Causal and constitutive explanation compared

Petri Ylikoski

Department of Social Research

University of Helsinki

petri.ylikoski@helsinki.fi

Abstract. This article compares causal and constitutive explanation. While scientific inquiry usually addresses both causal and constitutive questions, making the distinction is crucial for a detailed understanding of scientific questions and their interrelations. These explanations have different kinds of *explananda* and they track different sorts of dependencies. Constitutive explanations do not address events or behaviors, but causal capacities. While there are some interesting relations between building and causal manipulation, causation and constitution are not to be confused. Constitution is a synchronous and asymmetric relation between relata that cannot be conceived as independent existences. However, despite their metaphysical differences, the same key ideas about explanation largely apply to both. Causal and constitutive explanations face similar challenges (such as the problems of relevance and explanatory regress) and both are in the business of mapping networks of counterfactual dependence – i.e. mechanisms – although the relevant counterfactuals are of a different sort. In the final section the issue of developmental explanation is discussed. It is argued that developmental explanations deserve their own place in taxonomy of explanations, although ultimately developmental dependencies can be analyzed as combinations of causal and constitutive dependencies. Hence, causal and constitutive explanation are distinct, but not always completely separate forms of explanation.

Science is in the business of explaining things. Philosophers of explanation have rightly paid much attention to causal explanation, and some have even suggested that all scientific explanations are causal. However, most philosophers of explanation recognize that there is an important class of non-causal explanations, although it has received much less attention. These explanations are conventionally called constitutive explanations (following Salmon (1984), who introduced the distinction between etiological and constitutive explanations). Constitutive explanations explain how things have the causal capacities they have by appealing to their parts and their organization.

Until recently, very little systematic work has been done on the nature of constitutive explanation apart from Robert Cummins's (1975, 1983) pioneering work.[1] Luckily, Carl Craver's recent book (2007) has brought the issue of constitution to the center of attention of philosophers of explanations. The discussion (Couch 2011, Fazekas & Kertész 2011, Leuridan 2012, McManus 2012) around his proposals, however, shows that the relation between causation and constitution is not entirely clear. What is needed is a systematic comparison of causal and constitutive explanations. This paper aims to provide such a comparison.

I will begin by characterizing the *explananda* of constitutive explanations. I argue that constitutive explanations are primarily explanations of causal capacities, not explanations behaviors. The second section will discuss some important metaphysical differences between causation and constitution. Constitution is a synchronous and asymmetric relation between *relata* that cannot be conceived as independent existences. While there are some interesting relations between building and causal manipulation, causation and constitution are not to be confused. Causal and constitutive explanations have different *explananda,* and they track different kinds of dependency. However, I argue in the third section that despite their metaphysical differences, the same ideas about explanation still largely apply. Both explanations face similar challenges (such as the problems of relevance and explanatory regress) and the same theory of explanation can be used to deal with these problems. Ultimately both etiological and constitutive explanations are thus in the business of mapping networks of counterfactual dependence – i.e. mechanisms – although the counterfactuals are of a different sort. In the final section I take up the issue of developmental explanation. I make the case that developmental explanations deserve their own place in our taxonomy of explanations, although ultimately developmental dependencies can be analyzed as combinations of causal and constitutive dependencies. I also discuss gene knockout experiments that address a particularly interesting type of developmental dependency that intimately combines constitutive and causal elements. Hence, I argue, causal and constitutive explanation are distinct, but not always completely separate forms of explanation.

---

[1] Other important contributions to the understanding of constitutive explanations are Kauffman 1971, Bechtel & Richardson 1993, Simon 1996, Bechtel 2008, and Wimsatt 2007.

1. Causal capacities as *explananda*

The object of constitutive explanation is the causal capacity of a system (Harré & Madden 1975: 101-105, Cummins 1975, 1983: 14-22). The causal capacity tells *what the system would do in specified situations* (e.g. in certain enabling and triggering conditions). Before analyzing in greater detail the differences between causal and constitutive explanations, a few notes on the notion of a causal capacity are in order.

First, I use the term causal capacity to refer to a wide variety of dispositional notions: disposition, ability, power, affordance, liability, propensity, tendency, etc. There are subtle differences between these notions, but for the purposes of this paper they do not matter. What is crucial is that they all give an account what would happen if the entity to which the dispositional predicate is applied were to end up in a certain kind of causal setting. The dispositional predicates provide modal information; they do not state what is happening or has happened, but rather what could, or will, happen in the right kind of causal configuration.

Second, causal capacities, and other dispositional notions, are incomplete predicates (Prior 1985: 7-9). This means that the causal capacity's conditions of display (and often the precise nature of display) are left unspecified. The relevant background conditions are determined by pragmatic considerations. Thus when we say that a glass is fragile, but an iron bar is not, we are presupposing "normal" circumstances (like room temperature) and a "normal" triggering cause (like a human hitting the object with a hammer). Thus the fact that if the iron bar is cooled to a low enough temperature it probably breaks when hit, does not falsify the claim. There is nothing mysterious, or suspicious, about the pragmatics of dispositional notions, but not paying attention to them can cause problems under certain circumstances.

Third, it follows from the above that when a system S is said to have a causal capacity (or disposition) D, this claim always involves an implicit reference to a certain kind of environment E. Thus an attribution of a causal capacity takes the environment for granted. This does not, however, imply that the most interesting action is within the system. Even if an electric drill forgotten inside a nuclear reactor has an ability to cause a major disaster, it would be foolish to study this ability by looking at the inside of the drill. Note also, that the individuation of the system is partly a pragmatic

affair (Kauffman 1971, Cummins 1983, Craver 2007a, b), and could be done differently in different occasions.

To get a good grasp of the difference between causal and constitutive explanation, let us first consider a very simple example, the fragility of a glass. We can distinguish three different but related *explananda*:

> (1) How did the glass become fragile?
> (2) What makes the glass fragile?
> (3) Why did the glass break?

Questions (1) and (3) are causal, they deal with processes and events. The first is an inquiry about the processes that led the glass to be disposed to break when a certain kind of triggering cause is present (a high note, a strike by a hammer, etc.). The answer to this causal question will tell us about the causal history of the glass – it will specify the crucial features of the process that led to the object being fragile rather than robust. The latter is a query about the triggering cause, it asks about the event that made the glass to fulfill its potential to break. In contrast, question (2) is constitutive. It asks what gives the glass the causal capacity to break, and the answer will tell us something about the things that the glass is made of (e.g. about molecules and their bonds). The answer to this constitutive question does not focus on earlier events. It will detail the relevant aspects of the object's molecular structure that make it fragile. So while the explanation-seeking questions may look the same – both questions (1) and (2) can be phrased as "Why is the glass fragile?" – the request for explanatory information is quite different. Answering question (2) with an answer appropriate for (1) would be a category mistake. While as the answer to question (1) it would be correct, it would be providing the wrong kind of information as an answer to question (2). In other words, explaining how the object acquired its disposition is different from explaining what makes it fragile. Without a clear understanding of the differences between causation and constitution, confusion like this easily occurs. As I will argue later, the articulation of the contrast space for the intended *explanandum* is a practical way to tease out whether one is seeking causal or constitutive explanations. (Of course one can be, and people often are, interested in both kinds of questions. This does not imply that they are not conceptually distinct.)

