the recovery, we shall look for all the mechanical details involved in between the two events. To do this, in turn, we can wonder what would have happened if the patient had taken an amount x_1 or x_2 … of the drug, or if she had taken drug y_1 or y_2, and so forth. This all has to do with the content of the causal claim "taking an amount x of the drug y caused the recovery": testing a causal claim, i.e. saying under the variation of which features it would still hold, is to say something about its content.

Very recently, Ned Hall has claimed that causation, *"understood as a relation between events comes in at least two basic and fundamentally different varieties"* (Hall, 2004:225), one being dependence, i.e. counterfactual dependence, and the other production, i.e. an event C's bringing about an event E. Hall believes that two events can stand in a kind of causal relation which can be adequately explained by a counterfactual analysis, or they can stand in a completely different kind of causal relation which can be explicated by means of production. I have here maintained that a conceptual analysis of causation can be carried out by means of both a mechanical and counterfactual approach at the same time[11]. Counterfactuals let us know *that* a causal link holds between A and B, while mechanical accounts inform us about what exactly *goes on between* A and B; counterfactuals can be used to discover *what is causally relevant*, and mechanical accounts tell us *how* relevant entities perform their productive activities, i.e. how causal relevance translates into productive causality[12]. If associated with interventions, as Woodward suggests, counterfactuals can play an essential heuristic role and yet leave all the explanatory power of mechanisms unaffected. I

[11] These two by no means exhaust the possible approaches to causation. On causal pluralism and on the role of context, see also: Schaffer (2000); Hitchcock (2003); Cartwright (2004); Galavotti (forthcoming); Campaner and Galavotti (2007); Hitchcock (2007).

[12] See (Machamer, 2004:36).

believe Salmon would have approved of this, and so may the main contemporary supporters of the mechanical conception. Counterfactuals tell us neither everything nor enough about causation, yet they can help to open "black boxes" of nature. Although Psillos expresses a strong preference for counterfactuals, which he takes as more basic than mechanism, he also argues that

> if both, [counterfactual and mechanical], approaches work in tandem in practice, they can offer us a better understanding of aspects of Hume's secret connection, and hence a glimpse of it. (Psillos, 2004:291)

If we try to understand more and more deeply both how different views can genuinely work in tandem, and what role causation plays in the sciences in practice, we may realise that *"very often, the connexion is not secret at all."* (Glennan, 1996:68).

University of Bologna
Raffaella.Campaner@unibo.it

REFERENCES

Bechtel, W. and Abrahamsen, A. (2005), "Explanation: A Mechanist Alternative", in: Craver, C. and Darden, L. (eds.), Special Issue: "Mechanisms in Biology," *Studies in History and Philosophy of Biological and Biomedical Sciences,* 36, pp. 421-441.

Bechtel, W. and Richardson, R.C. (1993), *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Princeton: Princeton University Press.

Bogen, J. (2004), "Analyzing Causality: The Opposite of Counterfactual is Factual," *International Studies in the Philosophy of Science,* 18, pp. 3-26.

Bogen, J. (2005), "Regularities and Causality; Generalizations and Causal Explanations", in: Craver, C. and Darden, L. (eds.), Special Issue: "Mechanisms in Biology", *Studies in History and Philosophy of Biological and Biomedical Sciences,* 36, pp. 397-420.

Campaner, R. (forthcoming), "Causal Medical Reasoning. Therapies and Pre-emption", in: Galavotti, M.C., Scazzieri, R. and Suppes, P. (eds.), *Reasoning, Rationality and Probability*, Stanford: CSLI.

Campaner, R. and Galavotti, M.C. (2007), "Plurality in Causality", in: Machamer, P. and Wolters, G. (eds.), *Thinking about Causes*, Pittsburgh: University of Pittsburgh Press, pp. 178-199.

Cartwright, N. (2004), "Causation: One Word, Many Things," *Philosophy of Science*, 71, pp. 803-819.

Craver, C. (2001), "Role Functions, Mechanisms and Hierarchy," *Philosophy of Science,* 68, pp. 53-74.

Craver, C. (2002), "Interlevel Experiments and Multilevel Mechanisms in the Neuroscience of Memory," *Philosophy of Science,* 69, pp. S83-S97.

Craver, C. (2005), "Beyond Reduction: Mechanisms, Multifield Integration, and the Unity of Neuroscience," in: Craver, C. and Darden, L. (eds.), Special Issue: "Mechanisms in Biology," *Studies in History and Philosophy of Biological and Biomedical Sciences,* 36, pp. 373-397.

Craver, C. and Darden, L. (2001), "Discovering Mechanisms in Neurobiology: The Case of Spatial Memory," in: Machamer, P., Grush, R. and McLaughlin, P. (eds.), *Theory and Method in the Neurosciences* (pp. 112-137), Pittsburgh: Pittsburgh University Press.

Craver, C. and Darden, L. (2005), "Introduction: Mechanisms Then and Now," in: Craver, C. and Darden, L. (eds.), Special Issue: "Mechanisms in Biology," *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36, pp. 233-244.

Darden, L. (2002), "Strategies for Discovering Mechanisms: Schema Instantiation, Modular Subassembly, Forward/Backward Chaining," *Philosophy of Science,* 69, pp. S354-S365.

Darden, L. and Craver, C. (2002), "Strategies in the Interfield Discovery of the Mechanism of Protein Synthesis," *Studies in the History and Philosophy of Biology and Biomedical Sciences,* 22, pp. 1-28.

Dowe, P. (1992), "Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory," *Philosophy of Science,* 59, pp. 195-216.

Dowe, P. (1995), "Causality and Conserved Quantities: A Reply to Salmon," *Philosophy of Science,* 62, pp. 321-333.

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

Faye, J. (1984), "Causal Beliefs and Their Justification," in: Faye, J., Scheffler, U. and Urchs, M. (eds.), *Logic and Causal Reasoning* (pp.141-168), Berlin: Akademie Verlag.

Galavotti, M.C. (forthcoming), "Causal Pluralism and Context", in: Galavotti, M.C., Scazzieri, R. and Suppes, P. (eds.), *Reasoning, Rationality and Probability*, Stanford: CSLI.

Glennan, S. (1996), "Mechanisms and the Nature of Causation," *Erkenntnis* 44, pp. 49-71.

Glennan, S. (2002), "Rethinking Mechanical Explanation," *Philosophy of Science,* 69, pp. S342-S353.

Glennan, S. (2005), "Modeling Mechanisms," in: Craver C. and Darden, L. (eds.), Special Issue: "Mechanisms in Biology," *Studies in History and Philosophy of Biological and Biomedical Sciences,* 36, pp. 443-464.

Hall, N. (2004), "Two Concepts of Causation," in: Collins, J., Hall, N. and Paul, L. A. (eds.), *Causation and Counterfactuals*, London: The MIT Press, pp. 225-276.

Halpern, J.Y. and Pearl, J. (2005a), "Causes and Explanations: A Structural-Model Approach. Part I: Causes," *The British Journal for the Philosophy of Science,* 56, pp. 843-887.

Halpern, J.Y. and Pearl, J. (2005b), "Causes and Explanations: A Structural-Model Approach. Part II: Explanations," *The British Journal for the Philosophy of Science,* 56, pp. 889-911.

Hitchcock, C. (1995), "Discussion: Salmon on Explanatory Relevance," *Philosophy of Science,* 62, pp. 304-320.

Hitchcock, C. (2003), "Of Humean Bondage," *The British Journal for the Philosophy of Science,* 54, pp. 1-25.

Hitchcock, C. (2007), "How to Be a Causal Pluralist," in: Machamer, P. and Wolters, G. (eds.), *Thinking about Causes*, Pittsburgh: University of Pittsburgh Press, 200-221.

Humphreys, P. (2006) "Review Symposium. Invariance, Explanation and Understanding," *Metascience* 15, pp. 39-44.

Lewis, D. (1986), "Causation," in: *David Lewis's Philosophical Papers*, vol. II (pp. 159-213), Oxford: Oxford University Press.

Machamer, P. (2004), "Activities and Causation: The Metaphysics and Epistemology of Mechanisms," *International Studies in the Philosophy of Science,* 18, pp. 27-39.

Machamer, P., Darden, L. and Craver, C. (2000), "Thinking About Mechanisms," *Philosophy of Science,* 67, pp. 1-25.

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

Pedersen, S. A., Øhrstrøm, P. and Elvang-Gøransson, M. (1984), "The Structure of Causal Reasoning in Medicine and Engineering," in: Faye, J., Scheffler, U. and Urchs, M. (eds.), *Logic and Causal Reasoning* (pp. 231-253), Berlin: Akademie Verlag.

Psillos, S. (2004), "A Glimpse of the Secret Connexion: Harmonizing Mechanisms with Counterfactuals," *Perspectives on Science,* 12, pp. 288-319.

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

Salmon, W. (1989), "Four Decades of Scientific Explanation," in: Salmon, W. and Kitcher, P. (eds.), *Scientific Explanation, Minnesota Studies in the*

Philosophy of Science, vol. 13 (pp. 3-219), Minneapolis: University of Minnesota Press.

Salmon, W. (1997), "A Reply to two Critiques," *Philosophy of Science,* 54, pp. 461-477.

Salmon, W. (1998), *Causality and Explanation*, Oxford: Oxford University Press.

Schaffer, Jonathan (2000), "Trumping Pre-emption," *Journal of Philosophy,* 97, pp. 165-181.

Tabery, J. (2004), "Synthesizing Activities and Interactions in the Concept of a Mechanism," *Philosophy of Science,* 71, pp. 1-15.

Woodward, J. (1993), "Capacities and Invariance," in: Earman, J., Janis, A.I., Massey, G.J. and Rescher, N. (eds.), *Philosophical Problems of the Internal and External Worlds* (pp. 283-328), Pittsburgh: Pittsburgh University Press.

Woodward, J. (1997), "Explanation, Invariance and Intervention," *Philosophy of Science,* 64, pp. S26-S41.

Woodward, J. (2000), "Explanation and Invariance in the Special Sciences," *The British Journal for the Philosophy of Science,* 51, pp. 197-254.

Woodward, James (2001a), "Law and Explanation in Biology: Invariance is the Kind of Stability that Matters," *Philosophy of Science,* 68, pp. 1-20.

Woodward, James (2001b), "Probabilistic Causality, Direct Causes and Counterfactual Dependence," in: Galavotti, M.C., Suppes, P. and Costantini, D. (eds.), *Stochastic Causality* (pp. 39-63), Stanford: CSLI.

Woodward, J. (2002), "What is a Mechanism? A Counterfactual Account," *Philosophy of Science,* 69, pp. S366-S377.

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

Woodward, J. (2004), "Counterfactuals and Causal Explanation," *International Studies in the History and Philosophy of Science,* 18, pp. 41-72.

Woodward, J. (2006), "Review Symposium. Invariance, Explanation and Understanding. Author's Response," *Metascience,* 15, pp. 53-66.

Woodward, J. and Hitchcock, C. (2003a), "Explanatory Generalizations, Part I: A Counterfactual Account," *Noûs,* 37, pp. 1-24.

Woodward, J. and Hitchcock, C. (2003b), "Explanatory Generalizations, Part II: Plumbing Explanatory Depth," *Noûs,* 37, pp. 181-199.

Philosophica 77 (2006) pp. 45-68

CAUSATION, PLURALISM AND RESPONSIBILITY

Francis Longworth

ABSTRACT

Counterfactual theories of causation have had difficulty in delivering the intuitively correct verdicts for cases of causation involving preemption, without generating further counterexamples. Hall (2004) has offered a pluralistic theory of causation, according to which there are *two* concepts of causation: counterfactual dependence and production. Hall's theory does deliver the correct verdicts for many of the problematic kinds of preemption. It also deals successfully with cases of causation by omission, which have proved stubborn counterexamples to physical process theories of causation. Hall's theory therefore appears to be a significant improvement on extant univocal theories of causation, both physical and counterfactual. In this paper I present a series of counterexamples to Hall's theory. I also describe cases in which our causal judgments appear to be sensitive to moral considerations. It does not seem likely that conventional theories of causation, which attempt to situate causation in an objective metaphysical picture of the world, will ever accord with our intuitions in such cases. Finally, the notion of responsibility is considered, but rejected as an illuminating primitive for analyzing causation.

1. Introduction

Univocal theories of causation have struggled to account for cases of causation involving preemption, and cases of causation by omission. And attempts to refine the basic theories in order to give the right results in these cases frequently introduce new counterexamples. Hall (2004) has suggested that causation is not a univocal concept. According to his pluralistic theory, causation comes in two varieties: production and dependence. This account delivers the correct verdicts for the preemption cases and cases of causation by omission that have plagued extant univocal theories and therefore constitutes a major advance. While I am

broadly sympathetic to Hall's pluralistic approach, I will show that his theory faces a number of counterexamples.

I will proceed via an examination of a series of candidate counterexamples to Hall's analysis and relevant rival univocal analyses. In section 2, I begin with what has seemed to many philosophers to be the most promising approach to analyzing causation, based on the idea that effects depend counterfactually on their causes. I will show that such counterfactual theories of causation are subject to a variety of counterexamples. One response to these counterexamples would be to abandon the search for a counterfactual theory of causation and pursue some sort of local or intrinsic theory of causation. But such an approach also faces seemingly insurmountable difficulties: counterexamples involving causation by omission. In section 3, I provide a brief exposition of Hall's pluralistic theory, which attempts to circumvent these canonical counterexamples. In section 4, I present several counterexamples to Hall's theory. In section 5, I examine the thesis that our causal judgments take into account moral facts, and suggest that is unlikely that traditional metaphysical theories will ever deliver the intuitively correct results in such cases.

2. The failures of univocal theories of causation

2.1. Naive dependence

David Hume, in the *Enquiry Concerning Human Understanding* (1748), pointed out a link between causation and counterfactual dependence:

> [W]e may define a cause to be *an object followed by another, and where all the objects, similar to the first are followed by objects similar to the second*. Or, in other words, *where, if the first object had not been, the second never had existed*. (1748, Section VII, Part II).[1]

The first sentence in the above quotation expresses Hume's familiar constant conjunction theory of causation. The second (which, despite

[1] Hume's italics.

Hume's claim, is *not* logically equivalent to the first) expresses a counterfactual theory of causation: if, running counter to actual fact, the first object had not been, the second would not have existed. The second object therefore depends, counterfactually, on the first.

Following Hume, an initial counterfactual analysis (Naïve Dependence, ND) can be formulated as:

(ND) C is a cause of E if and only if E *counterfactually depends* on C. In other words, if C had not occurred, E would not have occurred.[2]

It is well known, however, that effects do not always depend counterfactually on their causes. Consider:

> Trainee and Supervisor: Trainee and Supervisor are on a mission to kill Victim. Trainee shoots first and Victim bleeds to death. Supervisor, observing that Trainee has fired, does not shoot. If Trainee hadn't shot, however, Supervisor would have stepped in and done so, again resulting in Victim's bleeding to death.[3]

Although Victim's bleeding to death would have depended on Trainee's shooting in the *absence* of Supervisor, Supervisor's presence breaks this dependence. Such cases, in which the actual cause *preempts* some redundant backup, are known as cases of 'preemption'. Preemption therefore presents a problem for those who wish to base an account of causation on counterfactual dependence. Let us call the lack of dependence in cases of preemption the "preemption problem". The preemption problem has caused great difficulties for the counterfactual analyst of causation; indeed much of the literature on counterfactual theories of causation is concerned with attempts to get around the preemption problem by adding further conditions that deliver the intuitively correct theoretical verdict that preemption *is bona fide* causation, but without thereby introducing any new counterexamples.

---

[2] In keeping with common current practice, let us take the relata in (ND) to be occurrent events rather than Hume's 'objects'.

[3] Adapted from Hitchcock (2001).

Let us look at the first notable attempt to solve the preemption problem: David Lewis's appeal to the thesis that causation is always *transitive*.

2.2. Counterfactual dependence and transitivity

Lewis's counterfactual theory, presented in his seminal "Causation" (1973), was the first significant advance on (ND). His theory relied heavily on the assumption that the causal relation is transitive. Call this thesis 'Transitivity'.

> Transitivity: Causation is a transitive relation; that is, if C causes D and D causes E, then C is also a cause of E.

It seems intuitively quite plausible that causation is transitive: think of a line of dominoes toppling one after the other: the first causes the second to fall, the second causes the third to fall, and it seems correct to say that the first domino's falling is also a cause of the third domino's falling. It doesn't seem unreasonable to expect that transitivity would hold generally. In fact, Transitivity may be one of our central 'platitudes' concerning causation. The core of Lewis's analysis of causation can be summarized as:

(L) C is a cause of E if and only if there is a chain of intermediate events D1…Dn between C and E such that E counterfactually depends upon Dn, Dn counterfactually depends upon Dn-1,… and D1 counterfactually depends upon C.

The truth conditions of the counterfactuals are given in terms of the similarity of possible worlds to the actual world. Lewis stipulates that counterfactuals must not *backtrack*: if we are considering a world in which some event D_n, in a chain of dependency $D_1 \ldots D_n$, did not occur, D_{n-1} would still have occurred; so too would D_{n-2}, and all the other intermediate events stretching back to (and including) C. This is because such a world is *closer* to the actual world than a possible world in which C, $D_1 \ldots D_{n-1}$ do not occur, according to Lewis's similarity metric for possible worlds. We are to understand the non-occurrence of D_n, Lewis

says, as a "minor miracle": D_n is to be cleanly excised from the causal history of E, with *no* disruption of prior events.

Transitivity enables (L) to get the right result for our case of preemption, Trainee and Supervisor. Trainee's shot *is* linked to Victim's dying by a chain of dependence. Consider the flight of Trainee's bullet through some particular intermediate point *en route* to Victim (call this event 'B'). Event B depends counterfactually on Trainee's firing; if Trainee had not shot, B would not have occurred. In addition, Victim's dying depends counterfactually on B. To see why this is so, note that if B had not occurred, *Trainee would still have shot* (the 'no backtracking' rule). Hence Supervisor would not have shot, and Victim would not have died. Hence Trainee's firing causes B and B causes Victim's death. Invoking Transitivity, Trainee caused Victim to die. Does this appeal to the supposed transitivity of causation solve the preemption problem? Unfortunately not. There are other varieties of preemption for which this strategy does not work. Consider:

> Billy and Suzy: Billy and Suzy each throw a rock at a bottle. Suzy's arrives first and the bottle shatters. Billy's rock arrives a split-second later, encountering only flying shards of glass.

It is intuitively obvious that Suzy's throwing rather than Billy's caused the bottle to shatter, but in this case, there is neither simple counterfactual dependence between Suzy's throwing and the bottle's shattering, nor a chain of counterfactual dependence between them. In contrast to early preemption, we *cannot* say that if Suzy's rock had not been at some intermediate position *en route* to the bottle, the bottle would not have shattered, because *Billy would still have thrown*. Billy's throwing is independent of Suzy's throwing. Hence we cannot use the 'no backtracking' rule to argue that if Suzy hadn't thrown, Billy would not have thrown. Billy and Suzy is therefore a counterexample to Lewis's theory (L).

In Trainee and Supervisor, the backup process is cut short by Trainee's shot (the actual cause), early on. In Billy and Suzy, however, the backup process (the approach of Billy's rock) is only terminated at a very late stage, by the occurrence of the effect itself (the bottle's shattering). For this reason, these two cases are instances of what are referred to as *early* and *late* preemption respectively. While Lewis was

able to deal with early preemption counterexamples, late preemption counterexamples stalled the counterfactual research program for many years.

There are a number of responses that one might make to the problem of late preemption.

2.3. Causation as a local intrinsic relation

One response has been to attempt to define causation as a spatiotemporally local or intrinsic relation. Lewis (1986a), in his discussion of 'quasi-dependence' focuses on this approach. Suzy's throw, according to this approach would count as a cause of the bottle's shattering in virtue of the spatiotemporally continuous local and intrinsic relation that exists between the two events (corresponding to the trajectory of Suzy's rock).[4] In a somewhat similar fashion, one might describe the relation between Suzy's throwing and the bottle's shattering in terms of physical processes, perhaps involving transfer or exchange of energy, momentum or some other physical quantity. Fair (1979), Sober (1984), Salmon (1984, 1994) and especially Dowe (1992, 2000) have explored such approaches. While this seems a very intuitive and promising solution to the problem of late preemption, as a general analysis it faces considerable difficulties. There appear to be many relationships that we intuitively call causal which are neither intrinsic nor local. The major class of counterexamples is causation by omission. Consider, for example:

> Gardener: My plants died when I was away on vacation. If my gardener had watered them, as he was supposed to have done, they would not have died.

It seems correct in this case to say that the gardener's failure to water the plants was a cause of their death – perhaps even *the* cause. Yet there is no obvious spatiotemporally continuous series of events that connects the gardener to the plants; we may assume that the gardener was never in the vicinity of the plants, and our intuition remains the same. Gardener is

[4] See Hall (2004, p.235-257) and Menzies (2001) for further discussion, and attendant problems, of this approach.

therefore a counterexample to univocal theories based on intrinsicness or locality.

After abandoning quasi-dependence, Lewis (2000, 2004) returned to a purely counterfactual approach, redefining causation as the ancestral of "influence", a more fine-grained version of counterfactual dependence. This approach offers a potential solution to the problem of late preemption. I will not discuss the details here, but see Hall (2004) and Menzies (2001) for convincing objections.

2.4. Holding fixed

Further attempts to solve the problem of late preemption within a counterfactual framework have recently been advanced, which involve the notion of *holding fixed* certain facts or events. Notice that in the cases of early and late preemption above (Trainee and Supervisor and Billy and Suzy respectively), while the effects do not depend on their causes *simpliciter*, they *do* depend on their causes if we hold fixed certain facts. Victim's bleeding to death *does* depend on Trainee's shooting if we hold fixed the (actual) fact that Supervisor doesn't fire. Similarly, the bottle's shattering *does* depend on Suzy's throwing, if we hold fixed the fact that Billy's rock does not hit the bottle. By holding the right facts fixed, we are thereby able to reveal the latent dependencies between cause and effect that are hidden by the presence of the preempted backups. One simple candidate formulation of a "holding-fixed" counterfactual theory is:

(HF) C is a cause of E if and only if E counterfactually depends on C, while holding fixed some fact G.

(HF) bears a close relation to familiar epistemic methods for discovering causes in science. The Galilean notion of experiment involves trying to reveal causal relationships by manipulating some candidate cause and looking for an anticipated effect, while holding fixed any potentially interfering factors. (HF) therefore has some initial plausibility, and is currently a popular strategy in the causation literature; accounts giving a central place to some version of (HF) have been proposed, most notably, by Hitchcock (2001) and Yablo (2002, 2004), and also by Pearl (2000), Halpern and Pearl (2001, 2005) and Woodward (2003).

The 'holding fixed' approach seems to introduce two particularly problematic new types of counterexample, however: 'switches' and 'self-canceling threats.' Consider the following example of switching:

> Two Trolleys: Two parallel rail tracks ('left' and 'right') run alongside one another towards a movable section of track that is connected to a single main track. The moveable section can be positioned so that it either connects the right or left subtrack to the main track (the movable section is initially connected to the right subtrack). Two trolleys are hurtling along (one on each subtrack) towards the movable section of track. If a switch is flipped, the left subtrack will be connected to the main track, and the trolley that was traveling down the left subtrack will continue its journey along the main track. If the switch is not flipped, the left trolley will derail, but the trolley that was traveling down the right subtrack will continue onto the main track. Victim is strapped to the main track just beyond the flipping point. As the trolleys are approaching the flipping point, Suzy flips the switch, which takes the left trolley onto the main track; the right trolley derails. The left trolley hits Victim, who is crushed. Had Suzy not flipped, the right trolley would have continued onto the main track and Victim would still have been crushed.

Intuitively, Suzy's flipping, which makes no difference whatsoever to Victim's fate, is not a cause of Victim's crushing. Yet according to (HF), Suzy's flipping *is* a cause. For if we hold fixed the actual fact that the trolley on the right subtrack does derail, Victim's crushing *does* depend on Suzy's flipping the switch, since if she does not do so, the trolley on the left subtrack would derail, and Victim would not be crushed.[5] Hence Two Trolleys is a counterexample to (HF).

The following case is an example of a self-canceling threat:

> Two Assassins: Captain and Assistant are on a mission to kill Victim. On spotting Victim, Captain yells "Fire!" and Assistant shoots at Victim. Victim overhears the order, and although the bullet almost hits him, he ducks just in time and survives

[5] Hall (2000) discusses some potential replies to a similar kind of switching counterexample. In my view, however, these replies are unconvincing; I do not have space to provide arguments here.

