# The theory of emergence, social structure, and human agency

David John Elder-Vass

Birkbeck College

University of London

# Abstract

The question of structure and agency has long been regarded as a central problem in sociological theory. In addition to the traditional approaches offered by methodological individualism and methodological collectivism, recent decades have seen a number of attempts to transcend these two positions, most prominently by Giddens and Archer. However, this remains a highly contested subject, and none of these proposed resolutions has yet prevailed.

This thesis investigates whether a more satisfactory understanding of structure and agency can be developed by applying the theory of emergence.

It engages with several branches of the literature on emergence, including the early twentieth century emergentists, complexity theory, critical realism, and the philosophy of mind, in order to clarify the theory of emergence itself and to show how it provides a viable alternative to extreme varieties of reductionism and dualism. Having done so, it examines whether human agency and social structure can be understood in emergentist terms, and whether theorising them in this way enables us to produce a more satisfactory account of their relationship to each other and to social events.

This thesis seeks to complement and extend Archer's existing analysis of structure and action in terms of emergence. While it is critical of the sociological ontology of Giddens and others, it also seeks to demonstrate that a wide range of sociological *theory*, including much of theirs, is not only compatible with but also complementary to an emergentist account of structure and agency. Thus, for example, Bourdieu's concept of the *habitus* is one-sided both ontologically and theoretically, yet can become part of a more balanced theory of human action by integrating it within an emergentist framework.

Emergence, the thesis concludes, provides a strong foundation for clarifying, and indeed transcending, previous sociological understandings of social structure and human agency.

For Alisa, Hazel, Jasmine, and Gerald

# Contents

# List of Figures

# Acknowledgements

I would like to express my gratitude to Jason Edwards, my supervisor, for his careful reading and constructively critical comments on every part of this thesis, for pointing me in the direction of some of the essential debates in the literature, not least the work of Bhaskar and other critical realists, and for his positive and supportive attitude throughout this project.

I would also like to thank:

- Diana Coole for her useful comments on several chapters and for her continuing support.
- All those people who have read and commented on various drafts of various parts of this thesis, some in the form of conference papers or articles. These include Margaret Archer, Filipo Artoni, Eric Kaufman, Martin Lipscomb, and Geoffrey Thomas.
- Sam Ashenden and Sami Zubaida for their practical support.
- The participants at various conferences and seminars for their comments on my papers, including Roy Bhaskar, Bob Carter, Martyn Hammersley, John Holmwood, Jamie Morgan, Tobin Nellhaus, Caroline New, Wendy Olsen, and John Scott.
- The anonymous referees who reviewed the papers mentioned below.

My greatest debt, however, is to my wife Alisa, for her unfailing support throughout a process in which I have sacrificed security and prosperity for the sake of a dream. And of course my children Hazel, Jasmine, and Gerald have provided a constant distraction which has made a significant contribution to maintaining my sanity and sense of perspective throughout the process.

Parts of chapters two and three have been published in the following papers:

# 1      Introduction

This thesis argues that the theory of emergence, properly understood, can provide the foundation for an explanation of human agency and social structure that improves on those offered by existing social theory. It aims to make an original contribution to two distinct but inter-related areas: on the one hand, to the general theory of emergence, and on the other, to our understanding of the sociological problem of structure and agency. These two aims are complementary in that the general theory of emergence advanced in the early chapters of the thesis (chapters 2 and 3) is applied to the social world (and thus illustrated in some depth) in the later chapters. Here it is used to construct an analysis of social structure (chapters 4 and 5), of human agency (chapters 6 and 7), and of the relation between the two.

The first section of this introductory chapter will describe the scope and objectives of the thesis, outlining the subject of its core argument, identifying some of the issues that it is *not* able to cover, and introducing the range of literature with which it will engage. The second section will identify its primary substantive themes and define their central terms. The third and final section will discuss the methodological issues implicit in this theoretical thesis and position its argument with respect to some key questions in the philosophy of the social sciences.

**Scope and objectives**

*Applying emergence to the problem of structure and agency*

The concepts of structure and agency are central to sociological theory, yet there remains widespread disagreement about what they mean and about how (or even whether) they can be causally effective social forces. The causal role of social structure is particularly contentious. On the one hand, the concept remains implicit in, and indeed essential to, much sociological theorising, yet on the other hand many sociologists seem to mistrust the existing theoretical accounts of its role. Structure is often taken for granted, not because the concept is clearly understood and uncontroversial, but because addressing the theoretical issues seems so problematic. There are, of course, exceptions - thinkers who have sought to address this issue (see, for example, Archer et al., 1998; Bhaskar, 1998b; Giddens, 1984). Yet none of their solutions are widely accepted.

Agency is equally problematic. Although few writers today would deny that human agency can be causally effective, the questions of just what agency means and how it works remain contentious. There is a world of difference, for example, between Archer's stress on human reflexivity – the conscious prioritisation of concerns and translation of those concerns into life projects – and Bourdieu's emphasis of the role of the *habitus* – a body of socially acquired and physically embodied dispositions that seems to drive our behaviour with little conscious reflective input.

In this thesis I propose to address these problems by applying the theory of emergence to develop a viable concept of social structure and the outline of a complementary theory of human action. While, like Margaret Archer, offering an emergentist account of structure from a critical realist perspective, this thesis aims to develop a deeper understanding of emergence itself and show that this provides a stronger foundation for social ontology. Its early chapters consider the general problem of finding a viable middle way between reductionism and full-blown ontological dualism and seek to show that a particular, relational, version of the concept of emergence can provide such a route. This, I claim, provides a general approach to emergence that can be applied to the social domain to build an adequate foundation for understanding the nature of social structure, human agency, and the relationship between them.

Equipped with such an understanding, the thesis will criticise a variety of perspectives on both general and social ontology. The early chapters, in arguing for the general theory of emergence, argue *against* reductionisms, holisms, and extreme dualisms whilst arguing *in favour* of a scientific understanding of the relations between different ontological levels. The later chapters criticise more specifically social doctrines of explanation such as methodological individualism, methodological collectivism, and what Archer has called central conflationism. They go beyond ontology, however, to specifically theoretical questions, most notably in developing the outline of a theory of action that can reconcile some of the seemingly contradictory impulses in Archer and Bourdieu's conceptions of human agency.

### *Limitations and omissions*

This outline of the scope of the thesis, however, must be complemented with some recognition of its boundaries. In the time and space available, it has not been

possible to cover everything that would be addressed in an ideal project on these subjects. In terms of the general theory of emergence, I have ignored a number of issues that are sometimes considered important in discussions of emergence but are not significant for the approach developed here – notably the questions of supervenience and internal relations. As far as the analysis of the social world is concerned, there are two interrelated groups of limitations and omissions from this thesis, both of which I intend to rectify in my future research programme.

The first group concerns the role of language, meaning, and culture in the social world, and their inter-relationships to the problem of structure and agency. For example, social conditions have a strong causal influence on the development of culture; human action is inherently meaningful; and the functioning of social organisations and institutions rests to some extent upon our beliefs about them and our ability to express these beliefs in linguistic form. A full account of structure and agency, then, would require a full analysis of the ontology and theory of language, meaning and culture and an explanation of the various significant ways in which these all interact. This is a research programme in its own right, and far beyond the scope of any single work. Hence the treatment of these questions in this thesis falls far short of a comprehensive analysis. They are not, however, completely neglected. Thus, for example, chapter four touches on the role of beliefs in sustaining organisations, chapter five tentatively addresses the ontology of institutions in its analysis of structuration theory, and chapter seven discusses the interpretativist argument that action is inherently meaningful in so far as this is taken to constitute an objection to a causal social theory.

Secondly, the thesis can only focus on a narrow range of social structures. Chapters four and five develop an account of how organisations can be emergent social structures, and argue that some structures that we may not be used to thinking of as organisations – such as married couples – can be analysed in the same way. But other types of social structure also exist, arising from the existence of other sorts of relations between the people who constitute them. Thus a full account of social structure would require a discussion of whether and how other sorts of social structure should be considered emergent. There is a broad range of putative types of social structure which are neglected here – such as socioeconomic classes, capital, markets, patriarchy, and many others – as well as one type that is touched on only lightly in this thesis – institutions. It remains an open question whether any or all of these can be explained in

emergentist terms similar to those I have applied here to organisations, and this is another question I intend to pursue further in the future.

In deciding how to handle these cases, the principle I have sought to apply is that this thesis should establish the argument that at least some social structures are indeed emergent in the sense it describes. If it can do this, then this should be sufficient to establish that the most prominent conceptions of methodological individualism, methodological collectivism, and central conflationism rest in principle upon ontological errors. Whether some other types of social structure can be eliminatively reduced, and quite what are the relationships between particular types of structure and particular meanings, are important questions, but questions that can be deferred for the purposes of establishing the primary argument of this thesis. Similarly, if this argument can be established without reference to more general concepts like supervenience and internal relations, then the critique of those concepts can also be set aside for some other occasion.

### *The literature*

As a work of theory, this thesis develops original ideas through a dialogue with the existing literature on its objects of study. Those objects of study, however, are inherently diverse. Emergence in particular is a concept that is potentially relevant to a vast range of intellectual fields, and it has been employed by thinkers in many of them. To engage with the literature on emergence, then, is inevitably to engage with a broad variety of disciplines and their dialects. Similarly, an emergentist approach to structure and agency is inherently one that may invoke arguments from the full range of the social sciences and indeed the humanities. This thesis thus touches on an unusually diverse range of disciplines, and hence engages with a variety of different disciplinary literatures.

Given the range of literature potentially implicated in this project, and the continuous need to refer my theoretical arguments back to an exposition and critique of the literature appropriate to their particular topics, it would have been unwieldy and impractical to structure this thesis as a literature review followed by subject-oriented chapters. Instead, the thesis engages with the literature throughout, generally discussing the literature relevant to each section as part of the process of developing that section's argument.

In the chapters on emergence and cause, the primary literatures employed are those of the philosophy of mind and science (e.g. the work of Kim), the early

emergentist philosophers (e.g. Lloyd Morgan and Broad) critical realism (notably Bhaskar and Collier), and complexity theory (including Holland and Gell-Mann, among others).  When the thesis turns to the questions of structure and agency, the balance of the literature used shifts strongly towards social theorists (e.g. Durkheim, Giddens, Archer, Bourdieu, and many others) although there are also references to a variety of related fields, such as the philosophy of social science (Winch, Sawyer), the philosophy of mind (Davidson, Searle, Dennett), the sciences of the brain (Freeman), phenomenological psychology (Merleau-Ponty), and evolutionary psychology (Tooby & Cosmides).

I would defend the interdisciplinarity of this work on a number of grounds. First, it is implicit in the subject of the thesis, which applies an inherently interdisciplinary concept – emergence – to its central problem. Second, a central plank of the response to the problem of structure and agency advanced here is that we can only understand these fully once we understand how entities that are studied in different disciplines – sociology, psychology, neuroscience – relate to each other. But thirdly, I also suggest that the analysis advanced here has potential across the somewhat artificial boundaries that divide the social sciences from each other. It is far from clear, in particular, that the division of economics and political science from sociology can be justified in terms of any ontological difference of their objects of study, and I would suggest that the arguments about structure and agency advanced in this thesis might be applied profitably in social sciences other than sociology.[1]

**Substantive themes**

The substantive themes of this thesis are the sociological question of structure and agency, which provides the 'problem' to be addressed, and the theory of emergence, which is investigated as a potential solution to this problem. This section aims to position these two themes for the reader, making clear what is understood by them in this thesis. As in the subsequent chapters of the thesis, I begin here with the theory of emergence, on the grounds that it is necessary to understand how emergence works before one can see how it could help with the problem of structure and agency.

---

[1] Sawyer advocates a particularly trenchant version of this argument, suggesting that by becoming the science of social emergence, sociology can and should turn back the tide of disciplinary imperialism and take over much of what is currently covered by economics (Sawyer, 2005, pp. 225-9).

### The theory of emergence

*Emergence* is the idea that a *whole* can have *properties* (or powers) that are not possessed by its *parts*. Such properties are called *emergent properties*, and an entity possessing one or more emergent properties may be called an *emergent entity* (Stephan, 1992, p. 27). Perhaps the commonest illustration of emergence in the literature is the example of water, which has been used to illustrate this point as far back as John Stuart Mill (Mill, 1900, p. 243). The properties of water are clearly very different from those of its components oxygen and hydrogen when these are not combined with each other in the specific form that constitutes water. We can, for example, put out a fire with water, but the outcome would be very different if we tried to do the same with oxygen and hydrogen (Mihata, 1997, p. 31; Sayer, 1992, p. 119). Hence it would usually be argued that water has emergent properties.[2]

The value of the concept of emergence lies in its potential to explain how *higher-level* entities (such as water in this example) can have a causal impact on the world in their own right. Such an impact would be irreducible to the causal impacts of its parts, but also consistent with the scientific world view which rejects mystical explanations of the irreducibility of higher-level properties. From its earliest origins, the concept of emergence has been used in this way. Thus, for example, the philosopher C.D. Broad used it in the 1920s to counter both 'mechanistic' and 'vitalist' explanations of life (Broad, 1925, pp. 44-61). The nineteenth-century mechanists were reductionists who believe that life had an entirely physical explanation, whereas the vitalists' argument was that life was to be explained by the presence in living organisms of a mystical substance called entelechy as well as their physical parts. Emergentists sought (and still do) to offer a middle way between these two extremes.

However, there are different versions of the concept of emergence, and they do not all offer a viable solution to this problem. Indeed, many philosophers believe that none of them do, but this thesis argues that there is a version of emergentism that *does* deliver a viable middle way between reductionism and dualism. This version is able to accommodate scientific explanations of the relationships between entities at different levels of the natural and social world while resisting the reductionist conclusion that this entails the redundancy of higher levels, whether in terms of the causal efficacy of higher level entities or the need for a higher level science.

---

[2] Although these properties would not be emergent according to some 'strong' definitions of emergence – see chapter 2 below.

It would be premature to set out the full argument here, but there is one aspect of the concept of emergence that does need clarifying at the earliest possible stage. This is the distinction between *synchronic* and *temporal* conceptions of emergence.

The basic definition outlined above describes the synchronic conception of emergence. This is concerned with the relation between a whole and its parts at a given moment in time. In this view, wholes may have emergent properties as a consequence of the characteristic way in which those parts are related or organised *in the present* to form this kind of whole.

By contrast, the everyday *temporal* definition of emergence is concerned with the development of something over a period of time – with the process which leads to the first appearance of that thing. This is by the far the commonest use of 'emergence' outside the philosophical literature, but it is *not* the same thing as synchronic emergence. Granted, anything that is synchronically emergent must have emerged temporally at some point in its past (unless it has *always* existed), and it may be interesting and important to understand how this occurred. But temporal emergence and synchronic emergence are two different concepts. Something that has 'emerged' temporally need not have synchronically emergent properties, and only synchronic emergence offers us a response to reductionism. Hence it is a variety of synchronic emergence that is advocated in this thesis.

Unfortunately, a number of thinkers have failed to distinguish these two senses.[3] In many cases, the authors concerned demonstrate a good grasp of the synchronic concept in parts of their writing, but then seem to conflate it with the historical version in other places. Holland, for example, talks of "the emergence of good play" as a checkers playing program learns to improve (Holland, 1998, p. 56). Bunge, similarly, cites "the emergence of a plant out of a seed" as an example of emergence (Bunge, 2003, p. 3). Neither seems to be a reference to the synchronic sense of emergence although both are advanced as referring to a single unified concept of emergence. Woolly thinking like this is confusing for the reader, and certainly makes it more difficult to explain emergence clearly. While I shall confine my usage of the word *emergence* to its synchronic sense, we must be constantly aware that others often conflate the synchronic and temporal senses.

---

[3] Perhaps encouraged by the ambiguous treatment of 'novelty' as a criterion of emergence in early emergentist writings.

*Structure and agency*

If we are to understand why we need a theory of structure and agency at all, we must begin by looking at what the concepts of structure and agency refer to, and then consider why the relationship between them is problematic.

Despite its widespread usage in sociology, *social structure* is a term whose meaning is "strikingly nebulous and diverse" (Lopez and Scott, 2000, p. 1). As Lopez and Scott point out, "there is little consensus over what the word means, and it is all too easy for sociologists to be talking at cross purposes because they rely on different, and generally implicit, conceptions of social structure" (Lopez and Scott, 2000, p. 1). Chapters four and five explore the meaning of social structure in some detail by considering two different typologies of structure, which reflect the different meanings assigned to social structure in the different ontological traditions which contend over its explanation. In this introductory chapter, rather than trying to seek common ground between these definitions, or to gloss over the differences between them, it seems more appropriate to offer an approach that reflects my own position.

In this approach, social structures are emergent entities in the same sense that people or other natural objects are – they are wholes made up of parts, and possessing properties that are distinct from those of their parts.[4] In the case of social structures, those parts are primarily people, although in theory at least they could also have other parts that are not people. The most obvious case of a social structure, on this definition, is a group of people organised by definite inter-relationships of some type – such as formal organisations of all kinds, communities, families, and cultural groupings. This is the type of structure that will be considered most thoroughly in this thesis, although as pointed out earlier there are potentially other types too.

Similarly, there are a number of different definitions of agency, and we must distinguish at the outset between two groups of these – the concepts of *political agency* and *individual agency*. Political agency is the possession of "the power to bring about effective change in collective life" (Coole, 2005). Political agency, however, may potentially be exercised by things other than individual human beings, whereas this thesis is concerned with agency in its second sense: the specific powers of human

---

[4] The word *structure* is somewhat ambiguous, in the sense that while we sometimes say that entities *are* structures, we may also say that entities *have* structure. Collier calls the latter 'structure' and the former 'structurata', which is perhaps clearer than my usage, but I believe we can use the term structure for both while avoiding ambiguity by taking care to make the meaning clear from the context (Collier, 1989, p. 85).

individuals. Individual agency can be defined as "the power of actors to operate independently of the determining constraints of social structure. The term is intended to convey the volitional, purposive nature of human activity" (Jary and Jary, 2000,p. 9). The study of agency is therefore concerned with whether and how human individuals can be causally effective; in other words with the theory of human action.

Perhaps the most contentious question in this field, however, is whether *social structures* can be causally effective. This is the core problem of structure and agency: the question of whether, and how, social entities like organisations and institutions can be causally effective in their own right, as opposed to merely epiphenomena of the behaviour of the human individuals who are their parts. Without an adequate answer to this question, it is impossible to justify treating social collectivities as 'actors' in social explanations. And the presumption that there are collective social actors (even if only such mundane ones as 'the state' and 'corporations'), although commonly denied, is central to causal explanations throughout the social sciences. Any account which can demonstrate a valid basis for the concept of social structure, then, has a key role to play in social theory - and all the more so if it enables us to distinguish between circumstances in which it is valid to apply the notion of structure and those in which it is not.

In applying the theory of emergence to this question, this thesis will offer an argument for treating at least some social structures as causally effective in their own right, with powers that are distinct from the causal effects of human individuals, while recognising the contributory role that human individuals make to the functioning of social structures. In a parallel argument, it will develop a theory of action that shows how human individuals themselves can be causally effective in their own right, with powers that are distinct from the causal effects of both their biological parts and their social context, while recognizing the contributory roles of both biological parts and social context to the causation of human behaviour.

This solution is therefore distinct from methodological individualist positions, which deny causal effectiveness to social structures, and from methodological collectivist positions, which deny causal effectiveness (at least as regards the causation of social facts) to human individuals. It is also distinct from 'central conflationist' positions like that of Giddens, which seeks to bridge these other two positions by treating structure and agency as ontologically inseparable. Although this thesis follows Archer in rejecting such attempts to transcend the earlier two traditions at the level of

ontology, it recognises that in starting to theorise the interplay between structure and agency they have created materials that may be useful at the level of social theory. As Mouzelis has argued,

> the 'transcendence' strategies of Giddens and Bourdieu have, for a variety of reasons, proved unsuccessful. At the same time the concepts they have put forward (duality of structure, habitus) do provide a foundation for a conceptual restructuring that leads not to another type of transcendence, but to a closer *rapprochement* between the objectivist and subjectivist camps (Mouzelis, 2000, p.747).

Although I will argue that Giddens' "duality of structure" is irredeemably conflationist at the ontological level, it is accompanied by attempts to theorise the relation between structure and agency that could perhaps be lifted out from this conflationist ontology and reused. I am more confident of the argument that Bourdieu's habitus can be reused independently of his ontology, and chapter six will outline a way of combining Bourdieu's habitus with Archer's stress on reflexivity as part of an emergentist theory of action that helps to clarify the relations between structure and agency. More generally, this argument implies that an emergentist ontology may be consistent with a variety of theoretical arguments initially advanced in the context of other ontological positions. This brings us to the relationship of theory to metatheory, and hence to the methodological themes of this thesis.

## Methodological themes

This section positions the arguments of this thesis in relation to a variety of methodological and philosophical issues. The first part of it discusses the epistemological context through an examination of the distinction between theory and metatheory, and of how the arguments of this thesis relate to that distinction. The second stresses the important role of causality in this thesis, and the implications for the question of naturalism. The third positions this work with respect to the contemporary school of thought known as critical realism.

