ISSN 1748-7595 (Online)

Kent Business School

Working Paper Series

Explanatory Mechanisms: The Contribution of Critical Realism and Systems Thinking/Cybernetics

John Mingers
Kent Business School

Working Paper No. 241

February 2011

Explanatory mechanisms: The contribution of critical realism and systems thinking/cybernetics

John Mingers

Kent Business School

j.mingers@kent.ac.uk

Abstract

In recent years the philosophy of science has been moving from the traditional deductive-nomological, covering law model of explanation towards one centred on the key concept of explanatory mechanisms. The purpose of this paper is to contribute to our understanding of mechanistic explanation by bringing in theoretical ideas from two traditions which have had the idea of mechanisms at their core for many years. These are the philosophy of science known as critical realism and the discipline of systems thinking/cybernetics. After briefly outlining their respective literatures, this paper will cover issues such as: ontic versus epistemological explanations; generative causality; non-physical mechanisms including social and cognitive structures; functionalist explanation; localisation; and absences and omissions as causes.

Key Words: causality**,** critical realism, cybernetics, mechanisms, mechanistic explanation, systems thinking

Introduction

For many decades, the concept of explanation within the philosophy of science was presumed to revolve around the idea of universal laws. Events were to be "explained" in terms of being instances deduced from general covering laws, which were themselves developed through some form of induction from observable empirical examples. This was formalised in what became known as the deductive-nomological (D-N) model developed by Hempel (1965). However, in more recent years the limitations of this approach have become ever more obvious and an alternative has been generating much interest and support. This approach eschews universal laws in favour of particular "mechanisms" that causally generate the phenomena of interest to the scientist (Gerring 2007, Glennan 1996, Glennan 2002, Machamer 2004, Machamer, et al. 2000, Salmon 1998b, Symons 2008). Apart from avoiding many of the problems besetting the D-N model, especially concerning induction, the idea of mechanisms fits much better with the actual practices of scientists (Bechtel and Abrahamsen 2005) and, as we shall see, with explanations in everyday life.

As might be expected with a major new development, there are many issues to be debated, and indeed arguments about how the term mechanism should be conceptualised. The purpose of the paper is to contribute to the understanding of mechanistic explanations by bringing in

theoretical ideas from two domains that are highly relevant to this approach, and yet have so far been little discussed. These domains are those of systems thinking and cybernetics, and the philosophy of critical realism. The former is an obvious choice. Many of the papers on mechanistic explanation explicitly couch their models in terms of complex systems in which a variety of component parts interact with each other to form a mechanism that then has particular behavioural or emergent properties. However, they tend to make no reference to the huge literature on systems thinking and cybernetics that developed, originally in biology, ecology, and information science, from the 1920's onwards. This provides a strong body of conceptual and theoretical work on which the mechanistic viewpoint can be built.

Less well known, perhaps, is the philosophical approach called critical realism (CR) that has been developing as an alternative to both positivism and interpretivism (in social science) since the 1970's (Archer, et al. 1998, Bhaskar 1978, Bhaskar 1979, Bhaskar 1993). This is a comprehensive and sophisticated post-positivist paradigm that has at its heart the idea of generative causality via causal structures or mechanisms which possess powers or tendencies to behave in particular ways. The actual and empirical events that occur in the world are then seen as resulting from the interactions and interplay of these structures and mechanisms. We will show that many of the issues and problems of the mechanistic view of causation have already been encountered, and to some extent addressed, within critical realism. We should also note that CR itself uses many systemic and holistic concepts (Mingers 2011).

The approach of this paper is firstly to outline briefly the literature of critical realism and systems thinking. It then goes on to consider a range of issues and debates about mechanistic explanation, giving an indication, in each case, of the contribution that systems thinking and critical realism can make to our understanding of the issue.

Deduction, induction and abduction: the logic of mechanistic explanation in critical realism

Our starting point in explaining critical realism's view of mechanisms is actually the work of the pragmatist philosopher C. S. Peirce (1931-1958). One of Peirce's many contributions was the development of the logic of "abduction" or "retroduction" as opposed to deduction or induction (Psillos 2009).

Consider the following syllogism (an example from Peirce 2.623[1])

General law:	All beans in this bag are white
Particular case:	All these beans come from this bag
Conclusion:	All these beans are white

This syllogism, which can be classified as AAA-1 (Barbara) is valid and is an example of deductive inference. Given a general law or rule we can deduce a particular consequence from it.

This can be re-arranged as follows:

[1] References to Peirce's Collected Works are in the form (vol.para)

Context:	All these beans come from this bag
Empirical observations:	All these beans are white
General law:	All beans in this bag are white

This syllogism can be said to capture the logic of induction – from particular instances we induce a general conclusion. In terms of pure logic it is invalid since there could still be beans in the bag that are not white but were not selected, but it obviously has utility as a practical mode of inference.

It can also be re-arranged as follows:

(Unexpected) observation:	All these beans are white
Possible cause:	All beans in this bag are white
Explanatory hypothesis:	All these beans come from this bag

This syllogism is also not valid in a logical sense – some other reason could explain why all the beans are white – but it has quite a different character from the previous two. Peirce called it "abduction" or "retroduction". In his 5th Lecture on Pragmatism, Peirce said, "Abduction consists in studying facts and devising a theory to explain them" (5.145) and in the 6th Lecture, he said "abduction is the process of forming an explanatory hypothesis" (5.171). So, with the process of abduction we begin with some particular occurrence or event, usually one that is unexpected or does not conform to current theories; and we then take an imaginative leap to think of some theory or explanation which might account for the event. This is neither an induction from the examples nor a deduction from the rule, but rather an explanatory or exploratory hypothesis as to why the situation might have occurred.

Abduction is the point where novelty, innovation and creativity enter the scientific method, as indeed they must. With deduction, we get nothing more than the consequences of the premises – but where did they come from? With induction, we just get a generalisation form the observations we have made – but how do we know they are all that matters? However, with abduction we get explanation and the possibility of new knowledge.

Peirce (2.781) recognised that actually all three modes were necessary for successful science. We begin with an unusual or puzzling phenomenon (C) and try to hypothesise something (A) which would account for the existence of the phenomenon, this is abduction. Then, second, we explore the consequences of A. If A is in fact the case, what other consequences would follow that we might be able to observe or test? This is deduction. Finally, we try and observe whether these consequences do in fact happen, which would thus confirm our explanation. This is a form of induction.