> unharmed... If Captain hadn't yelled "Fire!", Assistant would not have shot, and Victim would still have survived. If Victim had not ducked, however, he would have been hit by the bullet, and would not have survived.[6]

We do not intuitively feel that Captain's yelling "Fire!" is a *cause* of Victim's survival, yet holding fixed the fact that Assistant fired, if Captain hadn't yelled "Fire!", Victim would not have ducked, and consequently would not have survived. Hence, Victim's survival depends on Captain's yelling "Fire!", holding fixed Assistant's shooting, and therefore (HF) rules that Captain's yelling "Fire!" is a cause of Victim's survival.[7]

Self-canceling threats have the following structure: *C* introduces some threat to *E*, but at the same time also initiates some countermove that is successful in canceling the threat to *E*, and *E* consequently occurs. In Two Assassins, Captain's yelling "Fire!" poses a threat to Victim's survival, but at the same time, alerts Victim to the threat posed. Victim ducks, thus canceling the threat to his survival.[8]

Note incidentally, that the naïve dependence theory (ND) delivers the intuitively correct theoretical verdicts for Two Trolleys and Two

---

[6] Originally due to McDermott, but extensively discussed by Hitchcock (2003).

[7] Hitchcock (2003:9-11) reports (on the basis of informal surveys) that intuitions are either divided or unclear with regard to whether or not Captain's yelling "Fire!" is a *cause* of Victim's survival. I consider this intuition to be simply mistaken. In my experience, as soon as one reminds respondents that Assistant would not have fired had Captain not yelled "Fire!", they generally reverse their initial judgment. Such mistaken intuitions arise from a failure to take on board the stipulated facts of the case.

[8] One might attempt to reply to this counterexample by arguing that the intuition that Captain's yelling "Fire!" does not cause Victim's survival is mistaken. One could suggest, as Lewis (2004) has done, that *in general* assassination orders do not cause survivals, but that in this *particular* case, the order (the yell) did cause the survival. The mistake, Lewis argues, is a confusion of singular causation with general causation. I do not find this objection convincing; my intuition with regard to this particular case is still firm, even when taking note explicitly of Lewis's warning.

Assassins. Suzy's flipping is not a cause since Victim's crushing doesn't depend on the switching. Similarly, Victim's survival doesn't depend on Captain's yelling "Fire!". As is often the case, introduction of further technical conditions introduces new counterexamples that the simpler theory already dealt with satisfactorily.

To summarize, (ND) falls to cases of early preemption. (L), while delivering the correct verdict for early preemption, delivers the wrong verdict for late preemption. (HF), while delivering the correct verdicts for early and late preemption, introduces new counterexamples involving switches and self-canceling threats. Lastly, attempts to characterize the causal relation in intrinsic or local terms fall to cases of causation by omission.

Given the repeated failures of these initially promising univocal theories, a few philosophers, such as Hall (2004), Godfrey-Smith (forthcoming), Hitchcock (2003), and Cartwright (1999) have recently begun to explore pluralistic approaches to causation. In the next section, I focus on Hall's dualistic theory, which I consider to be the most fully developed and best defended of these approaches.

3. Hall's two concepts of causation

Hall (2004) proposes that there are *two* concepts of causation: production and dependence:

(TC) C is a cause of E if and only if (E depends on C) or (C produces E).[9]

Each disjunct is given a different analysis. Dependence is just counterfactual dependence (though without Lewis's addition of Transitivity). Hall does not attempt a definitive analysis of production, but says that "we evoke it when we say of an event *C* that it helps generate or bring about or produce another event E." Whatever production is, it is a local, intrinsic relation, which, Hall claims, will also

[9] Godfrey-Smith (forthcoming) has also emphasized the distinction between the 'difference-making' and productive aspects of causation.

turn out to be transitive. Hall tentatively advances the hypothesis that the producers of E are those events that are *minimally sufficient* for *E*, in appropriate circumstances, given the laws of nature. Sober (1984) suggests that this sort of productive relation might be usefully analyzed in terms of energy-momentum transfer. Dowe's 'conserved quantity exchange' is an alternative candidate for production. I will not pursue these possibilities here, but will assume that the notion of production is sufficiently intuitive for the purposes of this paper, and that we can recognize it when we see it.

Both disjuncts of (TC) are *sufficient* for *C* to be a cause of E. Note that dependence and production are frequently *co-instantiated*. For example, in paradigmatically causal billiard ball collisions, the motion of the second ball is both produced by the motion of the first *and* depends on it. We might call this relation 'productive dependence.'

It is worth noting that causation, as defined by Hall, is not ambiguous in the sense in which words like 'bat' and 'bank' are ambiguous. In these cases, the two disjuncts (e.g. river bank and savings bank) are generally not co-instantiated in the same particular; their extensions do not overlap: there are no individuals that are both river banks and savings banks. Production and dependence, on the other hand, very often *are* co-instantiated in the same particular, as in the billiard ball case. It appears to be an accident that we use the same word 'bat' for both the nocturnal flying mammal and a piece of sports equipment. The two meanings of word 'bat' are not related in any interesting way, and the two types of bat share few significant properties. 'Bat_1' and 'bat_2' are merely homonyms.[10] In Dutch, two different words are used: 'vleermuis' and 'knuppel' respectively.[11] In the case of causation, however, the production and dependence senses *do* seem to be closely related, and related in interesting ways. For example, they are both able to play similar roles in explanation, prediction, agential control, and so on. It is no accident that the same word 'causation' is used for both production and dependence. Causation exhibits *polysemy* rather than *homonymy*.

[10] Their shared properties are rather uninformative and do not seem central to the meaning of either homonym. For example, they both share the property of being physical objects.

[11] Thanks to an anonymous referee for this point.

The senses are so closely related that only trained philosophers, such as Hall, might think to tease them apart.

How does (TC) fare will regard to the canonical counterexamples to the major univocal theories? I will not attempt to provide a comprehensive survey here and will instead restrict myself to pointing out some of the major advantages of Hall's dualistic theory. (TC) takes care of both early and late preemption with impressive ease. In Trainee and Supervisor, Trainee's shot is a cause of Victim's death because of its local, productive relationship (via Trainee's bullet) with Victim's death, despite the absence of dependence. It is extremely plausible that when making intuitive judgments about preemption, it *is* this local productive relation that we pay attention to. This is a very natural psychological diagnosis of our intuition-forming process. Early and late preemption count as cases of causation *not* in virtue of any dependence of the effect on the cause (while holding fixed the redundant backup) as the counterfactualist would have it; rather, C is a cause of E in virtue of the *productive relation* between the two. (TC) handles our late preemption counterexample Billy and Suzy in exactly the same fashion.

(TC), since it does not need to appeal to the holding fixed strategy in order to deliver the intuitively correct verdicts in cases of preemption, has the great advantage of not thereby ruling in switching and self-canceling threats such as Two Trolleys and Two Assassins respectively. In switching and self-canceling threats, there is no dependence *simpliciter* between the putative cause and effect, and the latent dependencies that would be revealed by holding fixed certain facts *remain* hidden, as we desire.

(TC) also deals straightforwardly with causation by omission counterexamples such as Gardener that beset the physical process theories of Salmon (1984, 1994) and Dowe (1992, 2000). My gardener's not watering my plants caused their death in virtue of the *dependence relation* that links the two, despite the absence of any local physical process linking the two events.[12]

---

[12] Strictly, the death of the plants does not depend on the gardener's not watering them. The plants, being mortal, would have died sooner or later. In order to make the counterexample work, some other detrimental effect on the plants (due to their not being watered) should be chosen as the effect. Alternatively, we could precisify the effect (e.g. the plants' death at time t).

(TC) is thus an important advance on univocal counterfactual theories. In addition to (TC)'s success with recalcitrant counterexamples to univocal theories, Hall provides a more general argument in favor of splitting our concept of causation in two: it enables us to preserve what he takes to be several of our important platitudes about causation: locality, intrinsicness and transitivity. While locality, intrinsicness and transitivity do not apply in cases of omission, they always apply in cases of production.

4. Counterexamples to (TC)

It appears, however, that there are several counterexamples to Hall's theory; there are cases that exhibit neither production nor dependence, but which we intuitively judge to be causation, and cases that exhibit production and/or dependence, which we intuitively judge not to be causation. We therefore have reason to suspect that if causation *is* a non-univocal concept, Hall's non-univocal analysis is not quite the right one.

4.1. Causation with neither production nor dependence

Hall himself provides the following counterexample to (TC):

> Second Escort: Suzy is piloting a bomber on a mission to bomb a particular target. She is escorted on this mission by Billy in a second plane, and Mary in a third plane. Enemy's fighter approaches, intending to shoot down Suzy's bomber. Billy shoots before Enemy does, however, and Enemy's plane goes down in flames. Suzy proceeds to the target and completes the mission. If Billy hadn't shot down Enemy, Mary would have. (Hall, 2004).

Billy's action prevents Enemy from preventing Suzy's bombing (which Hall calls 'double prevention'). Even though no local process connects Billy's shooting with Suzy's Bombing, Hall claims that Billy's shooting

is intuitively a cause of the bombing.[13] This example is an instance of 'preempted double prevention'; Billy's prevention of Enemy's attempted prevention preempts Mary's prevention of it. In virtue of the lack of locality, there is no production, and, in addition, the introduction of the backup preventer Mary breaks the dependence between Billy's shooting and the bombing. Hence we have causation with neither production nor dependence, and Second Escort is therefore a counterexample to (TC). Hall leaves this type of example as important 'unfinished business' for his account.

Second Escort illustrates an important general point. If we make just *one* link in a transitive causal chain non-local, there will be no productive relation between C and E. By adding in a redundant backup, we can also remove any dependence. We can thus generate counterexamples to (TC) at will. Here is a similar counterexample:

> Victim's Plants: Trainee shoots Victim, who bleeds to death. If Trainee hadn't shot Victim, Supervisor would have done. Victim, having bled to death at the hands of Trainee, is now unable to water his plants, which subsequently die.

The death of the plants does not depend on Trainee's shot, since Supervisor would have shot Victim had Trainee not done so. There is no productive relation between Trainee's shot and the death of the plants either, since there is no local connection between them. Yet Trainee's shooting seems intuitively to be a cause of the plants' death.

It seems plausible that what we are doing psychologically when we make our intuitive judgments in these cases is the following: we naturally break the cases down into two discrete steps. In Second Escort, the first step consists of Billy's shooting down Enemy's plane. The second step consists of the absence of Enemy's attack on Suzy and the subsequent bombing. Intuitively, each of these constituent steps is clearly causal. We then implicitly link these two causal steps together in a chain and conclude that Billy's shooting was a cause of the bombing. We can tell a

---

[13] I must confess that I am not entirely confident of Hall's intuition regarding this case. If one is skeptical, the force of this counterexample is reduced. There are, however, several other clear counterexamples of similar form such as Victim's Plants.

similar story for Victim's Plants. This may seem a plausible account of our judgment processes in these cases, but it will not do as an analysis. Transitivity is *false*; there are counterexamples in which C causes D, and D causes E, but C is *not* a cause of E. In Two Trolleys, for example, Suzy's switching causes the left trolley to travel down the main track, and that trolley's travelling down the main track causes Victim's crushing. Yet we do not want to say that Suzy's switching causes Victim's crushing. Hence Transitivity fails.

Similarly, in Two Assassins, Captain's yelling "Fire!" causes Victim to duck, and his ducking causes him to survive. Hence, according to Transitivity, Captain's yelling "Fire!" causes Victim to survive. Again, this verdict is highly counterintuitive.

Moreover, even if there were some way of rescuing Transitivity, we could easily generate related counterexamples in which we could not appeal to this chaining strategy. This could be done by starting from an ordinary case of early preemption such as Trainee and Supervisor, and making the productive link non-local. For instance:

> Action at a Distance Guns: Trainee and Supervisor are armed with action-at-a-distance guns. Trainee shoots first and Victim vaporizes. If Trainee hadn't shot, Supervisor would have, and Victim would have been vaporized in exactly the same manner.

In this case there is neither production nor dependence, yet our intuition that Trainee's shooting is the cause of Victim's vaporization remains solid. Neither can we point to a chain of intuitively causal links.

A second general method of generating such counterexamples is to begin with an omission, and add in a redundant backup omission:

> Patricidal Brothers: Jack and Bobby are tired of waiting to inherit their father Joe's money and independently decide to do away with him. They both decide that the best way to kill Joe is to withhold his medication. Joe must take two pills every day (one red, one green) in order to keep him alive. Every evening, before going to bed, Jack leaves a red pill on the kitchen table for Joe, and Bobby leaves a green pill. One evening, Jack, unable to wait any longer for his inheritance, decides not to leave his red pill on the table, and retires for the evening. A few minutes later, Bobby, who has decided on the same course of action, notices that his brother Jack

> has not left his pill on the table. Bobby, not wanting to risk being incriminated for his father's death, leaves his green pill on the table as usual. But if Jack had left his red pill on the table, Bobby, wanting to guarantee Joe's demise, would not have left his green pill. Joe, deprived of his full dosage, dies shortly thereafter.[14]

It is perfectly clear that Jack's omission is a cause of Joe's death. Yet there is no productive relation between these two events. There is also no dependence, since Jack's omission merely preempted Bobby's omission. If Jack had left the red pill, Bobby would have withheld the green pill, and Joe would still have died. This case is interesting in that it challenges Hall's hunch that "there could be nothing more to causation by omission than counterfactual dependence."[15] In virtue of what then, does Jack's omission count as a cause, if not dependence? One is initially tempted to answer that it is counterfactual dependence, but holding fixed the fact that Bobby *did* deliver the green pill. Note that (HF) would also deliver the intuitively correct verdicts for Second Escort, Victim's Plants, Action at a Distance Guns. If we hold fixed the fact that Mary (the Second Escort) doesn't shoot, then the Bombing depends on Billy's shooting; if we hold fixed the fact that Supervisor doesn't shoot, the death of Victim's plants depends on Trainee's shooting; if we hold fixed the fact that Supervisor doesn't shoot his action-at-a-distance gun, Victim's being vaporized depends on Trainee's shooting. Yet we have seen that (HF) drags in switching and self-canceling threats as *bona fide* types of causation (e.g. Two Trolleys and Two Assassins). So where does this leave us? It appears that there is *something* right about (HF), yet as a univocal theory, it falls frustratingly short of universality.

4.2. Production and/or dependence without causation

Two Trolleys, as we have seen is a counterexample to Transitivity; Suzy's flipping of the switch is not a cause of Victim's crushing. However, it seems clear that there is an intrinsic, spatiotemporally

---

[14] Cases with an analogous structure can be constructed in which an individual refuses to vote for a certain proposition for which unanimity is required.

[15] Hall (2004).

continuous relation between Suzy's flipping and Victim's crushing, corresponding to the motion of the trolley from switch to Victim. Hence Two Trolleys is a counterexample to the sufficiency of production for causation. Note also that this example indicates that production (conceived of as a local, intrinsic relation) is not transitive, contrary to Hall's claim. Suzy's flipping causes the left trolley to travel down the main track, and this trolley's traveling down the main track causes Victim's crushing, yet Suzy's flipping is intuitively not a cause of Victim's crushing.

Dependence does not seem to be sufficient for causation either. Consider:

> Queen Elizabeth: My plants died when I was away on vacation. If Queen Elizabeth had watered them, they would not have died.

Clearly Queen Elizabeth's failure to water the plants is not a cause of their death. Yet if she had watered the plants, they would not have died. Hence the plants' death does depend counterfactually on whether or not the Queen waters them. Woodward (2003) has suggested that we do not judge Queen Elizabeth's omission to be a cause of the plants' death because we do not take it to be a *serious possibility* that (counterfactually) she would have watered them. This seems too strong, however. Consider the case of the chronically unreliable gardener, who has *never* remembered to water my plants. We would still want to say that his failure to water my plants was a cause of their death, even though, after a certain point, we would no longer take serious the possibility of his watering them.[16]

5. Causal judgments, moral facts and responsibility

Perhaps we judge the unreliable gardener's failure to water the plants to be a cause of their death because we consider him to be *morally to blame* for their death, perhaps in virtue of having violated some gardener's

[16] Beebee (2004) makes a similar point.

contract, for example. The idea that causal facts might be judged partly on the basis of moral facts seems to be supported by examples such as:

> Automobile Accident: Billy is driving steadily down a deserted highway when suddenly, without warning, a truck ploughs into the side of his car. It is later revealed that the driver of the truck was heavily intoxicated and had run a red light.

Was Billy's driving steadily down the deserted highway a cause of the accident? One instinctively replies 'no.' (Note that this example demonstrates that Hall's production and dependence are *not even jointly sufficient* for causation). If we change the example and replace the car and the truck by billiard balls, however, we *would* say that each ball's motion was a cause of the collision. Hence our causal judgments appear to be sensitive to the presence or absence of moral agents. It might be objected that billiard ball variant merely indicates that we have conflated causation with moral responsibility in Automobile Accident: that although Billy is not *morally* responsible for the accident, his driving steadily down the highway *is* a cause of the accident. Perhaps this is the correct judgment, although the intuition that Billy's driving down the highway is *not* a cause of the accident does appear to be quite robust. Why not, therefore, take this example to indicate that (human) causation is in part a moral concept?

Others have also proposed that our causal intuitions are sensitive to considerations of moral responsibility. For example, psychologists have found empirically that our causal judgments are influenced by moral facts (Alicke, 1992; Knobe, MS). And for certain kinds of omission, Beebee (2004) has argued convincingly that moral responsibility plays a role in our causal judgments. Obviously causation cannot involve moral responsibility if moral agents (humans) are not involved. For example:

> Automatic Watering Machine: My plants died when I was away on vacation. If my automatic watering machine hadn't broken down, they would not have died.

In this example, it is perfectly acceptable to say that the machine's breaking down is a cause of the plants' death, yet we cannot attribute any *moral* responsibility to the machine (at least not in the strict sense). Note,

however, that it does seem acceptable to claim that the machine's breaking down was *to blame* for the plants' death, at least in some loose sense. On this subject, Hitchcock (MS) has recently written:

> [I]n fault analysis in engineering, or in performing an autopsy, one is trying to discover which component of a complex system is *responsible* for the failure of that system to function; this type of responsibility is not literally moral, but is broadly normative in character. Given this role, it is not altogether surprising that our judgments of token causation are influenced by normative considerations.[17]

Theories of causation that attempt to fit causation into an objective metaphysics will struggle to deliver the intuitively correct results in cases where judgments of (moral) responsibility seem influential. It is hard to see what objective metaphysical difference there could be, for example, between my gardener's failure to water my plants and the Queen's. Hence it would seem that theories with this goal will have to be revisionary to some extent.

The examples discussed in this paper are summarized in the table below. For each case, the table displays whether or not the putative cause in question is intuitively a genuine cause (Int), whether or not the putative cause produces its putative effect (Prod), whether or not the putative effect depends on its putative cause (Dep), and the respective theoretical verdicts of (TC) and (HF). In addition, whether the putative cause is responsible (Resp) for the relevant effect, in the sense described above, is also presented. Hall's (TC) only delivers intuitively correct theoretical verdicts for Trainee and Supervisor, Billy and Suzy, Gardener, Two Assassins, Unreliable Gardener and Automatic Watering Machine. The other examples are all counterexamples to (TC). Counterexamples to the various theories are indicated by asterisks.

[17] My italics.

Ex	Int	Prod	Dep	(TC)	(HF)	Resp
Trainee & Supervisor	Yes	Yes	No*	Yes	Yes	Yes
Billy & Suzy	Yes	Yes	No*	Yes	Yes	Yes
Gardener	Yes	No*	Yes	Yes	Yes	Yes
Two Trolleys	No	Yes*	No	Yes*	Yes*	No
Two Assassins	No	No	No	No	Yes*	No
Second Escort	Yes	No*	No*	No*	Yes	Yes
Victim's Plants	Yes	No*	No*	No*	Yes	Yes
Action at a Distance Guns	Yes	No*	No*	No*	Yes	Yes
Patricidal Brothers	Yes	No*	No*	No*	Yes	Yes
Queen Elizabeth	No	No	Yes*	Yes*	Yes*	No
Unreliable Gardener	Yes	No*	Yes	Yes	Yes	Yes
Automatic Watering Machine	Yes	No*	Yes	Yes	Yes	Yes
Automobile Accident	No	Yes*	Yes*	Yes*	Yes*	No

Given the failure of (ND), (L), (HF), local process theories and (TC), what further options are available to the causal analyst? One might try to retain (TC) and attempt to find alternative means of ruling out Two Trolleys, Queen Elizabeth and Automobile Accident, and ruling *in* the *bona fide* cases of causation that display neither production nor dependence (Second Escort, Victim's Plants, Action at a Distance Guns and Patricidal Brothers). But it is not at all obvious how one might go about doing this.

A second option would be to try to persist with (HF) and to try to rule out switching, self-canceling threats, Queen Elizabeth and Automobile Accident as genuine cases of causation. This option would obviously lead us away from pluralism, however, and back towards a univocal counterfactual analysis. One might try to go the Lewisian route and argue that ordinary intuitions are simply mistaken in all of these cases; but that seems like an uphill battle.

Third, notice from the rightmost column of the table above that responsibility is, *by itself*, sufficient to distinguish between the genuinely causal examples above, and the non-causal. Might then the following 'analysis' of causation be satisfactory?

(R) C is a cause of E iff C is responsible for E.

Is (R), finally, an analysis of causation that evades all of the canonical counterexamples presented above? Or have we 'cheated' in some way? Is responsibility *really* a suitable primitive on which to base an analysis of causation? Or is responsibility too closely synonymous with causation to provide any real illumination of what causation is? Perhaps all we have really done when asking "Is C responsible for *E*?" is ask "Is *C* a cause of *E*?" *Moral* responsibility is clearly distinct from the broader normative notion of responsibility, and from causation: one can be responsible for setting off a booby-trapped bomb (and can cause it to explode) without being *morally* responsible for its exploding. But it is not clear that such a weakening of the notion of moral responsibility to responsibility *simpliciter* (as we would *need* to do to all for non-human cases of causation such as Automatic Watering Machine) leaves us with anything more than a mere synonym for causation. Indeed, it is hard to think of any cases in which *C* can be responsible for *E* without C causing *E*, and *vice-versa*. To a degree, of course, this is what we want from an analysis.

But one suspects that responsibility just falls into the same category as very near-synonyms for causation such as 'bringing about', and consequently does not really provide an illuminating reductive analysis of the concept.

6. Conclusion

Hall's pluralistic theory of causation appears to be a significant improvement on univocal theories of causation, handling several canonical counterexamples with ease. There are, however, several clear counterexamples to Hall's theory. In addition, our causal judgments appear to be sensitive to considerations of moral responsibility, and it does not seem likely that objective metaphysical theories of causation will ever accord with our intuitions in such cases. Such theories will therefore need to be somewhat revisionary. Lastly, the notion of responsibility is considered, but rejected, as an illuminating primitive for analyzing causation, since it appears to provide only an unenlightening synonym.

University of Birmingham
f.longworth@bham.ac.uk
Ohio University
longwort@ohio.du

REFERENCES

Alicke, M. (1992), "Culpable Causation," *Journal of Personality and Social Psychology*, 63, pp. 368-378.

Beebee, H. (2004), "Causing and Nothingness," in: Collins, J., Hall, N. and Paul, L. (eds.), *Causation and Counterfactuals* (pp. 291-308), Cambridge (MA): MIT Press.

Cartwright, Nancy (1999), *The Dappled World*, Oxford: Oxford University Press.

Dowe, P. (1992), "Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory," *Philosophy of Science,* 59, pp. 195-216.

Dowe, P. (2000), *Physical Causation*, New York: Cambridge University Press.

Fair, D. (1979), "Causation and the Flow of Energy," *Erkenntnis,* 14, pp. 219-250.

Godfrey-Smith, P. (forthcoming), "Causal Pluralism," in: Hitchcock, C., Beebee, H. and Menzies, P. (eds.), *Oxford Handbook of Causation*.

Hall, N. (2000) "Causation and the Price of Transitivity," *Journal of Philosophy,* 97, pp. 198-222.

Hall, N. (2004) "Two Concepts of Causation," in: Collins, J., Hall, N., and Paul, L.A. (eds.), *Causation and Counterfactuals* (pp. 225-276), Cambridge: MIT Press.

Halpern, J. and Pearl, J. (2001), "Causes and Explanations: A Structural-model Approach — Part I: Causes," *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence* (pp. 194-202), San Francisco: Morgan Kaufmann.

Halpern, J. and Pearl, J. (2005), "Causes and Explanations: A Structural-model Approach — Part I: Causes (expanded version)," *British Journal for the Philosophy of Science*, 56, pp. 843-887.