### *Theory and metatheory*

This thesis contains no primary empirical work. In the terms that are usually applied to doctoral work, it would therefore be described as a theoretical thesis. However, I would like to position it with reference to a more differentiated model of enquiry. This model distinguishes not only between theory and its application, but also between domain-specific or substantive theory, and metatheory: theory about theory. Merton's reference group theory would be a good example of the former – it is a

sociological theory about the behaviour of human individuals and is thus specific to the social domain (Merton, 1968). Metatheory, on the other hand, is concerned with the conditions of possibility of theory itself, and thus applies more generally across many or all of the more specific domains of theory – to theories about physics, biology, and literature, for example, as well as to social theory. Typically such metatheory falls within the scope of philosophy. It would include, for example, generalised work on ontology and epistemology. However, domain-specific ontologies which identify the sorts of entities that populate a given domain (what Benton and Craib call *regional ontologies*, and what Bhaskar calls *scientific ontologies*) would fall into the domain theory category and not the metatheory category (Benton and Craib, 2001, p. 5; Bhaskar, 1986, p. 36).

We can illustrate the relation between these categories of work in a simple diagram:
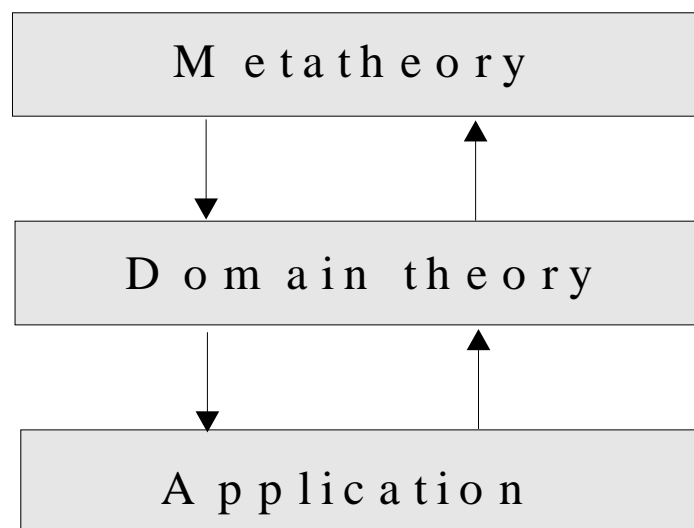


**Figure 1.1 – The role of metatheory[5]**

In these terms, this thesis is concerned both with metatheory and with the domain theory of the social world. It is also profoundly concerned with the relation between the two. There is a strong analogy between this relation, on the one hand, and that between domain theory and its application, on the other. Doctoral theses are usually

---

[5] Although I have transferred them here to a different area of thought, the graphical and conceptual structure of this model are drawn from the work on information systems strategy of Ian Page, as adapted for Dixons Stores Group by Mark McCormack. Since drafting this section and diagram, I have also discovered a similar diagram applied to the relationship between different levels of ontology by Rob Stones, although he conceives of this relationship somewhat differently (Stones, 2005, p. 77).

focussed on this latter relationship, either seeking to test some piece of domain theory through its empirical application (which is represented by the downward arrow between the two corresponding boxes in Figure 1.1), or to derive some piece of domain theory as a result of conducting and interpreting an empirical study (represented by the upward arrow between these two boxes). It is through the gradual accumulation of such results that domain theories acquire (or lose) our confidence – as they are confirmed or disconfirmed by a series of attempts to connect them to the empirical world and vice-versa. This model privileges neither theory nor its application, but sees the two as mutually interdependent. Any development at either level may have implications for the adjacent level, and testing the validity of these implications is part of the process of validating the original development.

Precisely the same logic, I suggest, applies to the relation between domain theory and metatheory. Generalised ontological and epistemological work is tested by the production of domain theory that conforms (or fails to conform) to it – the downward arrow. However, we can equally well take existing domain theory and assess whether our metatheory is consistent with it, or even derive metatheory from the analysis of successful domain theories – the upward arrow. The validity of the metatheory is thus confirmed or disconfirmed, not only by its own internal coherence, but also by its consistency with good domain theory. And the 'goodness' of domain theory is itself tested, of course, by the process described in the previous paragraph. The epistemology of metatheory is thus deeply implicated with the epistemology of theory itself. Once again, the model privileges neither layer. On this view, then, philosophy is neither underlabourer nor master builder to the social sciences, but both are mutually interdependent.[6] To continue the building-trade analogy, metatheory, domain theory, and application are like different trades working in a co-operative partnership.

In these terms, the theory of emergence developed in this thesis, being a claim about ontology in general, is a piece of metatheory, whereas the problem of social structure and human agency is a matter of domain theory. In applying the theory of emergence to the problem of structure and agency, then, I am both validating the metatheoretical claim by testing whether it yields useful domain theory, and also seeking to develop domain theory that is valuable in its own right, and capable of validation by application to real problems. This thesis does not go on to address the

---

[6] The building-trade analogies come from Locke via Bhaskar (Bhaskar, 1998b, pp. 16-17) (Benton and Craib, 2001, p.1).

lower level of Figure 1.1 – the link to the empirical – in any depth. Yet we cannot neglect the empirical entirely, and so I have aimed to follow C Wright Mills' dictum "never write more than three pages without at least having in mind a solid example" (Mills, 1970, p. 246). Thus the thesis focuses on the relation between domain theory and metatheory while keeping in view the application of theory through the use of illustrative examples.

There is a clear parallel here with the more usual doctoral approach, which focuses on the relation between theory and its application while at most acknowledging the role of metatheory in a methodology chapter. Of course, if the metatheoretical foundations of the social sciences were firmly and uncontroversially established, then the usual approach would be more valuable. But this thesis proceeds from the general recognition that they are not – that the question of structure and agency remains problematic precisely because the ontological foundations of social structure and human agency continue to be disputed. In this context, I believe, developing an improved approach to the relation between metatheory and domain theory is a more urgent and important task than working on the link to the empirical.

### Causality, explanation, and naturalism

One of the metatheoretical assumptions that is fundamental to this thesis is the belief that the events that occur in the social world are caused. Emergence is important because it offers a metatheoretical account of how it is possible for something to have a causal impact; outside the framework of a causal approach to the world the concept of emergence adopted here would have no value.

The belief that social events are caused, however, does not entail that they demonstrate observable empirical regularities. Chapter three will adopt and develop Bhaskar's view that events are co-determined by a variety of interacting causes. This view implies that where the set of interacting causes is complex and varies from case to case, causality does not entail predictability. I will be rejecting the view, therefore, that the unpredictability of social events justifies a belief that they are not causally determined.

What that unpredictability *does* entail is that it may be very hard, and often in practice impossible, to reliably *explain* social events. The claim that these events are causally determined entails that they are *in principle* explainable, but in practice we may often be unable to give more than a very tentative and partial explanation of the set of interacting causes that produces any particular social event. Nevertheless there are

patterns in social events – what Lawson calls *demi-regularities* or *demi-regs* – and these may provide the opportunity to discover some of the causal influences at work in the social world (Lawson, 1997, pp. 204-13). The more such causal influences can be discovered, then the richer and more plausible will be our attempts to provide partial explanations of individual social events.

Perhaps the two foremost objections to the causal perspective on the social world arise from a belief in the free will of human agents, and from the claim that the social world may only be interpreted and not explained. Chapter seven assesses both of these claims and argues that they do not stand in the way of a causal social science. Nor, however, need a commitment to a causal social science lead to a rejection of the need for interpretative work. Like Bhaskar, for example, I accept that social action is indeed inherently meaningful, and hence also accept the consequent need to adopt some of the methods of hermeneutics (Bhaskar, 1998b, p. 38). Like Sayer, however, I do not believe that this prevents us from looking at social life in causal terms: "the subject-object interactions merely become more complicated, and the realist proposition of the intransitivity of social phenomena as objects of social research stands" (Sayer, 1999, pp.33-4). Specifically, I argue that the meaningfulness of social action does not prevent us from identifying social actions clearly (although fallibly), and that it does not render causal explanations redundant or impossible.

Now, in arguing that social events are caused, but that they may also require interpretation, I have already opened up the question of naturalism: to what extent do, or should, or can, the social sciences resemble the natural sciences? And indeed, I have already started to offer the response that they do and should resemble the natural sciences in some respects – the provision of causal explanations – but not in others – the need to interpret human actions. More generally, this sort of combination of naturalistic and anti-naturalistic claims is implicit in the adoption of an emergentist framework.

Thus, for example, the sorts of things that we will find in a regional ontology of the social world, like social organisations, institutions, cultural constructs and human artefacts, are ultimately emergent from the sorts of things we will find in a regional ontology of the natural world, like atoms, molecules, and cells. Hence we can see both as parts of an ontologically unified world in which everything that exists emerges as part of a single hierarchy of structures. At the same time, however, this is a differentiated unity; the sorts of powers and properties that belong to entities at each branch of the hierarchy vary, depending upon the varying properties of their parts, and

the varying ways in which these parts are organised. As a consequence of these differences in the objects of study of different sciences, the methods appropriate to their study also differ (Sayer, 1999, p. 19). As Bhaskar puts it, "the *predicates* that appear in the explanation of social phenomena will be different from those that appear in natural scientific explanations and the *procedures* used to establish them will in certain vital respects be different too (being contingent upon, and determined by, the properties of the objects under study); but the *principles* that govern their production will remain substantially the same" (Bhaskar, 1998b, p. 20).

To put the point in different words, emergentism implies what we might call 'ontological naturalism' – that both the natural and the social sciences are concerned with different parts of the same stratified reality – but not 'methodological naturalism', since the different structures of these different parts can have very different implications for their study.[7]

### *Realism and emancipation*

As should be clear from the sources and arguments cited in the foregoing section, this thesis adopts a critical realist perspective. In particular, it advocates a perspective on emergence and cause that is closely related to that developed by Bhaskar in *A Realist Theory of Science* (Bhaskar, 1978).[8] A number of other critical realist thinkers have also been important influences. Inevitably, given the theme of the thesis, the most significant of these is Margaret Archer, who has advanced the most substantial and impressive attempt to date to provide an emergentist account of structure and agency. As will unfold in the later chapters, I have some disagreements with the details of Archer's argument; nevertheless this work pursues the same objective as her own – the development of a realist and emergentist account of structure and agency – and most of her work on this subject remains consistent with my argument.

---

[7] This, of course, is not a statement that applies uniquely to the difference between the social and the natural sciences. The natural sciences are by no means homogeneous in their methods, for just the same reason as has been identified above: the different natural sciences differ in their objects of study, and these objects differ in their structures and properties. There are as a result many criteria that differentiate between different kinds of science, and on some of these criteria some or all of the social sciences fall into the same group as at least some of the natural sciences. We might, for example, contrast the 'experimental' and the 'historical' sciences, in which case we would find geology and evolutionary biology in the latter group along with the social sciences (Benton, 1985, p. 188; Machlup, 1994, p. 6)

[8] I have also suggested some constructive modifications of Bhaskar's approach to emergence and cause in (Elder-Vass, forthcoming, 2006) and (Elder-Vass, 2005a).

Contemporary critical realism, however, is a somewhat diverse school of thought, and my approach to it is selective and (constructively) critical. While I strongly endorse his critiques of empiricist and postmodernist thought, I am less than convinced by some of the other strands of Bhaskar's own thinking. I am at best ambivalent with regard to his theory of emancipatory critiques, and his advocacy of a dialectical turn in critical realism, and I am highly sceptical of the more recent spiritual turn in his thought.[9]

Despite my concerns about Bhaskar's treatment of emancipation, I remain committed to the need for social theory with an emancipatory intent. This thesis is not directly engaged with the normative element of social theory, but as Bhaskar has argued, there is an important connection between emergence and emancipation:

> It is only if social phenomena are genuinely *emergent* that realist *explanation* in the human sciences are justified; and it is only if these conditions are satisfied that there is any possibility of human self-*emancipation* worthy of the name. But, conversely, emergent phenomena require realist explanations and realist explanations possess emancipatory implications. Emancipation depends upon explanation depends upon emergence (Bhaskar, 1986, pp. 103-4).

I am rather more cautious than Bhaskar about the claim that "realist explanations possess emancipatory implications". For one thing, they may equally well possess the opposite sort. For another, I believe that emancipatory political proposals depend upon combining our understanding of the world with a clearly understood set of values, but that those values cannot be derived rationally and objectively from the facts of the world: they always depend upon our social experience and context (Sayer, 1997).

Yet I do believe that we cannot pursue an emancipatory politics without a good understanding of how the social world does work and how it could work differently. It is only if we can provide causal explanations of the social world that we can attempt to predict the consequences of a possible change. It is only if we are able to predict, at least in broad outlines, these consequences that we can assess whether that change offers progress in a normative sense. And it is only if we can do this that we can honestly advocate it as an emancipatory strategy. In this sense, at least, I endorse the first part of Bhaskar's claim: emancipation depends upon explanation. This thesis, however, is about the second part of Bhaskar's claim: it seeks to demonstrate that in the social world, explanation does indeed depend upon emergence.

---

[9] Critical realists are divided in their response to these turns in Bhaskar's thought. For a useful overview, see (Dean et al., 2005).

# 2      Emergence

Although emergence is an idea with roots reaching back into the nineteenth century, the level of academic interest in it has waxed and waned over the intervening period.[10] Recent years have seen a widespread revival in interest across a broad range of disciplines. Emergence has been endorsed and employed by philosophers (e.g. Searle and Kim), physicists (e.g. Gell-Mann), sociologists (e.g. Archer and Sawyer), biologists (e.g. Kauffman), and information scientists (e.g. Holland), amongst others (Archer, 1995; Gell-Mann, 1995; Holland, 1998; Kauffman, 1995; Kim, 1999; Sawyer, 2005; Searle, 1992). Perhaps the feature of emergence that has contributed most to this revival is its potential to explain how the various sciences and their objects relate to one another without succumbing to the reductionist claim that the success of any one science – and the causal effectiveness of its objects of study – implies the redundancy of others, or the dualistic belief that the objects of one science can be studied in isolation from those of others.  In the social sciences, a number of thinkers have identified emergence as a concept with the potential to reconcile the study of the social with the study of human individuals.

However, many of those who have invoked emergence have done so without careful and thorough analysis of the concept. Indeed, there are substantial divergences of opinion on what emergence means, let alone how it works (Corning, 2002, p. 54). Many philosophers of mind, most notably Jaegwon Kim, have adopted (and criticised) a strong concept of emergence drawn from the work of the early emergentist philosopher C D Broad, which seems to have extremely limited application (Broad, 1925; Kim, 1993). Complexity scientists and critical realists, on the other hand, have mostly adopted (and advocated) a weaker, relational, concept of emergence which I will argue is much more useful.

Before we can apply the concept of emergence to the question of structure and agency, we must first confront and resolve these divergences of opinion on what emergence is and how it works. The objective of this chapter and the next, therefore, is to distinguish these various conceptions of emergence from each other, evaluate their

---

[10] This chapter draws on a number of previous papers (Elder-Vass, 2005a; Elder-Vass, 2005c; Elder-Vass, forthcoming, 2006).

usefulness, and clarify the version of the concept that I will be applying to the social realm in later chapters. In the process, these two chapters will distinguish my account of emergence from others in the literature, and provide the theoretical materials that will be used in the later chapters to build the theory of social emergence.

This chapter will focus on the concept of emergence itself, developing an account of the various factors that must be covered by any theory of emergence, while the next one will complete my account of the general theory of emergence by examining the realist theory of causation in a world characterised by emergence. Both will contribute to the evaluation of the strong and relational versions of the concept, the first by explaining them, and the second by showing how the relational version can address the challenges of dualism and reductionism in the context of a realist understanding of cause.

This chapter will begin by looking briefly at the history of the concept of emergence, then it will engage more substantially with the meaning of the concept, which will lead into a discussion of the differences between the strong and relational versions. It will go on to introduce the concepts of morphostasis and morphogenesis, which are central to the realist understanding of emergence, but also to criticise some of the other claims made about emergence in the realist literature, notably the denial of its compositional basis and the attempt that is sometimes made to substitute relations for entities in a realist ontology. Finally, it will discuss the question of levels of stratification, which is often presented in a confused and confusing way.

## History of the concept

This first section, then, provides a brief overview of the historical development of the concept of emergence, in order to position my own account of emergence in its historical context. Arguably, there is as yet no definitive history of the concept, but there are several sources in the literature that provide more substantial accounts than the one below, and this account draws heavily on these sources: (Blitz, 1992; McLaughlin, 1992; Sawyer, 2005, ch. 3).

### *Origins*

The term 'emergent' was coined in 1875 by G H Lewes, along with the term 'resultant', in a development of Mill's distinction between 'homopathic' and 'heteropathic' laws, which several writers have identified as the root of the modern

concept of emergence (Lewes, 1874-9; Lloyd Morgan, 1923, pp. 2-3; McLaughlin, 1992, pp. 59-65; Mill, 1900, p. 244-245). Despite his use of 'laws' where a contemporary realist would write 'powers', parts of Mill's brief discussion bear some intriguing resemblances to recent realist accounts of emergence and causation:

> the component parts of a vegetable or animal substance do not lose their mechanical and chemical properties as separate agents, when, by a peculiar mode of juxtaposition, they, as an aggregate whole, acquire physiological or vital properties in addition. Those bodies continue, as before, to obey mechanical and chemical laws, in so far as the operation of those laws is not counteracted by the new laws with govern them as organised beings. When, in short, a concurrence of causes takes place which calls into action new laws bearing no analogy to any that we can trace in the separate operation of the causes, the new laws, while they supersede one portion of the previous laws, may co-exist with another portion, and may even compound the effect of those previous laws with their own (Mill, 1900, p. 245)

Sawyer points out that both Mill and Lewes were influenced by Comte, whose case for the new science of sociology rested on a denial of reductionism as applied to the social (Sawyer, 2005, p. 38). In a separate development of Comte's thought, Emile Durkheim developed an emergentist approach to sociology, which as Sawyer has argued, has been widely misunderstood (Sawyer, 2005, p. 100). Durkheim clearly thought in emergentist terms:

> Whenever certain elements combine and thereby produce, by the fact of their combination, new phenomena, it is plain that these new phenomena reside not in the original elements but in the totality formed by their union. The living cell contains nothing but mineral particles, as society contains nothing but individuals. Yet it is patently impossible for the phenomena characteristic of life to reside in the atoms of hydrogen, oxygen, carbon, and nitrogen (Durkheim, 1964 [1901] , p. xlvii )

And he clearly applied this logic to the social world: "We assert not that social facts are material things but that they are things by the same right as material things, although they differ from them in type" (Durkheim, 1964 [1901] , p. xliii ). Given that what Durkheim meant by 'social facts' has a great deal in common with what we mean today by 'social structures', this may be the first statement of the core argument of this thesis.

Durkheim, however, has largely been ignored by the philosophers at the centre of the recent resurgence of interest in emergence; they have been more influenced by the early twentieth century British school of emergentists, most notably C D Broad, C Lloyd Morgan, and Samuel Alexander. These thinkers turned to emergentism in an attempt to find a middle way between the doctrines of vitalism and mechanism in explaining the existence of life (Broad, 1925, ch. II; Stephan, 1992, p. 25). Vitalism asserted that physical bodies were alive because the physical elements were combined

with a non-physical vital spirit, commonly called entelechy, and it was the presence of this mysterious entelechy that accounted for life. Mechanism denied the existence of any such vital spirit, and insisted instead that life was nothing more than a consequence of the set of physical parts that made up a living body – a consequence that would ultimately be explainable completely in terms of lower-level laws. Vitalism, then, was an extreme ontological dualism, whereas mechanism was a species of what we would today call reductionism.

### *Revivals*

Recent philosophers of mind have investigated this thinking in a parallel attempt to find a middle way between reductionistic physicalism and Cartesian dualism in the study of the relationship between the body and the mind. They have tended to take Broad's 'strong' definition of emergence as their standard. This is unfortunate because Broad has an unusually restrictive conception of emergence, as should become clear below, where I examine Broad's definition of emergence in more detail (p. 33). Indeed emergentism fell into disrepute soon after Broad had published his major work on the subject, as new developments in physics showed that the examples upon which he had rested his case were not in fact emergent in the strong sense he advocated (Broad, 1925; McLaughlin, 1992).

Since then there have been a number of cycles of revival and neglect of the concept, which are well documented by Blitz (1992). The dating of these cycles is a little arbitrary, but the most recent cycle can perhaps be traced to work in the 1970s which has directly influenced today's emergentist thinkers (Blitz, 1992, ch. 13). Most pertinently to this thesis, Bhaskar's *A Realist Theory of Science* was founded on an emergentist approach to causal powers which he drew in part from the work of Harré (Bhaskar, 1978; Harré and Madden, 1975), while in parallel the relevance of emergence to the mind-body problem started to be investigated in neuroscience and the philosophy of mind (Sperry, 1969). More recently, emergence has become an important element in complexity theory, although here the intellectual influences can perhaps be traced back to von Bertalanffy and his general systems theory (Bertalanffy, 1971, pp. 53-4). Although there has no doubt been some cross-fertilisation, these three research programmes have largely proceeded in parallel with each other, with the result that there are different (although overlapping) conceptions of emergence and its foundations in each of them. One of the objectives of this thesis is to synthesise the most valuable

lessons from each of these research programmes into a single coherent view of emergence.

We must begin this task by examining the basic characteristics of the concept of emergence.

## What is emergence?

### *A basic definition*

Put at its simplest, emergence is operating when a whole has properties or powers that are not possessed by its parts.