Thus it is very important to be precise about the *explanandum* when characterizing the differences between causal and constitutive explanation. For example, much of the recent mechanism literature is potentially confusing, as it takes the *explanandum* of a constitutive explanation to be a behavior (or an activity) of the system (or the mechanism) (Craver 2007a: 128, compare Bechtel & Abrahamsen 2005). The problem with these notions is that they can refer to both causal capacities and events. This kind of vagueness is not conducive to an analysis of explanation, as the nature of the *explanandum* has clear consequences for the kinds of things that the *explanans* should be citing. Explaining why a glass broke is clearly different from explaining why it is fragile, and there is no reason to assume that the need for precision somehow diminishes when one is addressing more complex systems, such as those of cell biology or neuroscience. While I agree with Craver and others that the ultimate aim is to understand a system's behavior in its various aspects, it does not follow that there is no point in being more precise about the individual explanation-seeking questions the scientists are addressing. When building an understanding of a system's behavior, scientists pose more focused questions that abstract away some dimensions of the system's behavior. Constitutive questions are examples of these focused questions.

The constitutive questions abstract away from the behavior and orchestrated activities of the parts, and ask how the system has a capacity for this kind of behavior (Cummins 2000, 122). One could say that the question only addresses a synchronous time-slice of the system. Of course, this is a heavy abstraction, but it helps in articulating an important explanatory question. Constitutive questions are often theoretical: they are not ultimately about the causal capacities of a particular system, but more generally about causal capacities of certain kinds of systems. Of course, this does not imply that explanations of the causal capacities of particular systems cannot be given. Rather, just as in the case of causation, scientists are usually more interested in more general knowledge. So, for example, cognitive neuroscience is not in the business of explaining particular behaviors of individual animals (including humans), but attempts to understand more generally how it is possible for these organisms to have the capacity for a specific kind of information processing (perception, memory, inference, etc.). In other words, the aim of research is to understand how the organisms do it, and the answers provided will discuss the causal capacities of the component systems and their

organization. Ultimately this general understanding should be applicable to the explanation of behaviors of particular systems (of that kind), but applied questions like this are not the central focus of theoretical research.

In contrast to constitutive explanations, causal (or etiological) explanations are in the business of explaining events, that is, changes in the properties of an entity (or system).[2] Thus they can focus on origin, persistence, and changes in a system's properties. In contrast, constitutive explanations are in the business of explaining causal capacities, meaning the properties of entities. Thus the *relata* of causation and constitution are different – one deals with events, the other with properties. To make both the differences and similarities of the explanations clear, we can compare the following two canonical formulations. First we have the causal or etiological explanation:

(EE) e happened in circumstances R, due to events c_1, …, c_n and their (spatial and temporal) organization

Then we have the constitutive explanation:

(CE) System S has causal capacity k in circumstances T, due to S's components s_1, …, s_n and their organization O

When formulated in this way, the differences and similarities between these two kinds of explanations are more apparent. The *relata* are different, but the basic structure of the explanation seems quite similar. These similarities will be explored in section 3, but first I will look at the differences between these *relata*.

2. The metaphysics of constitution

Let us take a closer look at the metaphysics of constitution.[3] In constitution, the parts (and their respective causal capacities) as well as their organization – the basis, for short – constitute the

_____

[2] As any philosophical reader knows, a huge literature exists on the nature of events, but very little consensus. Readers might not accept my account of events, but I hope they are at least ready to recognize the legitimacy of the distinction I am making in this paper.

[3] In metaphysics and philosophy of mind, related issues have been discussed under topics of the causal basis of dispositions (Harré & Madden 1975, Prior 1985) and realization (Wilson & Craver 2007). However, these debates have rarely discussed such issues in the context of explanation or as topics of scientific research.

causal capacities of a system (or whole). Here are some quite generally accepted observations about this relation (see for example Craver 2007, Craver & Bechtel 2007, Bennett 2012).[4]

First, this relation is *asymmetric*: the system's causal capacities do not constitute its parts and their organization. This asymmetry reflects the direction of explanation in constitutive explanation, (I will return to the direction of explanation later in this section.) Second, constitution is an irreflexive relation, nothing constitutes itself. Third, the relation of constitution is *synchronous*: constitution does not take time. Similarly, if changes in the basis give rise to changes in the causal capacities of the system, these changes take place in the same instant. Thus it does not make sense to talk about processes in the case of constitution. While causal processes take time, constitution does not. Underlying this feature is the third important property of constitutive relation: *the relata are not independent existences*. In the case of causation it is possible to require that cause and effect are distinct existences, but the same idea does not work in the case of constitution. To have certain causal capacities is to have a certain constitution (i.e. certain parts organized in a certain way).

The last two characteristics could suggest the idea that constitution is just plain old identity, except that identity is a symmetric and reflexive relation. To clearly understand why constitution cannot be analyzed as identity, we must distinguish between the constitution of all causal capacities of the system and the constitution of an individual capacity. In the former case – the complete set of causal capacities of a particular system (at a particular time) – we could identify the causal capacities and their causal basis (the parts and their organization), and say that to have those causal capacities is to have that particular causal basis. In a way, in the case of a token object, all there is to having a set of causal capacities is to have a certain causal basis. However, things become more complicated when we turn our focus to the constitution of individual causal capacities. Here the identification of the capacity with its basis is blocked by *alternative constitution*: two objects can have the same causal

---

[4] In the social sciences, Alexander Wendt (1998, 1999: 77-88) is an early advocate of the idea of constitutive explanation. The problem with his discussion is that while his notion of constitution contains the idea of constitution of causal capacities, it also contains many other ideas. For example, he fails to see that many how-possibly explanations are causal: they are about necessary conditions for something happening. He also confuses the criteria for category membership with constitutive relations: asking what makes an entity a member of a certain kind is quite different from asking what makes an entity have a certain property. Finally, he associates the contrast between causation and constitution with that between explanation and understanding in a manner that is ultimately unhelpful. For constitution in the social sciences see Ylikoski 2012.

capacity due to different bases. This is what philosophers usually call multiple realization. It implies that we cannot say that to be fragile is to have a specified molecular structure, although we can say that the fragility of a particular glass is due to its having that specific molecular structure. Other objects can have the same causal capacity (in similar circumstances) due to different bases, and even the same object can retain its causal capacity while the basis of this capacity changes. The alternative constitution effectively blocks identity: if causal capacity is identical both to basis A and basis B, then by the transitivity of identity A and B should also be identical, which is absurd.