Hitchcock, C. (2001), "The Intransitivity of Causation Revealed in Equations and Graphs," *Journal of Philosophy,* 98, pp. 273-299.

Hitchcock, C. (2003), "Of Humean Bondage," *British Journal for the Philosophy of Science,* 54, pp. 1-25.

Hitchcock, C. (MS), "Prevention, Preemption, and the Principle of Sufficient Reason."

Hume, D. (1902), *An Enquiry Concerning Human Understanding*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press. First published 1748.

Knobe, J. (MS), "Attribution and Normativity: A Problem in the Philosophy of Social Psychology."

Lewis, D. (1973), "Causation," *Journal of Philosophy,* 70, pp. 556-567. Reprinted in: Lewis (1986b), pp. 159-172.

Lewis, D. (1986a), "Postscripts to 'Causation'," in: Lewis (1986b), pp. 159-213.

Lewis, D. (1986b), *Philosophical Papers*, Volume II. Oxford: Oxford University Press.

Lewis, D. (2000), "Causation as Influence," *Journal of Philosophy,* 97, pp. 182-197. Expanded version appears in Collins, J., Hall, N. and Paul, L.A. (eds.) (2004), *Causation and Counterfactuals* (pp. 75–106), Cambridge (MA): MIT Press.

Longworth, F. (2006), *Causation, Counterfactual Dependence and Pluralism*, Ph.D. Dissertation, University of Pittsburgh.

Menzies, P. (2001), "Counterfactual Theories of Causation," in: *Stanford Encyclopedia of Philosophy*.

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

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

Salmon, W. (1994), "Causality Without Counterfactuals," *Philosophy of Science,* 61, pp. 297-312.

Sober E. (1984), "Two Concepts of Cause," *PSA 1984*, vol. 2, pp. 405-424.

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

Yablo, S. (2002), "De Facto Dependence," *Journal of Philosophy,* 99, pp. 130-148.

Yablo, S. (2004), "Advertisement for a Sketch of an Outline of a Prototheory of Causation," in: Collins, J., Hall, N., and Paul, L.A. (eds.), *Causation and Counterfactuals* (pp. 119-137), Cambridge: MIT Press.

Philosophica 77 (2006) pp. 69-96

CAUSAL PLURALISM VERSUS EPISTEMIC CAUSALITY[1]

Jon Williamson

ABSTRACT

It is tempting to analyse causality in terms of just one of the indicators of causal relationships, e.g., mechanisms, probabilistic dependencies or independencies, counterfactual conditionals or agency considerations. While such an analysis will surely shed light on some aspect of our concept of cause, it will fail to capture the whole, rather multifarious, notion. So one might instead plump for pluralism: a different analysis for a different occasion. But we do not seem to have lots of different concepts of cause – just one eclectic notion. The resolution of this conundrum, I think, requires us to accept that our causal beliefs are generated by a wide variety of indicators, but to deny that this variety of indicators yields a variety of concepts of cause. This focus on the relation between evidence and causal beliefs leads to what I call *epistemic* causality. Under this view, certain causal beliefs are appropriate or rational on the basis of observed evidence; our notion of cause can be understood purely in terms of these rational beliefs. Causality, then, is a feature of our epistemic representation of the world, rather than of the world itself. This yields one, multifaceted notion of cause.

1. The indicators of causality

The indicators of causality are several and disparate. We base our causal claims on observed associations, observed independencies, temporal cues, known mechanisms, theoretical connections, experiments, controlled trials, other causal knowledge, intuitions about subjunctive conditionals, and more. In trying to understand the nature of causality it is reasonable to attempt to analyse causal connections in terms of one or

[1] This work was carried out as a part of the 'Mechanisms and Causality' project supported by the Leverhulme Trust. I am very grateful to Federica Russo and an anonymous referee for helpful comments.

other of these indicators. Thus we have a range of contemporary theories including mechanistic, probabilistic, counterfactual, and agency-based (the last three of which are often classed as *difference-making* accounts of causality, since they are based on the intuition that a cause should make a difference to its effects).

Unfortunately, these monistic theories of causality have great difficulty in accounting for the epistemology of causality. While such a theory does well at explaining how the particular indicator used in the analysis can be taken as evidence for causal claims, it has trouble explaining how other indicators can also count as evidence. Thus mechanistic theories (Salmon, 1998; Dowe, 2000) have little problem explaining how knowledge of mechanisms and physical theory can ground causal claims, but they have their work cut out explaining how, even when we know about the salient mechanisms, we seek further evidence, e.g., evidence of probabilistic dependencies or independencies. It is not enough to know of some chain of connections linking exchange rate and inflation, we want to know that exchange rate makes a difference to inflation before we claim that it is a cause. In contrast, probabilistic theories (see, e.g. Suppes, 1970) do well at accounting for probabilistic indicators, but poorly when it comes to mechanistic indicators. It was not enough to know that in samples the prevalence of smokers made a difference to the prevalence of lung cancer, we needed to know that the two are linked by a physical mechanism before the claim that smoking causes cancer could be accepted. Similarly, counterfactual (Lewis, 1973) and agency (Price, 1992) approaches struggle with respect to mechanisms. This epistemological problem is presented in more detail in Russo and Williamson (2007a).

Perhaps the main reason why we seek varied evidence for a causal claim is this. Causal claims have two uses: they are used for inference on the one hand and explanation on the other. The inferential use – making predictions, diagnoses and strategic decisions on the basis of causal claims – requires that a cause should typically make a difference to its effects, for otherwise information about the presence of a cause would tell us nothing about the presence of its effects and vice versa, and instigating a cause would not be a good strategy for achieving its effects. The explanatory use requires something more, namely some physical account of why the event in question happened. When asked for an explanation of an event, it is not enough to say that some other event

occurred and that there exists a difference-making relationship between the two, for that is no explanation at all – it leaves the question, why did the explaining event make a difference to the explained event? Inasmuch as we can answer 'why' questions at all, we do so at root by invoking physical theory, physical events and physical processes. If a causal story is to offer an explanation, it had better fit with physical theory and tell us a bit about the ultimate physical explanation. Hence causal claims need to say something about physical mechanisms as well as about difference-making.

The above epistemological problem for monistic accounts motivates the move to a less simplistic account of causality – an account that takes the full variety of causal indicators seriously. Pluralism is a step in this direction. However, I shall argue that it is the wrong step (section 2). Instead, I shall argue in section 3 that an *epistemi*c theory of causality offers the right way to handle the full range of indicators of causality. In section 4 I shall suggest that, in general, an epistemic theory of a complex concept can have more to offer than a simple-minded analysis of the concept in terms of a single indicator, or even a more sophisticated pluralist analysis. The appendix, section A, outlines a formal causal epistemology that forms a component of the epistemic theory of causality.

2. A plurality of causality?

The move to causal pluralism is often motivated by the inadequacies of contemporary monistic accounts of causality and incompatibilities between mechanistic and difference-making accounts. In section 1 I suggested that contemporary accounts lack a viable epistemology: mechanistic accounts make it a mystery as to why we back up our causal claims with evidence of difference-making *over and above* evidence of mechanisms, while difference-making accounts cannot explain why we seek evidence of mechanisms *as well as* evidence of difference-making. But there are other paths to pluralism that pick up on other inadequacies of monistic accounts. Hall (2004) argues that his favoured difference-making approach, the counterfactual theory, cannot account for basic features of causality (namely its transitivity, the spatio-temporal continuity of causal processes, and the causal character of a process

being determined by its intrinsic non-causal features), and that while a mechanistic approach can account for the latter properties it cannot account for counterfactual dependence being sufficient for causation, nor can it account for absences being causes and effects. Consequently – Hall (2004, section 6) claims – there are two concepts of cause, one which corresponds to counterfactual dependence and the other which corresponds to mechanistic production. Cartwright (2004, section 2) argues that contemporary accounts are incompatible with one another and that no individual account has universal applicability. She concludes that each account specifies a different kind of causal law.

Not only are there several paths to pluralism but there are also several varieties of pluralism. Pluralists agree that there is no single thing that is picked out by causality,[2] but that leaves plenty of scope for disagreement. Some (e.g. Psillos, 2006; Godfrey-Smith, 2008, section 6) do not think much more can be said about what causality is, while others (e.g. Hall, 2004) argue that there are distinct and coherent senses of cause and would like to understand each of these senses.

I take it that the former, nebulous variety of pluralism is a last resort. If one can't say much about the number and kinds of notions of cause then one can't say much about causality at all; this stance should only be adopted if there is no viable alternative. I do not think the latter, determinate variety of pluralism includes any viable alternative, for the reasons set out below. But I do think that there is a viable monistic account, as developed in section 3, so there is no need to resort to nebulous pluralism.

Of course those in the latter, determinate-pluralism camp differ substantially as to the number and nature of the senses of cause. They also differ as to the task at hand. One might think that it is enough to

---

[2] Note that under this conception, pluralism does not encompass the view that there are different *forms* of causal relationship – e.g., 'type cause' versus 'token cause', or 'component effect' versus 'net effect' (Hitchcock, 2001) – except where it is maintained that these different forms appeal to entirely different concepts of cause (Sober, 1985; Hitchcock, 2003); for a single concept of cause might be used in different ways to shed light on the various forms of relationship. Nor does this conception of pluralism encompass the view that there are different forms of causal *explanation* (Weber et al., 2005) – except where it is maintained that different forms of explanation require different notions of cause.

shed light on the alleged different notions of cause, e.g., by saying that there are two notions of cause, one mechanistic and the other difference-making. But one might want to go further by rendering the various notions precise, in order to explicate the notion of cause in the sense of Carnap (1950, section 2), or to provide a reductive analysis of cause. Further, one might want to delimit the proper zone of application of each concept of cause, e.g., by saying that mechanistic causality is appropriate in the natural sciences while difference-making causality is appropriate in the social sciences (a move analogous to the pluralism about probability of Gillies (2000, chapter 9)).

There are a number of problems with determinate pluralism. First, pluralism is not parsimonious – if, as I suggest in section 3, there is an adequate monistic account of cause, then arguably that account should be preferred purely on the grounds of parsimony. Second, while we have many words that are suggestive of causation, e.g., 'push' and 'pull' (Anscombe, 1971; Cartwright, 2004, section 3), we have only one word-stem 'cause' for the fully general notion.[3] If causality were a plural concept then one would think that we would have several word-stems, including one for each general notion. Perhaps, the pluralist might reply, this is just a case where our language has not adequately evolved to match our world. But if so, one would still expect some qualifiers (e.g., 'mechanistic', 'difference-making') to the word 'cause' to be routinely used to distinguish the types of cause under consideration. At the very least, one would expect clarificatory questions to be used to disambiguate uses of the word 'cause' (Godfrey-Smith, 2008, section 3). But all this is absent.

Third, it is clear that determinate pluralism will not do justice to the worry with which we began, namely the problem of accounting for causal epistemology. How can one explain the fact that there was excellent evidence that smoking and lung cancer sat in the right sort of difference-making relation, yet some suitable physical mechanism linking smoking and lung cancer was required before the causal claim could be established? How can one explain the fact that there is excellent

[3] Arguably 'prevent' is also fully general. However there is general consensus that 'cause' and 'prevent' are not different concepts. Either they are two sides of the same coin, causation being a positive relationship and prevention negative, or causation is taken to subsume prevention.

evidence of a mechanism linking smoking and breast cancer (involving the presence of breast cancer carcinogens in tobacco smoke), yet that causal claim has hitherto not been established because the evidence concerning difference-making is equivocal? The pluralist is stuck. If she says that some particular claim invokes a difference-making account of causality then she cannot explain the requirement of a mechanism. If she says that the claim involves a mechanistic account of causality then she can not explain the requirement of difference-making. If she says that the claim simultaneously involves two notions of cause – mechanistic and difference-making – then she is in danger of not being a pluralist at all, but of espousing a single concept of cause that has two necessary conditions, one mechanistic and one difference-making. The different aspects of causality – mechanistic and difference-making – are clearly connected, since a causal claim requires both as evidence. But pluralism sheds no light on this connection; if anything, it pushes these two aspects apart, viewing each as evidence for a different claim.

If determinate pluralism doesn't cut muster and contemporary monistic accounts also fail, then we need to go back to the drawing board. Our options are a more elaborate form of monism, or, as a last resort, nebulous pluralism. As a first attempt, as suggested above one might try to develop a monistic conception which takes a causal connection to require *both* an underlying mechanism *and* that the cause make a difference to the effect. Unfortunately, this will not do either, for the simple reason that not all of our causal claims have an underlying mechanism and not all of our claims reflect difference making – see, e.g. Hall (2004), who dismisses monism on these grounds. While we *seek* evidence of a mechanism as well as evidence of difference-making, such evidence is sometimes unattainable – this fact puts paid to a monistic analysis of causality in terms of one or other or both of these notions.

Perhaps, then, we must look to some less determinate account which appeals to a vague cluster of different notions that underlie a single concept of cause. This nebulous monism may be marginally more appealing than nebulous pluralism, but again not one to which we need resort, since, as we shall see, the epistemic theory offers a more determinate kind of monism.

3. The epistemic view of causality

In a sense causality is a very simple concept – it is just an asymmetric binary relation.[4] Therefore, it can only carry so much information. But we demand a lot of this relation. Causality is used throughout the sciences and in daily life for inference and for explanation: we represent the world causally so that we can make predictions, diagnose faults, make strategic decisions, explain events and apportion blame and praise. Thus we overload a simple relation with connotations both of difference-making and of mechanisms. As pluralists have observed, there are some tensions between the inferential and the explanatory uses of the causal relation. This explains the multi-faceted epistemology of causality and the apparent complexity of the notion of cause.

The *epistemic theory of causality* (Williamson, 2005a; Williamson, 2006a; Williamson, 2007a) takes causal epistemology as primary and builds up causal metaphysics from this epistemology. Arguably only by this process of reverse engineering can one address the epistemological problem of Section 1.

The epistemic theory takes an epistemology of *rational belief* as its starting point. The idea is that an agent's evidence constrains the range of *causal beliefs* that it would be rational for her to adopt. Some possible causal beliefs are incompatible with the evidence, others are suggested by the evidence; the agent should choose from the latter. These beliefs are just that – they are highly defeasible in the light of new evidence and nothing like as stable as causal knowledge. Nevertheless, this relation between evidence and rational causal belief is enough both to develop a full causal epistemology and to isolate the concept of cause itself.

Just what is this relation between evidence and rational causal belief? What causal beliefs should an agent adopt on the basis of her evidence? The answer to a question about what an agent should do

---

[4] Some have argued that causality is slightly more complicated than that – e.g., Schaffer (2005) argues that it is not binary, Mellor (1995) that it is not a relation, Spirtes (1995) that it is not acyclic – but these alternative conceptions are still simple enough to make the subsequent point, and the analysis of this section can be modified to take these complications into account if need be. Pluralists argue that causality is not a *single* relation, but I have argued in section 2 that pluralism is unacceptable.

hinges, of course, on what the action is intended to achieve – in this case, on the uses to which she will put her causal beliefs. The explanatory use of causality requires that the causal relation should typically fit with known physical theory and evidence of mechanisms: typically, some cause of an event will be invoked by its physical explanation. The inferential use of causality requires that the causal relation should typically fit with evidence of difference-making: typically, cause and effect should be probabilistically dependent, when intervening to fix the cause and controlling for the effect's other causes. There is also the rather general use of beliefs to systematise one's evidence: an agent's beliefs should typically be able to offer some kind of explanation of her experience and evidence. For example, if the agent discovers that two events are probabilistically dependent, and she knows of no non-causal explanation of this dependence (the events are not known to be overlapping, for instance) then she should (tentatively) believe that some causal connection between the events gives rise to the dependence, because dependencies between physical events are typically explained causally. This sketch involves a lot of 'typically's', because none of these features of causality hold invariably; if they did, a more straightforward analysis of causality in terms of one or more of these features might be possible; yet 'typically' is quite enough for causal beliefs to be useful from an inferential and explanatory point of view.

One way of making this sketch more precise proceeds as follows – a more detailed exposition is given in the Appendix and the motivation behind some of the assumptions can be found in Williamson (2005a). An agent's causal beliefs can be represented by a directed acyclic graph whose nodes are the variables of interest in her domain and whose arrows correspond to direct causal connections. As explained in Section 4, her *evidence* or *epistemic background*, β, which contains everything that the agent takes for granted in the context at hand, can be used to determine a probability function, p_β, over the variables in her domain – namely the probability function that satisfies constraints imposed by background knowledge but that is otherwise as non-committal as possible, i.e., that has maximum entropy. p_β represents the degrees of belief that the agent should adopt on the basis of β. The causal belief graph, $\mathcal{C}_\beta$, that the agent should adopt on the basis of β is determined as follows. First, $\mathcal{C}_\beta$ should be compatible with the constraints κ that are imposed by the agent's mechanistic and theoretical knowledge: e.g.,

causes should not occur after their effects; if physical theory treats two variables symmetrically then neither can be a cause of the other (for otherwise each would be a cause of the other, breaking the asymmetry of causality); if mechanisms indicate a common cause of two variables rather than a direct causal relation from one to the other then this should be reflected in the causal belief graph. Second, as long as it is not prohibited by the mechanistic-theoretical constraints κ, there should be an arrow $A \rightarrow B$ in the causal graph to account for each *strategic dependence* from A to B, i.e., whenever p_β renders A and B probabilistically dependent when intervening to fix A and controlling for B's other direct causes (i.e., whenever A and B are probabilistically dependent conditional on B's other direct causes and A's direct causes). Third, the agent's causal beliefs should otherwise be as non-committal as possible: there should be no arrows in $\mathcal{C}_\beta$ that are not warranted by evidence β.

Interestingly, standard methods can be used to determine $\mathcal{C}_\beta$:

THEOREM 3.1. $\mathcal{C}_\beta$ is a minimal graph that satisfies κ and the Causal Markov Condition (cf. Definition A.7), if there is such a graph at all.
PROOF: See Appendix.

There are a whole host of algorithms for finding minimal causal graphs that satisfy the Causal Markov Condition and some set of causal constraints (Korb and Nicholson, 2003, Appendix B). The system Hugin, for instance, offers a commercial implementation of a range of techniques (see Andersen et al, 1989; www.hugin.com). Thus these methods fit well with the above epistemology.

Once we have an epistemology that elucidates the relationship between evidence and rational causal belief, one can use this epistemology to determine the concept of cause itself via the following identity: the causal relation is just the causal belief graph of an omniscient rational agent (an agent whose evidence is exhaustive).

This identity can be understood in two ways. It could be thought of as a fact about a concept of cause on which we have an independent handle. For example, the proponent of a mechanistic analysis of cause might want to claim that, if we had full empirical evidence, our rational causal beliefs would coincide with this mechanistic relation. But this claim would be very hard to maintain, thanks to the epistemological

problem of section 1: it is implausible that, if we had full evidence of mechanisms, further probabilistic evidence should not alter our causal beliefs. Alternatively, one might think of this identity as constitutive of causality – there is no independent handle, causality just is a set of rational beliefs. It is this second understanding that forms the crux of the epistemic theory of causality. According to this epistemic view, the epistemology of causality is determined by the uses we put this relation to – inference and explanation – and causality itself is determined by this epistemology, and so is ultimately reducible only to its uses.

According to the epistemic theory, then, the causal relation is characterised by the causal beliefs that an omniscient rational agent should adopt. It should be clear in principle how this characterisation can overcome the epistemological problem that besets other accounts. The epistemological problem is that of developing an account of causality that fits with the following epistemological fact: in certain cases, one should not infer a causal connection solely on the basis of evidence of difference-making, or solely on the basis of evidence of mechanisms.[5] I.e., in such a case one would not be rational to hold the corresponding causal belief. If so, and if in such a case there were difference-making but no mechanism, or vice versa, then an omniscient rational agent would not hold the causal belief. Hence according to the epistemic theory there would be no such causal connection. On the other hand, if in such a case there were both difference-making and a mechanism then there would be a causal connection. So in such a case there is a causal connection if and only if there is both a mechanism and difference-making. Thus there is a tight fit between the epistemic theory of causality and the epistemological fact. The epistemological fact is accounted for by the uses of causal beliefs: the explanatory and inferential uses of causal claims require that, where possible, causal claims should coincide both with mechanisms and with difference-making.[6]

---

[5] As alluded to at the end of section 2, in other cases one kind of evidence is sufficient for the causal claim.

[6] A referee astutely pointed out that it is unclear how an omniscient agent could have any inferential needs for causal beliefs to satisfy. It is important to note that the uses of causal beliefs outlined here – explanation and inference – are not the uses to which every hypothetical agent puts such beliefs. Indeed they are not the

The epistemic view leaves us with an elaborate epistemology but a simple metaphysics: many indicators of causality but one concept of cause. That causal epistemology is pluralistic and somewhat elaborate is no news to anyone. That causality itself is monistic and rather straightforward, yet fits with this epistemology, is perhaps more surprising. Moreover, this is not nebulous monism. One of the advantages of this view is that it is somewhat easier to agree on an appropriate causal epistemology, elaborate though it is, than to agree on an appropriate understanding of cause, simple though it may turn out to be. Since the latter task can be reduced to the former, metaphysical progress becomes possible and a determinate monism is within reach. Indeed, as I hope the Appendix shows, a causal epistemology may be made very precise, in which case the monistic concept of cause is precisely defined too.

There is an interesting question concerning the objectivity of the causal relation under the epistemic view. How much choice does an agent have when deciding which causal beliefs to adopt? Clearly an agent can not choose just any directed acyclic graph as her causal belief graph $\mathcal{C}_\beta$. In fact, if there is a graph that satisfies κ and the Causal Markov Condition, and if p_β is faithful, then $\mathcal{C}_\beta$ must be chosen from a Markov equivalence class – a set of directed acyclic graphs that have the same independencies via the Causal Markov Condition (see Proposition A.17). Gillispie and Perlman (2002) carried out studies that suggest that on average a Markov equivalence class has four members, i.e., the agent will be able to choose the directions of two arrows in the graph, on average, and all other arrows will be determined by background knowledge. Thus causality is very highly determined on the epistemic account. This seems to be just what we want – by and large the causal relation is objectively determined, but there are cases (see e.g. Hitchock, 2003) that suggest that causality is not fully objective.

---

uses to which every real agent puts such beliefs, since there are agents that make no inferences or proffer no explanations, and many agents have ulterior (non-explanatory, non-inferential) motives for adopting certain causal beliefs. The uses emphasized here are the general epistemological uses to which the bulk of us (non-omniscient) humans put our causal beliefs.

4. Epistemic metaphysics

We have seen, then, that traditional monistic accounts of causality, which seek to analyse causality in terms of just one of its indicators, face a crucial epistemological problem, namely that of accounting for the variety of evidence required for causal claims. Determinate pluralism fares no better: it also succumbs to this epistemological problem. The epistemic account, on the other hand, provides a monistic theory that has causal epistemology at its base, and is not beset by this problem. If one can provide a determinate epistemology, such as that outlined in the Appendix, then this leads to a determinate monism about causality; there is no need to resort to indeterminate monism or indeterminate pluralism.

The epistemic theory of causality is an example of a general strategy for developing a determinate, monistic metaphysics that is true to the epistemology of a concept. Sometimes, attempts to explain a concept by positing a single mind-independent entity that corresponds to the concept meet fundamental difficulties, including counterexamples and epistemological problems. For example, probability faces an analogous epistemological problem: our probability judgements are based on knowledge of frequencies, knowledge of symmetries, indifference in the face of lack of knowledge and so on; if we try to analyse probability in terms of one of these indicators it is hard to explain the relevance of the others. More generally, mathematics faces an epistemological problem: our mathematical claims are based on a vast panoply of evidence, including proofs, patterns and pictures; monistic views such as platonism face well known difficulties in accounting for this epistemology (Benacerraf, 1973). In these cases, simple-minded realism offers a poor account of the concept in question and some other kind of account is needed.

An *epistemic theory* of concept X is a good strategy in such cases, one that can be used to provide an account of X that is truer to its epistemology and less prone to counterexamples. According to such a theory, X is to be interpreted in terms of a rational agent's epistemic state: rational X-beliefs are determined by an agent's epistemic background (and the uses to which X-beliefs are put); X-facts are characterised by those X-beliefs that an omniscient agent ought to adopt.