However, a number of clarifications and expansions of this claim are required. First, we must distinguish between synchronic and diachronic conceptions of emergence. As I pointed out in chapter one, lay usages of *emergence* generally refer to diachronic emergence, which denotes the first appearance or initial development of some new phenomenon. While this is important, and is certainly complementary to the synchronic conception, I shall generally *not* use the word *emergence* in this sense. Instead I shall discuss this aspect under the label *morphogenesis* (see p. 46 below) and focus on the synchronic sense of emergence, which is concerned with the relationship between the properties and powers of a whole and its parts at any single instant in time.

Secondly, we must clarify what is meant by *wholes* and *parts*. Both wholes and parts in the basic definition above are *entities*, and the terms *whole* and *part* therefore describe roles played by particular entities in particular cases (an entity that is a whole in one context can be a part in another). Entities are to be identified with objects or things, although this does not mean that they are necessarily *material* things – examples include atoms, molecules, cells, trees, human individuals, theories, business corporations, and armies. Any entity (except perhaps the most fundamental material particles, if there are such things) consists of a set of parts that is in some way structured, such that the relations between the parts are more than merely aggregative. There may therefore be collections of parts that do not form entities, such as relatively arbitrary constructs like "all the rice in China" (Collier, 1989, p. 193). I follow Laszlo in calling such unstructured collections of parts "heaps" (Laszlo, 1972, p. 28). Now, strictly speaking, all collections of parts, however arbitrary, have relations between them – all material entities, for example, exercise a gravitational force on each other – so to be strictly accurate, we must define heaps as collections of parts that lack

*significant* structure; I shall return to the meaning of *significant* in this context below. Furthermore, an entity must have the quality of persistence, in the sense that it must sustain its existence over a significant period or time. To summarise, an entity may be defined as *a persistent whole formed from a set of parts that is significantly structured by the relations between these parts*.

Thirdly, we must clarify what is meant by a *property* or *power*. A property is some intrinsic aspect of an entity that can have a causal impact on the world. I use *intrinsic* in order to exclude purely formal relations with other entities, such as 'larger than *x*', from the definition of properties.[11] *Properties* and *powers* may therefore be regarded as synonyms.[12]

Emergence occurs when a whole possesses one or more *emergent properties*. An emergent property is one that is not possessed by any of the parts individually or when they are aggregated without a structuring set of relations between them. Perhaps the commonest illustration of emergence in the literature is the example of water used in chapter one, which has been used to illustrate this point as far back as John Stuart Mill (Mill, 1900, p. 243). The properties of water are clearly very different from those of its components oxygen and hydrogen when these are not combined with each other in the specific form that constitutes water. One cannot, for example, "put out a fire with oxygen and hydrogen" (Mihata, 1997, p. 31). Hence water has emergent properties. Another illustration is provided by colour: "The collective structure of bulk matter reflects light at certain preferred wavelengths; those determine the color. Color is an emergent phenomenon; it only makes sense for bulk matter" (Cohen and Stewart, 1995, p. 232). Molecules – the parts of bulk matter – simply do not have the property of colour; hence this property emerges from their structured combination into larger wholes. These particular examples rest on the ability of the entities concerned to interact with other external entities (burning material and light), or in other words, these powers can only be exercised when and because there are other entities with corresponding liabilities, but this is not the case for all powers.[13] Stars, for example, have a power to emit radiation over a variety of wavelengths, including visible light, and this power does not significantly depend upon the liabilities of other entities (except their parts).

---

[11] See Sayer on 'formal relations' (Sayer, 1992, p. 88).

[12] I am adopting a causal powers approach to causation here, as opposed to a covering law approach. This is a question that will be discussed more explicitly in the next chapter.

[13] Liabilities are described in (Harré and Madden, 1975, p. 89) and below in chapter 3.

### *Resultant properties*

Emergent properties may be contrasted with *resultant properties* – these are properties of wholes that *are* possessed by its parts in isolation, or in an unstructured aggregation. The classic example of a resultant property is mass – the mass of a molecule, for example, is the sum of the mass of its constituent atoms.[14] Similarly, heaps (and entities) may have attributes like group size and average height that are not possessed by their component parts, but these are all aggregative or resultant as opposed to emergent properties, since they result from the simple addition of the properties of the parts. A property that is resultant at one level may be (and perhaps must be) emergent at a lower level.

As Mill pointed out (though in different terms), it is entirely possible – indeed it is normal – for entities to have a mixture of emergent and resultant properties. In addition to their (emergent) ability to douse flames, for example, water molecules have the resultant property of mass, which is a simple addition of the masses of their component atoms of hydrogen and oxygen. More generally, any whole that is composed of parts must have certain statistical properties arising from the grouping of its components, such as various measures of size. All wholes, therefore, whether emergent or merely 'heaps' possess resultant properties (except perhaps fundamental particles, if such things exist), and emergent entities are distinguished from heaps by having emergent properties too.

It is less obvious whether it is possible for an entity to possess *only* resultant properties. Ultimately this is a definitional question, since the answer rests upon whether we choose to define 'entity' to include a persistent collection of parts that does not have emergent properties as a whole – say, for example, a fence with a post box attached to it. This brings us back to the difference between a significant and an insignificant structural relation which came up in the earlier discussion of entities and heaps. In order to distinguish between the two, we had to define entities as possessing *significant* relations between the parts. If we make it the criterion of significance for a relation that it leads to the whole possessing a causal power not possessed by the parts without it, then the fence-and-post-box is a heap and not an entity. In this thesis I accept this strategy, and so it now follows by definition that all entities have emergent properties or powers.

---

[14] For a particularly thorough account of what it means for a property to be resultant, or aggregative, see (Wimsatt, 2000).

### *Relations and levels*

This does not mean, however, that entities can be dispensed with in favour of an account in terms of properties or powers alone. Properties are not free-floating phenomena; they always occur as the effects of a particular configuration of lower-level parts.[15] Mass, for example, can not exist except as a property of a particular thing. Now, admittedly, if a whole is an organised set of parts, and each part is itself an organised set of parts, then unless there is some lowest-level thing which is not *just* organisation, then any entity can ultimately be decomposed into a set of relations between relations. Nevertheless, an entity remains a *real* and persistent set of relations between relations, with causal powers that are irreducible to any of its lower-level decompositions (see the discussion of reduction in chapter three). This real set of relations is different from the properties that depend upon it, and any attempt to eliminate the entities from this picture obscures the nature of emergence.

It is often useful in discussing emergence, as for example in the previous paragraph, to use the concept of higher and lower levels, and I shall follow convention here in treating wholes as higher level entities than their parts or components, and vice-versa. The term *higher level entity*, however, is more general than the term *whole*, since as well as being at a higher level than its own parts, any given whole is at a  higher level than all other entities of a similar type to its parts or their parts, and so on recursively. Thus, a water molecule is a higher level entity than an atom of any kind, and also a higher level entity than an electron, a proton, a quark, and so on. Similarly, any given part will be considered to be at a lower level than all entities of a similar type to any whole(s) it belongs to, and so on. In places Roy Bhaskar seems to use higher and lower level in the opposite sense to that used here, (e.g. Bhaskar, 1978, ch. 3) but otherwise the usage I adopt here seems to be universal. The 'level' metaphor, however, can sometimes be less than helpful. As Andrew Collier points out, a 'tree' metaphor may be more useful, although even this oversimplifies the structure of emergence (Collier, 1989, p. 45). However, there are further ramifications of the question of levels, and I will return to these later in the chapter (p. 54).

First, however, we must discuss the differences between two distinct variants of the concept of emergence. Both variants are strongly represented in the literature, and

---

[15] Bhaskar clearly takes this view, e.g. 'Most things are complex objects, in virtue of which they possess an ensemble of tendencies, liabilities and powers,' in (Bhaskar, 1978, p. 51)

some of the confusion around the concept of emergence arises from a failure to distinguish between the two.

**Strong emergence**

I have already referred in chapter one to the need to distinguish between the synchronic and temporal senses of the concept of emergence; but there are also many variations on the synchronic version of the concept. In discussing the work of Sawyer in chapter five, for example, I will distinguish a functionalist variant; but the two variants that have dominated the literature are what I shall call *strong emergence* and *relational emergence*. Since the term *emergence* is often used without qualifying which of these senses is intended, misunderstandings can arise between thinkers who are using it in different senses without being aware that this is so. There is thus some value simply in clarifying the distinction. More importantly, though, I will argue that instances of strong emergence are rare, if they exist at all, and that the value of emergence rests upon adopting a relational sense of the concept.

*Broad's definition of emergence*

This section will begin the argument by describing the strong version. This is perhaps the original variant of the concept, most clearly described in the work of C. D. Broad:

> Put in abstract terms the emergent theory asserts that there are certain wholes, composed (say) of constituents A, B, and C in a relation R to each other; that all wholes composed of constituents of the same kind as A, B, and C in relations of the same kind as R have certain characteristic properties; that A, B, and C are capable of occurring in other kinds of complex where the relation is not of the same kind as R; and that the characteristic properties of the whole R(A,B,C) cannot, even in theory, be deduced from the most complete knowledge of the properties of A, B, and C in isolation or in other wholes which are not of the form R(A,B,C). The mechanistic theory rejects the last clause of this assertion (Broad, 1925, p. 61).

For Broad, then, a property of a whole is emergent if it cannot be explained from the properties of lower-level parts *and their substantial relations with each other*.[16] In this sense, a property can *only* be emergent if there is *no* way of providing an explanation of how it comes about as a result of the interaction of lower level entities and properties. Any property that was emergent in such a sense (if one existed) would

---

[16] I owe the term 'substantial relations' to Sayer (1992, p. 88); I use it here to exclude comparative relations between the parts – for A to be bigger than B, for example, plays no direct part in constituting them into a particular kind of whole.

not just be autonomous of lower levels; it is of the essence of this concept of emergence that *no* scientific explanation of the property would be possible. As Kim has pointed out, Broad and other early emergentists saw emergent properties as "not *explainable*, or *reductively explainable*, on the basis of their 'basal conditions', the lower-level conditions out of which they emerge" (Kim, 1999, p. 6). Horgan, similarly, writes that "there is no explanation for why emergent properties come into being, or why they generate the specific non-physical forces they do. These facts are metaphysically and scientifically basic… they are unexplained explainers" (Horgan, 2002, pp. 115-6). This variant may therefore be labelled *strong emergence* because, unlike the relational variant, it denies any possibility of explaining how any given case of emergence actually works. [17]

Broad illustrated his concept of emergence with examples drawn from chemistry, which he saw as irreducible to physics. Here he makes the point using the classic example of water:

> We will now pass to the case of chemical composition. Oxygen has certain properties and Hydrogen has certain other properties. They combine to form water, and the proportions in which they do this are fixed. Nothing that we know about Oxygen by itself or in its combinations with anything but Hydrogen would give us the least reason to suppose that it would combine with Hydrogen at all. Nothing that we know about Hydrogen by itself or in its combinations with anything but Oxygen would give us the least reason to expect that it would combine with Oxygen at all. And most of the chemical and physical properties of water have no known connexion, either quantitative or qualitative, with those of Oxygen or Hydrogen. Here we have a clear instance of a case where, so far as we can tell, the properties of a whole composed of two constituents could not have been predicted from a knowledge of the properties of these two constituents taken separately, or from this combined with a knowledge of the properties of other wholes which contain these constituents (Broad, 1925, pp. 62-3).

Any supposed example of strong emergence, however, is always vulnerable to the possibility that at some future time a reductive explanation might be found for it. At such a time, the claim to strong emergence would evaporate. Unfortunately for Broad's argument, this is exactly what happened to his examples from chemistry only a few years after the publication of the passage quoted here. As we have seen, the emergentist philosophy as a whole lost credibility as a result, and despite occasional attempts at revival, remained rather marginal until the 1980s (McLaughlin, 1992, pp. 54-5, 90).

---

[17] This usage of *strong emergence* is drawn from (Boogerd et al., 2005).

*Viability*

Perhaps the leading figure in the revival of interest in emergence amongst philosophers of mind has been Jaegwon Kim. Kim and the many philosophers who have been influenced by his work continue to define emergence in Broad's terms – as strong emergence. However, although he finds strong emergence to be a logically coherent concept, Kim is sceptical of the claim of strong emergence to provide a viable alternative to dualism or reductionism (Kim, 1993).

There are two key reasons for this scepticism. First, it seems unlikely that any properties at all genuinely are strongly emergent (Kim, 1999, p. 18). The whole tendency of modern science has been to provide more and more explanations of how higher level phenomena can be explained in lower level terms. Even where such explanations do not currently exist, it tends to be assumed within the scientific world view that this is due to gaps in our knowledge rather than to the inherent unexplainability of the phenomena concerned, and thus that suitable explanations will be found at some point in the future. Kim tentatively suggests one group of properties that may be unexplainable and hence strongly emergent: qualia (Kim, 1999, pp. 9, 18). But qualia are highly controversial properties in their own right, and it is not at all clear why they should not be explainable.

A contrary view is offered by Nancy Cartwright and John Dupré, who have challenged the presumption that all events are caused and thus explainable (Cartwright, 1983, p. 49; Cartwright, 1999, p. 32; Dupre, 2001, pp. 157-69). Their argument implies that events may be partially caused and partially uncaused, and some support for this is provided by the randomness (within strict parameters) of quantum events. This implies the possibility that there could be regularly appearing emergent properties that can not be explained strictly in terms of the effects of their parts and their relations to each other. In a sense quantum properties exhibit this feature: although there is a lawful regularity between causal factors and the probabilities of quantum events, the determination of the apparently random element of any given quantum event does not seem fully causally explainable. Thus Cartwright's denial of the ubiquity of cause could constitute an argument for the possibility of strongly emergent properties. This seems plausible when applied to random events, but it is more difficult to accept when there appears to be a regular empirical relationship between the properties of a whole and those of its parts and their relations to each other.

The second reason for Kim's scepticism is that, even if some properties were found that *were* strongly emergent, it is not clear that strong emergence constitutes a middle way between dualism and reductionism. Broad himself insisted that strong emergence could be an entirely natural phenomenon (Broad, 1925, pp. 67-8). However, natural or not, the existence of strongly emergent properties would seem to represent an ontological dualism. Strongly emergent properties can only exist when the relevant lower-level parts are present (the A, B, C of Broad's definition) in the relevant relations to each other (the R), but Broad denies that they are explainable by the interaction of the parts and their relations. It this is an epistemological claim, it is inevitably provisional and constantly awaits refutation by the progress of science. But if it is an ontological claim, as seems to be intended, then the assertion that something can exist without being caused in some way by the presence of and relations between its parts seems to imply that there is a realm of nature that is as distinct from its physical base as the Cartesian soul is from its body. Kim therefore suggests that strong emergence must inevitability collapse either into dualism – if there really are strongly emergent properties – or into reductionism – if and when those properties come to be explained (Kim, 1999, p. 5).

Cartwright's argument, however, also provides a potential response to this critique. Dualism, as much as reductionism, assumes a world in which everything is fully caused: this is why it is necessary for full-blown dualists to posit some extra force beyond those of the parts of a whole in order to account for higher level properties. Vitalism, for example, needs entelechy because it *both* denies that life can be explained by the effect of physical parts *and* feels the need to provide some alternative causal explanation for life. If the latter need is dispensed with, then so is the need for entelechy; if higher level properties are simply unexplainable because they are uncaused, there is no need for some extra causal factor beyond the impact of the physical parts. The possibility of strong emergence providing a middle route between reductionism and dualism in some particular cases cannot therefore be entirely ruled out. However, strong emergence would seem to apply in practice to relatively few part-whole relations (if any), and hence it cannot provide a *general* defence against reductionism.

### *Alternatives*

A considerable body of work in recent philosophy of mind has sought to find alternatives to strong emergence. Kim himself has paid a great deal of attention to the much weaker concept of supervenience. But if strong emergence is too strong, then

supervenience is widely felt to be too weak to provide a response to reductionism (Kim, 2002, p. 59; Savellos and Yalçin, 1995, p. 6). Kim himself seems to have ended up in a position that has a great deal in common with what I will discuss below under the heading of *relational emergence*, although he continues to think of 'emergence' as meaning strong emergence, and so does not seem to regard his own position as an instance of emergence at all (Kim, 1998, pp. 117-8). Meanwhile, John Searle has developed an approach to the philosophy of mind that explicitly advocates a relational variant of emergence (Searle, 1992; Searle, 1997).

There are also those, mostly it would seem philosophers of science, who have continued to pursue alternatives that fit the requirements of Broad's definition of emergence, and from time to time papers appear claiming that certain types of case are indeed strongly emergent (e.g. Boogerd et al., 2005; Newman, 1996). There are as yet, however, no well established examples.

Strong emergence, then, continues to be the prevalent variant of the concept of emergence in the philosophy of mind, largely because of the historical reference back to Broad and the early emergentists. But there are few cases of anyone outside the philosophy of mind adopting this definition, and it seems to have extremely limited potential as an alternative to reductionism and dualism.


## Relational emergence

Fortunately, there is a more promising variant of emergence. This is what I will call *relational emergence*. This is the predominant variant of the concept amongst those thinkers in the natural sciences – and most particularly amongst thinkers in the complexity tradition – who have invoked the concept of emergence. In this section I will describe the relational variety, but the response to reductionism it implies will be deferred until the next chapter.

### *The role of relations*

The relational approach to emergence argues that emergent properties arise *because of* the particular relationships that hold between the parts in a particular kind of whole. In other words, the source of emergence is the organisation of the parts: the maintenance of a stable set of substantial relations between the parts that constitute them into a particular kind of whole. Higher level entities are not just a simple aggregation of their component parts. A soup composed of the set of molecules that

previously made up a pile of vegetables, for example, is still a soup and not a pile of vegetables. Instead, the composition of higher-level entities is a *structured* one, in which particular characteristic relations must hold between the lower-level entities for the higher-level entity to exist. These particular molecules, for example, must be present in a particular spatial arrangement to constitute carrots, potatoes, and the like, and it is this set of *relations* between the components of a higher level entity that makes them more than the sum of the parts – that constitutes the emergent higher level entity from the lower level components.

The critical role of organisation as the source of emergent properties has been identified by authors in all the well-developed literatures on emergence. To illustrate the point, let me cite just a few of the many authors who have advanced this argument. In philosophy, for example, Emmeche and his colleagues have written: "What is 'more' about the whole is a specific series of spatial and morphological relationships between the parts" (Emmeche et al., 1997, p. 106). The neuroscientist Roger Sperry has argued "The emergent properties of the entirety and the laws for its causal interactions are determined by the spacing and timing of the parts as well as by the properties of the parts themselves" (Sperry, 1986, p. 266). In linking sociology to complexity theory, Smith has written "What defines such an emergent phenomenon is that it cannot be understood merely as an aggregative product of the entities or parts of the system but arises though their organization. Interaction often yields structures, forms that cannot be understood through simple linear decomposition of a system into its interacting parts" (Smith, 1997, p. 55). And complexity theorists like Holland have stressed this same point: "Emergence is above all a produce of coupled, context-dependent interactions. Technically these interactions, and the resulting system, are *nonlinear*" (Cilliers, 1998, p. 43; Holland, 1998, pp. 121-2). The crucial role played by the *relations between* the parts has thus been recognised in all the major literatures on emergence.

Relational approaches to emergence argue not only that higher level properties are co-occurent with particular organisations of parts, but also that these higher level properties can be *explained* by such organisation. As von Bertalanffy put it:

> The meaning of the somewhat mystical expression 'the whole is more than the sum of the parts' is simply that constitutive characteristics are not explainable from the characteristics of isolated parts. The characteristics of the complex, therefore, compared to those of the elements, appear as 'new' or 'emergent'. If, however, we know the total of parts contained in a system and the relations between them, the behaviour of a system may be derived from the behaviour of the parts (Bertalanffy, 1971, p. 54).

The relational argument for emergence, then, is that it is because a higher-level entity is composed of a *particular stable organisation* or configuration of lower-level entities that it may be able to exert causal influence in its own right. This does *not* mean that the emergence of a higher level property is *caused* by its parts or by their powers in the usual sense of the term; emergence is a synchronic relationship between a whole and its parts, whereas cause is a diachronic relation in which the powers of a group of entities at one moment *causally* determine the events which follow at the next. The point of emergence is that it is the way that a set of entities is related to each other at a *given* point of time that determines the joint effect they have on the world at that moment. Emergence, then, is a synchronic relation amongst the parts of an entity that gives the entity as a whole the ability to have a particular (diachronic) causal impact. The relation between a whole and its parts is thus a relation of composition, and not of causation.

### *Mechanisms and reduction*

Although this relation is not causal, we can nevertheless often explain how the relation between the parts produces the overall effect. Indeed, this is the *mechanism* that sustains the emergent property concerned. Returning to the case of water, for example, it is possible to explain why water has the property of being liquid at certain temperatures, why it has the property of being solid (ice) at others, and why its solid form is less dense than its liquid form (unlike most other materials), purely as a result of the properties of hydrogen and oxygen atoms and the sorts of bonds that form between them (see, for example, Gribbin and Gribbin, 1999, pp. 84-7).

From the account so far it might seem that relational emergence is thoroughly reductionist, given its acceptance of the possibility that higher level properties can be explained in terms of lower-level entities, their properties, and the relations between them. Stephan, for example, has somewhat pejoratively described this variant of emergence as *weak emergence* and argued that it is "compatible with reductionistic approaches without further ado" (Stephan, 2002, p. 79). As I shall show in the next chapter, however, this is far from being the case. There are certainly senses in which relational emergence can be described as reductionist, but there are many senses of reductionism, and I shall argue that relational emergent properties cannot be reduced in the sense that matters.