The distinction between the constitution of the full set of causal capacities and the constitution of a single causal capacity generates a continuum where the size (and the composition) of the causal capacities varies. This continuum might help to make sense of the debates in philosophy of mind and cognitive science about multiple realization (Bechtel & Mundale 1999, Shapiro 2000, Aizawa 2008). While the alternative constitution of a full set of causal capacities is a metaphysical possibility and the abundance of alternative constitution of single causal capacities is an often observed scientific fact, it is the case that as the size of the bundle becomes bigger the room for alternative constitution narrows (given the laws of nature and the constraints of evolution). Thus the debate about multiple realization could be helped if the participants were to articulate where they are located on the continuum by answering the question "multiple realization of what?" It might well be that the whole debate is an artifact of the loose manner in which multiple realization is talked about in philosophy of mind.

Multiple realization makes it natural to apply the notion of supervenience to the relation of constitution. We could say that the causal capacities of the system supervene on the causal capacities and organization of its parts (Beckermann 1992). This is correct, but not very informative with respect to the ontological nature of constitutive relation. Supervenience is merely a relation of property covariance, so it says little about the precise relation between causal capacities and their basis. As supervenience can be used to characterize various sorts of relations – between natural and aesthetic properties and between mental and physical properties, for example – this is not a great surprise (Horgan 1993, Kim 1993, McLaughlin 1995). As we make a distinction between causation

and correlation we should also distinguish between constitution and supervenience. In both cases the crucial element of dependence is missing.

Constitution can easily be seen as a relation of dependence: the causal capacities of the system depend on the causal capacities of its parts and their organization. And while the relation of causal basis can from a metaphysical point of view seem suspiciously close – after all it can be mistaken for identity – it is far from trivial from an empirical point of view. Much of science is focused on questions about constitution. Scientists wish to know what makes certain causal powers possible and how changes in the parts or their organization would change the causal capacities of the system. In fact, these questions are at the heart of causal enquiry.

The above suggests that constitution is a kind of building relation (Bennett 2012): the system is *made of* its parts and their organization. In other words, the system has its causal capacities because of its ingredients, i.e., the causal capacities of parts and their organization. The *made of* relation helps us to account for the asymmetry of constitutive explanation as it involves *the asymmetry of existence*: the parts can exist independently of the system, while the system cannot exist without its parts. Of course the system can exist without a particular part – remember the alternative constitution – but it needs *some* parts. The organization of parts is also crucial for the constitution of a system's causal capacities that are not mere aggregates of those of its parts. Organization has explanatory relevance, as it is a difference-maker with respect to the causal capacities of the system: if you change the organization, the causal capacities change as well (Kuorikoski & Ylikoski 201x).

The notorious notion of (synchronous) emergence is basically a (bad) placeholder for the role of organization in constitution. While the notion of emergence adds a touch of mystery to the issue (McLaughlin 1992, Kim 1999, Robinson 2005, Craver & Bechtel 2007), the role of organization is a matter that is open to empirical enquiry (Kuorikoski & Ylikoski 201x). Organization also has the advantage of dispelling the image of "downward" causation and replacing it with the idea of contextual causation: other parts of a system influencing a particular part. While the parts can be of different sizes and might be described at different levels of abstraction, there is no reason to talk about causation between levels (Ylikoski 2012).

Is the discourse about building too anthropocentric? (After all, most things in the world are not made by someone.) I do not think so. Consider the parallel with causation. While we cannot define causation in terms of manipulation and we do not restrict causation to cases of human agency, we can still think that a characterization of causation with the help of the notion of (possible) manipulation can be informative (Woodward 2003). It identifies an important class of dependency relations that we are interested in. We can leave it open whether or not causation could be identified with something at the most fundamental level; all we need is a set of objective relations of dependence that allow thinking in terms of intervention. This is sufficient to account for the direction of explanation. The same is true for constitution: the anthropocentric notion of building and its analogical extensions help us to identify relations that are objective and of interest to us. It is of no consequence whether ultimately all constitution is (physically) homogenous or not. What matters is that it constitutes an interesting class of dependencies. And these facts are the grounds for our explanatory preferences: we want to explain the system's causal capacities in terms of its basis and not the other way around.

Furthermore, the close proximity of manipulation and building – building involves manipulation – is not harmful. It emphasizes that these two things are similar, but not the same. They are not the same because the relation of constitution cannot be characterized by the notion of manipulation. The *relata* of constitution are not independent existences, so one cannot think of intervention on the basis that would not also be an intervention on the system. However, the experimental inquiry of constitutive relations involves causal manipulation (Craver 2007a: 144-152).[5] Constitutive explanation, furthermore, is about accounting for *causal* capacities, i.e. modal information about causal possibilities. The fact that causal and constitutive dependence are metaphysically different

---

[5] Craver's idea of mutual manipulation has generated much discussion (Fazekas & Kertész, Leuridan 2012, McManus 2012). I tend to agree with Crouch (2011) who argues that Craver primarily gives an account of how scientists find constitutive relations, not of the ontological nature of constitutive relations. In that case there is no conflict between his account and mine: I think Craver's discussion about the evidential requirements of constitutive claims is excellent. However, his discussion (2007a, 2007b) is a bit ambiguous and it is possible that the mutual manipulability is also intended to have metaphysical import. In that case my account which is based on the idea of asymmetry of existence as the basis of explanatory asymmetry, conflicts with his.

does not imply that they cannot be closely linked. Ultimately they are two sides of the same causal coin, not completely different things.

Finally, before I begin to discuss explanation, I want to raise an interesting possible ontological parallel between causation and constitution. This is a conceptual possibility that has not yet been explored by philosophers. Since the end of nineteenth century, philosophers have become accustomed to the idea of *indeterministic causation*. I suggest that something similar would apply to the case of constitution, i.e. *indeterministic constitution*[6]. Thus we would have a situation in which exactly the same components with exactly the same organization would produce different causal capacities of the system in exactly the same environment. However, the causal capacities of the system would still have a statistical distribution that could be found out by empirical enquiry, just as in the case of causation. (This would amount to a probabilistic version of supervenience.) Similarly, many of the conceptual problems – such as the explanation of single cases – created by indeterministic constitution would be similar to its causal cousin. This might seem nonsensical, but so did the idea of indeterministic causation for generations of philosophers. Now, I do not wish to make any specific claims about the possibility of indeterministic constitution here, but I think that a detailed exploration of this conceptual possibility will without doubt help to clarify the ontological nature of constitution and its relation to causation. For example, if we would have a principled argument about why indeterministic constitution is impossible while indeterministic causation is, we would have a much better understanding of their relation.