This mode of reasoning is also at the heart of critical realism (CR), which adopts an approach to causality that is known as "*generative causality*" (Mingers 2000). In distinction to positivism or empiricism, which adopts the impoverished Humean version of causality as nothing more than a constant conjunction of events, CR holds that there is a stratified external reality in which the occurrences and events we experience (the "*actual*") are the result of, or

generated by, the interaction of underlying structures and mechanisms (the "*real*"). These mechanisms may be physical, social or conceptual, and they may be observable or unobservable. CR's methodology (DREI), which is involves retroduction, is very much along the lines of Peirce's abduction (Hartwig, 2007, p. 195):

> (D) An unusual occurrence or anomaly is observed and Described
> (R) Retroduction is applied – putative causal mechanism(s) are hypothesised which, *if they existed*, would account for the occurrence or anomaly
> (E) Hypotheses are Eliminated where possible
> (I) The correct mechanism is Identified

The main difference with Peirce is that the account is couched in terms of mechanisms and structures that have causal powers or tendencies to bring about changes in the world. Bhaskar does not explicitly define what he means by mechanism, and sometimes he refers to them as structures, but they can be characterised by the following:

- Mechanisms exist in a real, ontological sense independently of how they may be known or described by observers. They are stratified, in the sense of depth or hierarchy, and they may be physical, social, or conceptual. They may be observable or unobservable. Their existence is judged by a causal rather than a perceptual criteria – i.e., that they have causal effects in the world.
- Mechanisms are relatively enduring in respect of the events that they cause but their absolute timescale may vary immensely. They have powers or tendencies, by virtue of their structural properties, to behave in particular ways or have certain effects. These powers may not be exercised all the time (perhaps needing to be triggered), or they may be exercised but have no effect because of the countervailing actions of some other mechanism. Through their interactions, mechanisms generate the actual occurrences and events of the world, only some of which are observed or noted empirically (Bhaskar 1979, p. 170). Thus a mechanism may be said to consist of a structure of inter-related parts together with the powers or tendencies that the structure possesses.
- Social structures or mechanisms have different properties or characteristics to physical ones (Bhaskar 1979). First, they only become manifest at all through the activities that they govern. That is, social structures cannot be directly observed, they exist only virtually as a set of practices or roles which govern or enable social activities – think of language as an example. Through these activities the structures become reproduced or indeed changed and transformed. Second, they rely to some degree on the knowledge and understanding of social actors who must be aware that they are doing a particular activity, and how to do it. Third, they are localised in time and space in the sense that they belong to particular cultures at particular times rather than being universal, apart perhaps from extremely general ones such as the human ability to use tools or language. Finally, social systems are inevitably open (rather than being able to be closed as in a laboratory experiment) and hence, in principle unpredictable.

We should also point out, and it will be discussed more later, that critical realism stresses the importance of the negative/absent as well as the positive/present as causally efficacious (Bhaskar 1993).

Parts, wholes and boundaries: the basis of systemic thinking

Systems thinking can be traced back to the Greeks, but in modern times it developed in the early twentieth century in fields such as biology, ecology and the then new discipline of cybernetics (the study of information and control in natural and man-made systems)[2]. The basic systemic insight is the anti-reductionist one that parts interact together to form wholes which have properties or powers that are emergent in that they cannot be reduced or explained purely in terms of the properties of the elements (Winther 2009). They only occur at the level of the system as a whole. Early biological organicists actually used the term 'system' and it was perhaps best articulated in Woodger's (1929) *Biological Principles*. Similar ideas developed in the Gestalt school of psychology (Wertheimer and King 2005), and ecology (Haeckel 1866, von Uexkull 1909), and even atomic physics, a bastion of reductionism, began to recognise wholeness at the very fundamental levels of subatomic particles which were not so much discrete particles but webs of interacting forces (Heisenberg 1963).

The central systemic idea – that the characteristics and behaviour of entities depended on the structure of relationships between components rather than the properties of the components themselves – carries with it several other concepts – *emergence*, *hierarchy* and *boundaries*. With emergence comes hierarchy. If we consider a system at a particular level it consists of components and relations. However, each component can itself be treated as a system and 'opened up' to reveal another set of components and relations. This process can in principle go on for an indefinite number of levels until we reach the bedrock of indissoluble forces. We can also go in the other direction from the initial system and see that it is only a component of a further hierarchy of wider systems.

The third concept is that of boundary. If emergent properties are attributed to a particular entity in virtue of its components and relations we must be able to demarcate the system that has the properties from its environment. This may seem relatively clear when we are dealing with physically discrete objects that have a single clear boundary, but becomes much more contentious when dealing with complex systems that may be physically diffuse; that may consist of different types of components some of which may not actually be physical (e.g., information or ideas); and above all when we deal with social systems (Mingers 2006b, Ch. 4).

Going beyond the structural aspects of systems, one of the founders of the systems movement Ludwig von Bertalanffy (1950) developed the concept of *open systems* as opposed to the closed systems of the laboratory, and also established general systems theory *(GST)*. based on the recognition that the systems concepts and principles we have described can be applied

[2] Good sources for overviews of the history of systems are Capra (1997), Checkland (1981), Hayles (1999), and Heims (1993) and there is an interesting and very detailed timeline at the (American Society for Cybernetics 2006)

irrespective of the particular nature or substance of the systems concerned. It is therefore possible to study systems relationships and organisations in the abstract and then apply them, as with mathematics, to particular domains.

Another major development was an entirely new discipline – *cybernetics* – the science of communication and control. The early cyberneticians, Weiner (1948), von Neumann (1958), Shannon (1949) and McCulloch (1943), were mainly mathematicians and engineers who were interested in the ways in which systems, both mechanical and biological, regulated and controlled themselves in a largely automatic way (Tamburrini and Datteri 2005). They recognised that the key to this was the concepts of *information* and *feedback*. Working initially on the design of self-controlling weapons, the ideas soon spread into modelling the functioning of the brain (Ashby 1952), developing the first digital computers (von Neumann 1958), anthropology (Bateson 1936) and psychiatry (Bateson 1973). Systems concepts were also applied extensively in sociology, for example Parsons (1951) whose work was criticised for being overly functionalist; Buckley (Buckley 1967) who emphasised the dynamic and processual aspects of systems; and Habermas (1987).