To summarise, it is the fact that a higher-level entity is composed of a *particular stable organisation* of lower-level entities that gives it the possibility of exerting causal influence in its own right. Only when this particular kind of parts is present in this particular set of relations to each other does the higher level entity exist, and only when this particular kind of parts is present in this particular set of relations to each other do they have the properties and causal powers that are characteristic of the higher-level entity.

Not all possible arrangements of parts, however, will necessarily constitute wholes with emergent properties. A higher-level whole is only emergent when it just so happens that, when a set of lower level entities is so organised as to create it, the resulting entity has a consistent causal impact that is not a simple summation of the impacts of its components. Now, the particular causal influences that any particular entity type may exert, and the way in which the presence of its parts in the required relations produce these higher level effects, are a matter for the particular science of the case – we cannot go any further at the philosophical level in explaining why particular cases of emergence work.

Having said this, we may be able to make some generalisations across broad types of emergence mechanisms. One interesting question is whether there are 'intrinsic' and 'extrinsic' types of emergence mechanism. The explanation so far has assumed that the behaviour of the higher level entity is the product of interactions between its lower level components. Now, in practice, the behaviour of all entities is not purely internally driven but arises from interactions with external entities. This does not, however, pose a problem for the theory of emergence, since it is *the way that the entity interacts with other entities* that is the product of its internal structure. Extrinsic emergence mechanisms, then, would seem to be ruled out by definition.[18]

Granted, an entity's internal structure will have arisen over time as a result of interactions with external entities. But the question of how an entity comes to exist in any particular form is a very different matter from the question of what emergent properties it possesses and how they can be explained in synchronic terms. Nevertheless, it is an extremely important matter, and it will be the subject of the following sections, which introduce the concepts of morphostasis and morphogenesis. Both of these important terms were coined by Walter Buckley (Buckley, 1967, pp. 58-

---

[18] Collier has sometimes advocated a non-compositional account of emergence; this is criticised below (p. 47).

9), and have been introduced to the realist literature on emergence and further developed by Margaret Archer (Archer, 1982; Archer, 1995).

**Morphogenesis and morphostasis**

So far, the argument has more or less taken the existence of wholes for granted, but in practice all wholes depend for their continued existence on the continuing maintenance of their parts in the particular set of relations that is required to constitute the whole from them. We can express this by saying that each type of entity has its own characteristic set of *compositional consistency requirements*. For example, for a string of DNA to exist, it must be composed only of certain sorts of molecules, arranged in a certain characteristic pattern. Although there are a huge number of possible variations of the arrangement of these molecules within this characteristic pattern, there are certain limits on the form that this pattern may take, and when a string of molecules falls outside these limits, it is not DNA and does not possess the characteristic set of properties of DNA. Every different type of entity has a different set of such compositional consistency requirements.

As a consistency requirement, composition is a logical rather than a causal relation, hence it is both synchronous and non-determinative – in this argument, we have not explained the existence of the whole, but merely assumed it, and asked what the corollaries of this might be. It is merely the case that the existence of a whole at a given time has as its logical corollary the simultaneous existence of a set of component entities that satisfy its compositional consistency requirements.

*The role of morphostatic causes*

However, there is more to composition than this. Entities and their properties do not simply exist at a given moment in time. They have continuity over periods of time, and explaining this continuity is an essential part of explaining the existence of these entities. The existence of an entity and its properties at any given time requires not just a logical but also a causal explanation. Causal explanation is diachronic rather than synchronic and it is a genuinely determinative relationship, rather than merely a consistency requirement: causal explanations show how a given state of affairs was produced as a consequence of the combination of causal mechanisms with previous states of affairs (cause will be discussed in much more depth in the next chapter). As such, there is no single causal explanation of any particular state of affairs, since this

will inevitably be a consequence, not of a single previous state, but rather of a series of previous states at different points in the past. It is a matter of judgement which previous states of affairs we consider most relevant in any particular case, although it is common to think in terms of the most recent change as the most relevant cause. We might say, for example, that a particular pen exists (or is made of a particular material, or has a particular weight, etc) because of the manufacturing process that was used to make it. This would be a *morphogenetic* causal explanation of the existence and properties of the pen – Buckley defines *morphogenesis* as "those processes which tend to elaborate or change a system's given form, structure or state" (Buckley, 1967, p. 58).

Behind this event, of course, lie a series of others, such as the previous design and manufacture of the machines that made the pen, the bringing of the materials to the factory, and so on, which we tend to ignore in most of our causal explanations. What is more important for the current argument, however, is that we also tend to ignore what happens *after* the most recent relevant change. Yet the existence of the pen at this moment is caused not only by its original manufacture, but also by the set of causes that have kept it in the form of a pen ever since. Hence for every entity that continues to exist for more than an instant, there must be some set of causal factors that maintains its stability. These factors provide a *morphostatic* causal explanation of the existence and properties of the entity concerned – Buckley defines *morphostasis* as "those processes in complex system-environment exchanges that tend to preserve or maintain a system's given form, organization, or state" (Buckley, 1967, p. 58).

As has already been suggested, morphostatic causes need not be purely internal to the entity concerned (i.e. they need not operate purely within and between its parts). Thus, for example, the continuing existence of a planet (certainly 'as a planet', and perhaps even as a continuing entity at all) depends upon the causal influence of the star within whose system it exists. Likewise the continuing existence of an animal depends upon the (internal) conscious activities of the animal, such as undertaking the activities necessary to feed itself, but it is equally dependent upon the continuing existence of a suitable environment – e.g. one with a suitable atmosphere, level of atmospheric pressure, and level of gravitational force. A particularly important special case of external causation in morphostasis is the need, implied by the laws of thermodynamics, for complex systems to draw energy from their environments (Laszlo, 1972, p. 37; Prigogine and Stengers, 1984).

Now, although it is clearly necessary for there to be a morphogenetic explanation of the coming into existence of any given higher level entity, there is a sense in which it is the morphostatic explanations of its continuity of structure that provide the critical basis for emergence. Any number of implausible combinations of lower-level entities may be brought about by a vast range of morphogenetic causes over the course of time, but it is only those combinations that have continuity of structure that persist. Furthermore, it is only those combinations that persist which are likely to have constantly repeated causal effects, and hence provide the empirical material that enables us to hypothesise the existence of an underlying causal mechanism. In particular, it is only those entities that persist that are likely to contribute to further levels of morphogenetic causation which bring about the next higher level of emergence. It is difficult to conceive of a higher level entity whose components do not themselves have a continuing existence of some sort.

It is an entity's morphostatic causes that ensure it continues to meet its compositional consistency requirements; this is simply another way of saying that they keep the higher level entity in continuous existence from moment to moment. Now at any time, it is possible that a more powerful morphogenetic cause may overcome these morphostatic causes for any given entity – such as the effect of heat if I throw my pen into a fire and it then melts and deforms. At this point, the emergence of the higher level entity is dissolved, and any point-in-time consistency requirements for the pen simply lose relevance. It is the contingent ability of morphostatic causes to resist such effects that sustains the existence of higher-level entities and hence any emergent properties they may have.

### *Varieties of morphostasis*

Morphostasis as I have described it in the previous section is an extremely general phenomenon, and of course it takes a different form for every different type of entity. However, there are some general types of morphostatic relationships that we can usefully identify, which help to illuminate some of the challenges involved in using the concept.

There are three typologies of entity structure that I suggest are relevant here. The first typology concerns the interchangeability of the parts. On the one hand, some entities may not be able to survive a change in which one token or instance of a given part replaces another (as when a human body 'rejects' an organ transplant, for example). But on the other, instances of some parts in some wholes may change repeatedly over

the course of their lifespan. The human body provides an example here too, in the form of the constant replacement of skin cells (and indeed many other cells) by other instances of the same type of cell throughout the life of the organism. Another example is human social institutions, say for example a democratic parliament, which remains a parliament despite the regular replacement of many of its members at election time. Holland illustrates the same type with an example from natural science: "the standing wave in front of a rock in a white-water river. The water molecules making up the wave change instant by instant, but the wave persists as long as the rock is there and the water flows" (Holland, 1998, p. 7).

The second typology concerns the *structural range* of any given type of entity. Some entity types may have only one narrowly defined viable form, with the consequence that all instances of the entity are identical (other than their positions in space and time) – such as atoms of a given element (although even here isotopes provide room for variation). These types have what we might call a *point range*. Others may have a more flexible range of viable forms. An oak tree, for example, may be large or small and may take on a broad variety of shapes, although there are certain limitations on how that shape can vary if it is indeed an oak tree. Such types have what we might call an *area range*. And others may arguably have a variety of forms that do not necessarily form a single continuous range – what we might call a *multiple area range*. This latter type has some similarity to the concept of a functional type, which we will come across again when we look at Fodor's and Sawyer's arguments against reductionism in chapter five.

And the third typology relates to the degree of structural stability of the entity. At one extreme of this typology, we have *rigid structures*, exemplified by certain human artefacts such as buildings – in this case, the continued existence of the building is usually implemented by maintaining its parts in a completely fixed set of relationships to each other over long periods of time (although even in this case I am abstracting from such changes as the opening and closing of doors and windows). At a slightly looser level of internal stability we have *cyclical structures* that maintain the same parts in changing but closely constrained sets of relationships. An example here might be an engine, where the same parts move in constrained patterns while still remaining an engine. Then, at the other extreme, we have *changeable structures*, in which the relations between the parts may change over time in a non-cyclical way,

while the whole remains a whole of the original type – as, for example, an oak tree remains an oak tree despite growing enormously over the course of its lifespan.

Note that changeable structures are only possible in entities with area ranges or multiple area ranges, since otherwise any change in the structure would take the entity beyond its structural range and it would cease to be a structure of this type. The converse, however, is not true: individual instances of entities with area ranges may nevertheless be fixed structures, although different instances of the entity type may vary from each other.

Now all of these typologies have implications for the concept of morphostasis. The first merely helps us to illustrate the concept: morphostasis has been sustained even when a part has been exchanged for another token of the same type, as long as this does not lead to some change in the properties of the entity as a whole (e.g. death). It is a contingent question whether any given type of entity can sustain token part exchangeability, but morphostasis is entirely consistent in principle with such exchangeability.

The second typology provides a degree of flexibility to morphostasis, the significance of which becomes clear when we consider the third typology: where an entity type has an area range of structural possibilities, rather than just a point range, it becomes possible for the continuing existence of the entity to be compatible with changes in the details of its structure over a period of time. We must make a definitional decision on whether or not this is to be considered morphostasis. On the one hand, we can read the *morpho* of morphostasis as referring to an exact set of structural relations, in which case morphostasis only occurs when the precise structure of an entity is maintained, and the entity may continue to exist despite a failure of morphostasis. On the other, we can read the *morpho* as referring to the range of variations of structure corresponding to a particular entity type, in which case morphostasis is compatible with changes in the precise structure that fall within the area of structural possibilities for that type of entity. One solution would be to label these as two distinct variants of morphostasis, which we might call *precise morphostasis* and *type morphostasis* respectively. Generally, however, I shall assume that morphostasis means the former of these: the maintenance of a precise structure over time. The consequence of this is that where entities are changeable, as most of those relevant to social theory will prove to be, their continued existence is not simply the product of morphostatic causes, but the outcome of an ongoing interplay between morphostatic and morphogenetic causes.

*Morphogenesis*

As we have seen, morphogenetic causes are those that bring about or change the form or existence of an entity. We have already touched on some aspects of morphogenetic cause – such as the difficulty of identifying a specific cause as *the* morphogenetic cause of an entity in a causal history that inevitably stretches back over a whole series of prior events, and the ongoing interplay or even conflict between morphogenetic causes that are tending to alter or destroy an entity and the morphostatic causes that are working to preserve it in its current form. This section will examine this latter interplay in a little more detail.

In general, morphogenesis encompasses processes which (a) contribute to the initial development or creation of any entity; and (b) contribute to the subsequent modification of its form within the structural range of the entity type. There is a continuum between the second of these and those processes which tend to take an entity's form beyond the structural range of its type. These are also processes that alter the form of the entity, and hence may be considered morphogenetic, but their effect is to bring the existence of the entity to an end. This may be entirely destructive, as when an entity is materially reduced to some aggregate of its parts, or it may be simultaneously creative and destructive, as when the entity is transformed into some alternative type of entity at a similar or higher level of organisation.

The structures that concern us in social theory are generally *dynamic structures* that maintain themselves, not in a stable internal relationship, but by constantly striking a balance between internal parts and relations that are in tension with each other. This is how Laszlo characterises social structures, which he says "adjust and adapt, maintaining themselves in a dynamic steady state rather than in one of inert equilibrium" (Laszlo, 1972, p. 46).[19] Such structures contain within themselves the potential for change; if their normal state is a dynamic one, then a change in their environment may lead them to adapt by moving to a new point of dynamic equilibrium, or indeed by moving without finding such a point at all. This may lead to one of a variety of outcomes: convergence on a variable but constrained pattern, continuing adaptive steps over a period of time, or collapse of the structure. In such systems, there is a constant interplay between morphostatic and morphogenetic causes.

As Buckley puts it,

---

[19] Although not all dynamic structures are social; most biological systems, for example, are also dynamic structures.

Thus, the complex, adaptive system as a continuing entity is not to be confused with the structure which that system may manifest at any time (a persistent error or ambiguity in Parsonian theory). Making this distinction allows us to state a fundamental principle of open, adaptive systems: *Persistence or continuity of an adaptive system may require, as a necessary condition, change in its structure* (Buckley, 1998, p. 86).

Buckley emphasises in particular that in such a system, variation or deviation is not abnormal and disruptive but normal and indeed essential to the continuing survival of the system (Buckley, 1998, p. 71).

The concepts of morphostasis and morphogenesis, then, are capable of elaboration and combination in ways that enable us to start describing complex adaptive systems that are reminiscent of social structures. In particular, such systems demonstrate an intriguing interaction between equilibrating and dis-equilibrating causal factors, which suggests that social theory based on the analysis of such systems may be able to overcome one of the problems typically attributed to Parsonian social systems theory – its focus on social stability to the point of denying mechanisms for social change. This is a benefit that is very clear from Archer's morphogenetic approach to emergent social systems (Archer, 1995; Archer, 1996a).

In the meantime, however, this analysis of morphostasis and morphogenesis provides us with the tools to address one misconception about emergence that can be found in the critical realist literature – the belief that emergence is not necessarily a higher-level product of the relationship between a whole and its parts.

## Emergence is based on composition

The composition of entities by their parts is central to the conception of emergence advanced in this thesis, as it has been from the earliest versions of the concept. McLaughlin, for example, tells us that,

> According to British Emergentism, there is a hierarchy of levels of organizational complexity of material particles that includes, in ascending order, the strictly physical, the chemical, the biological, and the psychological level. There are certain kinds of material substances specific to each level. And the kinds of each level are wholly composed of kinds of lower levels, ultimately of kinds of elementary material particles (McLaughlin, 1992, p. 50).

Most emergentists have continued to take the view, as I do, that the concept of emergence is inherently compositional. By this I mean that any higher-level entity (and its emergent properties) is dependent upon a collection of lower level entities in the sense that (a) they are the necessary component parts of the higher level entity; (b) each

emergent property of the higher level entity is dependent upon (but not eliminatively reducible to) the properties of these parts; and (c) each emergent property of the higher level entity, in the sense of a power or tendency, is not dependent upon the properties of other entities that are not such parts (although it may be so dependent for its realisation).

### Critical realists and composition

At times critical realists have adopted a compositional definition of emergence, yet at other times they have seemed to deny such a view. Let us begin with some examples of the compositional approach. Andrew Collier, for example, has written "Bhaskar is explicit about the reality of the complexity of complex entities, whose powers are 'emergent' with respect to their components" (Collier, 1989, p. 51). Elsewhere he writes:

> As against atomism and holism, Bhaskar's emergence theory allows us to conceive of real, irreducible wholes which are both composed of parts that are themselves real irreducible wholes, and are in turn parts of larger wholes, with each level of this hierarchy of composition having its own peculiar mechanisms and emergent powers (Collier, 1994, p. 117).

Yet elsewhere Collier has denied a compositional account of emergence, and Bhaskar has sometimes seemed ambiguous on the question. Let me discuss Collier's argument first. Elsewhere in *Critical Realism,* for example, he argues "that many (though not all) cases of rootedness-emergence relations are also relations of composition" (Collier, 1994, p. 116). The basis of this "not all", however, is clearer in his earlier work:

> A level of mechanisms depends unilaterally for its existence on lower levels – that is what it means to call it 'higher' and to call them 'lower'. *One* way in which a mechanism may so depend, is that it is a feature of structurata that [they] are composed of structurata governed by the lower-level mechanisms. But that is not the only way it may so depend. Language is composed of signs, not of people, but it is dependent on there being people, and people being governed by certain (biological, etc.) mechanisms. So there will be more strata in the hierarchy of vertical causality (dependence and emergence) than in that of composition…

> I think it is true that the human body is a structuratum with different elements from those of the mind (i.e. on the one hand, cells; on the other, cathected intentional objects) and that, in the hierarchy of dependence and emergence (vertical causality), there are two intervening levels (society, language), since these depend on the bodily organism, and are conditions of the emergence of mind (Collier, 1989, p. 99).[20]

Collier's argument appears to be that entities such as languages and minds do not emerge only from their parts but also from other things that are not their parts, since

---

[20] Note that 'structuratum' is synonymous with 'entity'. Also see (Collier, 1994, p. 133)

their existence (or perhaps that of their emergent powers) depends not only upon their parts but also upon these other things.

There are at least three possible interpretations of this argument, and we must consider each in turn.

First, he may be arguing that non-compositional lower levels may have been essential causes in the initial development – the diachronic emergence – of the higher level. Certainly it is true that the causal history of any given type of entity is likely to include a vast range of entity types that are not its parts. This, however, is completely irrelevant to the question of whether or not the entity currently possesses emergent powers and to how those powers emerge in the synchronic sense. If this is Collier's meaning, then he appears to be conflating diachronic and synchronic emergence in this account – or, to put it differently, conflating *morphogenesis* with the *mechanism* of emergence itself. The use of *is* in "language… is dependent on there being people", however, suggests that this is not his intention.[21]

Secondly – and this seems the most likely interpretation – he may be arguing that non-compositional levels may be essential causes in sustaining the continuing existence of the higher level. Again, it is true that the continuing existence of any entity is likely to depend causally on a vast range of entities that are not its parts. Indeed, such dependence relations are far more widespread than Collier's treatment suggests. Many biological organisms, for example, cannot exist without a breathable atmosphere. Rivers and oceans cannot exist without the gravitational force of a planet. Many people in the contemporary world could not continue to exist without the set of technologies that we depend upon for our food and other material needs. But once again, this is completely irrelevant to the question of whether or not an entity currently possesses emergent powers and to how those powers emerge. It is, once again, only if 'emergence' is given something like the diachronic sense rather than the synchronic sense, that it makes any sense to say that an entity or property or power emerges from the entities that contribute to sustaining its existence but do not interact to provide its powers. On this reading, Collier conflates the *morphostatic* causes of an entity with the mechanism of emergence.

The third possible interpretation is that Collier *is* thinking here of the role of non-compositional levels in the mechanisms of emergence themselves – although his

---

[21] Although it is arguable that this statement only makes sense as an account of diachronic emergence, since language could continue to exist in books, films, and the like even if people ceased to do so.

use of 'existence' in the first of the two quotations above seems to conflict with this. Here we could read his argument as an assertion that the emergent powers of a certain type of entity ($E$) can only appear when not only the parts of $E$, configured in the way characteristic of $E$, but also certain other entities ($F$), external to $E$, and in a particular set of relations to $E$, are also present. This is a different claim than the claim that these other entities are essential to the morphostasis of $E$, since the morphostatic relationship is a diachronic, causal one, whereas the relationship that I am now discussing is a synchronic one.

Such a claim is clearly true when the property or power in question is the ability of an $E$ to affect the $F$'s concerned – the power of water to put out a fire, for example, cannot be exercised unless there is a fire to put out. This sort of case is usually dealt with in critical realism via Bhaskar's argument that causal powers can exist unexercised if the conditions of their exercise are not met, and treating the presence of an $F$ as such a condition – and also by treating such cases as an interaction between the causal powers of an $E$ and the causal liabilities of an $F$ (where a liability is merely a passive type of power – a power to be affected in a certain way). Both of these approaches imply that an $E$ and an $F$ each have a relevant causal power, and that in cases of actual causation the effect that follows from these causal powers will only be realised if both the contributing powers are present. On this account, the real *existence* of the relevant emergent property of $E$ is *not* dependent on any external entity, but only its actual *exercise*, and hence this would seem to sustain a compositional account of emergence.

A more complex version of the argument results if we suggest that a property of $E$ might depend on the synchronic presence of an $F$ even though the exercise of the property does *not* affect the $F$ directly. This is a problem only if there really are such cases (and where these cannot be resolved by identifying the existence of a third entity, composed of an $E$ and an $F$, that is the real possessor of the property).[22] The role of society with regard to the emergent powers of our mind does not seem to be such a case – social entities clearly affect our behaviour, but they do so through the intermediate step of contributing causally to our knowledge or beliefs. This represents a morphogenetic effect on the structure of our brains/minds and not a synchronic dependence of our mental powers upon social entities.[23] It seems entirely possible, therefore, to see mental powers as emerging simply from human bodies (see, for

---

[22] The 'third entity' point was suggested by an anonymous reviewer for the *Journal of Critical Realism*.