3. Parallels in explanation

In the previous section I attempted to articulate the ontological difference between causation and constitution. While both are relations of dependence (Rosen 2010), they are metaphysically different. In this section I will show that explanations which attempt to track these dependences face similar challenges. I will begin by outlining how both kinds of explanation face the same fundamental problem, that of explanatory relevance, and how similar ideas about explanation help

---

[6] It should be kept in mind that we are not discussing indeterministic dispositions, i.e. propensies, but indeterministic constitution. Traditional propensities are assumed to be deterministically constituted.

to resolve it. Then I will proceed to show that the idea of mechanism-based explanation applies to both. Finally, I point out a number of parallel complications that both explanations face.

3.1. The problem of explanatory relevance

In the previous section I focused on the relation between the full set of causal capacities of the system and their basis. While this focus is natural for ontological considerations, it is not appropriate for analyzing ordinary explanations provided in science or everyday life. Usually we are not explaining all of the causal capacities of the system at the same time. Rather, we focus on a relatively small cluster of capacities and ask why the system has those capacities (instead of some other set of capacities). The same is true for causal explanation, as we are usually interested in explaining only some aspects of the *explanandum* event. This selectivity on the side of *explanandum* makes explanation challenging: we cannot assume that just because something is part of the system, it will have explanatory relevance with respect to the *explanandum* we have in mind. Again, the same point applies to causal explanation – just because something belongs to the causal history of an event, it does not imply that it will have a crucial explanatory role with respect to the intended *explanandum*. Thus we have *the problem of explanatory relevance*.

It could be claimed that the problem of explanatory relevance is the central challenge for any theory of explanation, especially because so many theories fail to provide a solution to it (Hesslow 1983, Hitchcock 1995, Ylikoski 2001, Craver 2007a, b). A satisfactory theory of explanations should be able to tell what determines explanatory relevance (for a given *explanandum*). The problem arises in causal cases because only some parts of the causal history are relevant and in constitutive cases because only some parts (or aspects of the organization) of the system are relevant (Craver 2007a, b; Couch 2011). The challenge is not only to find the right items, but also to describe them at the right level of detail.

Causation and constitution share the problem of regress as well. In the case of causation we talk about chains of causation. This is based on the idea that for every event that is a cause, there is another event that is its cause. In principle this chain of causation continues until the big bang, or if

no such thing exists, infinitely. A similar idea applies to constitution: we assume that all parts can be further decomposed into their own constitutive parts and their organization. We could call these chains of constitution. The chain of constitution will continue until we reach the level of fundamental entities, or if there are no such things, the chain will continue infinitely (Schaffer 2003).

We do not have to decide now whether there are ultimate causes or entities – we do not even have to accept that the question makes sense – but we should recognize the problem of regress created by these chains. An adequate theory of explanation should be able to make sense of chains of explanations (i.e. explanations having the *explanantia* of other explanations as their *explananda*) but avoid the automatic regress of explanation (i.e. the situation where in order to be an acceptable *explanans*, the thing itself must be explained) where we have not explained anything before we have an adequate explanation that refers to ultimate causes or entities. The problem is that if such things do not exist, we will never be able to explain anything. And it would be somewhat strange if our ability to explain at all would depend on the solution to this thorny and esoteric issue. The problem is solved when we recognize that an explanation presupposes that the *explanans* facts are the case, not that we must have an explanation for those facts. However, such a solution does not solve the problem of relevance; i.e. where in the chain (of causation or constitution) is the right *explanans* located for the given *explanandum*.

The similarities between causation and constitution do not stop at these very general challenges. There are parallel complications that an adequate theory of explanation should be able to deal with. The discussion of these complications is common in debates about causation, and I argue that they pose important challenges for both theories of causal and constitutive explanation. A satisfactory theory of explanation should have resources to deal with them, and it is an advantage if the same ideas can be applied in both cases.

First is *the parallel between multiple realization* and *alternative causes*. In both instances a different thing is responsible for the *explanandum* in different cases. In the case of causal explanation, alternative causes are not treated as a problem – they just make causal discovery more difficult –

but surprisingly, some people still have metaphysical problems with alternative realization. If the general point of this paper is correct, we should have an attitude towards multiple realization that is similar to what we have for alternative causes: whether something has an alternative realization is an empirical issue against which we should not have metaphysical prejudices. While alternative realization or causes are not in themselves very problematic, they are the building blocks for more troublesome situations.

Overdetermination and *preemption* (Bunzl 1979, Lewis 2000) are examples of such troublesome cases.[7] They are usually discussed in the context of counterfactual theories of causation, but instances of them can also be found in the sciences (for example, molecular biology). In cases of overdetermination and preemption, we have an abundance of causes for exactly the same effect: each of the causes is sufficient for the effect but none is individually necessary. The firing squad is a familiar example of overdetermination. In the case of strict overdetermination, the causes are simultaneous and the way they produce the effect is similar. For this reason, true cases of overdetermination are rare, and most candidates for overdetermination turn out to be cases of preemption, as its requirements for similarity of the manner of causation and its timing are less strict. The parallel for instances of overdetermination and preemption in the case of constitution is *systemic redundancy* (Craver 2007a: 156-157). In systemic redundancy the system has more than one constitutive mechanism that guarantees it will have a certain causal capacity. Thus eliminating one of these mechanisms would not prevent the system from having the relevant causal capacity. While systemic redundancy makes empirical inquiry more difficult, it also has a positive side: in cases of complex technological systems, for example nuclear power plants, it is a good idea to have systemic redundancy in safety mechanisms.

The third parallel can be found in the cases of absences, omissions, and other negative facts (Lewis 2000, Beebee 2004, Craver 2007a: 80-86). I will generally call these *negative explanatory factors*.

---

[7] Philosophers have distinguished various forms of preemption, for example early, late, and trumping preemption (Lewis 2000), but these differences are not relevant the the current discussion as it does not make sense to look for constitutive parallels for these temporal differences. More relevant is the idea of quantitative overdetermineion, where there the cause is so overwhelming that even a fraction of it would have been sufficient. A morbid example is eating a whole package of suicide pills when one would suffice. Parallels for this can be found in constitutive relations.

Negative causes figure quite often in both scientific and everyday explanations, but their ontological status is puzzling for some theories of causation. Negative facts are also relevant in the case of constitution (Craver 2007a: 201). It is quite natural to explain differences or changes in causal capacities by referring to an absence of certain components or to their malfunction. Similarly, the lack of organization explains why a mere aggregate of parts does not have the causal capacities of a properly organized system. Furthermore, philosophers have recently discussed the case of double prevention (Collins 2004), and we can find parallels to it in the case of constitutive explanation. In the simple case of prevention, a system has a part (A) that prevents it from having a certain disposition. In other words, if the part A were not there, the system would have the capacity in question. In double prevention, the system has another part (B) that prevents the part A from preventing the system from having the causal capacity. Again, a counterfactual claim would hold: if B were not there, A would prevent the system from having the *explanandum* capacity. But as B is there, the system has that capacity. As in the previous cases, an adequate theory of explanation should be able to deal with these complications.