Finally, an important realisation that came out of quantum physics, again in opposition to the prevailing positivist view of science, was the inevitable involvement of the *observer* in any observations or descriptions that we make of the world. Heisenberg's uncertainty principle showed that the results we might get could not be simply reflections of the external world alone but were always in part due to the very act of observation. As Heisenberg (1963, p. 75) put it, 'Natural science does not simply describe and explain nature … it describes nature as exposed to our method of questioning'. As we shall see, it is very much one of the important planks of systems thinking that the observer must be recognised as part of the system. This foreshadowed the development in the 1970's of what we might call interpretive systems thinking, based on the insights of phenomenology and interpretive sociology, and known as *2nd order cybernetics* (Maturana and Varela 1980) or *soft systems thinking* (Checkland 1981)[3].

Issues in mechanistic explanation

Illari and Williamson (2011) provide a useful overview of some of the issues involved in the philosophy of mechanistic explanation and we will use it to structure this section in which we try to show that systems thinking and critical realism can shed some light on these problems.

The nature of mechanistic explanation

Mechanistic explanation is clearly in contrast to the covering law model for several good reasons, and this has been one of the main arguments of CR against various forms of positivism and empiricism (Bhaskar 1978, Groff 2011). Positivism, in the form of the D-N model of explanation and resting on a Humean view of causality, involves a double reduction. It firstly reduces the domain of the real - enduring entities and structures that have causal powers - to the domain of the actual – particular events that actually occur (ignoring absences, i.e., events that might have occurred but for some reason did not). And then, it

[3] Some of the major systems works not referenced elsewhere are: Churchman (1968), Churchman (1971), Laszlo (1972), Weinberg (1975), Rapoport (1986), Klir (1991) and Open Systems Group (1981)

reduces the domain of the actual to that of the empirical, i.e., those events that happen to be observed and can be measured. From this impoverished base, it does no more than re-describe the data in the form of a mathematical law, with no greater concept of causality than constant conjunctions of events (Craver 2006).

In contrast, CR only begins with empirical observations, it then goes beneath the surface to try and explain what underlying mechanisms could, if they existed, produce the events that in fact occurred, or did not occur. CR has a stratified ontology – the *real* which consists of enduring mechanisms and structures, including human beings and social systems; the *actual* which are the events caused or precluded by the interacting mechanisms, and which themselves of course can have causal effects; and the *empirical* which is the subset of the actual experienced and observed by human beings. Bhaskar supports this view both on logical (transcendental) grounds that we will discuss later, and more pragmatic ones that echo those philosophers supporting a mechanistic approach. In particular that this properly provides an *explanation* for events rather than simply a redescription for them (Glennan 2002); that it accords with the actual practices of scientists (Bechtel and Abrahamsen 2005); and, again to be discussed later, that general or universal laws do not exist in many domains, especially the social sciences (Cartwright 1983). Chakravartty (2005) defends this view against several long-standing objections.

Illari and Williamson suggest that a second aspect of mechanistic explanation is a distinction in the literature between epistemic and ontic types of explanations. Salmon (1998b) originally drew this distinction suggesting that an epistemic form of explanation, such as that of Hempel (1965), was essentially an inferential argument to the effect that the events to be explained were to be expected given general laws and the initial conditions. On the other hand, an ontic explanation (sometimes called a physical explanation) is one which shows how the events have resulted from causal patterns and regularities. These may be of two types – constitutive, where the events result from the properties of a specific mechanism, and etiological where they are the outcome of a chain of events. More generally, an epistemic explanation is motivated by a desire to improve human understanding, and is therefore constrained by the knowledge and understanding of the audience. An ontic explanation concerns the actual causal mechanism and its effects whether or not it is properly understood. A similar distinction has been made between actual mechanisms in the world, and the scientists' descriptions and models of that mechanism (Bechtel and Abrahamsen 2005, Glennan 2002, Machamer, et al. 2000). As Illari and Williamson (2011, p. 823, my emphasis) say, "These differences exist because in the epistemic sense of explanation it is the *description* of the mechanism that explains, while in the physical sense, the *mechanism itself* does the explaining."

This distinction can be seen as part of a wider differentiation made by Bhaskar concerning what he calls the *transitive* (epistemic) and *intransitive* (ontic) dimensions of science (Bhaskar 1979, Ch. 1). It has long been argued, especially from within the sociology of science, that science is a human activity or practice much like any other, and therefore the results of science reflect such human practices as much as they do the object world. This argument can be taken to have significant sceptical conclusions, for example in the "strong"

sociology of knowledge programme (Bloor 1976) or various forms of post-modernism. Bhaskar accepts that indeed much of science is a human activity or production, which he calls the transitive dimension, but maintains strongly that there is also in intransitive dimension to science which consists of the objects of knowledge that are independent of us, or at least of how we describe or know them.

Thus, the transitive dimension involves all the human activity of producing knowledge or, perhaps better, transforming previously existing knowledge, and is therefore inevitably local, temporal and partial. We have to accept that knowledge can never be perfect, or even be "proved" to be correct. It is always fallible or epistemically relative but this does not mean that all theories are equal, or that there are not rational grounds for choosing between them. One reason for this is precisely the externality or ontological independence of the objects of knowledge in the intransitive domain. Such objects do not have to be physical, but can be social, cognitive or even linguistic. An academic paper is produced in the transitive dimension but, once published, becomes an intransitive object of knowledge able to be discussed or referenced.

Bhaskar also has a multivalent model of truth which is relevant in this context (Bhaskar 1993, Mingers 2008). This involves four levels or degrees of truthfulness. The lowest level (*normative-fiduciary*) is when one simply accepts the truth of what someone says on the grounds that they are a reliable or knowledgeable person who should "know" what they are talking about (e.g., a scientist or expert). Clearly this is very common in everyday life. The second level (*adequating*) is truth that is based on sound evidence or justification of some kind rather than mere belief. Both these levels are in the transitive domain, and thus relate to the epistemic approach above. The third level (*referential-expressive*) is like a weak correspondence theory relating theories or models in the transitive domain to their intransitive objects.