Objective Bayesianism provides an example of an epistemic theory of probability (see e.g. Williamson, 2005a, chapter 5). Here X-beliefs are

an agent's degrees of belief, which, as Bayesian argue, satisfy the axioms of probability. These *X*-beliefs are determined by an agent's background knowledge β as follows. Degrees of belief are used for inference, e.g., prediction and decision. Given the predictive use, one's degrees of belief should be calibrated with one's evidence. Thus knowledge of frequencies directly constrains degrees of belief (if you know just that 80% of days like today are followed by rain you should believe today will be followed by rain to degree 0.8). So does knowledge of symmetries (if you know that accepted physical theory treats the different possible values of a particle's spin symmetrically, you should believe a particle has spin up to the same degree that you believe it has spin down). Given the decision-making use of degrees of belief, they should not be susceptible to a Dutch book and should not be bolder than is warranted by evidence (for otherwise one opens oneself up to unnecessary risk – see Williamson (2007b)). Thus on a finite domain, an agent's degrees of belief are represented by the probability function p_β, from all those that satisfy constraints imposed by β, that is most non-committal (i.e., has maximum entropy). Probability facts are then determined by these probability beliefs. The probability facts at time *t* are characterised by the probability function that an agent with knowledge of everything up to time *t* should adopt as her belief function. Note that there are differences between this epistemic theory of probability and the epistemic theory of causality. In particular, since probability has been axiomatised and shown to have several models, a certain pluralism is inevitable. Thus objective Bayesianism may be used to provide an account of the probability of a single case while the frequency theory of von Mises (1928) may be used to explicate the probability attaching to an indefinitely repeated sequence of outcomes (see also Russo and Williamson (2007b) on this point).[7]

An epistemic theory of mathematics proceeds similarly (Williamson, 2006b). Like causal beliefs, mathematical beliefs are used for explanation as well as inference. The explanatory use requires that

[7] Under the epistemic view, *X*-beliefs are a *type* of belief (a directed relational belief in the causal case, a degree of belief in the probabilistic case) and are to be distinguished from beliefs *about X*, which are beliefs about which *X* -beliefs one should adopt if one had total evidence. An epistemic account understands *X* in terms of *X* -beliefs, not in terms of beliefs about *X* (the latter approach would lead to problems of circularity – see Williamson (2007a), section 7).

mathematical beliefs be justifiable by means of proofs and interpretations, and that mathematical beliefs account for the evidence and be non-committal in other respects. The inferential use also imposes constraints: e.g., if a proof of a proposition is available as evidence then the proposition should be believed. Under the epistemic view, mathematical facts are characterised by the mathematical beliefs that an agent with full evidence should adopt. This gives a grounding to mathematics that is radically different to contemporary accounts such as platonism, neologicism, structuralism and nominalism.

The epistemic formula may also be applied to other problematic *X*'s, e.g., logic, ethics. It is the failure of standard accounts of *X* which motivates the move to an epistemic account, not some global pragmatism or some modified criteria for adopting philosophical theories. In particular inference to the best explanation, a favoured mode of inference of the monistic realist, can be used to motivate an epistemic view of *X*. A realist conception of *X* may simply be untenable – prone to counterexamples or unable to account for the epistemology of *X*, for instance. In which case it does not offer the best explanation for our having the concept. There is thus room for an epistemic theory to provide the best explanation of our having *X*: we have concept *X* because of its utility (e.g., for inference and explanation), not because *X* corresponds to some single non-epistemic thing, just as we have hands because of their utility, not because of some kind of correspondence. Of course, the world must be such that *X* is a useful concept – just as the world must be such that hands are useful – so our having *X* says something about the world. It just does not say that there is something *X*-like in the world that is picked out by our concept. According to this stance, epistemic theories are to be judged by the same criteria as realist theories (see Williamson (2006a) for potential criteria). If a realist theory of *X* is viable, it may then be preferred over an epistemic theory of *X* on the grounds of simplicity. Thus we have the concept of table because there are tables that the concept picks out but we have the concept of cause because of its inferential and explanatory utility.

University of Kent
J.Williamson@kent.ac.uk

APPENDIX: A FORMAL EPISTEMOLOGY

This appendix provides the details of a formal epistemology that can be integrated with the epistemic theory of causality. See Williamson (2005a, chapter 9) for further discussion of the motivation behind this framework.

As in section 3, let β be the agent's background knowledge. We shall suppose that β can be represented by two components, π which is a set of probabilistic evidence, and κ which is a set of causal constraints, determined by the agent's other knowledge, including mechanistic and theoretical knowledge. For example, $\kappa = \{A \rightarrow B, A \not\rightarrow C, A \rightsquigarrow C, C \not\rightsquigarrow D\}$, where $A \rightarrow B$ signifies that A is a direct cause of B , $A \not\rightarrow C$ that A is not a direct cause of C, $A \rightsquigarrow C$ that A is a cause of C, and $C \not\rightsquigarrow D$ that C is not a cause of D. Let $p_{\kappa,\pi}$ be the probability function, from all those that satisfy constraints imposed by κ, π, that has maximum entropy (see Williamson (2005a, section 5.8) for an account of how causal knowledge constrains a probability function). All probability assertions will be made with respect to this probability function. For sets of variables X, Y, Z, $X \perp\!\!\!\perp Y \mid Z$ signifies that X and Y are probabilistically independent conditional on Z, while $X \not\perp\!\!\!\perp Y \mid Z$ signifies the opposite, that X and Y are probabilistically dependent conditional on Z. We are interested in determining $\mathcal{C}_{\kappa,\pi}$, a directed acyclic graph on the domain of $p_{\kappa,\pi}$ that represents the causal beliefs that the agent should adopt on the basis of κ and π (arrows in the graph correspond to direct causal connections). Any causal graph $\mathcal{C}$ will be assumed to be a directed acyclic graph (dag). With respect to such a graph, D_A is the set of direct causes of variable A and NE_A is the set of A's non-effects. The question arises first as to how the agent's probabilistic knowledge π constrains choice of causal graph: what set κ' of causal constraints is imposed by probabilistic knowledge π?

DEFINITION A.1. (STRATEGIC DEPENDENCE) There is a *strategic dependence* from variable A to variable B with respect to probability function p and causal graph $\mathcal{C}$, written $A \rightrightarrows B$, iff A and B are

probabilistically dependent conditional on B 's other direct causes and A 's direct causes, $A \rightleftharpoons B \mid D_B \backslash A, D_A$.[8]

The following definition and principle allow one to translate probabilistic constraints π into causal constraints κ':

DEFINITION A.2. (CAUSAL TRANSFER) Let $\kappa^* =_{df} \{A \rightarrow B : A \rightrightarrows B\}$. Given a causal graph $\mathcal{C}$ that satisfies κ, a *causal transfer* of π with respect to κ and $\mathcal{C}$ is a maximal subset κ' of κ^* such that $\mathcal{C}$ satisfies κ and κ' (i.e., $\mathcal{C}$ satisfies κ and $\kappa' \subseteq \kappa^*$, and there is no κ'' such that $\kappa' \subset \kappa'' \subseteq \kappa^*$ and $\mathcal{C}$ satisfies κ and κ''.

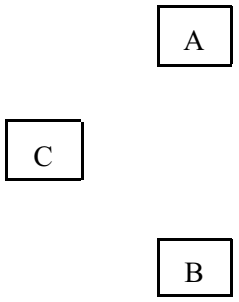

FIGURE 1: An empty graph.

[8] This definition is a bit simpler than that given in Williamson (2005a, section 9.5), but nothing very much hangs on the difference between the two definitions, other than the simplicity of some proofs.

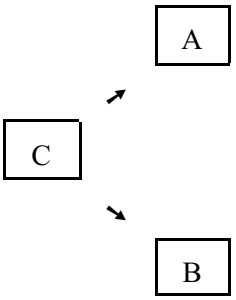

FIGURE 2: A common cause.

PRINCIPLE A.3. (PROBABILISTIC TO CAUSAL TRANSFER) $\mathscr{C}$ satisfies κ and π if and only if $\mathscr{C}$ satisfies κ and κ' , where κ' is some causal transfer of π with respect to κ and $\mathscr{C}$.

DEFINITION A.4. (EXPLANATORY RESIDUE) $\overline{\kappa} =_{df} \kappa^* \backslash \kappa'$ is the *(explanatory) residue* of $\mathscr{C}$, with respect to κ and π.

The smaller the residue the fewer the strategic dependencies that have no causal explanation in $\mathscr{C}$. Let $\mathbb{C}[\kappa,\pi]$ be the set of all causal graphs that satisfy constraints imposed by κ and π and that have smallest residue, $\mathbb{C}[\kappa,\pi] = \{\mathscr{C} : \mathscr{C}$ satisfies κ and π, $\mathscr{C}$ minimises $|\overline{\kappa}|\}$.

PRINCIPLE A.5. (RATIONAL CAUSAL BELIEF) An agent's rational causal belief graph $\mathscr{C}_{\kappa,\pi}$ should be chosen from the set $\mathbb{C}_{\kappa,\pi}$ of all minimal graphs in the set $\mathbb{C}[\kappa,\pi]$ of all minimum-residue causal graphs that satisfy constraints imposed by κ and π, $\mathbb{C}_{\kappa,\pi} = \{\mathscr{C} \in \mathbb{C}[\kappa,\pi] : \mathscr{C}$ has fewest arrows$\}$.

Thus $\mathscr{C}_{\kappa,\pi}$ is determined by first isolating the graphs that satisfy the constraints, then eliminating those that do not have minimum residue, then eliminating those that do not have the minimum number of arrows, then choosing one of the remaining graphs.

Here we have a minor point of departure from the approach of Williamson (2005a, section 9.5). There it was suggested that $\mathscr{C}_{\kappa,\pi}$ be determined simply by choosing a minimal graph from all those that satisfy the constraints. Here, we have an extra condition, namely that the

residue be minimised. This condition is motivated by the following example.

EXAMPLE A.6. Suppose the domain consists of three binary variables, $V = \{A, B, C\}$, with possible assignments a^0, a^1, for A , b^0, b^1, for B, and c^0, c^1, for C. Suppose that π $\{p(b^1 \mid a^1) \geq p(b^1) + 0.3\}$, and that $\kappa = \{A \not\rightarrow B, B \not\rightarrow A\}$. Then $A \rightleftharpoons B$ is the only dependence in $p_{\kappa,\pi}$, and the empty graph, Fig. 1, satisfies the constraints with residue $\overline{\kappa} = \{A \rightarrow B, B \rightarrow A\}$. On the other hand, the graph Fig. 2 also satisfies the constraints with no residue. Intuitively the latter graph is to be preferred, even though it has more arrows, because it includes an explanation of the dependence between A and B – it attributes the dependence to a common cause.[9] Thus residues should be taken into account.

Having isolated a rational causal belief graph, we turn to the question of how to find such a graph in practice. Clearly an exhaustive search through the space of all directed acyclic graphs will not be practical. Practical methods will make use of the following condition:

DEFINITION A.7. (CAUSAL MARKOV CONDITION) The *Causal Markov Condition* (CMC) is said to hold if each variable A in the domain is probabilistically independent of its non-effects, conditional on its direct causes, $A \perp\!\!\!\perp NE_A \mid D_A$.

---

[9] Two remarks are in order. First, it might be objected that background knowledge may include a non-causal explanation of the dependence, in which case one would not want any form of causal explanation of the dependence. But if that were the case then κ would rule out a common-causal explanation too, e.g., $\kappa = \{A \not\rightarrow B, B \not\rightarrow A, \ \forall X \neg(X \rightarrow A \wedge X \rightarrow B)\}$, and Fig. 2 would then not be adopted. Second, one might object that the common causal explanation doesn't really account for the dependence between A and B because the cause C is independent of these variables. But it must be remembered that this independence is with respect to rational degree of belief, i.e., with respect to current evidence. This leaves open the question of whether there are dependencies between C and A and between C and B with respect to their frequencies. If so, as evidence improves, rational degree of belief may be expected to reflect those dependencies; the explanation of the dependence then improves. Thus a causal picture is only by itself part of an explanation of a dependence – a full explanation would need to appeal to probabilities based on good evidence.

LEMMA A.8. For any causal graph $\mathcal{C}$, $\mathcal{C}$ satisfies $\kappa^* \Leftrightarrow$ satisfies CMC.

PROOF: [$\Rightarrow$] Suppose $\mathcal{C}$ satisfies κ^*. Suppose for contradiction that $\mathcal{C}$ does not satisfy CMC. Then there is some variable A and non-effect B such that $A \nrightarrow B \mid D_A$. This implies $A \nrightarrow B, D_B \mid D_A$ by the contrapositive of the Decomposition property of conditional independence (see, e.g., Williamson (2005a, section 3.2)), which in turn implies $A \nrightarrow B \mid D_A, D_B$ by the contrapositive of the Contraction property.

Since $D_B = D_B \backslash A$, $A \rightrightarrows B$. But this contradicts the assumption that $\mathcal{C}$ satisfies κ^*, since $A \not\dashrightarrow B$ in $\mathcal{C}$. Thus $\mathcal{C}$ does satisfy CMC after all.

[$\Leftarrow$] Suppose $\mathcal{C}$ satisfies CMC. Suppose for contradiction that $A \rightrightarrows B$ but that $A \not\dashrightarrow B$ in $\mathcal{C}$. There are four cases:
(i) If B is an (indirect) effect of A then CMC $\Rightarrow B \perp\!\!\!\perp A, D_A \mid D_B \Rightarrow B \perp\!\!\!\perp A \mid D_A, D_B$ (by the Weak Union property) which contradicts $A \rightrightarrows B$.
(ii) If A is an indirect effect of B then CMC $\Rightarrow B \perp\!\!\!\perp A, D_B \mid D_A \Rightarrow B \perp\!\!\!\perp A \mid D_A, D_B$ contradicting $A \rightrightarrows B$.
(iii) If A is a direct effect of B then $A \rightrightarrows B$ implies $B \nrightarrow A \mid D_B, D_A$ which is impossible since $B \in D_A$.
(iv) If neither is a cause of the other then CMC $\Rightarrow B \perp\!\!\!\perp A, D_B \mid D_A \Rightarrow B \perp\!\!\!\perp A \mid D_A, D_B$ contradicting $A \rightrightarrows B$.
Thus in each case we have the required contradiction. □

We come now to a restatement of Theorem 3.1.:

THEOREM A.9. Suppose there is some graph $\mathcal{C}$ that satisfies κ and CMC. Then $\mathcal{C}_{\kappa,\pi}$ is a minimal such graph.

PROOF: If $\mathcal{C}$ satisfies κ and CMC then by Lemma A.8 it satisfies κ and κ^*. Thus $\mathcal{C}$ has null residue. Hence $\mathbb{C}[\kappa,\pi] = \{\mathcal{C} : \mathcal{C}$ satisfies κ and CMC$\}$. The result follows by Rational Causal Belief, Principle A.5. □

DEFINITION A.10. (STRATEGIC CONSISTENCY) If there is a causal graph that has no residue with respect to κ and π (equivalently, if there is a

graph that satisfies κ and CMC) then κ and π are said to be *strategically consistent*.[10]

If κ is not strategically consistent with π then all is not lost. Suppose κ is consistent (i.e., there is some directed acyclic graph that satisfies κ), contains only direct-causal constraints (i.e., constraints such as $X \rightarrow Y$ or $X \not\rightarrow Y$ that involve only direct causal connections), contains only atomic constraints (i.e., no logically complex constraints such as $(X \rightarrow Y) \vee (Y \not\rightarrow Z)$, and contains no repetitions. Then we can write $\kappa = \kappa^+ \cup \kappa^-$ where κ^+ is the set of positive constraints in κ and κ^- is the set of negative constraints in κ. Consider the following algorithm:

ALGORITHM A.11.
Input: κ (consistent; atomic direct-causal constraints; no repetitions), π, $p_{\kappa,\pi}$.

1. Choose a maximal set of constraints λ such that $\kappa^+ \subseteq \lambda \subseteq \kappa$ and there is some $\mathcal{C}$ that satisfies λ and CMC.
2. Take a minimal such $\mathcal{C}$.
3. Remove arrows from $\mathcal{C}$ to satisfy the constraints in $\kappa \backslash \lambda$ and yield a graph $\mathcal{C}'$.

Output: $\mathcal{C}'$

THEOREM A.12. Suppose κ is consistent and contains atomic direct-causal constraints with no repetitions. Then $\mathcal{C}_{\kappa,\pi}$ can be taken to be $\mathcal{C}'$ produced by the above algorithm.

PROOF: First we need to show that such a λ exists. By Lemma A.8, $\mathcal{C}$ satisfies λ and CMC if and only if $\mathcal{C}$ satisfies λ and κ^*. There is such a λ because the complete directed acyclic graph is bound to satisfy κ^+ and κ^*.

By consistency of κ and construction of $\mathcal{C}'$, the graph $\mathcal{C}'$ satisfies the constraints in κ and π.

---

[10] This differs from the definition of strategic consistency given in Williamson (2005a, section 9.6) but the sentiment is the same – κ is strategically consistent with π if κ does not block the transfer of strategic dependencies to arrows in Principle A.3.

By maximality of λ and atomicity of κ, each constraint in $\kappa\backslash\lambda$ violates a single constraint in $\lambda \cup \kappa^*$. Since κ contains no repetitions, each constraint in $\kappa\backslash\lambda$ violates a *different* constraint in $\lambda \cup \kappa^*$. Since κ is consistent, each constraint in $\kappa\backslash\lambda$ must violate a different constraint in κ^*. Hence the size of the residue of $\mathcal{C}'$ (with respect to κ, π is $|\kappa\backslash\lambda|$.

$\mathcal{C}'$ must be a minimum-residue graph because λ is maximal.

Finally, $\mathcal{C}'$ must be a minimal minimum-residue graph. This is because the choice of maximal λ makes no difference to the size of $\mathcal{C}'$.

Note that if there is some graph $\mathcal{C}$ that satisfies κ and CMC, then the algorithm reduces to that of Theorem A.9. □

The assumption that κ contain only atomic direct-causal statements is quite restrictive: while we often know claims of the form '*X* is a cause of *Y*' or '*X* is not a cause of *Y*', it is rarer that causal knowledge takes the form '*X* is a direct cause of *Y*' or '*X* is not a direct cause of *Y*'. Thus it would be much more useful to be able to include atomic causal statements, so that κ contains atomic statements of the form $X \rightarrow Y$, $X \not\rightarrow Y$, $X \dashrightarrow Y$, $X \not\dashrightarrow Y$. Unfortunately the above algorithm is not guaranteed to succeed if we extend κ in this way. Suppose $\kappa = \{E \not\rightarrow A, B \not\rightarrow E, C \not\rightarrow E\}$, and the only minimal graphs that satisfy CMC are Fig. 3 and Fig. 4. Then $\lambda = \{B \not\rightarrow E, C \not\rightarrow E\}$ and $\mathcal{C}'$ is determined by removing arrows from Fig. 4 to give, e.g., Fig. 5, which has a residue of size 3. However, the rational causal graph is obtained by removing arrows from Fig. 3 to give Fig. 6 which has residue of size 2.

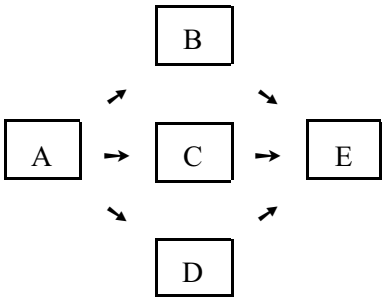

FIGURE 3: One graph satisfying CMC.

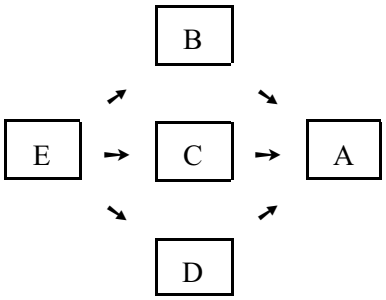

FIGURE 4: Another graph satisfying CMC.

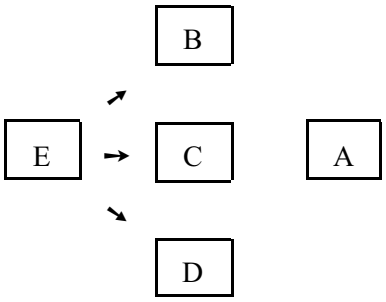

FIGURE 5: The result of the algorithm.

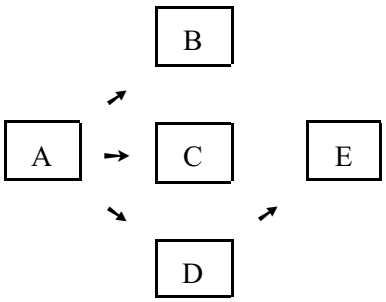

FIGURE 6: The rational causal graph.

The more general form of κ therefore requires a modified algorithm:

ALGORITHM A.13
Input: κ (atomic), π, $p_{\kappa,\pi}$.

1. Let $\mathbb{C} = \{\mathcal{C} : \mathcal{C}$ satisfies CMC and $\kappa^+\}$.
2. Set $\mathbb{D} = \varnothing$, $n_{\min} = \infty$.
3. For each $\mathcal{C} \in \mathbb{C}$,
   (a) remove as few arrows as possible from $\mathcal{C}$ to satisfy the constraints in κ, yielding graph $\mathcal{D}$,
   (b) let $n =$ number of arrows removed,
   (c) if $n < n_{\min}$, set $n_{\min} = n$, $\mathbb{D} = \varnothing$,
   (d) if $n = n_{\min}$ and $\mathcal{D}$ satisfies κ, add $\mathcal{D}$ to $\mathbb{D}$.
4. Let $\mathbb{E} = \{\mathcal{D} \in \mathbb{D} : \mathcal{D}$ is minimal$\}$.

Output: $\mathbb{E}$

THEOREM A.14 Suppose κ is atomic. Then $\mathbb{E} \subseteq \mathbb{C}_{\kappa,\pi}$, where $\mathbb{E}$ is produced by the above algorithm. Moreover, if κ is consistent then $\mathbb{E} \neq \varnothing$.

PROOF: Each $\mathcal{E} \in \mathbb{E}$ satisfies κ. Further, $\mathcal{E}$ is produced in step 3a by deleting as few arrows as possible from a graph that satisfies CMC, i.e., satisfies κ^*, so $\mathcal{E}$ satisfies π as well as κ. Step 3c ensures that E has minimum residue, while step 4 ensures that $\mathcal{E}$ is minimal. Thus $\mathcal{E} \in \mathbb{C}_{\kappa,\pi}$.

If κ is consistent then there is some graph $\mathcal{C}$ that satisfies κ. Any complete supergraph of $\mathcal{C}$ satisfies κ^+ and CMC, and hence is in $\mathbb{C}$. Thus $\mathcal{C}$ will result from step 3a and is a candidate for admission in $\mathbb{E}$; if $\mathcal{C}$ is not in $\mathbb{E}$ then that is because a graph that is smaller or has smaller residue is in $\mathbb{E}$ instead. Hence $\mathbb{E}$ is non-empty. □

Note that it may not be the case that $\mathbb{E} = \mathbb{C}_{\kappa,\pi}$, since different ways of carrying out step 3a may lead to different graphs in $\mathbb{C}_{\kappa,\pi}$, some of which are omitted from $\mathbb{E}$. The algorithm can be modified to output $\mathbb{C}_{\kappa,\pi}$ by altering step 3a so that the algorithm runs through the *set* of maximal $\mathcal{D} \subseteq \mathcal{C}$ that satisfy κ.

If κ is not atomic, then it may not be possible to divide κ into κ^+ and κ^-. The above algorithm can be modified to cope with this more general type of causal constraint just by letting $\mathbb{C} = \{\mathcal{C} : \mathcal{C}$ satisfies CMC$\}$ in step 1.

We see then that practical methods for finding a minimal graph satisfying CMC and a set of further constraints can be applied to the

problem in hand, namely determining $\mathscr{C}_{\kappa,\pi}$. Note that such methods invariably involve querying the probability function $p_{\kappa,\pi}$. However, it may be possible to construct $\mathscr{C}_{\kappa,\pi}$ directly from the background knowledge κ and π itself, without having to determine $p_{\kappa,\pi}$ as an intermediary, via the following algorithm:

ALGORITHM A.15.
Input: κ (atomic), π (strategically consistent)

1. Construct an undirected graph $\mathscr{G}$ on the variables in κ and π by linking each pair of variables with an edge if they occur together in the same constraint in π.
2. Find a minimal (in terms of fewest edges) triangulation $\mathscr{G}^T$ of $\mathscr{G}$.
3. Form the set $\Omega = \{\omega : \omega$ is a maximum cardinality ordering of the variables, ω is a causal ordering consistent with $\kappa\}$. (N.b. ω is a maximum cardinality ordering if each variable A_i is a variable from $\{A_j : j \geq i\}$ that is adjacent in $\mathscr{G}^T$ to the largest number of variables in $\{A_1, ..., A_{i-1}\}$.)
4. For each ordering ω form a directed acyclic graph $\mathscr{H}_\omega$ as follows:
(a) Let $D_1, ..., D_l$ be the cliques of $\mathscr{G}^T$, ordered according to highest labelled vertex.
(b) Let $E_j = D_j \cap (\cup_{i=1}^{j-1} D_j)$ and $F_j = D_j \setminus E_j$, for $j = 1, ..., l$.
(c) Take the variables as vertices of $\mathscr{H}_\omega$.
(d) Add an arrow from each vertex in E_j to each vertex in F_j, for $j = 1, ..., l$.
(e) Add further arrows, from lower numbered variables to higher numbered variables, to ensure that there is an arrow between each pair of vertices in D_j, $j = 1, ..., l$.
(f) Add arrows corresponding to the positive constraints κ^+ in κ.
5. Let $\mathbb{H} = \{\mathscr{H}_\omega : \mathscr{H}_\omega$ satisfies $\kappa, \mathscr{H}_\omega$ is minimal$\}$.