[23] This does not mean I am denying synchronic emergent powers to social entities – see (Elder-Vass, 2005b)

example, Archer, 1995, p. 102), or more specifically, as emerging from the configurations of neurons that underlie our knowledge and beliefs.

It is certainly true that establishing the hierarchy of emergence mechanisms that underlie language and mind (if the latter is a useful concept at all) is a challenge that still requires further work, and Collier's argument is clearly motivated by an attempt to meet this challenge. It is also possible that the composition relations that underlie these entities may be unusually complex. However, there is no necessary reason to believe at the moment that the resolution of these problems will require a non-compositional account of emergence.

### *Intrastructuration*

A second argument that appears at first sight to deny the compositional basis of emergence appears in Bhaskar's *Dialectic*, where he argues that emergence

> consists in the formation of one or other of two types of superstructure (only the first of which has generally been noted in the Marxist canon), namely, by the superimposition (Model A) or intraposition (Model B) of the emergent level *on* or *within* the pre-existing one – *superstructuration* or *intrastructuration* respectively (Bhaskar, 1998a, p. 599).

Although here he talks in diachronic terms of the formation of structures and thus in terms of morphogenesis, the argument clearly implies that these structures continue to exist and possess synchronic emergent powers. A similar argument is expressed in directly synchronic terms by Bunge:

> P is an *emergent* property of a thing b if and only if either b is a complex thing (system) no component of which possesses P, or b is an individual that possesses P by virtue of being a component of a system (i.e. b would not possess P if it were independent or isolated) (Bunge, 1996, p. 20).

The former case would seem to correspond to superstructuration, and the latter to intrastructuration. Bunge's formulation, however, has the merit of making clearer that there is still a compositional basis to intrastructuration: in such cases, the properties of an entity are altered as a consequence of it having become part of a particular type of whole. Here, these new properties of the part are still a consequence of the composition of the whole by its parts; all that is different from the usual case of emergence is that it appears to be the part that is exhibiting a different property rather than the whole. Bunge argues, for example, that atoms change their form when they become parts of a molecule, rather than simply being held together while retaining their previous form (Bunge, 2003, p. 12). A more significant example for the purpose of this thesis would

be the case of human beings who become parts of organisations, and whose causal powers are changed as a result of them adopting a role in the organisation.

Despite its advocacy by both Bhaskar and Bunge I have some doubts about the concept of intrastructuration. What is at issue here is the question of whether a property of an entity that is the consequence of it being part of a larger whole is really a property of the part at all, or whether it is just a property of the whole that happens to be localised in some respect within the part. What is seen in Bunge's account as a property of an atom which has become part of a molecule, for example, might be better represented as being a property of the molecule itself. This makes no difference, of course, to the argument that such changes in properties can occur, nor to the argument that they remain consequences of the composition of the whole by its parts.

Neither Collier's concerns, then, nor Bhaskar's, undermine the compositional account of emergence. And the compositional account also has a number of epistemological advantages over Collier's alternative. First, it maintains a clear relationship between emergence and its primary theoretical usage: to maintain the tenability of a stratified view of reality in the face of eliminative reductionist arguments. Second, it maintains a degree of simplicity and hence clarity to the concept of emergence that makes it easier for us to understand its theoretical role. And third, a more detailed compositional account of emergence makes it possible for us to understand how emergence works in practice, as discussed elsewhere in this chapter. None of these arguments, of course, constitutes a definite proof of the compositional account of emergence. But I believe this section has offered good grounds, both ontological and epistemological, for holding the compositional view.

**We cannot substitute relations for entities**

A related confusion in some accounts of emergence is the view that the significance of organisation or relations is so large that we should substitute relations for entities in our accounts of the world – or of the social world, at least. Bhaskar, for example, in his essentially emergent account of the world, seems to do this when he turns to the social world, citing Marx: "society does not consist of individuals [or, we might add, groups], but expresses the sum of the relations within which individuals [and groups] stand" ( from the Grundrisse, quoted in Bhaskar, 1998b, p. 26 – the internal comments are Bhaskar's). Collier writes of "societies (composed as they are of relations between people, and ramifications of those relations)" (Collier, 1994, p. 145). The claim

that societies are composed of relations rather than individuals, however, seems to me to confuse the issue. As we have seen, emergence implies that each higher-level entity is composed of lower-level entities, but not just of an unrelated collection of lower-level entities. Rather, that collection of lower-level entities is structured by the set of relationships between them. These relationships must exhibit systematic organisation of some type – structure – for emergence to be possible. As Collier himself says, "The latticework of relations constitutes the structure of 'society'" (Collier, 1994, p. 140). But it is one thing for the latticework of relations to constitute *structure* (i.e. the mode of organisation), and quite another for those relations to be seen as the *parts* of higher-level wholes.

It is worthwhile in this context revisiting the application of the same principles to natural science. Molecules, for example, are emergent from the organisation of atoms. Molecules are composed of atoms, but not random unrelated collections of atoms; they exist only as a result of stable and systematically organised inter-relations between the atoms that compose them. Those relations constitute the *structure* of the molecule, while the *parts* of the molecule are the atoms themselves. There is no obvious reason why we should not treat social entities in a parallel way.

Now it is true that Marx uses *relation*, at least sometimes, in a different sense to that I have employed here, and presumably Bhaskar is following him in this usage. As Ollman points out, Marx uses "the term 'relation' in two different senses: first, to refer to a factor itself, as when I call capital a relation, and also as a synonym of 'connection', as in speaking of the relation between different factors" (Ollman, 2003, p. 26). Ollman defends the first usage as follows:

> Most modern thinkers would maintain that there cannot be relations without things just as there cannot be things without relations. Things, according to this commonsense view, constitute the basic terms of each relation and cannot themselves be reduced to relations. However, this objection only applies to Marx if what he is doing is caricatured as trying to reduce the terms of a relation to that which is said to stand between them. But his is not an attempt to reify 'between' or 'together'. Instead… the sense of 'relation' itself has been extended to cover what is related, so that either term may be taken to express both in their peculiar connection (Ollman, 2003, p. 36).

Thus, in this usage, (I shall follow Ollman in capitalising the 'R' to distinguish this kind of Relation from the commonplace relation) a Relation includes all parties to it plus the connection between them. But my argument above implies that this is logically equivalent to an emergent entity at the next higher level of structure. In the example of the molecule, the Relation that includes its constituent atoms and the relations between

them is nothing more nor less than the molecule itself. Now, admittedly, we could construct Relations that did not constitute emergent entities, but merely arbitrary collections of things and their connections, such as 'the first three people in a bus queue' or 'the hydrogen atoms in my desk'. But such Relations have no substantial explanatory value, precisely because they do *not* constitute emergent entities. There therefore seems to be no value in distinguishing Relations from entities in an emergent account of reality: Relations as defined by Ollman, and as sometimes used by Marx and Bhaskar, simply *are* entities.

## Stratification and branching

Once we have recognised that our universe is populated with entities composed of parts, which are themselves in turn composed of parts, and so on down to the lowest possible level, it is natural to think of these entities in terms of higher and lower levels, with each level consisting of entities composed from the entities at the next lower level. It is then a common step to identify these levels with the different sciences that study them. In this conception, the universe as a whole can be seen as a nested set of domains characterised by emergence of the explanatory entities of one domain from those of its 'root' domain, and science as divided into distinct sciences to explain the behaviour of the entities in each domain.

Indeed, this is a widespread interpretation of the shape of an emergent universe. We have already seen that the early British emergentists thought in this way (see page 47 above) (McLaughlin, 1992, p. 50). Similarly, critical realists like Bhaskar and Collier talk explicitly in terms of nature being stratified in this way (Collier, 1989, p. 45).[24] Kim is repeating an accepted view in the philosophy of mind when he describes "a layered world, a hierarchically stratified structure of 'levels' or 'orders' of entities and their characteristic properties" (Kim, 1993, p. 9). And Buckley is reflecting the general systems theory from which contemporary complexity theory grew when he says that

> A systems view of reality allows one to see that it is made of successive layers of bonded elements, each layer with properties emergent from the previous one… The challenge of science is to understand the nature of these particular bondings or interrelationships, and that of systems science especially is to unravel the way in which new properties emerge at each level to constitute a dynamic whole able to act as a unit (Buckley, 1998, p. 78).

---

[24] Collier refers to a "'tree' of sciences" but then immediately provides a classification that is an ordered hierarchy of levels (Collier, 1989, p. 45).

Now, in many respects this is a useful and reasonable way to depict an emergent world (and I shall continue to talk in terms of levels from time to time), but the layered model does need some qualification. This section will propose some improvements to the simple 'layered' model, but I begin with a context in which a layered model does indeed make sense.

### Internal layering

One implication of emergence is that entities with emergent properties or powers are themselves composed of other such entities, which are in turn so composed, and so on.[25] A plant, for example, consists of cells, the cells consist of molecules, the molecules consist of atoms, and so on. Any given entity, then, can be seen as internally stratified into many different levels or layers, each level representing sets of parts that are combined into the entities at the next level up.

Now, for most purposes, when we discuss any given entity we are in the habit of ignoring the role of its parts. To treat an entity in this way is to take what I propose to call a *level abstracted* view of it – i.e. a view that considers the effects of the whole entity in isolation from the existence or effects of its parts. I argue, however, that for other purposes we sometimes need to treat a whole entity quite explicitly as a stratified ensemble of parts at various ontological levels. This is to take what I call a *downwardly inclusive* view of the entity. These two terms are illustrated in Figure 2.1 below.

Here, L1 represents the highest level of a whole – to continue the example, a plant. L2 represents the first decomposition of the whole into its parts – in this case, perhaps, the cells of the plant, and the relations between them that constitute them into a whole plant. L3 represents the next decomposition – here, the molecules that make up the cells and the relevant relations between them. And the pyramid may continue downwards, until its base is lost in the mists of our limited understanding of sub-quantum science. Of course, a plant is not made up of the whole plant *plus* its cells *plus* its molecules, and so on; each of these levels represents a different decomposition of the same whole; it is only our view of the plant that must sometimes encompass the recognition that the whole plant is simultaneously each of these different decompositions.

---

[25] It is not clear in the current state of science whether this nesting proceeds indefinitely or whether there is some lowest level of entity that will eventually be reached in this series of progressive decompositions. We can ignore this question for the purposes of the argument presented here.
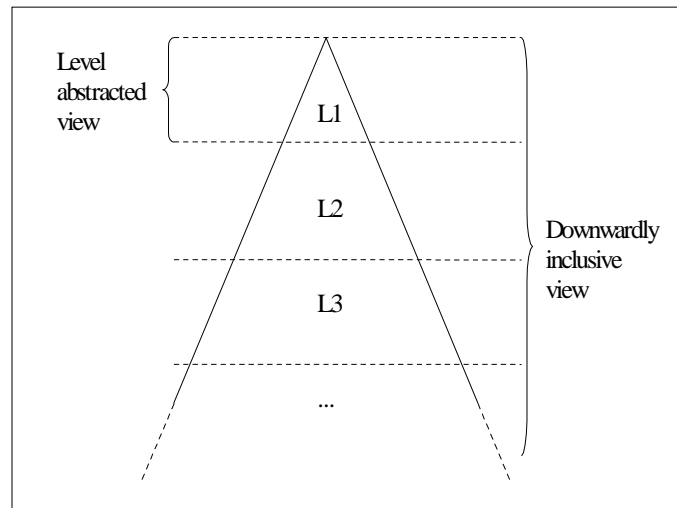
**Figure 2.1 – Internal stratification**

In considering any individual entity, then, it may be useful to represent its structure as a number of layers, each being a successive decomposition of the whole into its parts. However, we must take care with the layer metaphor even here. It may seem that we can divide the parts and their parts quite neatly into layers without any branching out into different entity classes as we go down. But it is quite possible for the components at the next layer down to belong to several different classes, so that there is a kind of inverse branching – rooting, perhaps – as we move down the hierarchy of an individual entity's parts and sub-parts. Granted, we may then find a recombination of these downward branches when we arrive at some of the more fundamental physical particles; nevertheless, the idea that there are unambiguous layers of structure even within a single entity must be treated as a useful tool rather than an article of faith.

*Branches not layers*

However, when we turn from individual entities to consider the whole set of entities that populates our universe, the idea of 'strata' or levels becomes even more misleading. The key problem is that above any given level it is possible that a variety of different classes of higher level entity may emerge. Each of these classes may behave in a significantly different way. Thus, for example, both meteorology and plate tectonics study entities that emerge from various types of aggregations of molecules – as does biology. It is therefore more accurate to see emergent reality as branching in a tree structure than as layered in homogeneous strata. To use a different metaphor, sometimes emergence is seen as producing a series of 'nested' domains – but it is not a 'Russian

doll' type of one-in-one nesting; sometimes there are a number of higher domains nested inside an individual lower domain.

We can make the argument a little clearer. Let me use the word 'classes' to refer to groups of entity types that all emerge by similar mechanisms from similar sets of lower-level entities. Such a class constitutes a branch of the tree of emergence, and is divided into types. Thus, for example, atoms are a class of entity, while hydrogen atoms and oxygen atoms are types within this class. In this model of 'stratification', there is a clear ontological basis for the existence and demarcation of each domain: each domain, each branch of the tree, corresponds precisely to a class of entities.

But it is common in stratificationist accounts to lump many different classes of entity together into a single 'layer' and to assume that each of these broad layers is a coherent emergent unit. The criteria for identifying what classes of entities belong in any given level are rarely specified or given any sort of theoretical justification. But despite this, some authors see emergence primarily as a relation between levels. Emmeche et al, for example, talk quite specifically of the emergence of primary levels and sublevels, and seem almost reluctant to admit that individual classes of entities also emerge (Emmeche et al., 1997, pp. 91-2, 106).[26] I argue, by contrast, that emergence is a relation between entities, not between levels, and hence that emergent domains, whether branches or levels, are merely collections of entities that share a common type of parts, and hence such domains are secondary and derivative concepts, to the extent that they are a useful concept at all.

By contrast, stratificationist logic leads to odd claims such as 'a physical effect may only be the result of a physical cause' – what Kim refers to as "the causal closure of the physical" (Kim, 1993, p. 192). This is phrased as an ontological restriction on causation that seems to mean that all causation is *horizontal*, i.e. it occurs within a particular level. I shall discuss this sort of claim more generally in the next chapter, but the point here is that there is no ontological basis for believing in such a restriction. If emergence as such implied, for example, that social institutions cannot affect human individuals (which it does not!), then it would also imply that human individuals cannot affect cells, cells cannot affect molecules, and molecules cannot affect atoms. But advocates of 'causal closure' seem to imply that entities in different branches within 'the physical' *can* affect each other, while entities in other levels cannot affect

---

[26] Occasionally Collier also seems to imply that it is the emergence of levels that is primary, e.g. (Collier, 1989, p. 102).

'physical' entities. There is no obvious ontological basis for such a claim, other than sheer metaphysical prejudice – a disease we will encounter in further guises in the discussion of reductionism in chapter three. But the practice of lumping a variety of emergent domains together into broadly defined levels such as 'the physical' encourages such confusions – and this is an extremely widespread habit.

### *Levels and disciplines*

Part of the motivation for this habit seems to be a desire to identify ontologically emergent levels or domains with particular scientific disciplines (see, for example, my quote from McLaughlin on p. 47 above, and (Collier, 1989, p. 45)). There is certainly some sort of relationship at work here, but it is clear that there is not a consistent one-to-one mapping between branches of the ontological tree and particular disciplines. More typically, a discipline, at least in the natural sciences, will address the explanation of a group of closely related entity classes, although sub-disciplines may map more closely onto individual classes. "Physics", for example, is extremely broad in this respect (ironically, Emmeche et al recognise this: 1997, p. 91). Thus, a single discipline will often relate to a number of entity classes, with a relatively arbitrary boundary defining the group of classes it studies. Indeed, it seems plausible to suggest that the real relationship between 'levels' and disciplines runs in precisely the opposite direction from that suggested by stratificationist thinkers: that so-called 'levels' are defined in practice by identifying the classes of entity that happen to be studied by a particular discipline, rather than on the basis of an ontological criterion.

Even if we accept the argument that disciplines sometimes study sets of entities that can be bounded on the basis of some objective ontological criterion, it seems that the sciences, like entity classes themselves, are divided into a branching pattern rather than a stratified pattern. As my earlier comment on meteorology, plate tectonics, and biology makes clear, the branching metaphor more accurately reflects the shape of our sciences.

The shape of the tree of sciences also diverges from the shape of the tree of emergent entities, though perhaps in a different way, when we arrive at the social. It is clear, for example, that anthropology deals with a group of classes of social entities that overlaps substantially with that addressed by sociology, while the set addressed by sociology overlaps in turn with the sets addressed by political science and economics. In the social sciences, then, the divergence between the two trees mostly arises from the substantial overlapping between the entity classes studied by different disciplines.

There is also a great deal of confusion over how best to draw the boundaries between domains or levels *within* the social. Two examples are instructive. Harvey and Reed (following earlier systems theorists) pile up rather arbitrary sections of social theory to identify no less than fourteen nested levels of social emergence (Harvey and Reed, 1995, p. 307). Brante, on the other hand, alternates levels composed of entity classes with levels composed of relations between entities – an approach that my earlier comments on Ollman's Relations demonstrates to be incoherent (Brante, 2001, pp. 178-180). What both accounts have in common is the absence of clear, theoretically coherent criteria for identifying what constitutes a level or a branch. I will argue, against both, that we can only construct a viable model of emergence in the social realm by identifying domains on the basis of the entity classes that emerge at each point in the structure. Furthermore, as in the more general case, a simple layered structure can not do justice to the nature of the interactions between the different classes of social entity and hence we will require a branching model.

## Conclusion

This chapter has elucidated the dominant conceptions of emergence in the existing literature by briefly examining the history of the concept and by explaining the distinction between strong and relational versions of the concept. Strong emergence, on the argument presented here, may occur in a restricted range of cases, although it may not occur at all. It certainly does not provide a *general* argument that can refute attempts to eliminatively reduce the social sciences and notions of social structure or indeed human agency. Relational emergence is more promising, but I have not yet presented a decisive explanation of how it provides a viable alternative to reductionism; such an account depends upon a more thorough analysis of cause than has been possible so far, and providing this will be a key objective of the next chapter.

What this chapter *has* been able to do is to clarify a range of analytical questions, both positive and negative. On the positive side, it has introduced important concepts like morphogenesis and morphostasis and clarified their relationship to emergence and to the dynamic nature of the sorts of entities we find in the social sciences. These concepts enable us to clarify the relationship between the synchronic and the diachronic aspects of emergence, which must be carefully distinguished to avoid serious confusion. On the one hand, entities with emergent properties can only exist as a consequence of a causal history, and thus depend diachronically on a range of

morphogenetic and morphostatic processes which create, develop, and sustain the web of relations that constitute the entity's parts into just this kind of whole. On the other hand, not all persistent wholes have emergent properties; where they do, such properties arise synchronically from the properties of the parts and their relations to each other. This is the causal mechanism underlying the emergent property or causal power concerned.

A full understanding of any given case of emergence therefore depends on being able to explain *both* the causal mechanism *and* the morphogenetic and morphostatic processes that create and sustain its existence. This recognition can be formalised as a methodological framework for analysing putative cases of emergence. In this framework, any claim that an entity possesses an emergent property must be supported by what I shall refer to as the *five pillars*. These five pillars are the answers to five questions: (i) what are its parts?; (ii) what are the relations between those parts that are characteristic of this particular type of entity?; (iii) what set of morphogenetic causes has produced the entity in its current form?; (iv) what set of morphostatic causes stabilises the entity and ensures its continued survival?; and (v) through what mechanisms do its parts and relations produce the specific properties of the entity?

On the negative side, this chapter has sought to clear away some misconceptions found in the existing literature, for example Collier's apparent denial of the compositional nature of emergence, various attempts to substitute relations for entities in the theory of emergence, and the limitations of the 'layers' metaphor that is often employed in discussing emergence.

The chapter has thus begun the task of providing a general understanding of emergence that we can apply to the analysis of the social world. This task remains incomplete, however, until we have examined the implications for the web of causal relations within which any given entity operates. In other words, we must examine the relationship between emergence and causation in more detail, and thus the question of whether and how emergence enables us to negotiate a viable path between dualism and reductionism.  This will be the subject of the next chapter.

# 3      Cause

Emergence matters because it provides the essential foundation for any understanding of how causal forces can (and do) operate in the world. Such an understanding is in turn an essential prerequisite for an adequate response to the errors of reductionism, dualism, and holism which plague the philosophy (whether explicit or implicit) of the social sciences. This chapter is dedicated to explaining the relationship between emergence and cause in general, so that the rest of this thesis can go on to show how this underpins the domain-specific ontology and theory of social structure, human agency, and their relationship to each other.

The chapter begins by showing how an emergent understanding of the world reveals a degree of complexity in the operation of cause that is generally ignored, building on the distinction between 'level abstracted' and 'downwardly inclusive' views introduced in the previous chapter. It then goes on to show how the critical realist model of real causal powers and actual causation offered by Roy Bhaskar provides a suitable framework for understanding this complexity (although it will also offer something of a qualification to Bhaskar's model). Relating this model of causation to the account of emergence developed in the previous chapter, the following section provides a response to reductionism, offering an argument that shows why it is impossible to eliminate emergent causal powers from causal explanations in favour of lower-level properties. Finally, this in turn provides a basis for understanding the question of downward causation (and the related question of diagonal causation), which is often considered problematic for emergentism (e.g. Kim, 1992).