The final level (*alethic*) is somewhat controversial (Groff 2000) for it moves the truthmaker entirely into the intransitive domain. There is no longer a correspondence between different domains, for it is the truth of things in themselves, and their generative causes, rather than the truth of propositions. It is no longer tied to language, i.e., it is no longer necessarily linguistic, although it may be expressed in language. It seems to be very akin to the ontic view of explanation that Illari and McKay described above. One could perhaps say that the experience of toothache generates its own alethic truth – we do not need to describe it or compare it with something else, we merely experience it to know its truth.

The reality of mechanisms

It may seem obvious, but if mechanisms are to be the core of scientific explanation, then it is necessary that mechanisms be "real", that is, at least some must have an independent existence and be responsible for the phenomena that they explain. Clearly there are extreme sceptical arguments that would question whether we can take anything to exist, including ourselves – the age old argument addressed by Descartes and Husserl. We will not consider those but there are nevertheless important issues that need to be addressed, particularly the

debate over “theoretical entities”, and the possibility of non-physical mechanisms such as social and cognitive systems.

First, theoretical entities – committed positivists and empiricists denied the legitimacy of unobservable, theoretical entities within scientific theories on the grounds that they were not observable or measurable, and so could not be assumed to exist. This is to adopt a perceptability criterion for existence. It is clearly against the practices of working scientists, who routinely hypothesise the existence of unobservable mechanisms and then set about trying to observe them or their traces (witness the billions spent on the search for the Higgs boson), and it is clearly against the possibility of mechanisms in the non-physical domain. Bhaskar (Archer, et al. 1998, Bhaskar 1978, Bhaskar 1989) has proposed several arguments against this view, and the related Humean view of causation as constant conjunctions of events. The main form of argument, which is also employed in a slightly different way by Cartwright (1999), is a transcendental argument a la Kant.

Transcendental arguments (Stern 2000) take the form:

Premise 1: There is something, X, that we experience or agree about.
Premise 2: X could not be experienced if Y were not the case.
Conclusion: Y must be the case.

For Bhaskar, the X that we experience and agree on is experimental scientific activity (within natural science), both its successes and failures. In conducting a laboratory experiment we, scientists, bring about (or fail to bring about) a particular effect under certain controlled conditions. Effectively, we engineer constant conjunctions of events which do not, in fact, happen regularly at all. Then, however, we find that these effects can also be brought about outside of the lab, in open rather than closed conditions. For this to happen it must be the case that (Y) causal laws are more than simply constant conjunctions – there must in fact be enduring structures or mechanisms that are distinct from the events they generate, and occur both inside and outside the laboratory.

> “On this view, laws are not empirical statements but statements about the forms of activity characteristic of the things of the world. And their necessity is that of a natural connection, not that of a human rule” Bhaskar, in Archer (1998, p. 34).

Clarke (2010) analyses the transcendental arguments of both Bhaskar and Cartwright concluding that while neither succeeds entirely, neither should be rejected out of hand. Bhaskar himself developed some of his ideas from previous work by Harré and Madden (Harre and Madden 1975) on the notion of causal powers[4] and it is also possible to relate it to the Aristotelian approach to causality (Pratten 2009).

The second issue concerns the reality of non-physical mechanisms such as social, psychological or informational systems. In fact, most of the philosophical literature on

[4] Although Harre has distanced himself from critical realism, especially in respect of social structures (Harrε 2002)

mechanisms restricts itself very much to the natural sciences such as biology and chemistry but the ontology of social structures has always been highly contentious in social science (Mingers 2004). We cannot cover such a debate here but we will highlight three related issues: i) whether we can accept non-physical and non-observable mechanisms (or systems, or structures) as having ontological reality; ii) whether there are social structures or mechanisms over and above the actions of individual people; and iii) whether the dependence of social mechanisms on peoples' understanding of them somehow compromises their reality. For critical realism, the answer is clearly yes to the first two and no to the third (Bhaskar 1979, Bhaskar 1989, Bhaskar 1997).

i) CR utilises a causal criterion for existence and a transcendental argument, as discussed above. Entities do not have to be physical or directly observable to be real; they only have to be causally efficacious. This means that concepts, ideas, rules and social practices, for instance, are no less real for being unobservable. We can also apply the transcendental argument to social structures. Our experiences of the social world cannot be explained purely in terms of individual actions – it must be the case that there are social mechanisms in operation that exist before, and over and above, particular individuals, and that necessarily enable our social activities. Some obvious examples are: language, the banking system and the use of money more generally, or the legal system. Considering language, it is something that pre-dates us, that we have to learn in childhood, and yet it then enables us to communicate with people we have never met. It is, in a general sense, a human construction for it would not exist without us, and yet none of us individually can change or develop it. For these experiences to occur it must be the case that language exists as an unobservable structure or mechanism separate from its embodiment in individual's nervous systems, or its instantiations in actual language use.

ii) For Bhaskar, society exists as an object in its own right, emergent from, but separate to, people and their activities, and with its own properties. Society always pre-exists individuals who do not therefore create it but only transform or (re)produce it. Nevertheless, society is necessary for social activity and it only *exists* in virtue of that activity. Society therefore conditions social activity and is either maintained or changed as an outcome of that activity. Equally, human action (praxis) is both a conscious production, i.e., intentional bringing about of purposes, and an unconscious (re)production of society. Society is an "ensemble" of structures, practices and conventions, where structures are relatively enduring generative mechanisms that govern social activities. Whilst emphasising the ontological reality of social structures, Bhaskar recognises that they have significantly different properties from physical mechanisms as was mentioned above. In particular: i) Social mechanisms do not *exist* independently of the activities they govern. ii) Social mechanisms cannot be *empirically identified* except through activities. iii) Social mechanisms are not independent of actors' *conceptions* of their activity. iv) Social mechanisms are *localised* to particular times and cultures. Despite these differences they are still suitable subjects for scientific theorising even if they lead to particular epistemological difficulties. The explicit use of "mechanism" as an explanatory device is growing in social science, see, for example, in history Steinmetz (1998), in organizational research Anderson et al (2006) and in politics Gerring (2007).

iii) A third issue raised by Illari and Williamson (2011) is that many of the writers on mechanisms (e.g., (Bechtel and Abrahamsen 2005, Craver 2007, Darden 2006, Glennan 2002)) presume that they must fulfil some *function* within a wider system so that a mechanistic explanation must be a functionalist explanation. This is problematic partly because of the long-standing debate about the validity of functionalist explanation, especially in the social sciences (Salmon 1998a), but also because the specification of a function would seem to depend on a description of the wider system which may in turn depend on the perspective of the observer.