Output: $\mathbb{H}$

THEOREM A.16. Suppose κ is atomic and strategically consistent with π, that $G = G^T$, and that π does not on its own imply any probabilistic independencies. Then $\mathbb{H} \subseteq \mathbb{C}_{\kappa,\pi}$.

PROOF: Each graph produced by the end of step 4e is acyclic and satisfies CMC with respect to $p_{\kappa,\pi}$ (Williamson 2005a, Theorem 5.6), hence so does each graph produced by the end of step 4f. Since π does not imply any independencies and $\mathcal{G}$ is already triangulated, the graphs resulting from step 4e include the minimal graphs satisfying CMC (Williamson 2005a, Theorem 5.4).[11] Thus if $\mathbb{H}$ contains a graph at all, it is a minimal graph satisfying κ and CMC, hence, by Lemma A.8 , it is a minimal graph satisfying κ and κ^*. Since there is no residue (by assumption κ is strategically consistent with π), it is a minimal graph satisfying κ and π and is in $\mathbb{C}_{\kappa,\pi}$. □

Note that $\mathbb{H}$ may be empty: in Example A.6, the smallest graph that satisfies κ and CMC, Fig.2, can not be obtained by removing arrows from the smallest graph that satisfies CMC (which has only one arrow, between *A* and *B*).

Computational considerations may motivate simplifications of this algorithm. In particular, the second step, finding a optimal triangulation, is NP-hard (Yannakakis, 1981). Thus rather than demanding that there be no smaller triangulation in step 2, one might demand instead that the triangulation be minimal in the sense that no subgraph is a triangulation – as Berry at al. (2004) show, this is much more feasible (see also Neapolitan (1990, section 3.2.3) for a fast triangulation algorithm). Similarly, one might want to stop step 4 when one directed acyclic graph has been found that satisfies κ. If such modifications are made, or if it is not known whether π implies any probabilistic independencies, then a resulting graph in $\mathbb{H}$ can be viewed as an approximation to $\mathcal{C}_{\kappa,\pi}$.

If κ is not strategically consistent with π , this algorithm can be combined with Algorithm A.11 or Algorithm A.13 to try to identify a rational causal belief graph.

Finally, the set of rational causal belief graphs satisfies some interesting properties:

PROPOSITION A.17 Suppose κ is strategically consistent with π.

1. If κ provides a causal ordering of the variables then $\mathcal{C}_{\kappa,\pi}$ is uniquely determined if and only if $p_{\kappa,\pi}$ is strictly positive.

---

[11] Such a graph can be used as the graph in a Bayesian net representation of $p_{\kappa,\pi}$. This is called an *objective Bayesian net* – see Williamson (2005b).

2. $[\mathcal{C}_{\kappa,\pi}] \subseteq \mathbb{C}_{\kappa,\pi}$, where $[\mathcal{C}_{\kappa,\pi}]$ is the Markov equivalence class of $\mathcal{C}_{\kappa,\pi}$.
3. If $p_{\kappa,\pi}$ is faithful then $\mathbb{C}_{\kappa,\pi} = [\mathcal{C}_{\kappa,\pi}]$.

See Williamson (2005a, §9.7) for the relevant definitions and proofs.

REFERENCES

Andersen, S., Olesen, K., Jensen, F. and Jensen, F. (1989), "Hugin—a Shell for Building Bayesian Belief Universes for Expert Systems," in: *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pp. 1080–1085.

Anscombe, G. (1971), *Causality and Determinism*, Cambridge: Cambridge University Press.

Benacerraf, P. (1973), "Mathematical Truth," in: Benacerraf, P. and Putnam, H. (eds.), *Philosophy of mathematics: selected readings* (pp. 403-420), Cambridge: Cambridge University Press, second (1983) edition.

Berry, A., Blair, J., Heggernes, P. and Peyton, B. (2004) "Maximum Cardinality Search for Computing Minimal Triangulations of Graphs," *Algorithmica*, 39(4), pp. 287–298.

Carnap, R. (1950). *Logical Foundations of Probability*, London: Routledge and Kegan Paul.

Cartwright, N. (2004), "Causation: One Word, Many Things," *Philosophy of Science*, 71, pp. 805–819.

Dowe, P. (2000), *Physical causation*, Cambridge: Cambridge University Press.

Gillies, D. (2000), *Philosophical Theories of Probability*, London and New York: Routledge.

Gillispie, S. B. and Perlman, M. D. (2002), "The Size Distribution for Markov Equivalence Classes of Acyclic Digraph Models," *Artificial Intelligence*, 141, pp. 137–155.

Godfrey-Smith, P. (2008), "Causal Pluralism," in: Beebee, H., Hitchcock, C. and Menzies, P. (eds.), *The Oxford handbook of causation*, Oxford: Oxford University Press.

Hall, N. (2004), "Two Concepts of Causation," in: Collins, J., Hall, N. and Paul, L.A. (eds.), *Causation and Counterfactuals* (pp. 225–276), Cambridge MA and London: MIT Press.

Hitchcock, C. (2001), "A Tale of Two Effects," *The Philosophical Review*, 110(3), pp. 361–396.

Hitchcock, C. (2003) "Of Humean Bondage," *British Journal for the Philosophy of Science*, 54, pp. 1–25.

Korb, K.B. and Nicholson, A.E. (2003), *Bayesian Artificial Intelligence*, London: Chapman and Hall / CRC Press.

Lewis, D.K. (1973), "Causation," in: *Philosophical papers, volume 2* (pp. 159–213), Oxford: Oxford University Press (1986).

Mellor, D.H. (1995), *The Facts of Causation*, London and New York: Routledge.

Neapolitan, R.E. (1990), *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*, New York: Wiley.

Price, H. (1992), "Agency and Causal Asymmetry," *Mind*, 101, pp. 501–520.

Psillos, S. (2006). "Causal Pluralism," www.phs.uoa.gr/ psillos/ Publications_files/TheNaturalCausal(Final).doc.

Russo, F. and Williamson, J. (2007a), "Interpreting Causality in the Health Sciences," *International Studies in the Philosophy of Science*, 21, pp. 157-170.

Russo, F. and Williamson, J. (2007b), "Interpreting Probability in Causal Models for Cancer," in: Russo, F. and Williamson, J. (eds.), Causality and Probability in the Sciences. Texts in Philosophy (pp. 217–241), London: College Publications.

Salmon, W.C. (1998), *Causality and Explanation*, Oxford: Oxford University Press.

Schaffer, J. (2005), "Contrastive Causation," *The Philosophical Review*, 114(3), pp. 297-328.

Sober, E. (1985), "Two Concepts of Cause," in: Asquith, P. and Kitcher, P. (eds.), *PSA 1984 volume 2* (pp. 405–424), East Lansing, Philosophy of Science Association.

Spirtes, P. (1995), "Directed Cyclic Graphical Representation of Feedback Models," in: Besnard, P. and Hanks, S. (eds.), *Proceedings of the 11th Annual Conference on Uncertainty in Artificial Intelligence (Montreal)* (pp. 491–498), San Mateo CA: Morgan Kaufmann.

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

von Mises, R. (1928), *Probability, Statistics and Truth*, second (1957) edition, London: Allen and Unwin.

Weber, E., Van Bouwel, J., and Vanderbeeken, R. (2005), "Forms of Causal Explanation," *Foundations of Science*, 10, pp. 437–454.

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

Williamson, J. (2005b), "Objective Bayesian Nets," in: Artemov, S., Barringer, H., d'Avila Garcez, A.S., Lamb, L.C. and Woods, J. (eds.), *We Will Show Them! Essays in Honour of Dov Gabbay, volume 2* (pp. 713–730), London: College Publications.

Williamson, J. (2006a), "Dispositional versus Epistemic Causality," *Minds and Machines*, 16, pp. 259–276.

Williamson, J. (2006b), "From Bayesianism to the Epistemic View of Mathematics," *Philosophia Mathematica (III)*, 14(3), pp. 365–369.

Williamson, J. (2007a), "Causality," in: Gabbay, D. and Guenthner, F. (eds.), *Handbook of Philosophical Logic*, volume 14 (pp. 95-126), Springer.

Williamson, J. (2007b), "Motivating Objective Bayesianism: from Empirical Constraints to Objective Probabilities," in: Harper, W.L. and Wheeler, G.R. (eds.), *Probability and Inference: Essays in Honour of Henry E. Kyburg Jr* (pp. 151-179), London: College Publications.

Yannakakis, M. (1981), "Computing the Minimum Fill-in is NP-complete," *SIAM Journal of Algebraic and Discrete Mathematics*, 2, pp. 77–79.

Philosophica 77 (2006) pp. 97-123

THE RATIONALE OF VARIATION IN METHODOLOGICAL AND EVIDENTIAL PLURALISM[1]

Federica Russo

ABSTRACT

Causal analysis in the social sciences takes advantage of a variety of methods and of a multi-fold source of information and evidence. This pluralistic methodology and source of information raises the question of whether we should accordingly have a pluralistic metaphysics and epistemology. This paper focuses on epistemology and argues that a pluralistic methodology and evidence don't entail a pluralistic epistemology. It will be shown that causal models employ a single rationale of testing, based on the notion of variation. Further, I shall argue that this monistic epistemology is also involved in alternative philosophical theories of causation.

1. Introduction

Different social sciences study society from different angles and perspectives. Sociology studies the structure and development of human society, demography studies variations in populations due to mortality, fertility and migration behaviours, economics studies the management of goods and services, epidemiology studies the distribution of disease in human populations and the factors determining that distribution, etc. In spite of these differences, the social sciences share a common objective: to understand, predict and intervene on society. In these three moments of the scientific demarche, *knowledge of causes* is in many cases a

[1] I wish to thank Daniel Courgeau, Robert Franck, Jon Williamson, and an anonymous referee for their helpful comments and suggestions. Financial support from the FSR-UCLouvain is also acknowledged.

necessary ingredient in order to provide an explanation of social phenomena as well as of individual behaviours.

This causalist perspective, however, is not always explicit. Causal vocabulary is sometimes replaced with more vague terms such as factor, determinant, risk, but not *cause* and *effect*. Also, it is said that in spite of the heavy formalism of modern models, the social sciences cannot establish *causal* relations but only make *associational* claims. However, if the social sciences merely *described* phenomena, it would be useless to design policies or prescribe treatments that rely on the results of research. Adopting an explicit causalist stance is motivated by two distinct but related objectives: cognitive and action-oriented. We pursue a *cognitive goal* in detecting causes and thus in gaining general causal knowledge of the causal mechanisms that govern the development of society, and such general causal knowledge is meant to inform and guide social policy, that is we also pursue an *action-oriented goal.*

In the social sciences, causal analysis takes advantage of a variety of methods and of a multi-fold source of information and evidence. In this paper I raise the question of whether such methodological and evidential pluralism also entails epistemological pluralism. In a nutshell, I shall give a negative answer and argue in favour of a monistic rationale of causality based on the notion of *variation*.

The paper is organised as follows. I first give an overview of methodological and evidential pluralism by presenting different causal models and the variety of types of evidence and of information used in causal analysis, and then spell out the question of whether this form of pluralism entails epistemological pluralism. Afterwards, I present the rationale of variation and support it with methodological arguments; I also offer a taxonomy of variations and discuss some possible objections. Finally, I show that this rationale is consistent with or even adopted in alternative philosophical accounts of causation.

2. Methodological and evidential pluralism

The first developments of quantitative causal analysis in the social sciences are due to the pioneering works of A. Quetelet (1869) and E. Durkheim (1895, 1897) in demography and sociology respectively. Significant improvements are due to H. Blalock (1964) and O. Duncan

(1975), and since then causal analysis has shown noteworthy progress in the formal methods of analysis. In the following, I shall just give some examples of different methods through which contemporary causal analysis is carried out.

2.1. Structural equation models

Arguably, structural equation models (SEM) are the most widespread methodology. Originators of SEM were mainly geneticists, such as S. Wright (1921, 1934), and economists, such as T. Haavelmo (1943, 1944) and T.C. Koopmans (1950). SEM consist of a set of equations, which can be used to determine a causal graph. SEM are designed in order to combine qualitative causal information with statistical data to provide quantitative assessment of cause-effect relationships among variables of interest. Other classes of models, e.g. covariance structure or hierarchical models, rely on SEM, so it is worth spending some time on their structure, assumptions and hypothetico-deductive methodology.[2]

To illustrate, we take a canonical econometric model relating price and demand through two equations[3]:

$$Q = \beta_1 P + \delta_1 I + \epsilon_1$$
$$P = \beta_2 Q + \delta_2 W + \epsilon_2$$

Q is the quantity of household demand for a certain product, P is the unit price of the same product, I is the household income, W is the wage rate for producing the product, ϵ_1 and ϵ_2 are the error terms, β and δ are the parameters. The first equation states that demand depends on – or is causally determined by – the unit price of the product and the household income, while the second states that the unit price of the product depends on – or is causally determined by – the demand and the wage rate.

---

[2] Nowadays, in the SEM-literature there isn't an unanimous consensus as to whether structural equations can be given a causal interpretation. However, as Judea Pearl argues at length (Pearl 2000, ch.5), the original interpretation of SEM was eminently causal and it is a trend of contemporary researchers to require extra ingredients for the causal interpretation.

[3] I borrow this example from Pearl (2000 : 27-28).

The basic idea underlying SEM is that in a system of equations we can test whether variables are interrelated through a set of relationships, by examining the variances and covariances of variables. Sewall Wright, as early as 1934, has taught us to write the covariance of any pair of variables in terms of path coefficients. The path coefficient quantifies the (direct) causal effect of a putative cause *X* on the putative effect *Y*; given the numerical value of the path coefficient β, the structural equation $Y = \beta X + \epsilon$ claims that a unit increase in *X* would result in β units increase in *Y*.

SEM rely upon a number of assumptions, some of which have mere statistical importance (for instance, normality) whereas others have a fundamental bearing on causality (for instance, the non-correlation of errors terms, covariate sufficiency or no confounding). Among causal assumptions the crucial one is the so-called invariance condition or structural stability. This condition states that parameters have to be stable across a large number of interventions or environmental changes. Stability of parameters is usually taken as *the* condition ensuring the causal interpretation of structural equations.

In SEM, the process of model building involves a continuous interaction between background knowledge and a sequence of statistical procedures for elaborating and testing hypotheses. This is the bulk of the hypothetico-deductive methodology (H-D). H-D methodology is a procedure that accounts for data obtained through observations and/or experimentation and that confirms or disconfirms a given causal structure by confrontation with empirical evidence. Empirical testing is performed through two stages:

(i) prior theorizing of out-of-sample information, including in particular the selection of variables deemed to be of interest, the formulation of a causal hypothesis, *etc.*;

(ii) iteratively:

a. building the statistical model;

b. testing the adequacy between the model and the data to accept the empirical validity or non-validity of the causal hypothesis.

Causal modelling requires accurate knowledge of the causal context: previous studies, well confirmed scientific theories or background knowledge are essential. The causal hypothesis states a hypothesized causal structure to be put forward for empirical testing. Thus, causality is a matter of confirmation, or borrowing the statistical

vocabulary, a matter of accepting or rejecting a given hypothetical causal structure. This strategy is hypothetico-deductive because the causal claim is not inferred from the data, as in inductive methods[4], but confirmed or disconfirmed in the given causal context and relative to the structural model. Elsewhere (Russo, 2005; Russo et al., 2006) I argued in more detail that several elements participate in justifying the causal interpretation of these models, for instance the causal context in which they are built, the specific set of assumptions having causal compass and their peculiar H-D methodology.

2.2. Covariance structure models

Covariance structure models (CSM) attempt to explain the relationships among a set of observed variables in terms of a generally smaller number of unobserved or latent variables.[5] Formally, CSM consist of an analysis of the covariances of the observed variables in two conceptually distinct steps. A *measurement* model links observed variables to unobserved variables and a *structural* model links unobserved variables. In turn, the measurement component of CSM consists of a confirmatory factor model explaining the covariations in a set of observed variables in terms of a smaller number of common factors. The idea behind CSM is that, although some variables of theoretical interest cannot be observed directly, information about them can be obtained indirectly from their effects on observed variables.

Thus, CSM consist in the simultaneous specification of the factor model and of the structural model. The task is to explain the interrelationships among the observed variables as indicated by the covariances among them, in terms of the relationships among the unobserved variables used in the structural equations. It is worth noting that what specifies causal relations in CSM is the structural equation model. On the other hand, the measurement model allows us to estimate latent variables from observed variables.

---

[4] An example of an inductive method, that tries to infer causal structures from data, is TETRAD, the algorithm developed by Spirtes, Glymour and Scheines (1993).

[5] For a very clear introduction see Long (1983).

2.3. Multilevel models

Recently, multilevel or hierarchical models[6] are used in a variety of disciplines, ranging from education to demography. Multilevel analysis is a methodology for the analysis of data with complex patterns of variability, the underlying assumption being that data shows a hierarchy that cannot be neglected in the analysis.

The object of a discipline does not straight specify the level of aggregation at which analyses have to be carried out. For instance, economics is interested in the production, distribution and consumption of wealth, however, there is no a priori specification of whether analyses have to concern individuals, markets, firms, or nations. Thus, multilevel analysis recognises the existence of a multiplicity of levels and tries – within the framework of a single model – to specify the relations holding among individuals and/or among different levels of aggregation. In other words, this approach recognises that the grouping of individuals introduces an influence of the group on its members, and, conversely, that members have an influence on the group's behaviour. Failure to recognise this twofold source of influence and variability may lead to two types of fallacy: the atomistic fallacy and the ecological fallacy.[7]

---

[6] Very good introductions to multilevel modelling are those of Goldstein (2003), Snijders and Bosker (2004), and Courgeau (2003) in which epistemological and methodological problems are also discussed in detail.

[7] The motivation for developing hierarchical models lies in the so-called ecological fallacy, firstly recognised by Robinson (1950). The ecological fallacy consists of inferring individual behaviours from aggregate measures. Robinson pointed out, for instance, that correlations between two characteristics measured on a binary basis among individuals (e.g. being black and illiterate in the US), or by proportions in regions (e.g. proportions of black and illiterate people in the population) were generally not identical and could even carry opposite signs. Conversely, the atomistic fallacy arises when, analysing individual behaviours, the context in which such behaviours occur is neglected.

2.4. Contingency tables

Causal analysis is also performed by means of contingency tables or cross tabs. When variables involved are non-metric, categorical data analysis (CDA) is more often employed instead. CDA has a very long history. It began in the early 1900s, when K. Pearson and U. Yule were debating on measures of associations, and two decades later CDA took advantage of significant contributions by R. Fisher. The first lucid exposition of the use of contingency tables in sociology is due to Boudon and Lazarfeld (1966) and in recent years, clear presentations and further improvements of CDA are available in the works of Hellevik (1984) or Agresti (1996).

A categorical variable is one for which the measurement scale consists of a set of categories. Categorical scales are very often used in the social sciences to measure attitudes and opinions on several issues. Categorical data consist of frequency counts of observations occurring in the response categories. Consider the simplest case, where only two variables X and Y are involved. X has i levels, and Y has j levels, according to the number of categories that X and Y involve. The ij possible combinations of outcomes are then displayed in a rectangular table having i rows and j columns. The cells of the table in fact represent the ij possible outcomes and contain frequency counts of outcomes. Tables thus construed are called contingency tables or crosstabulations. Such ordinary percentage tables can be analysed by adopting an explicit causal framework.

Such a causal framework is offered, for instance, by Hellevik (1984), where an explicit causal terminology is adopted. In the bivariate case, the independent variable Y represents the cause and the dependent variable X represents the effect. Different levels of X are then compared with regard to the proportion having a specific value on Y. The difference in proportion will then measure the degree of association of the two variables and, in this framework, it will be interpreted as the causal effect of Y on X. In contingency tables differences in proportions play the analogue of regression coefficients in SEM, giving highly similar results. Thus, to some extent, the causal framework for contingency tables rests on the same features as SEM, namely on background knowledge, choice of variables, issues of confounding and control, etc.

2.5. Data and evidence

By means of these different methodologies, social scientists try to make sense of observations and to infer causal relations between variables of interest with reasonable confidence. Observations, however, first have to be collected. Data comes from a variety of different sources: surveys, census, experiments, interviews, etc. Analogously, evidence of causal relations can come from different sources: previous studies, background knowledge, knowledge of mechanisms or of probabilistic relations, etc.

Previous studies often make it plausible to investigate a given causal relation in a different population or at a different time. Background knowledge gives a (causal) context to causal models, for instance by providing the socio-political context of a population or socio-demographic differences across different populations, etc. A different type of background knowledge is constituted by the iterated application of some methods for the analysis of a given relationship.

Evidence for causal relations can be of two different sorts. We infer causal relationships from probabilistic evidence: causes have to be statistically relevant for their effects, but we also require repetition of similar studies and coherence in their results, namely covariations among variables of interest have to show some stability. Yet, probabilistic evidence is not sufficient as correlations may be spurious, as is well known. To infer causal relationships we also have to exhibit a plausible mechanism.

It seems then clear that scientific practice in the social sciences takes advantage of a pluralistic methodology, evidence and source of information. Consequently, the following question arises: does this methodological and evidential pluralism entail ontological and epistemological pluralism? In other words, does the fact that in practice social scientists use different models, different sources of evidence and of information entail that there are different *concepts* or different *rationales* of causality? In the remainder of the paper I shall focus on epistemological pluralism and argue that a single rationale of causality, based on the notion of variation, is used in the social sciences.

3. Epistemological monism: the rationale of variation

3.1. The rationale of variation

Epistemology studies the origin, nature and limits of human knowledge. Epistemology of causality then wonders how we *know* about causal relations, what epistemic access we have to causal relations, under what conditions we can interpret correlations causally, whether specific conditions such as invariance under intervention, structural stability or the Markov condition in Bayes nets guarantee causality. More specifically, I address the question: what *rationale* of causality governs causal models in the social sciences? Is there a unique rationale or different ones depending on the model at hand?

A rationale is the principle or notion underlying some opinion, action, phenomenon, reasoning, model, or the like. A rationale of causality in causal modelling is then the principle or the notion that guides causal reasoning (in causal modelling). It is worth emphasising that a rationale is not a definition of what causality is. A definition is a description of a thing by its properties; thus, a definition of causality states what causality in fact is. Whilst providing a definition of causality is a job for metaphysics, the development of a rationale is a matter of epistemology. In this paper, I am concerned with the latter problem but not with the former.

The received view, an heritage of Hume, sees this rationale in the notion of regularity and this standpoint still pervades contemporary philosophy of science. As is well known, Hume believed that causality lies in the constant conjunction of causes and effects. In the *Treatise* Hume says that, in spite of the impossibility of providing rational foundations for the existence of objects, space, or causal relations, to believe in their existence is a "built in" habit of human nature. In particular, belief in causal relations is granted by experience. For Hume, simple impressions always precede simple ideas in our mind, and by introspective experience we also know that simple impressions are always associated with simple ideas. Simple ideas are then combined in order to form complex ideas. This is possible thanks to imagination, which is a normative principle that allows us to order complex ideas according to (i) resemblance, (ii) contiguity in space and time, and (iii) causality. Of the three, causality is the only principle that takes us

beyond the evidence of our memory and senses. It establishes a link or connection between past and present experiences with events that we predict or explain, so that all reasoning concerning matters of fact seems to be founded on the relation of cause and effect.