**Causal relations between emergent wholes**

Let me begin by examining some general characteristics of causation in a world of entities with emergent properties. Typically, theorists ignore the multi-layered nature of entities when they deploy their properties in causal accounts. This section will argue that this may be a valid strategy for some purposes, but it rests on a hidden practice of abstraction from the lower level composition of the entities and indeed the events concerned. When we are seeking to understand how the different levels interact in the causation of events, this practice ceases to be valid, and instead we must build an

understanding of cause that embraces the many causal interactions that may be occurring at many different levels.

### *Retroduction and retrodiction*

It is of the essence of the concept of cause that any given type of cause influences outcomes in a similar way across all relevantly similar cases. Thus, causality operates to determine individual events, but the causal factors that determine these events are generic in the sense that whenever they are present they will have an influence that is in some way consistent. Without such consistency, it would be quite impossible for us to disentangle the causal influences that affect our world, and quite pointless for us to speculate about general causal laws or mechanisms. With it, however, it becomes possible for us make useful generalisations across many similar instances.

At the first level, such generalisations are mere regularities. The 'covering law' model of causality interprets such regularities as exceptionless *laws* which enable us to deduce what will occur whenever the preconditions for the law to operate are present (Honderich, 1995, p. 170). However, as Bhaskar has argued, experimental science rests on the belief that laws whose existence is established through experimentation continue to operate in open systems when their operation does *not* produce exceptionless regularities. Hence the Humean idea of causality as 'constant conjunctions' of empirical experiences is untenable (Bhaskar, 1978, pp. 33-5). Yet some evidence of at least partial regularity must appear in the empirical record for us to be able to detect a causal influence at work and to prompt a search for what might be responsible for it. We may usefully follow Lawson in labelling these *demi-regularities* or *demi-regs* (Lawson, 1997, pp. 204-9).

At the second level, the analysis of demi-regs allows us to theorise the existence of underlying causal mechanisms that are responsible, subject to circumstances, for the observable degree of regularity – a process that critical realists have labelled *retroduction* (Lawson, 1997, p. 24). In practice, as Bhaskar has emphasised, such regularities are often masked by the operation of other causal mechanisms with conflicting effects (Bhaskar, 1978, p. 13). This means that we cannot interpret the patterns of events that we do detect as the consequences of exceptionless laws: "Theoretical laws are essential in calculating just what each cause contributes. But they cannot do this if they are literally true; for they must ignore the action of laws from other theories to do the job" (Cartwright, 1983, p. 12). In the experimental sciences, it is possible to create 'closed systems' in which the operation of such conflicting causes is

temporarily excluded and hence make it possible to observe full regularities (Bhaskar, 1978, p. 33). In many other sciences, however, we must make do with the observation of 'open systems' and infer causal relations from demi-regs.

Accounts of cause, however, must also encompass the opposite case, where, instead of deriving causal laws from the analysis of actual events, we explain particular events as the result of the causal factors at work – a process that critical realists have labelled *retrodiction* (Lawson, 1997, p. 221). In these cases, we assume that it is valid to take generalised causal factors and claim that they have had a determinative effect (generally in conjunction with other such factors) in a specific instance.

### *Causal laws assume an abstracted ontology*

Now it is common, though by no means necessary (as I will argue later in this chapter), for successful causal explanations to be framed in terms of sets of entities and properties that are all found at the same level of organization (what I have called in chapter two a domain). Thus, for example, we may offer mental explanations of mental events, or chemical explanations of chemical events. If, as it seems, we often find successful explanations operating at the same level as the type of event to be explained, then this suggests that same-level causation is a common feature of the world we live in, and therefore also a useful habit of thought in describing it.

This seems to work reasonably well for retrodiction – when we are formulating causal theories, in which we abstract from all the extraneous features of the many different instances across which we are generalising, and focus instead on the common features that provide a basis for lawlike generalisation. In such situations, we can work successfully with a *level abstracted ontology* which ignores the fact that each of the entities we are discussing is in fact composed of a variety of levels of lower entities. The composition of the entities we seek to explain (or use as causal factors) is simply one of the many things that we seem to be able to abstract from in formulating our theories. As a result it may appear in the resulting theories as if the entities which 'cause' and are 'caused' are autonomous of their component parts.

Such an abstracted ontology, which ignores the compositional structure of the entities it invokes, is perfectly usable and indeed positively useful for the purpose of describing individual causal mechanisms. It also works quite reliably in many practical *retrodictive* applications, both everyday and scientific, when same-level causation often seems to reflect what is going on well enough to provide us with reliable expectations. However, I argue that it is quite inappropriate for the discussion of what is happening

*over multiple levels* when we look more deeply at individual instances of causation. Ignoring the compositional structure of entities leads to one of three errors. The first is the belief that we can discuss causal relations between an entity and its components in particular instances as if the two were entirely independent of each other. This is the error of dualism.[27] The second error is the belief that once we have offered a causal account of the behaviour of entities of a given type at a given level of organisation, we can safely assume that the behaviour of their composing entities will follow along as a consequence: the higher level entities, as it were, dragging their tails behind them. This is the error of holism. The third is the belief that the causal impact of emergent higher level entities can be explained purely in terms of the impacts of their parts. This is the error of reductionism.

In the next section I shall offer an alternative ontological view, consistent with the account of emergence and composition so far, which allows us to provide a more viable account of causation across multiple levels in particular instances.

### Individual events are inherently multi-level

Although the sort of abstracted ontology that is implicit in causal generalisations is perfectly adequate for some purposes, I will argue in this section that we need different but complementary views of ontology for different purposes. In particular, when it comes to discussing how cause works over multiple levels in single instances of events, we need a *downwardly inclusive* view of emergent entities like that described in the previous chapter.

Let me begin by reviewing the nature of events themselves. And here the key is this: that our everyday (empirical) concept of an 'event', which we take to be the naturally-given subject of any explanation in science, is itself an analytical abstraction from reality. Thus, when we say, for example, 'the pen fell on the floor', we are already, in framing our reportage of the event, making an assumption about which abstraction from what was happening in a multi-level stream of interconnected happenings is the one that is relevant and requires explanation. We could have looked at the same happenings and chosen to explain the behaviour of the molecules or atoms involved, or the writing process or the world historical events or the social history of which the falling of the pen formed a part. But in selecting out one of these happenings from the

---

[27] There are, however, many varieties of dualism and indeed reductionism, and not all of them are errors. Here I am referring to the more extreme forms of dualism, of which the classic case is Descartes' view of the relation between mind and body.

rest as the thing to be explained, we create the illusion that this is an event that can be given an explanation in its own right, independently of its component events and of the larger events of which it forms a part: here we have a *level-abstracted view of the event*.

In seeing events as level-abstracted, we implicitly frame the question of how they are caused in a way that calls for explanations in terms of a particular level of organisation. But any causal account of an abstracted event forms only part of a larger picture. A more complete explanation can always be provided by re-integrating the event into the larger stratified picture of which it forms a part, and relating the explanation of this 'event' to the explanations of the other event abstractions in which it is inextricably implicated, either as subset or superset.

Now, I suggest that the way to make sense of causal explanations of individual events in this context, where an event is defined as the behaviour of a given entity at a given time, is to allow that in reality every event taken as an individual instance inescapably includes the behaviour of the composing lower level entities as well (Lloyd Morgan, 1923, p. 15). To view an event in these terms is to see it in *downwardly inclusive* terms. It might seem that we could also look at the event in *upwardly inclusive terms*, in which sense it would also include the behaviour of all the higher level entities of which the first entity is a part. But this seems inherently infeasible, given the indeterminate (and indeterminately large) range of higher level entities that may be part of this set, all the way up to the universe itself. There is no apparent reason why our interest in the falling of a pen, for example, should also require us to be interested in that complete set of higher level events, even for metaphysical purposes. We may, of course, be interested in some *particular* higher level event of which the falling of the pen is part, but if that is so we can take a downwardly inclusive view of that higher level event, which will include the behaviour of the pen. As a general rule, then, we need not take an upwardly inclusive view of an entity or event.

Hence, in explaining a downwardly inclusive event, we recognise, for example, that when a pen drops, it is inseparably part of this individual event that the components composing the pen remain in a set of relationships through which they constitute the pen, and behave in whatever ways are required for the pen to drop. This is the inevitable consequence of the set of morphostatic causes whose operation must be present for the pen to exist as such though the entire course of this event. Thus, the various material parts of the pen go through a series of events that forms part of the higher event, the molecules that compose those parts go through another series that also forms part of the

higher events, and so on through the atoms, subatomic particles, and so forth. Given that we do not have fully adequate understandings of the lower end of this spectrum, we must accept that only partial descriptions and hence only partial explanations are possible of the lower-level set of events that composes the higher level event. For most practical purposes we can and indeed must ignore the lower levels of this hierarchy, but for the purpose of understanding the ontology of events and causation we must recognise their significance. Actual events are downwardly inclusive and multi-levelled.

Like events, we are accustomed to perceiving entities in level-abstracted terms. But downwardly-inclusive events involve the behaviour of entities that are also defined in downwardly inclusive instead of level-abstracted terms. A downwardly-inclusively view of a pen includes its material components, its molecules, its atoms, and so on, and when we give a inclusive casual account of the dropping of the pen, we will be giving an account that presumes that the falling of the molecules, atoms, etc, is inherently part of that event, since these are inherently part of the entity that has been dropped. Both events and entities can be imagined now as pyramids, consisting of a single abstracted event or entity at the top, all of its components at the next level down, all of the components of those components at the next level down, and so on, at least as far as is permitted by our limited understanding of quantum and perhaps sub-quantum science.[28]

### *Single-instance causation requires an inclusive ontology*

Let me now use an example to show why level abstracted causal accounts are inadequate to the causal explanation of individual events over multiple levels. Consider the case of photosynthesis by a plant. In certain circumstances which need not detain us here, many plants 'convert' carbon dioxide from the atmosphere into oxygen. At the highest level of the event (i.e. a case of photosynthesis) we may simply say that it was caused by the power the plant has to photosynthesise. Many useful explanations may indeed rest on this power, and an empirical scientist could investigate, for example, the differential rates at which plants produce oxygen in different contexts without worrying about how photosynthesis worked at the cellular or molecular level.

But there are some parts of the event concerned that would inevitably remain unexplained by such an account. At another level (the molecular), the process of photosynthesis is a chemical reaction, and we could not explain either *how* photosynthesis works or *which* lower level parts of the entities involved are affected,

and in what way, without looking at this process at the molecular level. If we did examine it at the molecular level, this would not be an account of a different event, but a different account of the same event – one that is abstracted at a different level.

And yet, the lower level account still gives us only a partial account of the causal process at work here, because any explanation at only the molecular level will miss the key *higher level* causal factors which are also necessary for the event to occur. Thus, these molecules would not be configured in an arrangement that made this chemical reaction possible unless they had been organised into the form of the plant in the first place. Similarly, if we took the plant and blended it into a soup, we would still have the same set of molecules but they would no longer have the causal power of photosynthesis, which arises from their organisation into the form of a plant. The causal power of photosynthesis thus belongs to the plant and not to the molecules, but to provide a complete causal explanation of what happens when photosynthesis occurs we need a causal account that operates at multiple levels simultaneously, invoking both the causal powers of the plant and the causal powers of its molecules.

In other words, it is impossible to explain fully the causation of the event except as the outcome of a causal interaction between the whole 'pyramids' – between the entities concerned, viewed in downwardly inclusive terms – and not just the single points at the top – the same entities viewed in level abstracted terms.[29]

We can see why this is a useful way to look at causation if we consider the problem posed to level abstracted accounts by multiple realisability, i.e. cases where the higher-level outcome is consistent with a variety of different lower-level configurations. In these circumstances, level abstracted accounts are underdetermined, in that they can provide an account of the change that occurred at a higher level, but not an account of how the implicit lower-level changes occurred, thus leaving us without any explanation of how its components were brought to a state consistent with it. Downwardly inclusive accounts, by contrast, resolve this underdetermination since the whole range of states of all the component entities and sub-entities involved in the multi-levelled event are available to explain the causation of the lower-level changes.

Of course, each of the interactions at the lower levels can also be considered as inclusive events in their own right, so the higher-level event is at least partially

---

[28] Lloyd Morgan uses the 'pyramid' analogy in a partly similar way (Lloyd Morgan, 1923, pp. 14-16).

composed of a whole set of smaller pyramidal events. Now as a result of this, a reductionist might argue that the inclusive account suffers from the opposite problem to that discussed in the previous paragraph: it may seem to be overdetermined,[30] if we believe that the higher level entities are no more than the sum of their parts, and lower level explanations are available for the behaviour of each of those parts. In this case, it would seem that causes at the higher level are redundant to the explanation of the event, since the lower level causes do all the causing that is needed to produce it.  The refutation of this reductionist argument will be the subject of the third part of this chapter.

Now all this suggests that the abstracted explanations we commonly employ are massive simplifications of the real, multi-level causal processes. The causation of any individual event operates across the whole pyramid of entities and sub-entities, not at a single level of it. Causation as we generally apply it is an analytical abstraction from this, and applies to a single level. It is never truly 'independent' of what is happening at other levels in the individual instance; it is only analytically independent when generalised. Cause as we generally understand and apply it is therefore an attempt to simplify and extract from the impossible complexity of actual causation.

It is testimony to the consistency of the structure of the world that in many cases these simplifications work. Given that consistency, the evolution of this enormously simplified way of thinking about the causal interactions between inclusive entities has provided humanity with immense practical capabilities. Unfortunately, the value of simplifying causal interactions to a level our human brains can cope with (indeed have evolved to cope with), breaks down when it comes to analysing inter-level relationships within a given pyramid. I claim that many of the difficulties of existing approaches to emergence and reduction stem from the inappropriate application of a level abstracted ontology to this set of issues.

Now, this account of level-abstracted and downwardly-inclusive ontologies fits very comfortably with Bhaskar's *depth ontology*, and in particular with his division of ontology into the domains of the empirical, the actual, and the real. The next section

---

[29] This is a sub-case of the determination of events in the actual by a mix of many causes; and also a case of what Bhaskar calls multiple determination, which is discussed below (p. 72).

[30] I use 'overdetermined' here, not in Althusser's sense, but rather to indicate a logically impossible case – i.e. where the set of causally effective factors exceeds those required to explain the set of outcomes, with the result that they appear to mandate a set of outcomes that may be inconsistent with each other.

will examine this aspect of Bhaskar's work and how it can usefully extend the analysis so far.

## Emergence and depth ontology

This section will show how Bhaskar's *depth ontology* reinforces and enhances the view of causation that has been developed in the first section of this chapter. In Bhaskar's critical realist account of cause, there are two key elements – the concept of *real causal powers* and the combination of the causal powers of different entities to produce *actual causation*. This section will relate each of these in turn to emergence, then discuss Bhaskar's important account of *multiple determination*, concluding by showing how real causal powers and actual causation are products of a mutually interdependent interaction that is at the heart of an emergentist account of cause.

### *Bhaskar's ontological domains*

First, we must distinguish Bhaskar's conceptions of the real and the actual. In *A Realist Theory of Science*, Bhaskar argues from the intelligibility of experimental activity to the conclusion that "there is an *ontological* distinction between scientific laws and patterns of events" (Bhaskar, 1978, p. 12). Such laws depend upon the existence of 'natural mechanisms', and "it is only if we make the assumption of the real independence of such mechanisms from the events they generate that we are justified in assuming that they endure and go on acting in their normal way outside the experimentally closed conditions that enable us to empirically identify them" (p. 13). Similarly,

> events must occur independently of the experiences in which they are apprehended. Structures and mechanisms then are real and distinct from the patterns of events that they generate; just as events are real and distinct form the experiences in which they are apprehended. Mechanisms, events and experiences thus constitute three overlapping domains of reality, viz. the domains of the *real*, the *actual*, and the *empirical* (Bhaskar, 1978, p. 56).

The relationship between these domains is summarised in a table, reproduced below as Figure 3.1. Bhaskar clearly intends the domain of the empirical to be a subset of the domain of the actual, which in turn is a subset of the domain of the real (Bhaskar, 1978, Note to Table 1, p. 56; Bhaskar, 1993, p. 207).[31] Our interest here is in the relation between the real and the actual.

---

[31] I have discussed Bhaskar's domains and the elements they contain in more detail in (Elder-Vass, forthcoming, 2006).

|              | Domain of Real | Domain of Actual | Domain of Empirical |
| ------------ | :------------: | :--------------: | :-----------------: |
| *Mechanisms* | x              |                  |                     |
| *Events*     | x              | x                |                     |
| *Experiences*| x              | x                | x                   |

**Figure 3.1 – Bhaskar's three domains: populating entities (Bhaskar, 1978, p. 13)**


### *Real causal powers*

Let me begin with real causal powers: Bhaskar identifies these with "relatively enduring structures and mechanisms" that are "nothing other than the ways of acting of things" (Bhaskar, 1978, p. 14); or in other words, "the generative mechanisms of nature exist as the causal powers of things" (Bhaskar, 1978, p. 50).[32] These things "are complex objects, in virtue of which they possess an ensemble of tendencies, liabilities and powers" (Bhaskar, 1978, p. 51). Although this formulation does not directly invoke the concept of emergence, the relationship with emergence is clear: the powers and properties of an object or entity can be ascribed to the organisation of its parts into a particular kind of complex whole. In other words, real causal powers are emergent properties. This is why Bhaskar argues that "explanation depends upon emergence" (Bhaskar, 1986, p. 104). And Collier makes the connection still clearer:

> As against atomism and holism, Bhaskar's emergence theory allows us to conceive of real, irreducible wholes which are both composed of parts that are themselves real irreducible wholes, and are in turn parts of larger wholes, with each level of this hierarchy of composition having its own peculiar mechanisms and emergent powers (Collier, 1994, p. 117).

Similarly, Fleetwood argues that:

> when… one writes that *a mechanism has a tendency to x*, one is, in reality, referring to the ensemble of structures, powers, and relations: it is, strictly speaking, the ensemble that has a tendency to *x*. Once understood, however, there is no harm in shortening the phrase by omitting reference to structures, powers and relations (Fleetwood, 2001, p. 211).

We can translate this into the language of emergence by equating "ensembles" with higher-level entities whose components are lower-level entities and the relations between them. Fleetwood's argument thus translates into the claim that mechanisms are simply a level abstracted view of a multi-levelled entity. In considering the generalised powers of things, we can, as Fleetwood suggests, work successfully with an abstracted

---

[32] Cf. (Lawson, 1997, p. 21)

ontology that ignores the fact that each entity or thing is composed of a variety of levels of lower entities, and simply sees it as existing at a specific level of organisation.

### *Actual causation*

However, as Bhaskar himself recognises in more recent work, this technique is quite inappropriate for the discussion of what is happening over multiple levels when we turn to the second element of the critical realist account of cause. This is the combination of the causal powers of different entities to produce actual causation, in which actual events are produced by a complex interaction of the causal powers of the entities involved:

> unlike theoretical explanation in at least many of the natural sciences, viz. from explanatory significant structures to their higher-order structural explanation, applied explanation of concrete singulars, like changes in a particular structuratum, are a much messier affair. In a dialectical pluriverse an event e at a level L is as likely to be (multiply) explained by elements at the same and lower-order levels in addition to higher-order (deeper) ones, and/or even laterally, diagonally, tangentially by elements not locatable in the categorical or generic order at all (Bhaskar, 1993, p. 133).

Explanation of actual events is much "messier" because their causation operates across the whole pyramid of entities and sub-entities involved, not at a single level of it (note that Bhaskar is somewhat idiosyncratic in his use of 'lower' and 'higher' here, reversing the usual usage). Events, in all their multi-levelled glory, are the products of the combination of a variety of causal mechanisms operating on the prior state of the set of entities involved. In Bhaskar's account, this individual instance causation (which is of course interlinked with other individual instances of causation) occurs within the domain of the actual, but it is the consequence of the interaction of the real (but not actual) causal mechanisms or powers of the entities involved.

These interacting powers may belong to entirely distinct entities, but it is also important to recognise, as Collier does in the piece cited above, that the various entities that are the parts of the distinct higher-level entities involved also have causal powers. Any given higher level entity, then, can be seen as a pyramid of successively lower-level parts, and the causal impact of the higher level entity as a whole includes the causal impacts of those parts. At each level, the entities formed from the lower level parts have causal powers in their own right by virtue of how those parts are organised. The total causal impact of a higher-level entity conceived of in these pyramidal terms, then, includes the impact of all its lower-level parts as well as the causal powers that are emergent at its highest level.

These causal powers exist as emergent properties of the entities that possess them. Because they emerge at a specific level (e.g. the ability to photosynthesise belongs only to the plant as a whole; the molecules or cells of the plant couldn't photosynthesise if they were not organised into the form of a plant), then it is entirely reasonable to think of them in level abstracted terms. Nevertheless, they can only lead to actual events when they are combined with a multiplicity of causal mechanisms from other levels of the ontological strata. Thus real causal powers can be described in a level abstracted form, while actual causation always occurs in the form of multi-levelled events.