The fourth parallel concerns background conditions. An important feature of causal explanations is that they are made against the background of a certain *causal field*, or conditions (Anderson 1938, Mackie 1974, Ylikoski 2001: 36-37). Suppose an explosion occurred, and we begin to wonder about its cause. According to Mackie (1974: 35), the original question can be expanded to: "What made the difference between those times, or those cases, within a certain range, in which no such explosion occurred, and this case in which an explosion did occur?" The idea is that both cause and effect are seen as differences within a field, i.e. certain background conditions. Anything that is part of this field will automatically be ruled out as a candidate for the role of cause. In a philosophical sense we can think about the whole cause, the complete causal history of the event, but in typical everyday and scientific contexts we are interested in much more limited causes. The same point about the importance of distinguishing between explanatory factors and background conditions also arises in the context of constitutive explanation (Craver 2007a). Consider the case where you and your friend have similar laptop computers. The computers are of the same model and from the same manufacturer, but your computer is much more sluggish in some heavy computational tasks. Here,

as in Mackie's example, we are looking for variations that could explain the difference in the capacities of the computers. This search ignores all the similarities between the machines: their keyboards, hard drives, etc. These are treated as background conditions, in other words as a *constitutive field*. In contrast, any differences between the two laptops are candidates to be explanatory factors. The fact that one has a sticker on it does not make the difference, but the fact that your friend's computer has double the RAM memory provides an explanation that is supported by both well-known empirical generalizations and theoretical arguments. Clearly, both the challenge and the way we deal with it are quite analogical to the causal case. A typical constitutive explanation focuses only on part of a causal basis of a causal capacity, thus treating other elements of the causal basis as background conditions. Furthermore, constitutive explanations usually treat a system's environment as a given background. Thus we can regard the role of background conditions, the causal and constitutive fields, as being a legitimate parallel between causal and constitutive explanations.

The fifth parallel issue is *the direction of explanation.* In causal explanation it is important that causes explain effects, and not the other way around. This places an important constraint on a theory of explanation: it should be able to track the direction of metaphysical dependence. Not all theories of explanation are capable of meeting this challenge; the unification theory (Kitcher 1989) for example, has trouble with this particular issue. Similarly, certain counterfactual theories have difficulties in ruling out so-called backtracking counterfactuals in a principled manner. Theories of constitutive explanation face a similar challenge, as they should give a principled answer to the question of why the causal capacities of components and their organization explain the causal capacities of the system, but not the other way around.

3.2. Tracking mechanically mediated dependencies

In the previous section I argued that many parallels can be seen in the challenges that theories of causal and constitutive explanations face. This raises the possibility that the same ideas about explanation would apply in both contexts. The only crucial difference would be the kind of dependency they track. The applicability of the same explanatory ideas would then help us to understand why the difference between causal and constitutive explanation has remained hidden for

so long. Of course it is impossible to outline a full theory of explanation in one paper and to demonstrate how it applies to both cases. Thus I will only discuss three quite common ideas about explanation that can be fruitfully applied to both cases. This would not demonstrate that the same theory of explanation could take care of both cases, but it might at least establish that it makes sense to look for parallel solutions for both.

The first idea I have in mind is that of contrastive *explanandum*. This idea already implicitly came up above in the context of Mackie's idea of causal field. Explanations, while tracking relations of dependence that exist independently of us, can only relate to things described in a certain way. In other words, we always explain specific aspects of events or phenomena, not these events or phenomena themselves or as a whole. The idea that explanations are answers to contrastive questions (Garfinkel 1981, Hesslow 1983, Lipton 1991, Ylikoski 2007, Craver 2007) captures this idea naturally. We are interested in why things are one way rather than some other way; in other words, why the fact is the case rather than a foil, where the foil is an exclusive alternative to the fact. The contrastive idea also works in a comparative setting, where we are asking why an object has a certain property rather than being like an otherwise similar object that has a different property. Still often the contrast is imagined: we ask why an object has a particular property rather than a different property that we expected it to have or our theory predicted it would have (Hesslow 1983).

The idea of contrastive *explanandum* helps to make the explanation more explicit in an analytically fruitful manner (Ylikoski 2007). Spelling out the contrastive structure forces one to articulate what the object of the explanation is and in which respects we think the object could have been different. This kind of articulation is important if we want to know what we are explaining. Recall the comparison between laptops discussed above. The comparison with your friend's computer serves as a contrast that helps to fix what the precise *explanandum* is: why your computer is sluggish and not quick like your friend's. You are not interested in explaining all of the properties of your computer, you are interested in this specific question only and this interest guides your search for the explanatory factors. You want to find the things that constitute the specific differences between the two computers.

The contrastive articulation can also be used to determine whether we are interested in causal or constitutive information. While at the surface level (Why is this glass fragile?), causal and constitutive explanation-seeking questions might seem very similar or even identical, the articulation of the contrast class forces us to be clear about it. If the contrast space consists of alternative causal capacities, rather than events (like acquiring causal capacities), we know that the relevant explanatory facts are about constitution. Events and properties are not easily conceived as alternative values of the same variable.

Thus the contrastive thesis is not a claim about what people have in mind when they put forward an explanation-seeking question, but what they should have in mind (Ylikoski 2007). Quite often the original scientific research question, when articulated in contrastive terms, turns out to be a whole set of related contrastive questions. This is a good thing: smaller questions are something we can actually hope to answer by means of scientific enquiry. And of course, nothing prevents one from asking both causal and constitutive questions – and questions driving scientific research are often such – but the contribution of the contrastive idea is to make it possible to analytically distinguish questions that require separate answers.

The idea of the *explanans* as the difference-maker (Mackie 1973, Woodward 2003, Waters 2007, Strevens 2008) is a natural partner of the idea of contrastive *explanandum*. Together they provide a powerful heuristic of scientific research. First, you create, find, or imagine the difference to be explained, and then you proceed to find the differences between the cases. Then you test whether these candidates can really make the difference, by testing whether they can bring about the difference to be explained. If the procedure does not work, you can try to be more precise about the *explanandum*, and try again. While things like overdetermination can block this heuristic from working under a coarse-grained description of the contrastive explanation, they rarely are so tricky that they would not be detected by a more fine-grained description of the contrast.

The notion of difference-making also naturally fits the realist idea that explanations track objective relations of dependence. For example, in the case of causal knowledge, our goal is to know how changes in the causes bring about changes in the effects. Note that this dependence has a modal

aspect. The explanation is not about subsumption under empirical regularities, but about counterfactual dependence: X explains why Y if Y depends on X in the sense that if X had not happened, then Y would not have happened either (Woodward 1984; Woodward 2003; Ylikoski & Kuorikoski 2010).

In the difference-making approach the basic building blocks of explanations are claims about counterfactual dependence. If the *explanans* had been different, the *explanandum* would have been different as well. The difference between causal and constitutive explanation lies in the kinds of counterfactuals they appeal to. Causal explanations appeal to etiological counterfactuals, while constitutive counterfactuals are the material for constitutive explanations. Both counterfactuals are assumed to follow the direction of dependence, but their *relata* are different due to the kind of dependence they describe. Causal counterfactuals refer to antecedent events and/or their organization, while constitutive counterfactuals deal with properties of parts and/or their organization.