The causal connection is thus part of a principle of association that operates in our mind. Regular successions of impressions are followed by regular successions of simple ideas, and then imagination orders and conceptualizes successions of simple ideas into complex ideas, thus giving birth to causal relations. The famed problem is that regular successions so established by experience clearly lack the necessity we would require for *causal* successions (otherwise successions would be merely casual). Hume's solution is that if causal relations cannot be established *a priori*, then they must be grounded in our experience, in particular, in our psychological habit of witnessing effects that regularly follow causes in time and space.

My proposal is, instead, to depict the rationale of causality as the *measure of variation or change*. The study of change is the study of factors which produce change. Thus, measuring variations conveys the following idea: to test causal models means to measure suitable changes or variations. Causal models apply a H-D methodology: causal hypotheses are first formulated, and then put forward for empirical testing. In turn, empirical tests are designed to assess the presence of a variation, and to assess whether this variation satisfies certain conditions. Therefore, the point I want to make is twofold: (i) causal modelling is not governed by a rationale of regularity but by a *rationale of variation*, and (ii) there is only one rationale. In other words, methodological and evidential pluralism do not imply epistemological pluralism.[8] The

[8] This paper is mainly concerned with the scientific literature on causal modelling. In Russo (2005) I also show that a number of authors in the philosophical literature employ or presuppose the notion of variation in their accounts. For instance, in Woodward's account (Woodward, 2003) causal generalisations are "*change*-relating" or "*variation*-relating" and variation is a necessary condition for interventions. In Hausman's account (Hausman, 1998), modal invariance presupposes that intervening on the putative cause will produce a *variation* on the putative effect, this variation being possible to compute, or at least possible to estimate.

monistic epistemology I offer turns around a rationale of causality which is based on the notion of variation.

3.2. Methodological arguments

Let us consider structural equation models (SEM) first. Recall that the basic idea of SEM is that in a system of equations we can test whether variables are interrelated through a set of linear relationships, by examining the variances and covariances of variables, and that, given the numerical value of the path coefficient β, the structural equation claims that a unit increase in *X* would result in β units increase of *Y*. This means that β quantifies the *variation* on *Y* accompanied by the *variation* on *X*. The equality sign in structural equations does not state an algebraic equivalence; jointly with the associated graph, the structural equation is meant to describe the causal relationship implied by the data generating process. The path coefficient β, in turn, is meant to quantify the (direct) causal effect of *X* on *Y*. β quantifies the *variation* on *Y* produced by the *variation* of *X*, hence the structural equation does not merely describe a regular concomitant occurrence of *Y* and *X*, but how the dependent variable *varies* depending on the variation of the independent one.

It is commonly agreed that, in structural equations, variations in the independent variables, i.e. the *X*s, explain the variation in the dependent variable, i.e. *Y*. Witness, for instance, Haavelmo:

> In other words, we hope that, for each variable, y, to be "explained", there is a relatively small number of explaining factors the *variations* of which are practically decisive in determining the *variations* of y. (Haavelmo, 1944:23) (My emphasis)

So, one might suggest that the explanatory power of a causal model is given by the inverse of the unexplained variance in the dependent variable. Namely, because the dependent variable is determined to a certain extent by variations in the independent variables, the more we can account for those variations, the higher the explanatory power of the model. What is not determined by variations in independent variables, depends on the errors. In fact, the squared path coefficient r^2 measures the portion of the variance in the dependent variable the independent

variable is responsible for, or, differently put, the squared path coefficient represents the proportion of common variation in both the dependent and in the independent variable. Again, reasoning about the squared path coefficient involves the notion of variation rather than regularity.

CSM are governed by a rationale of variation too. By way of reminder, CSM have two models: a measurement model and a structural model: the covariance matrix in the measurement model is explained by the (causal) relations as indicated in the structural model. On the one hand, to analyse covariances means exactly to measure (joint) *variations*. The covariance matrix is the matrix of the covariances between elements of a vector and represents the natural generalization to higher dimensions of the concept variance of a scalar-valued random variable. The variance measures how much a single variable varies around the mean, and the covariance measures how much two variables vary together. On the other hand, the structural component of CSM is not meant to explain regularities, but joint *variations* and, as we just saw, is regimented by the variation rationale.

Analogously, as they are based on structural equations, multilevel models too rely on the rationale of variation. Furthermore, variations are measured not only at one level of aggregation, but also across different levels. For instance, a multilevel model can assess how educational achievement varies among students in the same class and across classes in a school.

Contingency tables also rely on the rationale of variation. The question is, in fact, the extent to which each of the independent variables contributes to the *variation* between the categories in the dependent variable. In the scheme of reasoning behind contingency tables it is not hard to recognize the *statistical-relevance model* (S-R) of explanation advanced by W. Salmon (1971, 1984). And in fact, not surprisingly, the rationale of variation is clearly involved in the S-R model.

According to Salmon, to explain a fact, i.e. to identify its causes, one has to identify the correct cell in the reference class the fact to be explained belongs to. Consider Salmon's example (1984:37). To understand why a particular individual – say, Albert – committed a delinquency – say, stealing a car, we first consider the broadest reference class Albert belongs to (American teenager); then, this class is partioned into subclasses based on the number of (all and only) relevant factors.

Sociological theories suggest taking gender, religious background, marital status of parents, type of residential community, socioeconomic status and education, etc. into account. We will thus obtain a large number of cells, each of which will be assigned a probability of the degree of delinquent behaviour. This fact, i.e. why Albert committed a delinquency, will be explained once the narrowest class Albert belongs to is identified, e.g. male and parents divorced and living in a suburban area and low education …

As the name suggests, statistical *relevance* relations are used in the S-R model in order to isolate *relevant* causal factors. Let me explain how it works – the rationale of variation will then become apparent. Let A denote American teenagers, and B_i various degrees of juvenile delinquency. What we are interested in is not just $P(B_i|A)$, that is, the probability of committing a delinquency in the population of American teenagers, but in a more specific probability, say $P(B_i \mid A \,\&\, C_j \,\&\, D_k \,\&\, E_n)$, where C_j, D_k and E_n are all *relevant factors*, for instance gender, religious background, marital status of parents, etc. The crucial point is that if conditioning on a further factor, say F_m, does not *change* the previous conditional probability, then F_m is not a relevant factor and hence should not be considered in the explanation. So all factors entering the S-R model are statistically relevant, i.e. *responsible for variations*, in the probability of the fact to be explained.

3.3. Varieties of variations

So far I argued that a monistic epistemology, based on the notion of variation, regiments different types of causal models. However, the rationale of variation as just described is still very general. In practice, social scientists may look for different types of variations depending on the case at hand. A taxonomy of variations can be sketched according to the following criteria:[9]

1. variation across time;
2. variation across individuals;
3. variation across characteristics;
4. counterfactual and control group variations;

[9] These taxa are detailed and fully exemplified in (Russo, 2005).

5. observational vs. interventional variations.

We might be interested in whether the same characteristic, say unemployment rate, varies across time (taxon 1) – e.g. in two successive surveys, or across individuals (taxon 2) – e.g. individuals in the survey may show radically different employment histories, or across characteristics (taxon 3) – e.g. unemployment rate may be different according to different levels of education. In observational studies we can model counterfactual variations (taxon 4), for instance the individual probability of finding a job given certain characteristics, or, in experimental studies, we can check whether variations hold between the test and control group. Finally (taxon 5), variations can be merely observed – when we deal with observational data, or can be the result of interventions – if we can manipulate and operate directly on data.

3.4. Objections, or further evidence

The rationale of variation is not exempt from possible objections. The first doubt might be that *variation* means something stronger, namely *dependence*. If so, Pearl has argued instead that *in*dependence is more basic:

> […] conditional independence is not a 'restrictive assumption' made for mathematical convenience; nor it is an occasional grace of nature for which we must passively wait. Rather, it is a mental construct that we actively create, a psychological necessity that our culture labours to satisfy. (Pearl, 1988:385)

In other words, *in*dependence is an essential feature for causality. Nonetheless, a few pages later, Pearl seems to hold quite a different view, when he draws some conclusions about causal poly trees. He addresses the old question: causation or covariation? According to him, the threshold is in the notion of *control*: causal directionality between *X* and *Y* can only be tested through the introduction of a third variable *Z*. This is because by introducing *Z* we test whether:

> by activating *Z* we can create *variations* in *Y* and none in *X*, or alternatively, if *variations* in *Z* are accompanied by variations in *X* while *Y* remains unaltered. […] the construct of causality is merely a tentative, expedient device for encoding complex structures of

> dependencies in the closed world of a predefined set of variables. It serves to highlight useful independencies at a given level of abstraction, but *causal relationships undergo change* upon the introduction of new variables. (Pearl, 1988:397) (my emphasis)

In the ultimate analysis, Pearl did use the rationale of variation, and this rationale seemed to precede the notion of independence, contrary to what he himself stated, i.e. that independence is the basic notion for causal learning.[10]

Another obvious objection to the rationale of variation is that this rationale is nothing but a reformulation of Humean regularist accounts. This is only partly true. Let me deal with the non-true part first. The crucial step in Hume's argument is significantly different from the rationale I propose. My claim is that we look for variations, not for regularities. Once variations are detected, a condition of invariance or structural stability (among others) is imposed on them. What does structural stability give us? Not logical or nomic necessity, nor mere constant conjunction. Structural stability is a condition imposed on joint variations, in order to ensure that the model correctly specifies the data generating process and to ensure that the model does not confuse accidental variations with causal ones. Although the invariance condition is the most important one, other conditions – e.g. no confounding – grant the causal interpretation of statistical models. The true side of the objection is that in observational studies attention is mainly directed to variations that happen to be regular, at least regular enough not to be accidental.

---

[10] It is worth noting that Pearl changed his mind about causality between 1988 and 2000: "Ten years ago, when I began writing *Probabilistic Reasoning in Intelligent Systems* (1988), I was working within the empiricist tradition. In this tradition, probabilistic relations constitute the foundations of human knowledge, whereas causality simply provides useful ways of abbreviating and organizing intricate patterns of probabilistic relationships. Today, my view is quite different. I now take causal relationships to be the fundamental building blocks both of physical reality and of human understanding of that reality, and I regard probabilistic relationships as but the surface phenomena of the causal machinery that underlies and propels our understanding of the world." (Pearl, 2000:xiii-xiv).

Let me underline why so much importance is bestowed on the notion of variation. Variation is conceptually a *pre*condition. If causality is not set in the notion of variation, then it will be lodged in the invariance condition, which is conceptually misleading. The same holds for regularity. In both cases there is a further question to ask: invariance *of what*?, regularity *of what*? The answer is, in both cases, *of a variation*. In particular, invariance – the queen of the causal conditions – only makes sense *within* a causal model, whereas variation is exactly what motivates testing invariance. In other words, without variation, invariance is devoid of meaning. This is why variation conceptually precedes invariance.[11]

To provide a rationale of causality means, to put it otherwise, to give the bottom-line concept – namely, variation – and the constraints to put on this variation – namely, invariance and regularity. Neither invariance nor regularity are apt to accomplish that task. But variation is. Hume inferred causation from regularity, whereas my claim is that we infer causation from variation because variation conceptually and empirically comes before regularity. Of course, both notions – regularity and variation – don't guarantee a straight causal interpretation, but the rationale of variation puts us on the right track because it makes causality an empirical issue rather than a psychological fact or a mere reduction to statistical conditions to be satisfied.

The Humean paradigm of regularity still dominates contemporary philosophy of science. Regularity views of causation are clearly an heritage of the Humean account. Defenders of regularist accounts claim, roughly, that to assert a causal relation between two events x and y means to assert the existence of a regular succession such that every time an event of type X occurs, then an event of type Y will invariably follow. Humeans like J.S. Mill or J. Mackie have advanced more sophisticated versions of the regularist view (e.g. the I.N.U.S. condition) and tried to characterise the kind of regularity that can underpin causal relations by tying causation to laws of nature. Even Lewis' counterfactual approach uses regularities as a means to capture the conditions under which counterfactual assertions are true. Also, the probability raising requirement often advocated in probabilistic accounts, has been backed

[11] For a thorough discussion of variation as a *pre*condition, see Russo (2005).

by a ceteris paribus condition – i.e. a condition of homogeneity – that makes things regular enough to let the cause raise the probability of the effect.

The variation rationale profoundly breaks down this received view. The variation rationale argues, in the first place, that this emphasis on regularity is not well-founded, for regularities themselves require a prior notion, which is variation. Secondly, the difficulty or impossibility of establishing causal laws in the social sciences is usually taken as a structural weakness or even as an intrinsic impossibility for the social sciences to reach the kingdom of "hard sciences". The variation rationale is the first step for a radical change in the dominant paradigm: if, in the ultimate analysis, causal modelling aims at measuring variations rather than establishing regularities, this might be due to the fact that the regularist rationale is not, after all, well founded as empiricists claim since Hume. This calls for a change of paradigm in causal modelling, rather than giving up our endeavour to establish causal claims or questioning the rigorous scientific character of the social sciences.

4. The rationale of variation in philosophical accounts of causation

A number of accounts of causation have been proposed in the last decades. The probabilistic, mechanist and counterfactual approach take causal relations to be objective, in the sense that causality is defined independently of the agent, the first relying on statistical relevance, the second on the notion of physical process and interaction, and the third on counterfactual logic. Agency theories, instead, define causality in terms of an agent's ability to operate on causal relations; manipulability theories try to get rid of anthropomorphism and to regain objectivity by developing a notion of intervention that fits causal modelling. A different attempt to give causality an objective character is epistemic causality, where objectivity is understood as "non arbitrary" rather than "mind-independent". The goal of this section is to disclose how the rationale of variation is consistent with or (more or less explicitly) adopted in those accounts.

4.1. Variation in probabilistic theories

Probabilistic theories (PT) of causality have been developed in slightly different manners by different philosophers in the last decades.[12] In spite of the significant differences in these accounts, a core of agreement can be found in the pioneering works of Good and Suppes that, roughly speaking, turn around the probability raising requirement: *ceteris paribus*, causes make their effects more probable. Prima facie, C is a cause of E if, and only if, (i) C occurs before E and (ii) C is positively, statistically relevant to E, that is $P(E|C) > P(E)$.

Thus PT focus on the difference between the conditional probability of the effect given the cause $P(E|C)$ and the marginal probability of the effect $P(E)$. To compare the conditional and marginal probability means to analyse a statistical relevance relation. The underlying idea is that if C is a cause of E, then C is also statistically relevant for E. To evaluate a statistical relevance relation exactly means to measure a *variation,* in particular, a variation in the conditional probability of E given C with respect to the marginal probability of E. That is to say, the change hereby produced by C in the effect E will be detected because the conditional and the marginal probability *differ*.

4.2. Variation in mechanist accounts

The mechanist approach[13] takes physical processes and interactions between them to be the fundamental concepts for causation. Simply put, the Salmon-Dowe theory is based on three definitions: (i) causal interaction, (ii) causal process, and (iii) causal transmission.

First, a *causal interaction* is an intersection of world-lines which involve exchange of a conserved quantity. In this definition, the meaning of *exchange* is worth stressing: at least one outgoing process manifests a *change* in the value of the conserved quantity and the exchange is governed by the conservation law. Second, a *causal process* is a world-line of an object that transmits a nonzero amount of an invariant quantity

---

[12] See for instance (Good, 1961), (Good 1962), (Suppes, 1970), (Cartwright, 1979), (Cartwright, 1989), (Eells, 1991).

[13] (Salmon, 1984), (Salmon, 1994), and (Dowe 1992).

at each moment of its history (each space-time point of its trajectory). Last, the principle of mark transmission is formulated as follows: a process transmits an invariant (or conserved) quantity from A to B ($A \neq B$) if it possesses this quantity at A and at B and at every stage of the process between A and B without any interactions in the half-open interval $(A,B]$ that involves an exchange of the particular invariant (or conserved) quantity.

The notion of variation plays a fundamental role in the definition of causal interaction. In fact, an exchange of invariant or conserved quantities between processes actually produces a *modification* or *variation* in them and this is what makes the interaction causal. Of course, the way in which the rationale of variation is here employed differs from the quantitative one depicted in causal modelling, but the qualitative claim still holds: the bottom-line concept of causality is in the concept of variation, not in regularity, stability or invariance.

4.3. Variation in counterfactuals

D. Lewis (1973) is the main proponent of the counterfactual theory of causation. Causal relations are analysed in terms of subjunctive conditionals, also called counterfactuals: "*A* caused *B*" is interpreted as "*B* would not have occurred if it were not for *A*". Counterfactuals are subjunctive conditionals where the antecedent is known or supposed to be false and are regimented by a possible-world semantics.

Possible-world semantics rest on the assumption of the existence of a plurality of worlds, among which there is also our actual world.[14] Worlds are compared with each other on the basis of their similarity or closeness. The relation of comparative over-all similarity among possible worlds is taken as primitive and we say that one world is closer to actuality than another if the first resembles our actual world more than the second does. The truth of a counterfactual is then ascertained by an "inspection" of what happens in other possible worlds. Given any two propositions A and B, the counterfactual $A \mathrel{\square\!\!\rightarrow} B$ reads: "if A were true, then B would also be true". The counterfactual operator $\mathrel{\square\!\!\rightarrow}$ is defined by the following rule of truth:

[14] This position is also know as modal realism.

The counterfactual $A\Box\rightarrow B$ is true (at a world w) if, and only if:
(i) there are no possible A-worlds[15], or
(ii) some A-world where B holds is closer to w than is any A-world where B does not hold.

The second case is the interesting one, for in the former the counterfactual is just vacuously true. Causality comes in because by asking whether a counterfactual is true, we wonder whether the occurrence of A is the *cause* of the occurrence of B. So, the counterfactual, if true, states that if the cause had not occurred, the effect would not have occurred either.

How the rationale of variation is involved in the counterfactual approach will become apparent once the motivation for its development is spelled out. Lewis wants to go beyond standard regularity theories as they failed to notice the second definition of cause Hume gave in the *Enquiry*:

> We may define a cause an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second had never existed. (Hume, 1748: section VII)

If the cause had not been, the effect had never existed. For Lewis, this is not just a restatement of the regularist view, but a clear input and encouragement to take up the counterfactual path. In Lewis' words:

> We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it. Had it been absent, its effects – some of them, at least, and usually all – would have been absent as well. (Lewis, 1986:160-161)

Causes are supposed to make a difference, i.e. causes are responsible for *variations*. The bottom-line concept is, again, not in regularity, uniformity, or invariance, but is in difference, change, and variation.
Agreed, it would be misleading to maintain the counterfactual approach *relies* on the rationale of variation; yet, the variation idea is definitively

[15] In Lewis' theory "A-world" means "the world in which A is true".

consistent with it, as Lewis' words clearly show. The rationale of variation here involved is not quantitative, as is the case in causal modelling. Instead, a *qualitative* notion of variation is here at stake. Surely Lewis' account is of little help in testing causality over large data sets, but counterfactuals do grasp, at least, our intuitions about how the causal relation works: *ceteris paribus*, if the cause had not occurred, the effect had never existed either. In other words, we expect the cause to be responsible for the *change* leading to the effect.

4.4. Variation in agency-manipulability theories

Agency theories[16] analyse causal relations in terms of the ability of agents to achieve goals by manipulating causes. In a nutshell, *C* is said a cause of *E* if bringing about *C* would be an effective way to bring about *E*. To explain what counts as an effective strategy, Menzies and Price invoke the means-end relation, which is characterised in terms of agent probabilities. Agent probabilities are conditional probabilities assessed from an agent perspective as follows: the agent probability of *E* conditional on *C* is the probability that should enter in the calculations of a rational agent, whose abilities consist in the capacity to realise or prevent *C*, and whose goals entirely concern *E*. So a strategy to bring about *E* is effective if a rational decision theory prescribes it as a way of bringing about *E*. That is to say, agents probabilities are defined in terms of their role in rational decision-making and this is why they embody a basis for a formal analysis of the means-end relation.

This account is consistent with the central idea of the probabilistic theory of causality. In fact, *C* constitutes a means for achieving *E* only in the case that the agent probability $P_C(E)$ is greater than $P_{not\text{-}C}(E)$, where P_C denotes the agent probability that *E* would hold, were one to realise *C*. Because the agency theory turns out to be consistent with probabilistic theories of causality, *a fortiori* it is consistent with the variation rationale.

---

[16] (Price, 1991), (Price, 1992), (Menzies and Price, 1992).

D. Hausman and J. Woodward[17] propose a manipulationist account of causation. Their purpose is to overcome the objection of anthropomorphism raised against the agency theory, by developing a notion of intervention which is not agency-dependent. In their manipulationist or interventionist account, causal relations have essentially two features: (i) they are potentially exploitable for purposes of manipulation and control, and (ii) they are invariant under intervention. Everything turns around the specification of the notions of intervention and invariance.

Briefly put, an intervention on *X* with respect to *Y* changes the value of *X* in such a way that, if any change occurs in *Y*, it occurs only as a result of the change in the value of *X* and not from other sources. On the other hand, the notion of invariance is closely related to the notion of intervention and takes advantage of the notion of generalization. A generalization *G* is invariant if it would continue to hold under some intervention that changes the value of *X* in such a way that the value of *Y* would change. "Continue to hold" means that *G* correctly describes how the value of *Y* would change under this intervention. For Hausman and Woodward, equations in SEM are correct descriptions of the causal relationship between *X* and *Y* if, and only if, were one to intervene in the right way to change the value of *X*, then *Y* should change in the way indicated by the equation.

Invariance, or structural stability, is then a necessary ingredient for avoiding equations describing contingent or spurious relations. Nonetheless, it is worth asking – and here is where the rationale of variation emerges, *what* exactly remains invariant. The answer is: invariance of a detected *variation*. As we have seen earlier, structural equations describe how variations in *X* accompany variations in *Y* and structural parameters quantify the causal effect of *X* on *Y*. Eventually, in characterising causality as invariance under intervention, manipulability theories rely on the variation rationale in an essential manner.

---

[17] (Hausman, 1998), (Woodward, 2003), (Hausman and Woodward, 1999), (Hausman and Woodward, 2004).

4.5. Variation in epistemic causality

According to Williamson's epistemic theory of causality[18], causality is an objective mental construct. Causality is mental because it is a feature of an agent's mental state, as opposed to physical causality which is a feature of the world "out there", and it is objective because if two agents differ as to causal structures, then at least one of them must be wrong. Causal relations belong to an agent's representation of the world, more precisely, epistemic causality deals with *causal beliefs*. It is convenient to represent the world in terms of causes and effects because such causal representations, if correct, enable accurate predictions, diagnosis, decisions and interventions.

This metaphysical stance about causality is also accompanied by an account of the epistemology of causality: causal relations are discovered by an hybrid of the inductive and hypothetico-deductive method. Because the variation rationale belongs to epistemology, the question is whether or not the rationale is compatible with Williamson's learning strategies for epistemic causality, which involve four stages:
(i) hypothesise;
(ii) predict;
(iii) test;
(iv) update.

The *first* stage – the inductive one – requires a procedure for obtaining a causal graph from data and standard artificial intelligence techniques allow us to induce a minimal causal graph that satisfies the Causal Markov Condition. In the *second* stage, predictions are drawn from the induced graph and those predictions will be tested in the *third* stage. By renewed information or by performing experiments, predictions will be confirmed or disconfirmed. Finally, the *fourth* step represents a radical change in the hypothetico-deductive method: in case predictions fail, we do not start from the very beginning at step one, but we update the causal graph according to new evidence and information gathered.

The variation rationale permeates Williamson's learning strategy in the same sense as it permeates causal modelling. Causal models are

[18] (Williamson, 2005).

tested by measuring suitable variations among variables, and this is exactly what happens in the test stage. Witness Williamson:

> If, for instance, the hypothesised model predicts that *C* causes *E*, and an experiment is performed which shows that *intervening to change the value of C does not change the distribution of E,* controlling for *E*'s other direct causes, then this evidence alone may be enough to warrant removing the arrow from *C* to *E* in the causal model. (Williamson, 2005:149) (my emphasis)

Again, the rationale for testing causal relations is based on variation, not on regularity nor invariance.

5. Conclusion

The social sciences perform causal analyses by means of a variety of methods and rely on several sources of information and of evidence. This methodological and evidential pluralism raises the question of whether ontological and epistemological pluralism ought to be accordingly adopted.

This paper focused on the epistemological side and argued in favour of a monistic epistemology based on a rationale of variation. The rationale of variation is shown to be the basic notion employed in causal reasoning in different causal models, e.g. structural equation models, covariance structure models, multilevel analysis and contingency tables. I have argued that the variation rationale goes against the regularity view and constitutes the bottom-line concept of causality because it is a *pre*condition both for regularity and invariance. Finally, I also showed how this rationale is consistent with or even adopted in various philosophical accounts, from probabilistic theories to mechanist approaches, from agency-manipulability theories to epistemic causality.