### *Multiple determination*

Bhaskar himself addresses this question of the contribution of causes operating at different levels through a concept which he calls "dual control", "multiple control", or "multiple determination". In considering actual natural and social events, he argues, we must accept that different causal mechanisms and the interactions between them account for different aspects of the events concerned, and that no single law "determines" the whole result:

> Laws leave the field of the ordinary phenomena of life at least partially open... To say that laws situate limits but do not dictate what happens within them does not mean that it is not possible to completely explain what happens within them. The question 'how is constraint without determination possible' is equivalent to the question how 'can a thing, event or process be controlled by several different kinds of principle at once?' To completely account for an event would be to describe all the different principles involved in its generation. A complete explanation in this sense is clearly a limit concept. In an historical explanation of an event, for example, we are not normally interested in (or capable of giving an account of) its physical structure (Bhaskar, 1978, pp. 110-111).

Bhaskar's argument does not relate *only* to the relations between causal powers at different levels of a given multi-layered entity; he is also concerned with the interaction of causal powers between entirely distinct entities, at whatever level they exist. But the same framework does apply equally well to the interaction between causal powers at different levels of the same entity, and Bhaskar makes the link to stratification explicit in a more recent work: "Emergence makes possible the important phenomena of *dual* and *multiple control*" (Bhaskar, 1994, p. 75).

It is precisely because "the [actual] ordinary phenomena of the world" are inherently multi-layered, that we need to deploy accounts of different [real] causal mechanisms, each of which emerges at a specific level, to explain different aspects of them. Thus explanation at each level, in the "area of autonomy" left by the incomplete

explanations at other levels, requires a "putatively independent science" of that level (Bhaskar, 1978, p. 114). And it is in combining all these level-specific explanations of the different levels of a particular event that we "completely account for an event". Although, of course, because we do not have viable sciences of every level, we can only produce incomplete subsets of the "complete" multi-layered account, which is why such a complete account can be seen only as "a limit concept". And in practice, we will not be interested in such complete accounts: we may be perfectly happy to explain an event at a given level while ignoring its lower-level ramifications.

To put this in my terms: in decomposing the behaviour of a downwardly-inclusive entity across its ontological levels, it is the organisation that appears at each level, the set of relations between the relevant lower-level entities, that is the "extra" piece of explanatory information that appears at that level; and this is what makes the "multiple determination" approach viable. We attribute a portion of the causal influence on a particular event to the level of organisation at the topmost level, a portion to the organisation at the next level down, and so on. This allows us to construct causal accounts of multi-levelled single instance causation in which all the levels of the prior situation can have an appropriate influence on the various levels of the outcome. In this model, any insistence on explanatory priority for any particular level becomes nothing more than an ontological prejudice.

It is worth noting that this conception of multiple determination is also required if we are to make any sense of experimental science. The most obvious causal regularity in experimental situations is the causal impact that the intervention of the experimenter has on the results of the experiment. Clearly there is a sense in which the experimenter causes the results of the experiment (Bhaskar, 1978, p. 33). It is only when we have a concept like multiple determination that allows different mechanisms at different levels to contribute to the determination of a multi-layered event that there is room for any other sort of cause to operate in experimental conditions as well as the causal input of the experimenter. Since experimental science works on the assumption that such other causes are in fact at work in experimental situations it also assumes that multiple determination is a feature of the world.

### Similarities of real causal powers and actual causation

Actual causation, then, is a process in which the real emergent causal powers of a variety of entities interact to produce events. Now, as we have seen in chapter two, the sustained existence of any individual entity and hence of its emergent causal powers is

the outcome of a set of interacting morphostatic causes. This interaction is itself a process of multiply determined actual causation. Thus, at the same time as actual causation is a process of combining multiple instantiations of real causal mechanisms, those instantiations of real causal powers are themselves the result of a process of actual causation. Once we take full account of emergence, then, we can see causation as a tightly interwoven interplay between 'real' causal mechanisms and 'actual' causation.

The workings of real powers and actual causation, however, also intersect in another interesting way. One way to understand this is to revisit the five pillars of emergence outlined earlier. Alert readers may have noticed that one of these five pillars is not a strict prerequisite for the possession of *sui generis* properties by a collection of entities taken as a 'whole'. While an entity, by my definition above, must be a *persistent* whole, there is no logical reason why a group of entities that forms a *fleeting* or temporary whole should not have causal powers *as a group* that are not possessed by any particular entities in the group. If emergent properties are the consequence of the existence of a particular set of entities organised in a particular way, then the presence of emergent powers need not depend on that configuration persisting for an extended period. Of course, this fleeting whole would possess those powers only for the few instants during which the particular set of parts and relations required to sustain those powers was in existence, but for this brief moment this whole would possess pseudo-emergent powers. Let me call these *fleeting emergent powers*. More mundanely, a collection of entities may interact causally in a strictly summative way, in which case we might argue that the implicit fleeting whole would possess fleeting resultant powers.

Now there is a clear analogy here between fleeting interactions between groups of entities and the process of actual causation. Indeed, the process of actual causation simply *is* a fleeting interaction between groups of entities and their causal powers. The outcome of that interaction may be determined by the simple addition of the causal powers involved (as in the classic Newtonian parallelogram of forces), or it may be determined by a more complex non-linear interaction between them. These two cases correspond exactly to the ideas of fleeting resultant powers and fleeting emergent powers. [33]

---

[33] There is therefore a sense in which at least some cases of actual causation are fleetingly emergent from the combination of real causal powers of the 'lower level' entities involved. Here I should acknowledge Tobin Nellhaus's suggestion in a personal communication that Bhaskar's domains might be emergent from each other, although I'm not sure whether he intended it in quite this sense.

To put the same point in a different way, any given event is the outcome of the actual interaction between the real causal powers of those entities causally involved in it, and the net outcome of these interactions depends upon the (purely temporary) relations in which these entities stand to each other at the time. This is directly analogous to the generation of the causal powers of a particular type of entity, which is the outcome of the interaction between the causal powers of its parts. The primary difference is that in the first case, the relations between the entities concerned are contingent and temporary, whereas in the second, the same set of significant relations is maintained over time as a result of the operation of morphostatic causes that maintain the structural stability of the entity, and hence there is a level of consistency in these causal powers over time.

Real and actual causation both therefore appear to be consequences of the same generic type of structural relation: the (diachronic) causal consequences that flow from a given set of entities existing (synchronically) in a given set of relations to each other. Actual causation, then, depends upon four of the five pillars of emergence: (a) a set of parts; (b) the relations between them; (c) the mechanisms resulting from the combination of these parts in these relations; and (d) the morphogenetic causes that bring this configuration of parts into existence at the moment of causation. Only the fifth pillar of emergence – the morphostatic causes maintaining the existence of this set of parts in this set of relations – is absent.

Nevertheless, there remains an important methodological distinction between real and actual causation: real causal powers always need to interact with other causal powers to produce an actual event, whereas actual causation by definition involves a complete set of the causal powers involved in an event's explanation.[34] This difference underpins the essential usage of the real vs. actual distinction in critical realist theory: the use of real causal powers as building blocks in the construction of explanations of actual events. It is inconceivable that we could produce viable explanations of events in terms of the unique configurations of entities at every level of stratification involved in each case unless there was some way of analysing such situations into interacting component parts. Now this argument may seem unattractive to critical realists, since it proposes an epistemological reason for preferring an ontological distinction. But it is an epistemological reason that is itself ultimately grounded in an ontological distinction: it

---

[34] I have discussed and dismissed some other apparent differences between real and actual causation in (Elder-Vass, 2005a, pp. 335-6).

is because of the persistence of entities and their consistent re-occurrence that their causal powers can be differentiated from the ongoing stream of actual causation, and hence it is because of that persistence and re-occurrence that it is valid to use those powers as building blocks in explaining actual causation.

The distinction between a 'persistent' or recurring whole and a fleeting one, is of course a matter of degree (and hence, on this argument, so is the distinction between 'real' and 'actual' causal potential). Although most cases may well be fitted easily enough into one category or the other, there is at least one interesting and important class of intermediate cases. These are the cases where a particular configuration with causal properties occurs first as an isolated incident – an apparently fleeting combination of entities – but subsequently acquires a set of morphostatic causes and hence is transformed into a persistent whole, an entity with real emergent causal powers. Where the acquisition of morphostatic causes is itself a path-dependent outcome, there may be no obvious inevitability to this transition.

This would seem to be the case, for example, in the development of certain social institutions. When the first boy picked up a football and ran with it at Rugby School for example, there was no institution of 'rugby football' that standardised this practice and made its reproduction likely. But today such an institution exists. What was a fleeting configuration of causal entities has been transformed by the creation of a set of supporting institutions into a social practice with real emergent causal powers – the game of rugby football. Many social institutions may have developed in this way – consider, for example, the origins of the insurance industry in a coffee house in London – although in many cases the first 'actual' interactions are lost to history and so we see only the fully developed 'real' form.

Real causal powers and actual causation, then, may be more similar than Bhaskar's division of them into distinct ontological domains would seem to suggest. Nevertheless, the distinction between the two is of considerable methodological value, and the account of cause that Bhaskar offers strongly complements the account of emergence offered in this thesis. Real causal powers *are* emergent properties, and Bhaskar's model of actual causation and multiple determination provides a framework for constructing causal explanations which recognise the complementary contributions of emergent properties at a variety of different levels.

## Reductionism

The theory of emergence is important to the explanation of social structure and agency because it shows how higher level and lower level entities can each contribute to the causation of events, and thus how *both* social structure *and* human agency can be causally significant in the social sciences. In other words, I claim that emergentism provides a viable alternative to a variety of false domain ontologies of the social sciences. Perhaps the most important and pervasive of these false ontologies is methodological individualism – a variety of reductionism which denies that social structures can be causally effective, insisting that all effects of social structures are derivative from the effects of human individuals and hence that social structures can be eliminated as causal factors from social explanations. In the coming chapters I will examine this claim specifically in relation to the social world, but in the present section I will be constructing a more generic refutation of reductionism, which will then be deployed in my response to reductionism in the social domain.

Even at the generic level, there are many varieties of the concept of reductionism, and I must begin by distinguishing between two of these: *eliminative reductionism* and *explanatory reductionism*. This section will go on to argue that emergent properties cannot be eliminatively reduced, and explain why. It will, by contrast, accept that emergence is compatible with explanatory reductions, and indeed argue that explanatory reductions merely confirm the causal effectiveness of emergent properties, rather than undermining it.

### *Eliminative reductionism*

Eliminative reductions, as described most famously by Ernest Nagel, occur when a higher-level theory is shown to be logically equivalent to a lower-level theory, with the result that the higher-level theory can be dispensed with entirely (Nagel, 1998). Now, although eliminative reductionism is often, as here, couched in terms that relate to a covering law conception of cause, the argument can be extended to a causal powers approach. In terms of causal powers, eliminative reductionism argues that the effects of a higher-level mechanism are nothing more than a summation of the effects of lower-level mechanisms, with the consequence that all properties and events at these higher levels can be fully explained by reference to properties of lower-level entities. Thus eliminative reductionists deny both the causal effectiveness of the higher-level entities

and their properties, and the need for (or value of) any science conducted in terms of these higher level properties.

I will contrast this below with the notion of an explanatory reduction; in an explanatory reduction, the higher level theory is explained by showing how it arises from lower level elements and the relations between them, but this is not taken to entail that the higher level theory can be eliminated. Emergentism is compatible with a generalised explanatory reductionism, but not with a generalised eliminative reductionism, since the point of emergence is that it explains the causal effectiveness of higher level entities that eliminative reductionism denies.

### Where does it stop?

Now so far in this description I have been somewhat vague about the dividing line between those higher level theories and entities that are to be eliminated and those lower level ones that are supposedly to reduce them. This reflects the fact that there are many different places where this line could be drawn. Eliminativists in the philosophy of mind, for example, draw this line between the mental and the physical. They argue that all causal value of mental entities or properties arises from their physical components and that explanations in terms of the mental can always in principle be reduced to explanations in terms of the physical, thus making explanations in terms of the mental ultimately redundant (Kim, 1993, p. 210). Analogously to this argument, methodological individualists in the social sciences claim that we should draw the line between social entities and human individuals, such that all social explanations can in principle be reduced to individual ones. And it is clear that a variety of other eliminative reductionisms could be advanced, drawing the line at any point in the hierarchy of entities and properties. Some versions of the argument imply that *no* entities or theories at a higher level than that to be eliminated can be sustained, while others simply seek to collapse two domains into one, leaving those above and below intact.

All versions of this argument depend on the belief that it is possible to justify the claim for the causal effectiveness of entities at *some* levels of a multi-levelled structure, while rejecting that of entities at other levels. Reductionists commonly deny emergentist claims for the latter levels, but do not seem to offer any positive argument to sustain their belief in the causal effectiveness of the levels they favour. Without such an argument, of course, reductionism is logically incoherent. A coherent reductionism must not only *dismiss* arguments for the causal effectiveness of higher levels, but also *establish* some for the causal effectiveness of lower levels.

Ironically, the only sort of argument available is the emergentist one itself. While it is conceptually possible that an emergentist argument could be made for reductionism, it would require a criterion of when emergence is valid and when it is not valid, based on an understanding of how emergence comes about, and a demonstration that this particular criterion does indeed explain why emergence fails at the claimed point. Yet eliminative reductionists sometimes argue as if they can make a generalised argument against emergence while still maintaining the irreducibility of their favoured ontological categories (see, for example, my discussion of King in chapter five below).

Thus, for example, the argument that all causal impact of mental properties can be ascribed to their physical components does not appear to be premised on any peculiarity of the relationship between the mental and the physical levels. Now, as I have argued in chapter two above, 'the physical' is not itself a single level, but rather can be broken down itself into multiple domains. The mental, it seems, is composed at the next level down of neurological entities, which in turn are composed (eventually) of molecules, which in turn are composed of atoms, and so on 'all the way down' to an undefined (and perhaps undefinable) bottom layer. Now, if a generic critique of emergence could be advanced to support the claim that there can be no mental causation (which I deny), it would also seem to substantiate similar claims that there could be no neurological causation, no molecular causation, no atomic causation, and so on all the way down. It could not therefore dispose of the claims of psychology without, at the same stroke, disposing of the claims of the whole of the biological, chemical, and physical sciences too, with the possible exception of the science of some presently murky fundamental level (see Humphreys, 1997, pp. 3-4 for another version of this argument ). Any case that is made against emergence *in general* undermines the ontological basis of lower-level explanations (except at some presently unknown fundamental level) just as much as that of higher-level explanations: these eliminative reductionists are merrily sawing off the branch upon which they sit.[35]

In principle, eliminativist arguments could avoid this reduction *ad absurdum*, since they could advance a theory that implies that emergence beyond a certain level is impossible, while remaining valid below that level. There is some such belief *implicit* in the eliminative position in the philosophy of mind – the belief that emergence is viable when the emergent entity is strictly 'physical' or 'material', but not otherwise. But there

---

[35] Durkheim made much the same point over a hundred years ago (Durkheim, 1974 [1898], pp. 28-9).

is no theoretical foundation for this belief; merely the metaphysical prejudice that only the material can be real. Methodological individualism in the social sciences often seems to imply another such prejudice – in this case, the belief, presumably founded in our personal experiences of agency, that human individuals *must* be causally effective agents. This is an argument I will return to in the chapter on agency.

Reduction *ad absurdum*, however, is only the first of eliminative reductionism's problems. The second is a practical one: when we replace a theory expressed in terms of a small number of higher-level elements with one expressed in terms of an inevitably larger number (and often very much larger) of lower-level elements, then any application of the theory will require the modelling of a very much larger set of interactions between those elements. If the higher level theory is viable, then the lower level theory redundantly requires knowledge of the initial states of the many lower level entities that make up the higher level ones, and modelling of their many interactions. In many situations, this becomes impractical, even given the huge increases in available computing power achieved over the last few decades (Holland, 1998, p. 118). As James has argued, "individualist explanations will clearly be enormously cumbersome" and their "sheer complexity might defeat the goal of explanation" (James, 1984, p. 53). Hence,

> if simplicity and applicability are allowed to be among the criteria for a good explanation, then it is by no means clear that individualism will surpass its rival. A realistic sense of what such reductions involve may lead us to question the point, as well as the feasibility, of the reductionist enterprise. If we can understand the social world in holist terms, why should we bother to reduce it to individualist ones? (James, 1984, p. 53).

This second argument, incidentally, applies to resultant properties as well as to emergents. It is usual practice, for example, to calculate the effects of gravity by assuming that objects have a certain total mass that can be treated as concentrated at a single 'centre of gravity' rather than by taking full account of the position and mass of each fundamental particle making up the object. The problem here is a practical one of calculability. This is an epistemological and not an ontological obstacle to reduction, unlike the first argument.

The third problem, though, is the most significant, and this will be addressed in the next section.

### Throwing out the baby

The biggest problem with eliminative reductions is that where the higher level is genuinely emergent, then any attempted eliminative reduction would eliminate elements

that are essential to a successful explanation (Laszlo, 1972, p. 20). To see why, let us consider the case in which it is claimed that the causal effect of an emergent property of a whole has been reduced by explaining it in terms of the properties of the parts, and the relations between them. For example, the liquidity of water over a certain range of temperatures can be explained as resulting from the way that its molecules try to bond with each other, which in turn is a consequence of their sub-molecular structure and their degree of movement at the energy levels corresponding to these temperatures (Gribbin and Gribbin, 1999, pp. 84-6).

This is the sort of "mechanistic" explanation that is incompatible with Broad's conception of strong emergence. It explains a property of a higher level entity (a body of water) in terms of the properties of its parts (hydrogen and oxygen atoms) and the way that they are related to each other when they take the particular form of water molecules (let us call this '$H_2O$ molecular bonds' for the purpose of this argument). It was because supposedly reductive explanations like this could be made of Broad's candidates for the title of emergent property that his emergentism fell into disrepute.

But this does not constitute an eliminative reduction at all. It would only be an eliminative reduction if the property of the whole could be explained purely in terms of the properties of the parts, ignoring any connective relations between them. Thus, for example, the mass of this same body of water is a simple sum of the mass of the hydrogen and oxygen atoms that compose it, irrespective of whether or not they are organised into water molecules, and we can therefore eliminate the entity 'water' and its distinctive properties from an explanation of this mass.

But when we seek to explain a property of water in terms of 'hydrogen and oxygen atoms and $H_2O$ molecular bonds', we have *not* eliminated the entity 'water' from the explanation of the property, for the simple reason that 'hydrogen and oxygen atoms and $H_2O$ molecular bonds' just *is* water. In such an explanation we have not replaced the higher level entity in our explanation, we have merely re-described it. If the higher level is to be explained by the lower level entities *and* the relations between them, we have covertly reintroduced the higher level back into the explanation, since it is nothing but the addition of these relations as an ongoing feature that distinguishes the higher level entity from the mere collection of lower level parts.

This is what I will call *the redescription principle*. This is the principle that if we explain a causal power in terms of the parts of an entity *x* plus the relations between those parts that pertain only when they are organised into the form of an *x*, then because

we have explained the power in terms of a combination – the parts and relations – that exists only when an *x* exists, we have not eliminated *x* from our explanation. All we have done is redescribed *x*.

To put the point another way: since an emergent entity is nothing more than its parts and their organisation, any explanation that depends upon the properties of its parts and on the characteristic way that they are related within this type of higher level entity is in fact an explanation in terms of the higher level entity. A resultant property *can* be explained without reference to the relations between the parts of the higher level entity. But emergent properties *depend upon* the existence of particular sets of relations between the parts of the entity possessing the property and so the higher level entity cannot be eliminated by any reductionist strategy from causal accounts that depend upon the exercise of its powers. Any attempted eliminative reduction of an emergent property will suffer from a loss of relevant structure – it cannot succeed without invoking a particular *configuration* of lower level entities as the relevant causal factor, but it cannot do so without reintroducing the higher level entity into the analysis.

Truly eliminative reductions, then, must take a different form – they must replace the higher level explanation with a lower-level explanation made in terms of the parts (and their powers) alone. They cannot depend on those relations between the parts that are characteristic of the whole whose powers are to be replaced in the reduction.

### In favour of explanatory 'reductions'

The argument so far in no way denies that we may be able to explain the relationship between higher and lower levels. It is not the attempt to explain higher levels that is eliminative reductionism's flaw; it is the belief that such explanations entail elimination. Even if they can be explained, emergent higher level properties are still causally effective in their own right.[36]

Although emergent properties cannot be eliminated by reduction, this does not mean that they cannot be *explained* in terms of their lower level parts *and their interactions* (Laszlo, 1972, p. 29). Holland, for example, despite dismissing simplistic reductions in terms of the parts alone, does nevertheless advocate an approach to emergence based on what he calls 'reduction'. But this is a form of reduction that no longer claims to eliminate the higher level in favour of the lower. What it does seek to

---

[36] In his recent work, even Kim has accepted this argument: "macroproperties can, and in general do, have their own causal powers, powers that go beyond the causal powers of their micro-constituents" (Kim, 1998, p. 85).

do is to explain how the higher level comes about, how it comes to be emergent, by examining its parts and their relations to each other. This is what I call explanatory reduction. The use of the term 'reduction' at all in this context is perhaps misleading, given the eliminative connotations it often seems to carry, but so many approaches to reduction take this form that it would seem idiosyncratic to refuse to use the term (in addition to the two quotes that follow, see (Campbell, 1974) for a classic statement of non-eliminative reductionism).

The point has been put superbly by both Fodor and Gell-Mann:

It seems to me (to put the point quite generally) that the classical construal of the unity of science has really misconstrued the *goal* of scientific reduction. The point of reduction is *not* primarily to find some natural kind predicate of physics co-extensive with each natural kind predicate of a reduced science. It is, rather, to explicate the physical mechanisms whereby events conform to the laws of the special sciences (Fodor, 1974, p. 107).