A natural consequence of combining the contrastive and difference-making is that explanation becomes doubly contrastive, as both *explanandum* and *explanans* are contrastive (Woodward 2003, Schaffer 2005). If we represent the contrastive claim "**a** rather than **a***" as **a** [**a***], the explanation has the following basic structure:

(BE) **a** [**a***] explains **b** [**b***] in the context U

In causal explanation **b** and **b*** will be about origin, persistence or changes in a system's properties, and **a** and **a*** will be about antecedent events (including their location and timing). The context U, or *the causal field*, will contain facts about both the system's environment and the system's other properties. Fixing the causal field has traditionally been a major challenge for counterfactual theories of causal explanation. There has been much fuss about pragmatic factors, without much of an idea about the principles by which this contextual relevance is fixed. However, when the idea of difference-making is combined with that of contrastive *explanandum*, this problem is solved rather naturally. The contrast helps to fix the causal background in a very definite way: all the elements of the causal history that **b** and **b*** share are assumed to belong to the causal field, and the concern is

with their differences. One could say that what is shared between the causal histories of **b** and **b***
belongs to the conditions, and the differences can be identified as the potential causes.

In constitutive explanation, **b** and **b*** will be about the system's causal capacities, while **a** and **a***
will be about the causal capacities of the system's parts or about their organization. The context U
will again contain both facts about the system's environment and its other properties. The same
contrastive idea can be applied in the case of constitutive explanation: all the parts that systems
having **b** or **b*** share are assumed to belong to *the constitutive field*, and the concern is with their
differences.

The third idea is that in both cases we can distinguish a simple explanatory claim identified by (BE)
and more ambitious explanatory claims that will tell us why the counterfactual dependence holds
and how it depends on background conditions that (BE) just takes for granted (the context U). We
could call the more extensive explanatory answer *mechanical explanation*. In the case of
mechanical explanation we have more information about the relevant relations of dependence – we
could say that mechanical explanation tracks networks of counterfactual dependence[8]. We rarely –
maybe never – have the full mechanical explanation, but the notion still refers to greater
explanatory understanding. If we know why the core explanatory dependence (between **a** [a*] and **b**
[b*]) holds and how the changes in the assumed background conditions modulate it we have a
significant increase in our ability to make a what-if inferences – that is, in our understanding
(Ylikoski & Kuorikoski 2010).

I hope the above sketch makes it plausible that the same basic ideas about explanation apply to both
cases. As the challenges are similar, one can expect that similar solutions will apply. Thus if the
combined ideas of difference-making and contrastive *explanandum* make it possible to solve the
problems of explanatory relevance and background conditions in the case of causal explanation, one
can use the same idea in the case of constitutive explanation. Similarly, if a closer look at the
relevant network of counterfactual dependencies helps to sort out cases like alternative causes,

---

[8] This characterization of mechanism is intended to be compatible with many other more detailed definitions
of mechanism found in the literature (Woodward 2002, Glennan 2002, Craver 2007a, Ylikoski 2011,
Hedström & Ylikoski 2010).

negative causes, and overdetermining causes, one can expect that a similar strategy would be helpful also in the case of constitution. Thus it can be expected that cases like multiple realization, missing parts, and systemic redundancy can be made sense of by carefully articulating the relevant network of counterfactual dependencies and by giving up the idea that the explanatory relationship can be simplified to a simple claim about counterfactual dependency. [9] Finally, if in the case of causation the asymmetry of manipulation provides a natural rationale for the direction of explanation – the connection between agency and understanding being the reason why we prefer explanatory information going in the direction of causal influence – one should not be surprised if similar considerations would bind the direction of constitutive explanation to the asymmetry of made-of relations.

4. The mixed case: development

I have now argued that despite their similarities, causation and constitution are different kinds of explanations as they track different sorts of dependence. Let us call this *the distinctness thesis*. It is important to recognize that I am not claiming all explanations to be exclusively either causal or constitutive explanations. Let us call this *the separateness thesis*. The separateness thesis does not follow from the distinctness thesis. It is very important to recognize that most scientific inquiries aim to track both kinds of dependencies, although it makes sense to see that individual explanation-seeking questions (when full articulated) often address either causal or constitutive questions.

However, there are also explanation-seeking questions that combine causal and constitutive elements. An example of such explanations are the cases where, the *explanandum* and *explanans* are causal capacities, but the dependence is clearly causal. These are generally called developmental explanations. The existence of developmental explanations provides an additional argument against the separateness thesis. In the remainder of this section I will analyze a particularly interesting example of developmental explanation and use it to establish that developmental explanations are

---

[9] It is notable that alternative causes, negative causes, and overdetermining causes provide counterexamples to simplistic definitions of cause, but their causal structure is not difficult to understand. (This is why they can serve as counterexamples.) Their structure can be articulated by carefully outlining the network of dependencies. I believe that the same idea applies to constitutive relations. The idea of simple constitutive counterfactual dependence is a building block for understanding more complex relations of dependence, not a straightjacket into which we should force all explanatory claims.

an interesting hybrid kind of explanation which deserves a separate treatment although, strictly speaking, developmental dependence is metaphysically a derivate notion.[10]

While a constitutive explanation relates the system's capacities synchronously, a developmental explanation involves both time and significant changes in the system's causal capacities. The aim of such explanations is to figure out the pathway by which the system's initial causal capacities are realized and transformed into new causal capacities that characterize the system's later phases. For example, explanations in developmental biology track the process by which a fertilized egg would develop to an adult organism. Another example is the developmental account of language acquisition. It begins with the perceptual and cognitive capacities of an infant and describes the pathway by which these capacities develop into a full ability to produce and understand expression in some natural language. As developmental explanations involve processes, they are subspecies of causal explanations. However, one important feature distinguishes them from ordinary causal explanations. In contrast to usual causal explanations, which take the system – or more broadly, the causal configuration – for granted as well as how certain changes in conditions, or triggering causes, modulate the effects, developmental explanations involve substantial changes in the causal setting (McManus 2012). In other words, the developmental process involves changes in the parts of the system and their relations. For this reason, developmental explanations should be conceived as tracking chains of causations that involve reorganization of the causal setting.

Typical developmental explanations explain a system's capacities by referring to its capacities at earlier phases of development or by significant earlier developmental events (like external shocks or deprivation of crucial developmental resources). However, some developmental explanations that involve a more significant constitutive element. The *explanantia* of such explanations are the system's parts or their organization in the initial phases of development. These explanations could be called hybrid explanations, as both causal and constitutive dependencies play an important role

---

[10] Space limitations do not allow me to explore other possible hybrids. For example, a referee suggested that evolutionary explanations could provide one. I am skeptical about this claim in the cases of straightforward evolutionary explanations that track changes in the composition of the population. I would analyze them as straightforward causal explanations. Still, some explanations in evolutionary developmental biology (*evo-devo)* could turn out to be very interesting hybrids.

in them. A good illustration of developmental dependencies of this sort is provided by gene knockout experiments.