We have seen that the inheritance of the received view is a rationale of regularity. The rationale of *variation* profoundly breaks down this conception. The philosophical gain in adopting the rationale of variation is fourfold. *First*, causality is not merely lodged in a psychological habit of observing regular successions of events. Agreed, we do experience such regular sequences but, I argue, it is not because of regularity that we interpret them causally. Instead, this is because certain

variational relations hold. *Second*, causality is not reduced to statistics either. Correlation, as is well known, does not prove causation. Further, to claim that variation is a *pre*condition for regularity and invariance has the undoubted advantage of not confusing the rationale of causality with the conditions that allow to interpret variations causally. *Third*, along the same line, the adoption of the rationale of variation avoids confusing (i) what causality is (metaphysics) with the notion employed in testing (epistemology) and (ii) with the conditions – e.g. invariance – to impose on the variation to interpret it causally (methodology). *Last*, the rationale of variation is a first step in redeeming the social sciences as *sciences*. In fact, as individuals and societies are too mutable, the social sciences cannot establish universal and necessary regularities as physics does. For this reason their scientific status has been often questioned. This indictment is ill-founded because the social sciences aim at establishing causal variations rather than regularities.

University of Kent
f.russo@kent.ac.uk
Université Catholique de Louvain
federica.russo@uclouvain.be

REFERENCES

Agresti, A. (1996), *An Introduction to Categorical Data Analysis*, New York: Wiley.

Blalock, H.M. (1964), *Causal Inference in Nonexperimental Research*, Chapel Hill: University of North Carolina Press.

Boudon, R. and Lazarfeld, P. (1966) (eds.), *L'analyse Empirique de la Causalité*, Paris: Mouton&Co.

Cartwright, N. (1979), "Causal Laws and Effective Strategies," *Noûs*, 13, pp. 419-437.

Cartwright, N. (1989), *Nature's Capacities and their Measurement*, Oxford: Clarendon Press.

Courgeau, D. (2003) (ed.), *Methodology and Epistemology of Multilevel Analysis. Approaches from Different Social Sciences*, Dordrecht: Kluwer.

Dowe, P. (1992), "Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory," *Philosophy of Science*, 59, pp. 195-216.

Duncan, O.D. (1975), *Introduction to Structural Equation Models*, New York: Academic Press.

Durkheim, E. (1895-1912), *Les Règles de la Méthode Sociologique*, Paris: Libraire Félix Arcan, 6th edition.

Durkheim, E. (1897-1960), *Le Suicide*, Paris: Presses Universitaires de France.

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

Goldstein, H. (2003), *Multilevel Statistical Analysis*, London: Arnold, 3rd Edition.

Good, I.J. (1961), "A Causal Calculus I," *British Journal for the Philosophy of Science*, 11, pp. 305-318.

Good, I.J. (1962), "A Causal Calculus II," *British Journal for the Philosophy of Science*, 12, pp. 43-51.

Haavelmo, T. (1943), "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, 11, pp. 1-12.

Haavelmo, T. (1944), "The Probability Approach in Econometrics," *Econometrica*, 12, pp. iii-vi + 1-115.

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

Hausman, D. and Woodward, J. (1999), "Independence, Invariance, and the Causal Markov Condition," *British Journal for the Philosophy of Science*, 50, pp. 521-583.

Hausman, D. and Woodward, J. (2004), "Modularity and the Causal Markov Condition: a Restatement," *British Journal for the Philosophy of Science*, 55, pp. 147-161.

Hellevik, O. (1984), *Introduction to Causal Analysis*, London: Allen & Unwin.

Hume, D. (1739-1740), *A Treatise of Human Nature*, edited by Norton, D.F. and Norton, M.J., Oxford/ New York: Oxford University Press, 2000.

Hume, D. (1748), *An Enquiry Concerning Human Understanding*, edited by Beauchamp, T.L., Oxford/New York: Oxford University Press, 1999.

Koopmans, T. C. (1950), "When is an Equation System Complete for Statistical Purposes?," in: Koopmans, T.C. (ed.), *Statistical Inference in Dynamic Models* (Cowles Commission Monograph n.10), New York: Wiley.

Lewis, D. (1973), "Causation," reprinted with postscripts in Lewis D. (1986), *Philosophical Papers II* (pp. 159-213), Oxford: Oxford University Press.

Long, J. S. (1983), *Covariance Structure Models*, Beverly Hills: Sage.

Mackie, J.L. (1980), *The Cement of the Universe: A Study on Causation*, Oxford: Clarendon Press.

Menzies, P. and Price, H. (1993), "Causation as a Secondary Quality," *British Journal for the Philosophy of Science*, 44, pp. 187-203.

Mill, J.S. (1843), *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*, New York: Harper & Brothers, 8th edition.

Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems*, San Mateo (Cal.): Morgan Kaufman.

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

Price, H. (1991), "Agency and Probabilistic Causality," *British Journal for the Philosophy of Science*, 42, pp. 157-176.

Price, H. (1992), "Agency and Causal Asymmetry," *Mind*, 101, pp. 501-520.

Quetelet, A. (1869), *Physique sociale. Ou Essai sur le Developpement des Facultés de l'Homme*, Bruxelles: Muquardt.

Robinson, W. (1950), "Ecological correlations and the behaviour of individuals," *American Sociological Review*, 15, pp. 351-357.

Russo, F. (2005), *Measuring variations. An epistemological account of causality and causal modelling*, PhD Thesis, Université catholique de Louvain, Belgium.

Russo, F., Mouchart, M., Ghins, M., and Wunsch, G. (2006), "Statistical Modelling and Causality," Discussion Paper 0601, Institut de Statistique, Université catholique de Louvain, Belgium.

Salmon, W.C. *et al.* (1971), *Statistical Explanation and Statistical Relevance*, Pittsburgh: Pittsburgh University Press.

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

Spirtes, P., Glymour, C. and Scheines, R. (1993), *Causation, Prediction, and Search*, New York: Springer-Verlag, 2nd Edition, MIT Press (2001).

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

Snijders, T. and Bosker, R. (2004), *Multilevel analysis. An Introduction to Basic and Advanced Multilevel Modelling*, London: Sage Publications, 4th edition.

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

Wright, S. (1921), "Correlation and Causation," *Journal of Agricultural Research*, 20, pp. 557-585.

Wright, S. (1934), "The Method of Path Coefficient," *Annals of Mathematical Statistics*, 5(3), pp. 161-215.

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

Philosophica 77 (2006) pp. 125-150

CAUSAL PLURALISM AND SCIENTIFIC KNOWLEDGE: AN UNDEREXPOSED PROBLEM[1]

Leen De Vreese

ABSTRACT

Causal pluralism is currently a hot topic in philosophy. However, the consequences of this view on causation for scientific knowledge and scientific methodology are heavily underexposed in the present debate. My aim in this paper is to argue that an epistemological-methodological point of view should be valued as a line of approach on its own and to demonstrate how epistemological- methodological causal pluralism differs in its scope from conceptual and metaphysical causal pluralism. Further, I defend epistemological-methodological causal pluralism and try to illustrate that scientific practice needs diverse causal concepts in diverse domains, and even diverse causal concepts within singular domains.

1. Introduction

Causal pluralism is currently a hot topic in philosophy. However, the consequences of this view on causation for scientific methodology and scientific knowledge are heavily underexposed in the present debate. The current literature in defence of causal pluralism seems to focus particularly on conceptual causal pluralism. Conceptual causal pluralists are convinced that our everyday notion of "causation" cannot be described univocally. No single current theory of causation seems

[1] I am greatly indebted to Erik Weber and the referee for their useful suggestions and insightful comments on earlier versions of this paper. The research for this paper was supported by the Research Fund of the Ghent University through research project BOF2001/ GOA/ 008 and by the Fund for Scientific Research - Flanders through research project G.0651.07.

sufficient to cover our notion of "cause" in all its diversity, which convinces conceptual causal pluralists that we will have to combine several causal theories in one way or another to get a grip on all these ways the notion of cause is used in our everyday causal talk. Another reason to break with conceptual causal monism is the finding that contextual factors influence the way we interpret the notion of "cause". Our interests, human limitations, and specific goals in asking for a cause, can affect the causal concept used and the causal selections made. Only a few authors overtly argue in one way or another that causal pluralism is not only a conceptual matter, but can also be defended on the basis of metaphysical considerations. The conviction is then that causation is also "plural" in reality, apart from how we perceive it and reason about it. One can, for example, argue that several empirical relations can be discerned which can all be labelled "causal", or that each level of organization has its specific causal characteristics which can only be captured by means of different approaches to causation in the world. However, one does not need to be metaphysical causal pluralist to be conceptual causal pluralist. One can for example argue that there is in fact only one kind of causal relation in the world - possibly at the fundamental organizational level - but that we need several causal concepts in everyday causal reasoning for pragmatic reasons. I am convinced one should discern epistemological- methodological causal pluralism as a still different approach, stating that we need different concepts of causation to gain scientific knowledge. Surprisingly, the latter kind of causal pluralism is currently hardly debated. If scientific examples are referred to in the debate on causal pluralism, they are used to defend a certain conceptual or metaphysical approach to causation, but hardly ever is it recognized that taking a conceptual and/or metaphysical pluralistic stance bears consequences for our scientific methodology and scientific causal knowledge, and that this in turn implies specific philosophical questions with respect to our scientific notion of causation.

My aim in this paper is to argue that an epistemological-methodological point of view (to which I will further refer by the contraction epimethodology) should be valued as a line of approach on its own, and to demonstrate how epimethodological causal pluralism differs in its scope from conceptual and metaphysical causal pluralism. In section 2, I discuss how an epimethodological approach is different from, but meanwhile related to, one's conceptual and metaphysical causal point

of view. In section 3, I comment on Phil Dowe's argumentation in *Physical Causation* (Dowe, 2000) to make clear that different scientific domains lean on different approaches to causation in their attempt to gain appropriate causal knowledge, and to illustrate further that a metaphysical approach has to be distinguished from an epi-methodological approach to causation. Additionally, I strengthen my point by indicating the applicability of the totally different approach of J. L. Mackie (1974) to the engineering sciences. Section 4 demonstrates that different approaches to causation can be even useful within a single domain of science by referring to the biomedical sciences. Section 3 and 4 show the need of a pluralistic epimethodological approach. I further substantiate this claim in section 5, by arguing that we should subscribe to epimethodological causal pluralism if we aim at accurate and adequate causal conceptions to base our scientific methodology and knowledge on. Section 6 contains my final conclusions.

2. Epimethodological causal pluralism vis-à-vis conceptual and metaphysical causal pluralism

2.1. Differences

Philosophers working on the topic of causation often use scientific examples as similar to everyday examples, and hence to underpin the conceptual point of view defended. Although our notion of causation in science is not at all totally different from our everyday notion of causation, this treatment does not recognize that science has specific aims and interests which can affect the way we reason about causation within scientific practice. A scientist will, for example, often be looking for causal relations on which general *policy* decisions can be based, while in everyday causal reasoning, one will rather be looking for singular factors *responsible* for certain specific effects. This example shows why it is not justifiable to interpret an epimethodological approach to causation as completely similar to a conceptual one.

On the other hand, science is often interpreted as a unique point of reference for finding out what causation is in the world within the literature on causation. In other words, science is used as a means to develop a metaphysical approach to causation. Phil Dowe describes his

approach in *Physical Causation* (Dowe, 2000) precisely as one that leans on the results of science to find out what the language-independent entity called "causation" is:

> [...] there are procedures for investigating such an entity, namely, the methods of science, which is in the business of investigating language-independent objects. Empirical philosophy can draw on the results of science, and so can investigate such concepts, in this case causation 'in the objects'. (Dowe, 2000:7)

He further argues that this "empirical analysis" is not a conceptual analysis of the way this term is used by scientists, but rather a conceptual analysis of the concept inherent in scientific theories. This way, Dowe hopes to get rid off false intuitions on what causation is. However, his approach is still problematic. First, science itself is not entirely free of (causal) intuitions. The weight of these intuitions is reduced to minimal portions in the acquisition of scientific knowledge, but science can nevertheless not do without a minimum of intuitions as Timothy Williamson (2004) argues. Further, there is no way to ensure ourselves that these "scientific" intuitions, or these intuitions on which science is based, are better than our everyday intuitions. All the more because the former are in line with the latter.

Secondly, the assumption that there is an objective, univocal concept of causation implicit in our scientific theories can be questioned, as is done by Daniel Steel[2]. Steel argues that any empirical analysis of causation will inevitably be a substantive thesis over and above what is given by the theories from physics or any other scientific discipline, since the term "cause" is never explicitly defined within these theories. Further, since the aggregate of scientific theories enables us to interpret "causation" in widely divergent ways, the choice for a single approach as the most adequate will always appeal to intuitions on the proper usage of the term "cause".

---

[2] This is argued for by Daniel Steel in a draft version of a book chapter entitled "Causal Structures and Causal Mechanisms" which I received through personal communication and which is to appear in (Steel, forthcoming).

Additionally, it can be argued that causal knowledge in general is not perpective-free, as is extensively defended by Huw Price in (Price, 2001) and (Price, 2007):

> I don't want to eliminate causation altogether from science, but merely to put it in its proper place, as a category that we bring to the world — a projection of the deliberative standpoint. Causal reasoning needn't be bad science, on my view. On the contrary, it's often an indispensable construct for coping with the situation we find ourselves in, as enquirers and especially as agents. It is bad science to fail to appreciate these facts, but not bad science to continue to use causal notions, where appropriate, having done so. Some perspectives simply cannot be transcended. (Price, 2007: 290)

Huw Price does not want to claim that causality is ontologically subjective, in that the existence of causal relations would depend on the presence of an observator. In his pragmatist approach, Price disregards this problem of ontological realism, namely whether there are real causal relations in the world when there are no agents observing them. Price's pragmatic causal view states that causality is practice-subjective. Practice-subjectivity is no ontological matter but neither a psychological matter in that talk of causation would be talk purely about agents or agency, and not about the world. Price's practice-subjective pragmatism about causation is tied to the conviction that we cannot understand the notion of "causation" if we do not understand its origins in the lives and practice of the agents using this notion. According to this view, the concept of "cause" is essentially tied to the experience of agents. This implies that our notion of cause is developed from a specific perspective, which cannot be transcended. Hence, an adequate philosophical description of causality should refer to the role of the concept in the lives and practice of these agents. This does nonetheless not justify the claim that there would have been no causation in the world if there had not been any agents observing them, nor that our notion of "cause" cannot tell us anything about the world. To the contrary, it tells us a lot about the world, but it does this from a specific perspective which we cannot escape from.

In conclusion, the inevitable influence of scientific intuitions, the absence of an objective univocal concept inherent in our scientific

theories, and the perspectivalism of (scientific) causal knowledge demonstrate that it is not justifiable to treat an epimethodological approach as similar to a metaphysical one.

2.2. Relations

Notwithstanding that an epimethodological approach to causation should be discerned from a conceptual and metaphysical one, such an approach should be compatible with the conceptual and metaphysical position taken toward causation. One should situate an epimethodological approach in between a conceptual and metaphysical approach. As has already been said, our scientific notion of causation lies in line with our everyday notion, but will on the other hand be affected by the specific goals, interests, etc. of the scientific domain which is involved. On the other hand, although scientists aim at a precise description of causal relations in the world, their views will not be perspective-free, will be determined by human restrictions, and built on their specific scientific interests and convictions. These features of scientist's causal view on their domain have not to be interpreted as a problem, but might just be necessary to be able to gain scientific knowledge at all, and hence to reach a description which is as precise as possible but meanwhile also useful in practice. Just like Huw Price argues with respect to perspectivalism:

> [...] unmasking the perspectival character of a concept does not lead to simple-minded antirealism — we may continue to use the concept, and even to affirm, in a variety of ways, the objectivity of the subject-matter concerned, despite our new understanding of what is involved (of where we 'stand') in doing so. Nevertheless, there is a tendency to think that perspectivity is incompatible with good science, in the sense that science always aims for the perspective-free standpoint, the view from nowhere. (Price, 2007:253)

I will focus here on the relations between an epimethodological and metaphysical approach to causation. Taking a certain metaphysical position will entail specific questions concerning scientific causal knowledge. These kinds of questions become even more significant if one takes a pluralistic metaphysical position toward causation. It are

these kinds of questions which characterize this separate line of approach to causal pluralism.

Let me illustrate the specific character of an epimethodological approach by way of a biological example of James Woodward (2003):

> As an illustration, consider the lac operon model for E. coli due to Jacob and Monod, which was widely regarded as a seminal discovery in molecular genetics. When lactose is present in its environment, E. coli produces enzymes that metabolize it, but when lactose is absent, these enzymes are not produced. What determines whether these enzymes are produced? According to the model proposed by Jacob and Monod, there are three structural genes that code for the enzymes as well as an operator region that controls the access of RNA polymerase to the structural genes. In the absence of lactose, a regulatory gene is active which produces a repressor protein which binds to the operator for the structural genes, thus preventing transcription. In the presence of lactose, allolactose, an isomer formed from lactose, binds to the repressor, inactivating it and thereby preventing it from repressing the operator, so that transcription proceeds. Biologists describe this as a case of "negative control". Unlike "positive control," in which "an inducer interacts directly with the genome to switch transcription on" (Griffiths, Miller, Suzuki, Lewontin, and Gelbart 1996, p.550), the inducer in this case, allolactose, initiates transcription by interfering with the operation of an agent that prevents transcription. [...] A causal relationship is clearly present between the presence of allolactose and the production of the enzymes, and the former figures in the explanation of the latter, but there is no transfer of energy from, or spatiotemporally continuous process linking, the two. (Woodward, 2003:225-226)

Woodward in fact used this example to underpin his view that one should resist a proliferation of concepts of causation in favour of a monistic conceptual approach. Specifically, he was arguing against the distinction made by Ned Hall (2004) between causation as dependence (which can be grasped in terms of counterfactuals) and causation as production (which needs another than a counterfactual approach). According to Hall, some causes can be causes in, for example, the dependence sense, but not in the production sense. Allolactose in the example above would form an instance of such a cause. In contrast, Woodward defends a

manipulationist account following which allolactose should straightforwardly be taken as a cause, without making any distinctions with regard to those inducers involved in what biologists refer to as "positive control". Biologists also seem to refer to both of these causes as just straightforward causes, and this convinces Woodward that one does not need Hall's distinction. I think Woodward is too fast in making this conclusion. In fact, Woodward's example conflicts with his own conclusion, since biologists do make a distinction. Although they refer to both kinds of influence of inducers as "causes", they introduce the labels "positive control" and "negative control" to discriminate between both. These labels clearly refer to two discernable ways of causing the transcription.

The answers to metaphysical questions will bear consequences for our notion of causation in the sciences and will lead to specific epimethodological questions. Suppose we accept that causality in the world consists of a single kind of empirical relation, but that we nonetheless find that a distinction between production causes and dependence causes is frequently used in biology. In that case we should be able to justify this distinction in biology and to relate it to a univocal metaphysical approach. Typical epimethodological questions that will follow are: Why do we need this distinction in scientific practice? How are they related to the univocal metaphysical account? Does one of both refer to real causation, and the other to a kind of quasi-causation[3]? Alternatively, are they both subconcepts of the univocal metaphysical concept? In that case, can all relevant causal relations within the biological domain be captured by way of these two subconcepts, or do we need to make further conceptual distinctions? Etc. Suppose, on the other hand, that we accept that the distinction between production and dependence refers to a real distinction between two kinds of empirical relations in the world. This conviction would, for example, lead to the following question: Are both kinds of causal relations present at all levels of organization and are they by consequence of equal importance for our causal knowledge of all domains of science?

---

[3] The term quasi-causation is introduced by Phil Dowe to make a distinction between real causes according to his theory in terms of conserved quantities, and, on the other hand, omitters and preventers. (Dowe, 2000; Dowe, 2004)

All these questions form typical epimethodological questions, related to, but not equal to, metaphysical questions regarding causation. Given the importance of causality in science, it is important to be aware of the specific character of the epimethodological approach. Epimethodological questions become even of crucial importance if one takes a pluralist position toward causation. However, specific epimethodological questions are hardly tackled in the current debate on causal pluralism. I am nevertheless convinced that we should appreciate the epimethodological approach as important on its own within the debate on causal pluralism. Generally spoken, such an approach should investigate whether scientists reason in a pluralistic way in their search for knowledge; which convictions lie at the basis of pluralistic scientific causal reasoning; what the consequences are of causal pluralism for scientific methodology; whether one can find differences between different domains in the answers to these questions; and if so, what these differences are.

To concretize the concerns of the epimethodological approach to causal pluralism, I tackle in the following sections two general epimethodological questions of central importance, namely whether we need diverse concepts of causation in diverse scientific domains (section 3), and whether we need diverse concepts of causation within single scientific domains (section 4). The examples will make clear that a positive answer should be given to both of these questions.

3. Diverse concepts in diverse scientific domains

3.1. Conserved quantities in the physical sciences

In section 1, I indicated that Phil Dowe (2000) describes his approach to causation as one that leans on the results of science to find out what the language-independent entity called "causation" is. In this section, I will argue against the line of argumentation of Dowe (2000) to make clear that different scientific domains lean on different approaches to causation to gain appropriate causal knowledge, and to demonstrate that a metaphysical approach has to be distinguished from an epimethodological approach to causation.

In fulfilling his goal, Dowe rejects one by one all of the "*major theories of physical causation*" (Dowe, 2000:12), such as Lewis' counterfactual theory, Hume's regularity theory, Suppes' probabilistic account and Salmon's process theory. For example, in the introduction to the second chapter, Dowe announces:

> First, the Humean deterministic accounts are rejected on the grounds that science yields examples of indeterministic causation, and second the probabilistic accounts of causation, including Lewis's counterfactual probabilistic theory, are shown to fall to a well-directed example of chance-lowering causality. (Dowe, 2000:14)

In both cases, the examples used are mainly taken from the domain of physics. For Dowe even a single physical counterexample forms a reason to reject a whole theory:

> In particular, probabilistic theories, taken as aiming to provide an empirical account of singular causation, *fall to an important counterexample from subatomic physics.* (Dowe, 2000:40) (my italics)

Subsequently, transference accounts are rejected, mainly because they cannot cope with immanent causation, a kind of causation which Dowe wants to be able to deal with because it is necessary to understand certain physical processes such as a spaceship's inertia as a cause of its continuing motion:

> there are a number of difficulties with the transference account. These concern problems of the identity over time of the transferred quantities, and the direction of causation. Further, I argue that there is a kind of causation, immanent causation, or causation as persistence, which is neglected by the transference accounts. (Dowe, 2000:41)

Subsequently, Dowe proclaims Salmon's theory as "superior" to the others discussed, though still not adequate. The destructive criticism on the alternative approaches paves the way for the defence of Dowe's own process theory of causation. In view of the foregoing chapters, it is clear that Dowe will conceive his own theory as the single appropriate theory

of "physical causation". In a nutshell, this theory comes down to the following:

> 1. A *causal interaction* is an intersection of world lines which involves exchange of a conserved quantity.
>
> 2. A *causal process* is a world line of an object which possesses a conserved quantity.
>
> A "conserved quantity" is any quantity which is universally conserved, and current scientific theory is our best guide as to what these are. Thus, we have good reason to believe that mass-energy, linear momentum, and charge are conserved quantities. (Dowe, 1995:323)

Now, what is wrong with this approach of Dowe? My criticism here does not concern the content of Dowe's theory and the "internal" relation with alternative ones. What I think of are criticisms about the way he presents his own theory and opposes the others. First of all: Why basing his approach almost exclusively on examples from physics? What is precisely meant by "physical causation"? The term is simply thrown into the arena without further clarification. Nowhere in his book, nor in his articles, Dowe makes explicitly clear which domain(s) of application he is talking about. His reference to all other theories of causation as theories of "physical causation" gives the impression that the word "physical" has to be interpreted very broadly since these alternative theories are meant to be broadly applicable. However, when taking a closer look at Dowe's argumentation, his description of a conserved quantity and the examples used in his book, it appears to concern a theory of physical causation in a rather narrow sense. It is in fact quite clear that his theory is not at all apt to get a grip on, for example, the concept of cause in the social and behavioral sciences. The interests of researchers in those domains of science are not in line with a physical approach to causation focussing on the conservation of momentum, energy, etc. This is not the kind of knowledge which social scientists search for and use in their explanations. For example, how applying Dowe's theory to the way scientists explain the fluctuation of quotations on the stock exchange as an effect of political or social incidents, or the mental condition of people as an effect of life events? Dowe's theory is

clearly not able to give an adequate description of the way scientists come to these kinds of causal explanations, while other approaches to causation are better apt to get a grip on the kind of causal reasoning involved when investigating such kinds of events.