Our view is that insisting that mechanisms must fulfil some function in order to be a mechanism is neither necessary nor legitimate outside of the domain of humanly-designed systems. Certainly modern systems theorists do not accept it. For example, the biologist and neurophysiologist Humberto Maturana (1970, 1975), who developed the concept of autopoiesis (self-producing systems) to explain the fundamental nature of living entities, was clear that his use of mechanistic explanation was non-teleological; as was Antony Giddens (1984) with his concept of structuration as a theory of the reproduction of social systems. Mechanisms arise historically through some particular chain of events, perhaps involving chance. They have effects, and it may well be that these effects, often in combination with the effects of other mechanisms, may give rise to behaviour that is self-sustaining or contributes to a wider system (another mechanism at a higher level). In this sense, they may be seen, *after the event and by an observer*, to play some functional role. However, the actual operation of the mechanism still occurs in terms of its own structure and local interactions:

> "The organization of a machine, … only states relations between components and rules for their interactions and transformations, in a manner that specifies the conditions of emergence of the different states of the machine which, then, arise as a necessary outcome whenever such conditions occur. Thus the notions of purpose and function have no explanatory value in the phenomenological domain which they pretend to illuminate" (Maturana and Varela 1980, p. 86).

Not only is functionalist explanation unnecessary, but it may be incorrect since mechanisms may have effects which are dysfunctional or non-functional as far as some wider system is concerned. Obvious examples are cancerous cells or insurgency, both of which are still the result of organised mechanisms. So it is right and proper to characterise mechanisms in terms of their components, relationships and emergent powers and properties, and then observe their behaviour in interaction with other kinds of mechanisms. The issue about a mechanism playing different roles in different systems will be discussed in the section on localisation.

Must mechanisms be local?

If we move, now, to the actual nature of mechanistic explanations, or more precisely the nature of the mechanisms that are postulated in such explanations, we find another potential problem in the literature – that such mechanisms are generally said to be "local" in a spatio-temporal sense (Bechtel 2001, Craver 2007, Hall 2004). Illari and Williamson (2011) identify three possible difficulties – the functional individuation of mechanisms; the existence of non-physical mechanisms; and omissions or absences, but still conclude that "it is a genuine

feature of mechanisms that they are local" (p. 833). The argument of this section, from a system perspective, is that whilst many mechanisms (especially physical ones) are indeed local, in a physical space sense, it is not a necessary characteristic of a mechanism. Rather, a mechanism or system operates within a state space dependent on the characteristics of its constituting components, which may or may not be physical.

There seems to be a general agreement amongst those advocating mechanistic explanation (e.g., (Bechtel and Abrahamsen 2005, Gillett 2007, Glennan 1996, Wimsatt 1994, Woodward 2002) that the postulated mechanisms are what we have called above "systems", i.e., groups of component parts (or mechanisms) that interact together to create a particular effect or phenomenon which is to be explained. They generally form hierarchies, with emergent properties at each new level. The implication of this conception, particularly in the case of physical phenomena which is what most of these authors discuss, is that the mechanism and its effects are localised in a physical sense. The size of the locality is hugely variable, from quantum to astronomical scale.

From a systems viewpoint, we would translate the idea of localisation into one of boundary. In order to identify a system (or mechanism, using the two equivalently) we need to distinguish what elements constitute the system as opposed to its environment, and this means specifying the system's boundary. This is not a straightforward task because of the variety of different types of system, as can be seen from the following points from Mingers (2006a).

A boundary is that which separates or demarcates that which is part of a system from that which is not. It may be physical or non-physical, actual or conceptual. In the case of physical systems, we can distinguish:

- Edges and surfaces that are the limit or extent of a system, e.g., a pool of water
- Enclosures, where there are specific boundary components that keep in that which is included, and/or keep out that which is excluded, e.g., a football or a fence or a membrane such as the cell wall.
- Demarcations, where the system is physical but not necessarily contiguous in space, e.g., the solar system or a central heating system

In general, systems can be conceptualised in different ways, generating different boundaries; and the components of a system may be part of multiple systems. For example, a central heating system could be conceptualised as a flow of water system (pipes, radiators, valves, water supply), a flow of energy system (gas, boiler, water, air), or a flow of information system (difference between actual and desired temperature, difference in thermostat setting, difference in degree of heating).

These examples show that systemic thinking involves more than the simple recognition of individual objects. It begins with a particular phenomenon to be explained or purpose to be achieved. It then requires a degree of conceptualisation, rather than mere perception, on the part of an observer to characterise an appropriate system in terms of components, relations and boundary. The boundary may in part have a material embodiment but generally it will

simply represent a distinction or demarcation between that which has been selected as part of the system and that which is not. This does not mean that the boundary is purely arbitrary, or is wholly a construction of the observer. It rests on the components and relations that exist independently in the intransitive domain even though it is selected by the observer. This is demonstrated by the fact that the observer may *get it wrong*. Knowledge is always fallible and the real world will soon let us know if our choices of components, relations and boundaries do not in fact yield the appropriate behaviour.

So far we have considered primarily physical systems where the idea of spatial localisation may be seen to be necessary, although even there non-physical elements such as information are often involved. However, as we move away towards conceptual or social systems we need to characterise them in terms of the space of interactions of the system itself, which may well not be physical space.