In gene knockout experiments (for recent reviews of this experimental technique, see Barbaric *et. al.* 2007, Guan *et. al.* 2010) one of an organism's genes is made inoperative ("knocked out"). This technique, for which Mario R. Capecchi, Martin J. Evans and Oliver Smithies were awarded the Nobel Prize in medicine in 2007, can be used to learn about functions of a gene. Research on knockout mice has greatly extended scientific understanding of normal developmental and disease processes, and has also helped in the identification of new avenues for medical therapy. I here skip over the intricate details of the experimental intervention involved in knockout experiments (such as the differences between gene targeting and gene trapping) and discuss these experiments only on an abstract – "metaphysical" – level. Thus I will ignore, not only all technical issues, but also all difficulties related to causal inferences like this. Developmental processes are complicated, and most gene knockout experiments do not provide clear causal information. However, here our concern is with a kind of dependence that an ideal experiment like this could identify. I am therefore after the simplest sort of explanatory dependence that the developmental explanations are trying to track.

In a typical gene knockout experiment, the target of the experimental intervention (making the specific gene non-functional) is a part of the system (an embryonic stem cell). While the intervention affects the causal capacities of the embryonic stem cell, the scientists do not focus on these properties. They are not looking for a constitutive explanation of the fertilized egg's causal capacities. Rather, they are interested in the phenotype – i.e. the appearance, behavior, metabolism etc – of an adult organism. In other words, they are interested in the causal capacities of an adult organism that are developmental consequences of the changes in its genetic material. The explanatory setting of this type of experiment is explicitly contrastive: the question is what kinds of differences in an adult phenotype would a change in this particular gene explain. An explanatory claim one can derive from these experiments is the simple explanatory claim (BE) discussed above. While the experiment shows the difference the gene makes, it leaves the details of the developmental process open, e.g. it does not specify the causal mechanism. However, the

knowledge gained from gene knockout experiments often provides crucial clues for understanding in greater detail the developmental process, that is, the networks of counterfactual dependence that characterize the developmental pathway.

Note also the artificial setting of gene knockout experiments: the developmental environment is kept the same in both conditions. This is unrealistic if we are considering the development of behavioral capacities. The gene might influence the mouse's behavior in such a way that the mouse would seek a different environment than its normal counterpart. If these environmental features then influence the mouse's development, we have a case of so-called active gene-environment correlation (active rGE) that would make it impossible to infer about the gene's full influence (or lack of it) in natural environments. This is an important feature, as it sets developmental constitutive explanation apart from synchronous constitutive explanation. In the latter, the choice of boundary between the system and its environment is largely a pragmatic issue. One identifies the system, and then one figures out how parts and their organization give it its capacities *within* that environment. While the environmental features that make it possible for that system to have a certain capacity might be more interesting than what is happening within the system, it is fully legitimate to focus on internal enabling features. In the case of developmental constitutive explanation, this is not a legitimate move – one has to take into account what happens when the system interacts with its environment. The assumption of a stable environment would be a signal of reductionist fallacy that ignores crucial interactions.

My central claim here is that the developmental explanations I have described above combine in an interesting way both causal and constitutive elements. The explanation is about a causal process, but both the *explanans* and *explanandum* are causal capacities. Furthermore, the *explanans* refers to parts (or their organization) of the developmental process, while the *explanandum* refers to the causal capacities of the system that is an outcome of this process. These explanations deserve their own place in our taxonomy of scientific explanations. However, while their structure requires special attention, they are not an ontologically distinct form of explanation. By this I refer to the fact that developmental dependence can be analyzed in terms of causal and constitutive processes. The existence of these explanations challenges the separateness thesis, but does not provide grounds

24

for an extended distinctness thesis that talks about three fundamentally distinct relations of dependence.

5. Two sides of the same coin

In this article I compared causal and constitutive explanation. I attempted to make two claims plausible. The first is that it makes sense to clearly distinguish between these two kinds of explanation. While scientific inquiry usually addresses both causal and constitutive questions, and some explanations combine both kinds of relations – like developmental explanations discussed above – making the distinction is crucial for a detailed understanding of scientific questions and their interrelations. These explanations have different kinds of *explananda* and track different sorts of dependencies. The second claim is that similar key ideas about explanation apply to both kinds of explanation. This helps to maintain the hope that a general theory of explanation could be possible and, if true, makes developing a detailed account of constitutive explanation much easier. While it took several decades for philosophers of science to reach the current understanding of causal explanation, one can hope that developing an understanding of constitutive explanation will not take as long. The second claim also makes it easier to understand why so many philosophers have missed the difference between causal and constitutive explanation. Only by looking at the details of the *relata* of explanatory relation it is possible to recognize the crucial ontological differences.

Naturally, there is room for further work on constitutive explanation. For example, we need a better understanding of constitutive counterfactuals, a more detailed ontological framework for discussing relations between causation and constitution, and more case studies of constitutive and developmental explanations in different sciences. However, I hope this article has shown that further work in these directions makes sense and is necessary for the development of a fully satisfactory account of scientific explanation.

References

Aizawa, K. 2008: Neuroscience and multiple realization: a reply to Bechtel and Mundale. *Synthese* 167: 493–510.

Anderson, John 1938: The Problem of Causality, Australasian *Journal of Psychology and Philosophy* 16: 127-142.

Barbaric, I., Miller, G., & Dear, T. N. 2007: Appearances can be deceiving: phenotypes of knockout mice. *Briefings in Functional Genomics and Proteomics* 6: 91–103.

Bechtel, William. 2008. *Mental Mechanism. Philosophical Perspectives on Cognitive Neuroscience*. London: Routledge.

Bechtel, William and Richardson, Robert C. 1993: *Discovering Complexity. Decomposition and localization as strategies in scientific research*. Princeton University Press. Princeton.

Bechtel, W., & Mundale, J. 1999. Multiple realizability revisited: Linking cognitive and neural states. *Philosophy of Science* 175–207.

Bechtel, W., & Abrahamsen, A. 2005: Explanation: a mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 421–441.

Beckermann, Ansgar 1992: Supervenience, Emergence, and Reduction. In Beckermann, Flohr & Kim (eds.): *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*. Walter de Gruyter, 94-118.

Beebee, Helen 2004: Causing and Nothingness, in John Collins, Ned Hall, and L. A. Paul (eds.): *Causation and Counterfactuals*, The MIT Press, 291-308.

Bennett, K. 2011: Construction area (no hard hat required). *Philosophical Studies* 154: 79–104.

Bunzl, Martin 1979: 'Causal Overdetermination' *The Journal of Philosophy* 76: 134-150.