From the book (Dowe, 2000) and his 1992 article (Dowe, 1992) it becomes eventually clear that Dowe is reasoning from a strong reductionistic point of view combined with some kind of supervenience account of causes on conserved quantities:

> Another possible criticism concerns reduction. Fair's approach involves a commitment to reduction of all science to physics, which some may not like. Salmon avoids this with his vague notions of "structure" and "characteristic". These can apply to any area of science, whereas energy and momentum may not. The present suggestion does not share this advantage. One answer is that the generality of "conserved quantity" might allow this to be used as a testable conjecture in various fields of science. But it is unlikely that it would stand the test: conservation laws seem to be confined to the physical sciences. A more desirable option is to take a middle road and adopt a supervenience account such as that of Kim (1984) where causes supervene on conserved quantities (see Menzies 1988). (Dowe, 1992:214-215)

However, this view is not explicitly referred to as the basis for his line of argumentation, and neither is it thoroughly motivated. His whole theory is nonetheless strongly connected to this reductionistic presupposition and the presupposition itself is not at all self-evident. Not everyone will easily accept this reductionistic causal point of view on the world.[4] And even if Dowe's reductionistic point of view would be justified as a metaphysical approach, we are clearly not able, and probably even not interested, in studying causal processes at all organizational levels in terms of causal processes on the elementary organizational level, which would not make a difference for Dowe's approach given his metaphysical aims, but would be an important fact for an epimethodological approach.

---

[4] It has, for example, been criticized by Nancy Cartwright (1999) and by Stephen Webster (2003).

Hence, although Dowe claims that his approach is founded on the causal concepts inherent in scientific theories, his choice for physics is clearly based on a further metaphysical presupposition in favor of physics, which is crucial for his point of view but nonetheless not defendable on the basis of "the results of science" on which Dowe nonetheless claims to found his whole approach. Whether this metaphysical presupposition is justified, and whether the resulting metaphysical approach to causation is justified, is not the concern of an epimethodological approach to causation. What will concern such an approach is that the choice for another scientific domain than physics would have resulted in a totally different approach to causation. The causal concepts inherent in other than physical scientific theories will clearly not all be of the kind on which Dowe's approach is based.

Seen from an epimethodological point of view, and provided that one takes a pluralistic stance, there is no problem at all in the inadequacy of Dowe's process theory of causation for the characterization of causal processes in other domains such as the social sciences. From such a point of view, one can appreciate Dowe's approach as one that characterizes the concept "cause" prominent in the physical sciences. As Dowe in fact admits himself, the same characterization will not be applicable in any area of science. Apart from causal reasoning in the social sciences, conserved quantities do neither seem to play a prominent role in causal reasoning in, for example, the domain of engineering or biomedicine, which will be focussed on in the following two sections.

3.2. Sufficient and necessary conditions in the engineering sciences

Let me start with trying to find out which kind of approach to causation is prominent in causal reasoning in the engineering sciences.[5] Engineers have to think in terms of function. The physical structure of the things they design, should be a means to produce a certain effect Y. Hence, building an X has the function to produce Y. In other words, whenever a physical structure enables X to take place, Y should occur. Put still differently, if X is a good tool to produce Y, then Y will occur under the usual ceteris paribus condition that no disturbances occur. However,

[5] My analysis here is based on (Kroes, 1998) and (Kroes, 2003).

different physical structures can form a useful tool to perform the same function, leading to the same effect. Common to all these structures will be that they are designed with knowledge of some physical phenomena and with the purpose to take certain actions that lead to the desired effect. Hence, what engineers need to think of are designs of physical structures which have, within the normal background conditions, the necessary and sufficient characteristics to function in the production of a specific effect. I take an example from Peter Kroes (in a detailed way presented in (Kroes, 1998)), namely the Newcomen engines, which are one of the earliest types of steam engines. Those engines were used for pumping water, which is their function Y. An explanation of how the fulfillment of this function was brought about by the machine, will have to refer to certain physical phenomena (transforming water into steam increases its volume manyfold, cooling of steam in a closed vessel creates a vacuum, etc.), the design of the engineer (the steam engine consists of certain parts such as a cylinder, piston, great beam, boiler, etc.; the piston may move up and down in the cylinder, etc.), and a series of actions (after opening the steam valve the cylinder fills with steam and the piston moves up, closing of the steam valve and injection of cold water creates a vacuum in the cylinder; etc.) (Kroes, 1998:8). All these elements together explain how the function is fulfilled by the machine in normal circumstances.

This example indicates that the primary concept of causation within the engineering sciences can probably best be captured by John L. Mackie's approach to causation (Mackie, 1974). This approach is based on the approach of John Stuart Mill (Mill, 1973). Mackie takes Mill's "plurality of causes" as the basis of his own approach:

> It is not true that one effect must be connected with only one cause, or assemblage of conditions; that each phenomenon can be produced only in one way. There are often several independent modes in which the same phenomenon could have originated. One fact may be the consequent in several invariable sequences; it may follow, with equal uniformity, any one of several antecedents, or collections of antecedents. (Mill, 1973: Book III, Chapter 10, Section 1; cited in Mackie, 1974:61)

This is formally expressed by Mackie as " 'All (ABC or DGH or JKL) are followed by P' and 'All P are preceded by (ABC or DGH or JKL)' "

(Mackie, 1974:62). He ascribes the following characteristics to the elements of this definition:

> the complex formula '(ABC or DGH or JKL)' represents a condition which is both necessary and sufficient for P: each conjunction, such as 'ABC', represents a condition which is sufficient but not necessary for P. Besides, ABC is a *minimal* sufficient condition: none of its conjuncts is redundant: no part of it, such as AB, is itself sufficient for P. But each single factor, such as A, is neither a necessary nor a sufficient condition for P. Yet it is clearly related to P in an important way: it is an *insufficient* but *non-redundant* part of an *unnecessary* but *sufficient* condition: it will be convenient to call this (using the first letters of the italicized words) an *inus* condition. (Mackie, 1974:62)

The single factors of the disjunction of conjunctions can further also be negative conditions. $\bar{A}$ is then the absence of a counteracting cause A. Mackie recognizes these causes as real causes, contrary to Mill. Further Mackie also introduces the importance of a causal field (F), which forms the background of the causal event, but does not make part of it. An explosion in a block of flats will, for example, be attributed to a gas leak, while other factors such as the presence of the building and its gas pipes, and of people living in that building and lighting cigarettes from time to time, will rather be interpreted as the "normal" background situation within which the causal event arose.

Mill's description of "causation" is reflected in the notion of cause prominent in the engineering sciences, as described above. Note the following properties: its focus on deterministic causation, the incorporated idea of the importance of the plurality of single factors leading all together to a certain effect, and the plurality of different possible conjunctions of conditions all sufficient to lead to the effect, the inclusion of the absence of certain interfering elements as genuine causes and the importance of the causal field as the background within which the causal event arises. All these characteristics describe a notion of "cause" which lies much closer to the way one reasons about causation in the engineering sciences than, for example, in the physical sciences or maybe even clearer, in the biomedical sciences.

4. Diverse concepts within a single scientific domain: average effects in the biomedical and social sciences

While the interests of physical scientists lie in detailed descriptions of the causal processes and causal interactions leading up to some particular effect, social and biomedical scientists are interested in general overall causal patterns recurring in the population. These causal patterns are not necessarily exceptionless, as is presupposed in the case of the engineering sciences. The precise causal history leading up to one particular effect in singular cases is often very complicated and intractable and not interesting for the general purposes of the biomedical sciences. This led e.g. John Dupré (1993) and Ronald N. Giere (1997) to the following view with respect to probabilistic theories of causation:

> [...] causes should be assessed in terms of average effect not only across different causal routes, but also across varying causal contexts. (Dupré, 1993:199)

and hence:

> One interesting fact about these models is that it could turn out that **C** is causally irrelevant for **E** in the *population* U even though **C** is *not* causally irrelevant for **E** in all *individuals* in U. [...] Population models always average over individuals and, therefore, ignore what might be important differences among individuals. (Giere, 1997:204-205).

It is argued one should execute controlled experiments on fair samples to find statistically significant differences between experimental and control groups. These differences are claimed to be good standards for causal judgments about populations. Dupré opposes the rival view of e.g. Nancy Cartwright (1979), Paul Humphreys (1989) and Ellery Eells (1991). They hold on to the context unanimity condition. This condition maintains that:

> a genuine cause must raise the probability of a genuine effect of it **in *every* causal background context**. (Eells, 1991:94) (my bold)

More specifically, this would come down to the following:

> To use an example of Cartwright's (1979), ingesting an acid poison (X) is causally positive for death (Y) when no alkali poison has been ingested (~F), but when an alkali poison has been ingested (F), the ingestion of an acid poison is causally negative for death. I will argue that in a case like this it is best to deny that X is a positive causal factor for Y, even if, overall (for the population as a whole), the probability of death when an acid poison has been ingested is greater than the probability of death when no acid poison has been ingested (that is, even if $Pr(Y|X) > Pr(Y|\sim X)$. I will argue that it is best in this case to say that X is causally *mixed* for Y, and despite the *overall* or *average* probability increase, X is nevertheless not a positive causal factor for y in the population as a whole. (Eells, 1991:94)

As John Dupré (1993) argues, this kind of causal knowledge is not of interest in social and biomedical scientific practice. What he nonetheless does not recognize is that the context unanimity approach would form a more adequate approach than the average effect approach for finding out what the real causal structure of the world consists in. Dupré fails to discriminate here between a metaphysical and an epimethodological approach. He takes his own approach, which is clearly an epimethodological one, as a metaphysical approach. His own approach should then replace the unanimity approach, which he doesn't want to appreciate as a metaphysical one because of its impractibility:

> We should avoid metaphysical doctrines for which we neither have, nor possibly could have, empirical evidence of applicability. This is a methodological principle that the unanimity thesis fails dismally to satisfy. (Dupré, 1993:201)

Although the unanimity approach may not be practicable, it would form an adequate toolbox for metaphysical purposes. It would however not be adequate for the main interests of social and biomedical sciences, where the context unanimity condition is of little or no practical use. What the approach does, is demonstrating the limitations and presuppositions of our research methods. This seems precisely what a metaphysical approach should do. Being of practical use is to the contrary something one should expect from an epimethodological approach to causation.

Dupré (1993) convincingly argues that what one needs in the practical context of the biomedical and social sciences is an average effect approach rather than a context unanimity approach to causation. If we take a closer look at scientific practice within specifically the biomedical sciences, one will nevertheless rapidly be convinced of the importance of the search for causal mechanisms, next to the search for probabilistic evidence. Causal mechanisms are necessary for further confirming and explaining the causal relations derived on the basis of correlations.

This is clearly argued for by, among others, Paul Thagard in *How scientists explain disease* (Thagard, 1999). Thagard argues that knowledge of mechanisms is not strictly necessary, but nonetheless often searched for by medical researchers, in line with everyday causal reasoning:

> Whereas causal attribution based on correlation (covariation) alone would ignore mechanisms connecting cause and effects, ordinary people are like medical researchers in that they seek mechanisms that connect cause and effect. [...] Reasoning about mechanisms can contribute to causal inference, but it is not necessary for such inference. In domains in which causal knowledge is rich, there is a kind of feedback loop in which more knowledge about causes leads to more knowledge about mechanisms, which leads to more knowledge about causes. But in less well-understood domains, correlations and the consideration of alternative causes can get causal knowledge started in the absence of much comprehension of mechanisms. (Thagard, 1999:109)

Thagard (1999) and Ahn & Kalish (2000) indicate three ways in which causal mechanisms can contribute to our search for causal explanations. First, they can confirm the existence of possible causal relations supposed to be present on the basis of a correlation. Hence a mechanism can help as a confirmative tool. Secondly, when a causal connection between two variables is supposed on the basis of statistical information, but no plausible mechanism can be found that elucidates this connection,

we may be confronted with a spurious or a coincidental relation[6]. In this case the mechanism works as a disconfirming tool with respect to statistical evidence. Thirdly, a hypothetical mechanism points to possible causes that can be (dis)confirmed by other indicators. In this case, mechanisms work as a heuristic tool for the generation of hypotheses which are then further testable on the basis of, for example, statistical information. As Thagard points out, these characteristics constitute three ways in which mechanisms can enhance the explanatory coherence of causal explanations.

According to Thagard's analyses, biomedical scientific practice is based on at least two different causal approaches: a probabilistic approach in terms of average effects, and an approach in terms of causal mechanisms[7]. Since the idea of a causal mechanism encloses precisely the conviction that there is a process at an *underlying* organizational level connecting cause and effect, causal mechanisms will clearly not be of any practical use when trying to get a grip on the causal relations at the elementary organizational level[8], with which Dowe is concerned. Consequently, this example does not only illustrate that one will (often) need diverse concepts of causation to successfully gain causal knowledge within a single scientific domain, but adds as well to my conviction that one will need diverse causal concepts in diverse scientific domains.

5. Methodology for the epimethodological causal pluralist

The examples discussed in the previous sections demonstrate that one needs a pluralistic epimethodological approach to causation. Scientific practice cannot do with a single, univocal approach to causation.

---

[6] The usefulness of this method is nevertheless denied for the social sciences by Daniel Steel in (Steel 2004), by means of the argument that one can very easily imagine a plausible mechanism connecting nearly any two variables representing aspects of social phenomena.

[7] A defence of the need for causal pluralism within the social sciences can be found in Weber (2007).

[8] See, for example, Glennan (1996).

Although this is not yet a generally accepted idea, it seems not that strange. Let me refer to the words of G. L. Newsome to illustrate this:

> Theorists might integrate (these) approaches by assuming (1) that cognizers' conceptions of causality often vary as a function of their existing world knowledge, and (2) that the role played by different aspects of these conceptions may vary as a function of the situation in which the causal judgment is made. (Newsome, 2003:100)

This statement forms part of Newsome's plea to integrate a covariation approach with a mechanism approach to causation, but the basic idea can be generalized to all approaches to causation.

As is clear from the examples above, one will be confronted with (a) certain specific conception(s) of causality once one is focussing on a specific domain of science. This conception will depend on the specific kind of causal knowledge — related to the typical organizational level of reality — researchers in the involved domain are working with. Further, the type of research involved and/or the scientific (sub)domain itself form a typical situation or context in which the causal judgments are made — this includes a typical kind of interest one starts from in the investigation.

To make this clearer, I introduce the distinction between accuracy and adequacy, in line with Philip Kither's view on scientific knowledge (Kitcher, 2001). In his book, Philip Kitcher compares "making science" in general with the making of maps in that what is accurate information on a map is dependent on the needs and desires of the users, and hence does not need to be in full correspondence with reality:

> Like maps, scientific theories and hypotheses must be true or accurate (or, at least, approximately true or roughly accurate) to be good. But there is more to goodness in both instances. Beyond the necessary condition is a requirement of significance that cannot be understood in terms of some projected ideal - completed science, a Theory of Everything, or an ideal atlas. (Kitcher, 2001:61)

Translated to our framework, accuracy refers to the ability of a causal theory to describe the way scientists achieve justified knowledge of the causal relations in the world; adequacy refers to the ability of a causal theory to describe the way scientists achieve the kind of causal

knowledge requested in the involved situation. The latter is related to the way they look at the (scientific) information they have, the point of view from which they analyze the situation, the aim in questioning for the cause, the domain of science they are working in, ... A theory of causation adequately describes causal reasoning in a certain context if it shows us how scientists achieve the kind of information they are interested in, that is, which is useful, relevant, and accessible within the context they are reasoning from. Further, as Kitcher demonstrates, what is believed to be an accurate description (of "causation") is in turn dependent on what is believed to be an adequate description (of "causation") in the involved context:

> What counts as an omission or an inaccurate spatial representation depends on the conventions associated with the kinds of maps, and, in their turn, those conventions are in place because of the needs of the potential users. (Kitcher, 2001: 56)

Let me illustrate this by means of Kitcher's example, namely the map of the London Underground. Since this map enables ten thousands of people a day to successfully find their way through the underground network, it clearly provides them with an accurate description of this network. The map is nonetheless not meticulously describing the real situation of the network in that it does, for example, present all underground lines as straight lines, and denies the real distance between the underground stations in its representation of the situation. As travellers we nevertheless perceive the map as accurate, since it is accurate in view of its goal of informing us on how to travel from one underground station to another. In view of this goal, it should provide certain information and neglect other information, which precisely makes it an adequate map for users of the London Underground. The same map would however be totally inadequate for a building contractor planning public works in the London Underground, since for use in that context, it would be far too inaccurate with respect to, amongst others, the precise length of the underground lines and the precise position of the underground stations. This demonstrates that what is interpreted as an accurate description depends on what is perceived as adequate in the context.

When returning to the probabilistic versus the mechanical approach in the biomedical sciences, there seems no objective reason to prefer one analysis above the other, nor to demand that each accepted causal relation should be confirmed by both approaches. In fact, it will be the situation within which we need to decide on the acceptance of a causal relation which will be decisive for the approach to causation to prefer. As Thagard suggested, in a phase of the survey in which knowledge of specific mechanisms is lacking, statistical information can be taken as a sufficient proof for a causal relation, while it is on the other hand possible that in other situations one will require an explaining mechanism before accepting a certain causal relation. Similarly, if we really want to understand the specific process leading to a certain effect, we may focus on causal mechanisms, while focussing on probabilistic evidence if we are mainly searching for general tendencies. As Hitchcock (2003) claimed:

> When we are asked what causes what, we may pay attention to one of these relations in one scenario, to another of these relations in a different scenario. One of these relations may be a component of one philosophically significant concept, while another is a component of another. All of these relations are causal, in a broad sense, and worthy objects of study within a theory of causation. (Hitchcock, 2003:8)

Hence whether a certain causal theory is adequate for our purposes, depends on the context within which we are reasoning and the particular interests connected with it. Striving for a causal theory adequate for use in all or a lot of divergent possible contexts in which causal reasoning occurs, will be at the expense of the required accuracy since no single theory seems able to capture our notion of "causality" in all its diversity in an accurate way. Hence, it is the situation in which the causal judgment is made that determines which approach we will choose as accurate and adequate for our purposes, and consequently, which factor will be labelled as the cause.

The question that may follow is whether this does not inevitably result in a very particularistic approach? Do we need to analyze the usefulness of a certain approach to causation for each causal reasoning process on its own? Of course this forms no solution. Consequently, the

problem is to find a feasible approach that stands midway between a too monolithic and a too particularistic one.

Given that the requirement of accuracy and adequacy demands for a pluralistic approach to causation, my proposal is precisely to start thinking from domains of application when developing an epimethodological approach to causation. With this I come back to my conclusions from the plea of Newsome. To sum up, if we reason from a certain fixed (sub)domain of science, we are confronted with a fixed kind of knowledge and organizational level of the world researchers are reasoning from and with a fixed kind of research situation and some typical research interests. The former seem to ask for a proper form of accuracy, the latter seem to ask for adequacy, such that the demanded accuracy is attainable after all. Hence reasoning from specific (sub)domains offers a tool to fix the way in which one can reason accurately and adequately and hence justifies generalization over the concerned (sub)domain. Focussing on domains will give us the possibility to generalize over a broad area of application. The unity in the way of reasoning about causal relations within one such domain will justify these generalizations.

Remark that we have to do with a continuum ranging from generalizing over several contiguous domains of science (e.g. the social and behavioral sciences, the natural sciences) to specifying into particular subdomains (e.g. medical etiology). The more we specify, the better our characterization of the involved causal concept(s) will do and the easier it will be to increase the accuracy of our causal theories, but the more our characterizations can become fragmented. However, working this way will offer us anyhow some theories of causation that are much closer to the scientific reality than the ones who aim at an overall characterization of the concept of causation inherent in scientific theories.

6. Conclusion

I tried to convince the reader that one should discriminate between three approaches to causal pluralism: conceptual causal pluralism, metaphysical causal pluralism, and epimethodological causal pluralism. I have further tried to demonstrate what kinds of questions can be

answered if one takes an epimethodological approach as an approach of importance on its own. This led me to the defence of epimethodological causal pluralism, by demonstrating that science needs diverse causal concepts in diverse scientific domains, and even diverse causal concepts within singular domains. A lot of work is to be done yet to substantiate epimethodological causal pluralism by further investigating its characteristic research questions in relation to specific scientific domains. The resulting pluralistic approach to causation will anyhow be closer to scientific reality, and will not force one to adopt certain metaphysical positions which cannot be (directly) derived from it.

Universiteit Gent (UGent)
Leen.DeVreese@UGent.be

REFERENCES

Ahn, W. and Kalish, C.W. (2000), "The Role of Mechanism Beliefs in Causal Reasoning," in: Kiel, F.C. and Wilson, R.A. (eds.), *Explanation and Cognition* (pp. 199-225), Cambridge: MIT Press.

Cartwright, N. (1979), "Causal Laws and Effective Strategies," *Noûs*, 13(4), pp. 419-437.

Cartwright, N. (1999), *The Dappled World. A Study of the Boundaries of Sciences*. Cambridge: University Press.

Dowe, P. (1992), "Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory," *Philosophy of Science*, 59, pp. 195-216.

Dowe, P. (1995), "Causality and Conserved Quantities: a Reply to Salmon," *Philosophy of Science*, 62, pp. 321-333.

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

Dowe, P. (2004), "Causes are physically connected to their effects: Why preventers and omissions are not causes," in: Hitchcock, C. (ed.), *Contemporary Debates in Philosophy of Science* (pp. 189-196), Oxford: Blackwell.

Dupré, J. (1993), *The Disorder of Things. Metaphysical Foundations of the disunity of Science,* Cambridge/Londen, Massachusetts/ England: Harvard University Press.

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

Giere, R.N. (1997), *Understanding Scientific Reasoning*, Fortworth: Harcourt Brace College Publishers.

Glennan, S. (1996), "Mechanisms and the Nature of Causation," *Erkenntnis*, 44, pp. 49-71.

Hall, N. (2004) "Two Concepts of Causation," in: Hall, N., Collins J. and Paul, L.A. (eds.), *Causation and Counterfactuals* (pp. 225-276), Cambridge: MIT Press.

Hitchcock, C.R. (2003), "Of Humean Bondage," British Journal for the Philosophy of Science, 54, pp. 1-25.

Humphreys, P. (1989), *The Chances of Explanation. Causal Explanation in the Social, Medical and Physical Sciences*. Princeton: Princeton University Press.

Kitcher, P. (2001) *Science, Truth and Democracy*, Oxford: Oxford University Press.

Kroes, P. (1998), "Technological explanations: The relation between structure and function of technological objects," *Techné*, 3(3), pp. 1-13.

Kroes, P. (2003), "Screwdriver philosophy; Searle's analysis of technical functions," *Techné*, 6(3), pp. 22-34.

Mackie, J.L. (1974), *The Cement of the Universe. A study of causation*. Oxford: Clarendon Press.

Mill, J.S. (1973), "A System of Logic Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation," in: Robson, J.M. (ed.), *Collected Works*, Toronto: University of Toronto.

Newsome, G.L. (2003), "The Debate between Current Versions of Covariation and Mechanism Approaches to Causal Inference," *Philosophical Psychology*, 16 (1), pp. 87-107.

Price, H. (2001), "Causation in the Special Sciences: the Case for Pragmatism," in: Galavotti, M.C., Constantini, D. and Patrick Suppes (eds.), Stochastic Causality (pp. 103-120), Stanford: CSLI Publication.

Price, H. (2007), "Causal perspectivalism," in: Price, H. and Corry, R. (eds.), *Causation, Physics and the Constitution of Reality: Russell's Republic Revisited* (pp. 250-292), Oxford: Clarendon Press.

Steel, D. (2004), "Social Mechanisms and Causal Inference," *Philosophy of the Social Sciences*, 34, pp. 55-78.

Steel, D. (forthcoming), *Across the Boundaries: Extrapolation and Causality in the Biological and Social Sciences,* Oxford: Oxford University Press.

Thagard, P. (1999), *How Scientists Explain Disease*, Princeton (N.J.): Princeton University Press.

Weber, E. (2007), "Conceptual Tools for Causal Analysis in the Social Sciences," in: Russo, F. and Williamson, J. (eds.), *Causality and Probability in the Sciences* (pp. 197-213), London: College Publications.

Webster, S. (2003), *Thinking about Biology*, Cambridge: Cambridge University Press.

Williamson, T. (2004), "Philosophical 'intuitions' and scepticism about judgement," *Dialectica*, 58(1), pp. 109-153.
Woodward, J. (2003), *Making Things Happen. A Theory of Causal Explanation,* Oxford: Oxford University Press.