Let us consider as an example the nervous system as a system (or mechanism) in its own right, separate from although obviously part of, a body (Maturana 1970, Maturana 1980, Maturana, et al. 1995, Varela 1991). An organism without a nervous system, such as amoeba, acts in response to chemical changes in its immediate, local, environment. Its outer wall is essentially both its sensory and effector surfaces. However, in the nervous system cells have become specialised in two particular ways: first, they have developed lengthy extensions called dendrites that connect them to many other, sometimes distant, neurons. This leads to a separation of sensor from effector and the possibility of a transmission of difference or disturbance. Second, they have developed a generalised response medium – electrical activity – into which all forms of sensory/effector interactions can be translated. A third development is that of internal neurons that only connect to other neurons, and form the vast majority of developed brains. These effectively sever the direct connection between sensor and effector and are the basis for cognition, language and ultimately self-consciousness.

The effect of this is that the development of the nervous system opens up a whole new domain of interactions beyond the purely local physico-chemical ones of amoeba. They allow the organism to interact with the *relations* or *differences* between events, rather than simply with the events themselves. Neurons develop that are only triggered by particular combinations of other neurons, which represent particular configurations within the environment. Thus, although the nervous system is a physical system, and does have physical interactions, its domain of interactions *as a nervous system*, is states of relative neuronal activity triggered by relations and differences in its environment and acting back on that environment. Such interactions cannot be localised to the brain. For example, in a mobile phone conversation with someone physically distant, differences in sound are transmitted through the phone system to differences in sound at the receiver, differences in brain activity, differences in sound etc. etc. The whole forms a system in which spatial location is not a necessary or defining feature. This view is related to theories such as radical enactivism (Menary 2006) and the extended mind (Clark and Chalmers 1998).

The nature and boundaries of other non-physical systems, especially social systems, is controversial, and we do not have the space to discuss it here (Archer 1995, Bhaskar 1979,

Giddens 1984, Luhmann 1995, Mingers 2002, Mingers 2004), but we will just give one illustration. It is very common in the commercial world to talk about the "market mechanism" as a particular type of economic process and we would argue that it can, indeed, be seen as a social or economic mechanism. It is a particular form of (largely unregulated or controlled) trading where buyers and sellers interact fairly directly, and prices change in response to the balance between supply and demand. Historically, this mechanism actually did operate locally in a physical market, located in space and time, and lasting for a certain duration. However, today we can see that trading, especially in financial or commodity markets, is highly attenuated being conducted electronically throughout the world and 24 hours per day. It does not make sense to talk about its locality, but it is still important to delineate the boundaries of such a system – i.e., what constitutes the system and what constitutes its environment even if the answer may depend in part on the observer and their purpose.

Absences and omissions as causes

Finally, in this section we shall discuss the question of omissions or more generally absences as causal elements of mechanisms. Illari and Williamson (2011) suggest that absences may be a problem for mechanistic explanation on the grounds that they may not be local, a condition that they consider necessary as discussed in the previous section. Torres (2009) proposes a revision to the mechanistic models of Glennan (1996) and Machamer et al (2000) to more properly account for negative or absent causes.

We have already argued above that we do not consider localisation as a necessary condition for a mechanism, and so we do not consider the fact that absences sometimes cannot be localised to be a problem. Rather, we would argue strongly that varieties of absence often lie at the heart of causal mechanisms, and both critical realism and systems thinking, particularly cybernetics, supports this view.

In terms of CR, as Bhaskar developed his ideas towards the dialectical version in *Dialectic: The Pulse of Freedom* (Bhaskar 1993), he came to see absence as more and more important, indeed ultimately as more significant that presence. Against the prevailing worldview that deals only with what positively occurs or exists (especially in positivism and empiricism), Bhaskar maintains that it is the absent or the negative which has priority for it is only against this that the positive stands out or happens. Bhaskar suggests four categories of absence: i) simple or ontological absence, i.e., that some thing or event that is expected does not occur or does not exist. Such absences can have causal effect and therefore 'exist' in the same way as other things. He calls them 'de-onts'. The instrument that is not to hand, the bill that is unpaid, or the appointment that is missed all have causal effects[5]. ii) Absence as a verb, which could be absenting something or negating something, e.g., draining water or removing dirt; or which could be absenting an absence, e.g., removing a need or want by fulfilling it. Developing from these two are: iii) 'process-in-product', whereby a process (e.g., shopping) leads to an absence (e.g., money in the bank); and iv) product-in-process whereby an entity or

[5] It is important that the absence must have been expected or must normally occur, for it to be significant. At any instant there is an infinity of things that are not happening but the majority are not in any sense relevant. That I am not at the bus stop at this moment is of no importance if I was not intending to be.

structure (e.g., poverty, lack of money) exercises its powers in producing an absence (necessities of life).

This is interesting from a systems thinking point of view because it is *not* something that is generally discussed or considered in the modern literature and yet is clearly of great importance. In fact, its significance was recognised by some: it can be seen as the basis of cybernetic explanation as Bateson, one of the founders of cybernetics, observed:

> "Causal explanation is usually positive. ... In contrast to this, cybernetic explanation is always negative. We consider what alternative possibilities could conceivably have occurred and then ask why were many of the alternatives not followed, so that the particular event was one of those few which could, in fact, occur." (Bateson 1973, p. 375)

A similar idea is at the heart of Luhmann's (1990) theory of social communication in which a message acts as a trigger or selector from among the many responses or replies that could be generated – it selects that which is presenced from among all the other absent possibilities. We can also see the importance of absence in the idea of control by feedback. The feedback system (e.g., a thermostat) is always trying to close a gap (absent an absence) between the desired state of the system and the actual state of the system (Wilden 1977).

Conclusions

The purpose of this paper has been to demonstrate that many of the issues and debates within the recent philosophy of causal mechanisms have already been much discussed within the literatures of critical realism and systems thinking. And, moreover, that the conclusions reached there may well be useful in the philosophical discourse about mechanisms.

After a very brief review of the literatures of critical realism and systems thinking, a range of issues concerning mechanisms were discussed. In particular, it was argued that:

- We need to distinguish between the events that occur (and do not occur) that are to be explained (the actual) and the underlying, enduring structures and mechanisms (the real) that, through the operation of their powers in interaction, causally generate these events. We should also distinguish between the transitive domain of science (which is epistemic) in which theories and knowledge is humanly produced, and the intransitive domain of the independent objects of our knowledge (which is ontic).
- We can accept the reality of mechanisms (or systems more generally) may be non-physical and/or non-observable. The ontological criterion should not be perceptability but causal efficaciousness. Thus, social mechanisms (e.g., "society"), informational mechanisms (driven be information), or cognitive mechanisms (e.g., ideas or motives) all have causal effects and may thus be part of explanatory theories.
- Mechanistic explanation does not have to be a form of functionalist explanation.
- Mechanisms do not have to be localised in a purely physical sense although they need to be bounded, or able to be demarcated, within their space of interactions.