Collins, John 2004: Preemptive Prevention, in John Collins, Ned Hall, and L. A. Paul (eds.): *Causation and Counterfactuals*, The MIT Press, 107-118.

Couch, M. B. 2011: Mechanisms and constitutive relevance. *Synthese* 183: 375–388.

Craver, Carl. 2007a: *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. New York and Oxford: Clarendon Press.

Craver, Carl. 2007b: Constitutive explanatory relevance, *Journal of Philosophical Research* 32: 3-20.

Craver, C. F., & Bechtel, W. 2007: Top-down Causation Without Top-down Causes. *Biology & Philosophy* 22: 547–563.

Cummins, Robert 1975: Functional Analysis, *The Journal of Philosophy* 72: 741–765.

Cummins, Robert 1983: *The Nature of Psychological Explanation*. Cambridge, MA: Bradford/MIT Press.

Cummins, Robert 2000: "How does it work?" versus "what are the laws?": Two conceptions of psychological explanation, in F. Keil and R. Wilson (eds) *Explanation and Cognition*. Cambridge: MIT Press. pp. 117–144.

Fazekas, Peter & Gergely Kertész 2011: Causation at different levels: tracking the commitments of mechanistic explanations, *Biology and Philosophy* 26: 365-383.

Garfinkel, Alan 1981: *Forms of Explanation*. Yale University Press. New Haven.

Glennan, S. 2002: Rethinking mechanistic explanation. *Philosophy of Science* 69: S342–S353.

Guan, C., Ye, C., Yang, X., & Gao, J. 2010: A review of current large−scale mouse knockout efforts. *Genesis* 48: 73–85.

Harré, Rom & Madden E. H. 1975: *Causal Powers. A Theory of Natural Necessity*. Basil Blackwell. Oxford.

Hedström, Peter & Petri Ylikoski. 2010: Causal Mechanisms in the Social Sciences. *Annual Review of Sociology* 36: 49-67.

Hesslow, Germund 1983: 'Explaining differences and weighting causes', *Theoria* 49: 87-111.

Hitchcock, Christopher 1995: Discussion: Salmon on explanatory relevance, *Philosophy of Science* 62: 304–20.

Horgan, Terence 1993: 'From Supervenience to Superdupervenience: Meeting the Demands of a Material World'. *Mind* 102: 555-586.

Lipton, Peter 1991: *Inference to the Best Explanation*. Routledge. London.

Kauffman, Stuart 1971: Articulation of Parts Explanation in Biology and the Rational Search for Them, *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 1970: 257-272.

Kim, Jaegwon 1993: *Supervenience and Mind*. Cambridge University Press. Cambridge.

Kim, Jaegwon 1999: Making sense of emergence, *Philosophical Studies* 95: 3–36.

Kitcher, Philip 1989: 'Explanatory Unification and the Causal Structure of the World', in Kitcher & Salmon (eds.): *Scientific Explanation. Minnesota Studies in the Philosophy of Science vol XIII*. University of Minnesota Press. Minneapolis: 410-505.

Kuorikoski, Jaakko & Ylikoski, Petri 201x: How Organization Explains, in Dennis Dieks and Vassili Karakostas (eds.) *EPSA 2011 Proceedings*, Springer, forthcoming.

Leuridan, B. 2012: Three Problems for the mutual manipulability account of constitutive relevance in mechanisms. *British Journal for the Philosophy of Science* 63: 399–427.

Lewis, D. 2000: Causation as influence, *Journal of Philosophy* 97: 182–197.

Mackie, J.L. 1974: *The Cement of the Universe*. Oxford: Oxford University Press.

McLaughlin, Brian 1992: The Rise and Fall of British Emergentism, In Beckermann, Flohr & Kim (eds.): *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*. Walter de Gruyter. Berlin: 49-93.

McLaughlin, Brian 1995: Varieties of Supervenience, in Savellos & Yalçin (eds.): *Supervenience. New Essays*. Cambridge University Press. Cambridge: 16-59.

McManus, Fabrizzio 2012: Development and Mechanistic Explanation, *Studies in the History and Philosophy of Biology and Biomedical Sciences* 43: 532-541.

Prior, Elizabeth 1985: *Dispositions*. Aberdeen University Press.

Robinson, William S. 2005: Zooming in on Downward Causation, *Biology and Philosophy* 20:117–136.

Rosen, Gideon 2010: Metaphysical Dependence: Grounding and Reduction, in Hale & Hoffmann (eds.): *Modality. Metaphysics, Logic, and Epistemology*, Oxford University Press, 109-136.

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

Schaffer, J. 2003: Is there a fundamental level? *Noüs* 37: 498–517.

Schaffer, J. 2005: Contrastive causation. *Philosophical Review* 114: 327–358.

Shapiro, L. A. 2000: Multiple realizations. *Journal of Philosophy* 97: 635–654.

Simon, Herbert 1996: *The Sciences of the Artificial. 3rd ed*. MIT Press.

Strevens, Michael 2008: *Depth. An Account of Scientific Explanation*. Harvard University Press.

Waters, C. K. 2007: Causes that make a difference. *Journal of Philosophy* 104: 551–79.

Wendt, Alexander 1998: On constitution and causation in International Relations. *Review of International Studies* 24: 101-117.

Wendt, Alexander 1999: *Social Theory of International Politics*. Cambridge University Press.

Wimsatt, William 2007: *Re-Engineering Philosophy for Limited Beings. Piecewise Approximations to Reality*. Cambridge, MA: Harvard University Press.

Wilson, Robert & Craver, Carl 2007: Realization. In Paul Thagard (ed.): *Handbook of the philosophy of science, Vol. 12, Philosophy of psychology and cognitive science*. Elsevier. The Netherlands: 81–104.

Woodward, James 1984: A theory of singular causal explanation. *Erkenntnis* 21: 231–262.

Woodward, James 2002: What is a mechanism? A counterfactual account. *Philosophy of Science* 69: S366–S378.

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

Ylikoski, Petri. 2001. *Understanding Interests and Causal Explanation*. PhD thesis. University of Helsinki, May 2001. http://ethesis.helsinki.fi/julkaisut/val/kayta/vk/ylikoski

Ylikoski, Petri 2007: The Idea of Contrastive *Explanandum*. In Johannes Persson and Petri Ylikoski (eds.): *Rethinking Explanation*, Dordrecht: Springer: 27-42.

Ylikoski, Petri 2011: Social Mechanisms and Explanatory Relevance. In Pierre Demeulenaere (ed.) *From Social Mechanisms to Analytical Sociology*, Cambridge: Cambridge University Press, 154-172.

Ylikoski, Petri 2012: Micro, Macro, and Mechanisms, in *The Oxford Handbook of Philosophy of the Social Sciences* (edited by H. Kincaid), Oxford University Press, 21-45.

Ylikoski, Petri & Jaakko Kuorikoski 2010: Dissecting Explanatory Power. *Philosophical Studies* 148: 201–219.