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.