- Absences and omissions may be causes and thus may legitimately be part of mechanistic explanations.

This paper has only been able to skim the surface of the many possible connections between mechanistic explanation and systems thinking and critical realism, but this will hopefully demonstrate the value of such an engagement for both sides.

References

American Society for Cybernetics (2006).*A timeline for the evolution of cybernetics,* 2006, 14the November.

Anderson, P. , R. Blatt, M. Christianson, A. Grant, C. Marquis, E. Neuman, S. Sonenshein and K. Sutcliffe (2006). Understanding Mechanisms in Organizational Research, *Journal of Management Inquiry*, 15, 2, 102-113.

Archer, M. (1995). *Realist Social Theory: the Morphogenetic Approach*, Cambridge: Cambridge U. P.

Archer, M., R. Bhaskar, A. Collier, T. Lawson and A. Norrie (Ed), (1998).*Critical Realism: Essential Readings,* London: Routledge.

Ashby, W. Ross (1952). *Design for a Brain*, London: Chapman & Hall.

Bateson, G. (1936). *Naven*, Stanford: Stanford University Press.

Bateson, G. (1973). *Steps to an Ecology of Mind*, Hertfordshire: Granada Publishing.

Bechtel, W. (2001). The compatibility of complex systems and reduction: A case analysis of memory research, *Minds and Machines*, 11, 4, 483-502.

Bechtel, W. and A. Abrahamsen (2005). Explanation: A mechanist alternative, *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36, 2, 421-441.

Bhaskar, R. (1978). *A Realist Theory of Science*, Hemel Hempstead: Harvester.

Bhaskar, R. (1979). *The Possibility of Naturalism*, Sussex: Harvester Press.

Bhaskar, R. (1989). *Reclaiming Reality*, London: Verso.

Bhaskar, R. (1993). *Dialectic: the Pulse of Freedom*, London: Verso.

Bhaskar, R. (1997). On the ontological status of ideas, *Journal for the Theory of Social Behaviour*, 27, 2/3, 139-147.

Bloor, D. (1976). *Knowledge and Social Imagery*, London: Routledge.

Buckley, W. (1967). *Sociology and Modern Systems Theory*, Englewood Cliffs: Prentice Hall.

Capra, F. (1997). *The Web of Life: a New Synthesis of Mind and Matter*, London: Flamingo.

Cartwright, N. (1983). *How the Laws of Physics Lie*, Oxford: Clarenden Press.

Cartwright, N. (1999). *The Dappled World*, Cambridge: Cambridge University Press.

Chakravartty, A. (2005). Causal realism: Events and processes, *Erkenntnis*, 63, 7-31,

Checkland, P. (1981). *Systems Thinking, Systems Practice*, Chichester: Wiley.

Churchman, C. W. (1971). *The Design of Inquiring Systems*, New York: Basic Books.

Churchman, C.W. (1968). *The Systems Approach*, New York: Dell Publishing.

Clark, Andy and David Chalmers (1998). The extended mind, *Analysis*, 58, 1, 7-19.

Clarke, S. (2010). Transcendental realisms in the philosophy of science: on Bhaskar and Cartwright, *Synthese*, 173, 299-315.

Craver, C. (2007). *Explaining the Brain*, Oxford: Clarendon Press.

Craver, Carl (2006). When mechanistic models explain, *Synthese*, 153, 3, 355-376.

Darden, L. (2006). *Reasoning in Biological Discoveries*, Cambridge: Cambridge University Press.

Gerring, J. (2007). Review article: The mechanismic worldview: Thinking inside the box, *British Journal of Political Science*, 38, 161-179.

Giddens, A. (1984). *The Constitution of Society*, Cambridge: Polity Press.

Gillett, C. (2007).The metaphysics of mechanisms and the challenge of the new reductionism, In *The Matter of the Mind*, M. Schouton and H. de Jong (Ed.), Oxford: Blackwell, 74-100.

Glennan, S. (1996). Mechanisms and the nature of causation, *Erkenntnis*, 44, 49-71.

Glennan, Stuart (2002). Rethinking Mechanistic Explanation, *Philosophy of Science*, 69, S3, S342-S353.
Groff, R. (2000). The truth of the matter - Roy Bhaskar's critical realism and the concept of alethic truth, *Philosophy of the Social Sciences*, 30, 3, 407-435.
Groff, R. (2011).Getting past Hume in the philosophy of social science, In *Causality in the Sciences*, P. Illari and J. Williamson (Ed.), Oxford: Oxford University Press.
Habermas, J. (1987). *The Theory of Communicative Action Vol. 2: Lifeworld and System: a Critique of Functionalist Reason*, Oxford: Polity Press.
Haeckel, E. (1866). *Generelle Morphologie der Organismen*, Berlin: Verlag.
Hall, N. (2004).Two concepts of causation, In *Causation and Counterfactuals*, L. Paul, E. Hall and J. Collins (Ed.), Cambridge, MA: MIT Press, 225-276.
Harre, R. (2002). Social reality and the myth of social structure, *European Journal of Social Theory*, 5, 1, 111-123.
Harre, R. and E. Madden (1975). *Causal Powers: A Theory of Natural Necessity*, Oxford: Blackwell.
Hayles, N. K. (1999).The second wave of cybernetics: from reflexivity to self-organization, In *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*, N. K. Hayles (Ed.), Chicago: University of Chicago Press, 131-159.
Heims, S. (1993). *Constructing a Social Science for Postwar America: The Cybernetics Group 1946-1953*, Massachusetts: MIT Press.
Heisenberg, W. (1963). *Physics and Philosophy*, London: Allen and Unwin.
Hempel, C. (1965). *Aspects of Scientific Explanation*, New York: Free Press.
Illari, P. and J. Williamson (2011).Mechanisms are real and local, In *Causality in the Sciences*, P. Illari and J. Williamson (Ed.), Oxford: Oxford University Press, 818-844.
Klir, George J. (1991). *Facets of Systems Science*, New York: Plenum Press.
Laszlo, Ervin (1972). *Introduction to Systems Philosophy : Toward a New Paradigm of Contemporary Thought*, N.Y.: Gordon and Breach.
Luhmann, N. (1990).Meaning as sociology's basic concept, In *Essays in Self-Reference*, N. Luhmann (Ed.), NY.: Columbia University Press, 21-79.
Luhmann, N. (1995). *Social Systems*, Stanford: Stanford University Press.
Machamer, P. (2004). Activities and causation: The metaphysics and epistemology of mechanisms, *International Studies in the Philosophy of Science*, 18, 1, 27-39.
Machamer, Peter, Lindley Darden and Carl F. Craver (2000). Thinking about Mechanisms, *Philosophy of Science*, 67, 1, 1-25.
Maturana, H. (1970).The Neurophysiology of Cognition, In *Cognition: a Multiple View*, P. Garvin (Ed.), NY: Spartan Books, 3-23.
Maturana, H. (1975). The organization of the living, a theory of the living organization, *Int. Journal Man Machine Studies*, 7, 313-332.
Maturana, H. (1980).Biology of Cognition, In *Autopoiesis and Cognition: The Realization of the Living*, H. Maturana and F. Varela (Ed.), Dordrecht: Reidel, 1-58.
Maturana, H. and F. Varela (1980). *Autopoiesis and Cognition: The Realization of the Living*, Dordrecht: Reidel.
Maturana, H.M, J. Mpodozis and J. Letelier (1995). Brain, language and the origin of human mental functions, *Biology Research*, 28, 15-26.
McCullough, W. and W. Pitts (1943). A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics*, 5, 115.
Menary, R. (Ed), (2006). *Radical Enactivism: Intentionality, Phenomenology and Narrative*, Amsterdam: John Benjamins.
Mingers, J. (2000). The contribution of critical realism as an underpinning philosophy for OR/MS and systems, *Journal of the Operational Research Society*, 51, 11, 1256-1270.

Mingers, J. (2002). Can social systems be autopoietic? Assessing Luhmann's social theory, *Sociological Review*, 50, 2, 278-299.
Mingers, J. (2004). Can social systems be autopoietic? Bhaskar's and Giddens' social theories, *Journal for the Theory of Social Behaviour*, 34, 4, 403-426.
Mingers, J. (2006a).Observing systems: The question of boundaries, In *Realising Systems Thinking: Knowledge and Action in Management Science*, J. Mingers (Ed.), New York: Springer, 65-100.
Mingers, J. (2006b). *Realising Systems Thinking: Knowledge and Action in Management Science*, New York: Springer.
Mingers, J. (2008). Management knowledge and knowledge management: Realism and forms of truth, *Knowledge Management Research and Practice*, 6, 62-76.
Mingers, J. (2011). The contribution of systemic thought to critical realism, *Journal of Critical Realism*, forthcoming,
Open Systems Group (Ed), (1981).*Systems Behaviour*, London: Harper & Row.
Parsons, T. (1951). *The Social System*, Glencoe.
Peirce, C. (1931-1958). *Collected Papers of Charles Sanders Peirce (8 Volumes)*, Cambridge: Harvard University Press.
Pratten, Stephen (2009). Critical realism and causality: Tracing the Aristotelian legacy, *Journal for the Theory of Social Behaviour*, 39, 2, 189-218.
Psillos, S. (2009).An explorer upon untrodden ground: Peirce on abduction, In *Handbook of the History of Logic: Inductive Logic*, D. Gabbay, S. Hartmann and J. Woods (Ed.), 10, Amsterdam: Elsevier, 117-151.
Rapoport, Anatol (1986). *General system theory : essential concepts & applications*, Tunbridge Wells, Kent: Abacus.
Salmon, W. (1998a).Comets, pollen and dreams: Some reflections on scientific explanation, In *Causality and Explanation*, Oxford: Oxford University Press, 50-67.
Salmon, W. (1998b). *Causality and Explanation*, Oxford: Oxford University Press.
Shannon, C. and W. Weaver (1949). *The Mathematical Theory of Communication*, Illinois: University of Illinois Press.
Steinmetz, G. (1998). Critical realism and historical sociology. A review article, *Comparative Studies in Society and History*, 40, 1, 170-186.
Stern, R. (2000). *Transcendental Arguments and Scepticism: Answering the Question of Justification*, Oxford: Oxford University Press.
Symons, J. (2008). Computational models of emergent properties, *Minds & Machines*, 18, 4, 475-491.
Tamburrini, G. and E. Datteri (2005). Machine experiments and theoretical modelling: from cybernetic methodology to neuro-robotics, *Minds & Machines*, 15, 3/4, 335-358.
Torres, P. (2009). A modified conception of mechanisms, *Erkenntnis*, 71, 233-251.
Varela, F., Thompson, E. and Rosch, E. (1991). *The Embodied Mind*, Cambridge: MIT Press.
von Bertalanffy, L. (1950). The theory of open systems in physics and biology, *Science*, 111, 23-29.
von Neumann, J. (1958). *The Computer and the Brain*, New Haven: Yale University Press.
von Uexkull, J. (1909). *Umwelt und Innenwelt der Tiere*, Berlin: Springer.
Weinberg, Gerald M. (1975). *An Introduction to General Systems Thinking*, N.Y.: Wiley.
Weiner, N. (1948). *Cybernetics: or Communication and Control in the Animal and the Machine*, Ca. Mass.: MIT Press.
Wertheimer, M. and D. King (2005). *Max Wertheimer and Gestalt Theory*, London: Transaction Publishers.
Wilden, A. (1977). *System and Structure*, London: Tavistock.

Wimsatt, W (1994). The ontology of complex systems: levels, perspectives and causal thickets, *Canadian Journal of Philosophy*, 20, 201-274.
Winther, R. (2009). Part-whole science, *Synthese*, 1-31.
Woodger, J. (1929). *Biological Principles: a Critical Study*, London: Keegan, Paul and Co.
Woodward, Jim (2002). What Is a Mechanism? A Counterfactual Account, *Philosophy of Science*, 69, S3, S366-S377.


University of Kent