I know of no serious scientist who believes that there are special chemical forces that do not arise from underlying physical forces. Although some chemists might not like to put it this way, the upshot is that chemistry is in principle derivable from elementary particle physics. In that sense, we are all reductionists, at least as far as chemistry and physics are concerned. But the very fact that chemistry is more special than elementary particle physics, applying only under the particular conditions that allow chemical phenomena to occur, means that information about those special conditions must be fed into the equations of elementary particle physics in order for the laws of chemistry to be derived, even in principle. Without that caveat, the notion of reduction is incomplete... At each level there are laws to be discovered, important in their own right. The enterprise of science involves investigating those laws at all levels, while also working, from the top down and from the bottom up, to build staircases between them (Gell-Mann, 1995, p. 112).

One important consequence of this approach is that rather than eliminating higher-level theories, explanatory reductions do precisely the opposite: they provide extra justification for them by demonstrating that they are well-founded in the theory of the lower level, that they are consistent with other accepted bodies of theory, and indeed that they extend their explanatory power (Kitcher, 1998; Meyering, 2000, p. 181). In Gell-Mann's words, they are not eliminated but "cemented".

The causal powers conferred by relational emergence, then, can be explained in lower level terms, but they cannot be eliminated from scientific explanations in favour of lower level causal powers. Thus relational emergence succeeds where strong emergence fails: it provides a viable middle way between dualism and reductionism. It avoids dualism by allowing mechanistic explanations of higher level properties, and simultaneously avoids eliminative reductionism by showing why higher level entities and properties cannot be eliminated from scientific explanations

The primary conclusion of this section, then, is that eliminative reductionism can be rejected as a general thesis, since relationally emergent entities (as well as strongly emergent ones, if there are any) have causal powers in their own right. The implication is clearly that such entities can exert a causal influence on a variety of other entities. Some thinkers, however, have suggested that this is problematic when those entities are at a different ontological level, or in a different ontological domain. This is the question of *downward causation* and of *diagonal causation,* and the last section of this chapter will be devoted to showing why these 'directional' cases of causation do not constitute a problem for an emergentist theory of causation.

## The problem of 'directional' causation

The concepts of directional causation ('downward', 'horizontal', and 'diagonal') that this section discusses depend on a simplified representation of cause as operating between two entities. 'Horizontal' explanation, for example, describes the case where an entity at a given level affects another entity at the same level. In my account of emergence, this can be made more specific: an entity of a given class affects another entity of the same class – one molecule affecting another molecule, for example.

Now, as was argued earlier in the chapter, any given actual event is co-determined by a variety of interacting causes affecting a variety of ontological levels. The question of directional causation, then, is clearly an abstraction from the complexity of actual causation. Not only does it inherently assume a level-abstracted notion of the entities concerned, it also isolates a particular causal effect of one entity on another, neglecting other interacting causal mechanisms. In other words, when we are discussing directional causation, we are considering an argument that relates to the real causal powers of the entities concerned, and not to the more complex case of actual causation.

While focussing on causal mechanisms allows us to abstract from this complication, there are other complications we must recognise. For example, it may be misleading for some purposes to represent causal mechanisms as if they are purely the result of the powers of an 'affecting' entity. Causal effects are dependent not only on the ability of the 'affecting' entity to have an impact, but also on the ability of the 'affected' entity to be affected in this particular way. Planets can exert a gravitational force on objects, for example, but not on ideas. Critical realists usually represent this by saying that cause is a result of the combination of a power of the affecting entity with a *liability* of the affected entity (Harré and Madden, 1975, pp. 88-9). Rocks, for example, have a

liability to be eroded by wind and rain, whereas oceans do not. This, however, does not affect the argument that follows.

Now the default assumption in thinking about cause seems to be that causal mechanisms are horizontal. However, the meaning of 'horizontal' causation varies, depending upon whether we are assuming that emergence produces broad levels, or entity-based branching domains. Causation, for example, that is horizontal in the sense of operating between two entities that are both found within the broad level sometimes labelled 'the physical' – for example between an atom and a molecule – may well be upward, downward, or diagonal when we look at it in terms of more narrowly defined entity classes. One consequence of broad level-based approaches to stratification, then, may be to exaggerate the impression that causal mechanisms are typically horizontal by mis-labelling as horizontal mechanisms that are really instances of other directional types.

Whichever variation we are using, though, I argue that the assumption that all causal mechanisms are horizontal is an error. Indeed, I shall argue that this is another species of ontological prejudice, although perhaps a more subtle one than the desire to 'reduce' explanations to lower levels. This claim will be substantiated by making the case for the validity of other types of directional causation, beginning with diagonal causation.

### *Diagonal causation*

In general, 'diagonal causation' describes a causal mechanism in which an entity of one class affects another entity of a different class, but excluding cases where the two entities stand in a part-whole relationship to each other. There are a number of sub-types here, since the affecting entity may be a member of an entity class that is higher or lower in the emergence hierarchy than the affected entity, or indeed from a completely different branch of the emergence hierarchy. Thus, for example, there may be causal mechanisms in which a business corporation (higher) may affect a consumer; a free electron (lower) may affect an atom; or a weather system (different) may affect a rock formation. Note that I am including in 'diagonal causation' cases where the affecting and affected entities are from *classes* of entities that stand in part-whole relationships to each other, as long as the *instances* concerned are not in such a relationship. (Thus, for example, I do not regard the effect of a water molecule on an oxygen atom as a case of downward causation unless the oxygen atom is the one that is a part of that specific

water molecule). This is because the particular challenges posed by downward causation only occur when the specific instances concerned are in a part-whole relationship.

Now, the essential point of this section is that there is no difference between the logic of horizontal causation and the logic of diagonal causation. Diagonal mechanisms operate (in general terms) in just the same way as those that operate between two entities of the same class, and are just as valid. This is because in both cases the causal mechanism depends ultimately on the presence of the *level of organisation* represented by the 'causing' entity. In both cases, the operation of the higher level causal effect will depend on the causal effects of the parts, but as we have seen in the account of explanatory reduction, it is only when they are organised in the form of the 'whole' causing entity that they have this effect. There is no difference in the logical structure of this explanation between cases where the caused and causing entities are of the same type and cases where they are not.

Let me illustrate this for the horizontal case. Thus, for example, a molecule-molecule interaction depends on atom-atom interactions at lower levels, then on subatomic particle interactions, and so on. But the characteristic that allows us to validly label it a molecule-molecule causal mechanism is that it is only when those lower level parts are organised into the form required to constitute the molecule that they have the overall effect identified at the molecular level. Thus, the generalisation is only true between the molecules, whatever participation is required at lower levels to implement it. And the causal mechanism just *is* the intransitive referent of this generalisation.

Now this is equally true when the affecting entity belongs to a different class than the affected entity – say, when a person fires a gun. It seems reasonable to suggest that there are demi-regs that apply to such cases – e.g. 'when a person pulls the trigger of a loaded gun a bullet will usually be fired'. The fact that a whole series of lower level events are implicit in this statement is of purely supplementary significance – the generalisation depends on the characteristic properties of persons and loaded guns. The existence and behaviour of the cells, molecules, and so on at the lower level is implicit in the presence of the higher level entities, and is indeed a necessary element of the actual causal process, but it is not a sufficient element in the sense that without the *organisation* of those lower level entities that constitutes them into the higher level entities, they would not have this causal effect. Just as in the previous case, the mechanism can only operate when the entities 'person' and 'gun' are present, whatever participation is required at lower levels to implement it.

To assert the possibility of diagonal causation, then, is once again to deny a generalised eliminative reductionism. The reductionist response to this example would presumably be to claim that the 'real' causation was happening at some lower level. But this would immediately strip the explanation of all those characteristic features that depend precisely upon the entities being what they are – it would no longer be an explanation of a person firing a gun but only of some part of the process.  And there is no *quid pro quo* in the form of a superior ontological force for lower-level explanations, because they too are nothing more than statements of the contribution of a particular level of organisation to the causal process.

### *Downward causation*

I excluded the case of strict downward causation from the previous discussion primarily because of the significant attention that has been devoted to the problem of downward causation in the philosophical literature on emergence. The term *downward causation* was introduced by Donald Campbell (McLaughlin, 1992, p. 51), to describe the case of evolution by natural selection:

> Where natural selection operates through life and death at a higher level of organisation, the laws of the higher-level selective system determine in part the distribution of lower-level events and substances. Description of an intermediate-level phenomenon is not completed by describing it in lower-level terms (Campbell, 1974, p. 180).

As Klee points out, however, Campbell's example is not downward causation in the strict sense that I have adopted here, but rather an example of "determinative connections between two independently functioning systems" (Klee, 1984, p. 58) – in other words, diagonal causation. The most significant early advocate of true downward causation seems to have been Sperry, who illustrated the idea most graphically by arguing that in a wheel rolling down a hill it was the combined effect of gravity and the shape of the wheel that was responsible for the motion of an individual atom within the wheel (Sperry, 1969) (quoted in Klee, 1984, p. 57).  Kim has upped the ante by suggesting that emergentism logically implies downward causation, and therefore by implication the concept of emergence itself stands or falls depending upon whether the argument for downward causation can be sustained (Humphreys, 1997, p. 3; Kim, 1992, p. 121).

The argument, I suggest, *can* be sustained. Downward causation is merely a special case of diagonal causation, with the added twist that in true cases of downward causation the causal mechanism of a higher level entity is affecting its own parts. Before

dealing with the theoretical issues this raises, let me illustrate the principle with an example – the emission of photons (i.e. light) by a star.

To simplify enormously, the emission of photons by a star is the result of the extreme conditions of pressure and temperature in its core, and these in turn result from compression of the various nuclear particles that form the core of the star by the forces of gravity that are generated by the mass of the star itself (Gribbin and Gribbin, 1999, pp 189, 195). Now the point here is that the emission of photons can in a sense be accounted for by the interaction between the particles themselves, but that interaction itself presupposes a certain set of relationships between the entities concerned (proximity, temperature, etc) and that set of relationships only occurs as a result of the existence of the star. The same particles organised in some other way – e.g. distributed evenly across space – would not emit photons, hence the emission of photons can only be accounted for by combining the part played by the particles with the part played by the relationships between them, and the relationships between them are precisely what constitutes them into a star. It is only when those particles are arranged in that manner that a star exists, and only when they are arranged in that manner that photons are emitted. Thus the emission of photons from a set of particles that would not otherwise emit them must be accounted for by the level and form of organisation that constitutes them into a star.

The star, then, has a downward causal effect on the particles, causing them to emit photons, which is another way of saying that this is the effect which the group of particles, *organised as a star*, has on individual members of the group. We can thus offer an explanatory but not an eliminative reduction of this causal mechanism – one which recognises that the role of the higher-level structure cannot be eliminated from the story without doing violence to the causal account. Once again it would be pure ontological prejudice to insist that the real causal work is going on only at the lower level when both levels are necessary to the process concerned.

### Challenges to downward causation

Some authors, however, perceive an inconsistency between the idea of a higher level whole having a causal effect on one of its parts while the whole is itself constituted by that part (amongst others). Stephan, for example, criticises Sperry for his claim that "emergent phenomena … have a causal impact qua emergent phenomena on the very microstructure that determines the emergent phenomena" (Stephan, 1992, p. 44), and Klee criticises Sperry in similar terms (Klee, 1984, pp. 60-61).To Stephan, at

least, this seems to suggest a circularity, or an overdetermination, in which different and incompatible states of the lower level entity may be simultaneously mandated by the various causes working at different levels.

Now I suggest that this apparent problem comes from the neglect of the role of time, and in particular from the neglect of the different time-status of cause and composition.[37] In downward causation, a higher level entity causes a change in one of its parts over a period of time – cause is a diachronic relationship. But the composition relationship is a synchronic relationship – it is a logical statement of the relationships that must exist between a group of parts at a given moment in time (let us call this time *t*) for them to constitute a whole of a given type. Remember that composition is not in itself a relationship with any determinative force; it is only binding to the extent that the (diachronic) morphostatic causes that maintain it continue to do so. For the higher level entity to have a causal impact of any kind, the relevant morphogenetic and morphostatic causes must have led to the satisfaction of the entity's compositional requirements at time *t*. Hence the state of the system at time *t+1* is determined by the combination of these morphostatic causes, if they are still operating, with any other causal mechanisms that happen to be operating, including the downward causal mechanism generated by the state of the higher level entity at time *t*. The outcome, logically, may include changes in the parts that are consistent with the continuing existence of the whole, or changes in the parts that destroy the structural integrity of the whole, or indeed changes in the parts that transform the whole from one type of higher level entity to another. Thus the part played by a downward causal mechanism may even in some actual cases be the critical factor in destroying the entity possessing the mechanism – suicide, for example. None of these outcomes is inconsistent with the compositional consistency requirements that describe the *initial* conditions in which the whole will be formed by the parts, since it is always contingent whether these conditions will be maintained over time, and there is no reason why a causal power of a higher level entity at time *t* should not be a factor in affecting whether these conditions continue to exist at time *t+1*.

This picture of downward causation should enable us to clarify one last challenge that has been raised in the literature on downward causation. Kim argues that downward causation implies that "these 'higher-level' mental events and processes cause lower-level physical laws to be violated" (Kim, 1992, p. 120). There is no such

---

[37] My argument here is similar in some respects to Archer's critique of the role of time in Giddens' approach to social structure (Archer, 1982, pp. 466-471; Archer, 1998, pp. 358-60).

implication. The causal mechanisms arising from higher levels of organisation supplement those arising from lower levels, they do not violate them.[38] In the case of light being emitted from a star, for example, the star has a downward causal effect on the particles that it causes to emit photons as a result of bringing them into a relationship in which they exercise causal mechanisms that they already possessed, but would not have exercised had they not been organised into a star.

To put the point slightly differently, in this case *both* the lower level particles and the star taken as a whole have causal powers which contribute to the outcome. The outcome depends upon both being present, although there may also be other configurations in which a different higher level causal power combines with the same lower level causal power to produce a similar event. Thus, for example, a particle may be induced to emit a photon when it is not part of a star, but for this to occur, some other higher-level configuration must be created that has this effect. For example, a scientist may set up an experiment which induces a particle to emit a photon, but in such a case there is still a higher-level entity exercising a causal power to co-determine the outcome: the experimental apparatus itself.

Thus, the star example nicely illustrates the point that events may be co-determined by the causal powers of higher and lower level entities, even where the lower level entities concerned are parts of the higher level entity concerned, and even where the events which result are changes in those very same lower level entities. If this is the case in the natural world, then it may also be the case in the social world, and the application of this argument to the concepts of social structure and agency will be at the heart of the remainder of this thesis.

**Conclusion**

This chapter has constructed an account of cause in an emergent world and examined some of its implications. Following Bhaskar, it has shown that such an account depends on a careful separation of real causal mechanisms from the actual causation of events, so that we can see the latter as the outcome of an interacting set of mechanisms. When we apply this model in an explicitly emergent world, we must recognise that causal mechanisms are abstractions from the multi-levelled nature of actual entities and events, and that those mechanisms arise from the 'extra' *organisation*

---

[38] Bhaskar and Mill express this same view in the quotes cited earlier.

that appears with the emergence of each class of entity. Therefore, actual downwardly-inclusive events are to be explained as the outcome of an interacting set of level-abstracted real mechanisms. While I have argued that the distinction between real causal powers and actual causation is more methodological than ontological, it is nevertheless a useful methodological distinction for understanding cause in a universe built of entities with emergent properties.

With this account, we can tackle the questions of reductionism, diagonal causation, and downward causation. Given that causal mechanisms arise from the *organisation* that appears with each class of entity, there is no ontological reason why mechanisms at one level should not affect entities at another. The actualisation of such causal relations will involve further causal relations at lower levels; but these too are nothing more than the outcome of the organisation that appears at that level, so there is no reason to privilege them in the explanation. This argument is unaffected by the question of whether or not the affected entities are parts of the affecting entity, although careful explanation of the timing issues involved has been offered to make clear why this is so.

This conception of cause applies not only to the 'external' world, but equally to human social behaviour itself, and this is a theme we will return to as we move on now from the general theory of emergence to the application of that theory to social structure and human agency.

# 4        Social Structure

        The central claim of this thesis is that social structure is a product of emergence. This chapter thus presents the core argument of the thesis, at least in its basic form: that any viable theory of social structure must define and explain social structures in terms of the emergence of *social entities* and their properties. The chapter therefore moves us on from metatheory into the domain theory of the social world, as it begins the task of applying an emergentist ontology to the understanding of specifically social structures. It argues that social entities, composed primarily of human individuals, have emergent properties or powers in their own right, and that if 'social structure' is to have some explanatory value it must relate to these properties or powers.

        Social events, I argue, are produced by the interaction of the causal powers of a variety of social (and indeed natural) entities, just as natural events are produced by the interaction of multiple causal powers. Some of the complexities involved in understanding the process of social causation are therefore the same as those involved in understanding the process of natural causation, but others are different, arising from the unique characteristics of social entities. A secondary theme of this chapter will be to show how an emergentist theory of social structure can accommodate both kinds of complexity in its account of the social world.

        The chapter begins with a discussion of the concept of social structure, by looking briefly at the typology of competing concepts of social structure offered by Lopez and Scott, and by introducing the distinction between structure-as-whole and structure-as-relations, which is essential to understanding these existing concepts and how they can be accommodated within an emergentist scheme. It then moves on to begin the construction of an emergentist social ontology by considering what types of emergent social entities may be important in sociology, before focussing in on one such type – organisations. This section seeks to show that organisations are a type of social structure that has emergent causal powers, and goes on to argue that this model has much wider applicability to social entities than might at first be apparent. The chapter then moves on to the secondary theme, examining some of the complexities of emergent social structure. One section examines the sources of dynamism and complexity of entities like organisations, and the next returns to Lopez and Scott's typology of

concepts of social structure to show how the sorts of complexity it identifies can be accommodated in an emergentist model of social structure.

## The concept of social structure

Despite its widespread usage in sociology, *social structure* is a term whose meaning is "strikingly nebulous and diverse" (Lopez and Scott, 2000, p. 1). As Lopez and Scott point out, "there is little consensus over what the word means, and it is all too easy for sociologists to be talking at cross purposes because they rely on different, and generally implicit, conceptions of social structure" (Lopez and Scott, 2000, p. 1). This section will start to sketch out the range of such conceptions by summarising Lopez and Scott's own typology of concepts of social structure.[39] It will then move on to address another important preliminary conceptual question that arises when we try to fit existing concepts of social structure into an emergentist framework: are social structures relations or things?

### *Lopez and Scott: three facets of social structure*

Lopez and Scott argue that:

> the history of sociology shows the long-term coexistence of two different conceptions of social structure. On the one hand, there is that which we identify as the idea of *institutional structure*. Here, social structure is seen as comprising those cultural or normative patterns that define the expectations that agents hold about each other's behaviour and that organize their enduring relations with each other. On the other hand, there is the idea of what we call *relational structure*. Here, social structure is seen as comprising the social relations themselves, understood as patterns of causal interconnection and interdependence among agents and their actions, as well as the positions that they occupy (Lopez and Scott, 2000, p. 3).

They attribute the roots of both of these conceptions of structure to the work of Durkheim. On one hand, they see the idea of *institutional structure* as deriving from Durkheim's *collective representations* – from systems of shared norms, values, and ideas that shape social behaviour. As they say, "Social institutions have their basis in the culture that people share as members of a community or society. They are, at heart, cultural phenomena" (Lopez and Scott, 2000, p. 21). Institutional structure was most characteristically advocated by Parsons and the structural functionalists, and examples include both large-scale institutions like marriage, patriarchy, property, and contract, and also "the micro-institutions of day-to-day existence, such as those concerned with

---

[39] A useful history of approaches to social structure is also provided by (Crothers, 2002).

queuing, turn taking in conversations, dinner party entertaining, and gift giving" (Lopez and Scott, 2000, p. 23).

On the other, they argue that *relational structure* is based in Durkheim's *collective relationships*. Relational structure was most characteristically advocated by Radcliffe-Brown and structural anthropology, for whom social structure is "the sum total of all the social relationships of all individuals at a given moment in time" (Radcliffe-Brown, quoted in Lopez and Scott, 2000, p. 46).

Each of these schools of thought largely ignores the concept of structure implicit in the other, but there have also been other thinkers who seek to link the two (and as Lopez and Scott point out, Durkheim has also influenced many of these) notably Mauss, Levi-Strauss, Foucault, and Bourdieu. With the work of Giddens, Foucault and Bourdieu, though, there appears what Lopez and Scott see as a third conception of social structure (Lopez and Scott, 2000, pp. 17-18, 90):

> According to this point of view, patterns of institutions and relations result from the actions of individuals who are endowed with the capacities or competencies that enable them to produce them by acting in organized ways. These capacities are behavioural dispositions, and so social structure has to be seen as an *embodied structure*. Embodied structures are found in the habits and skills that are inscribed in human bodies and minds and that allow them to produce, reproduce, and transform institutional structures and relational structures (Lopez and Scott, 2000, p. 4).

For Lopez and Scott, despite the past disagreements between advocates of these different conceptions, they represent not mutually exclusive approaches, but rather potentially complementary facets of "the basis of social order" (Lopez and Scott, 2000, p. 92). By implication, this argument rests on the belief that embodied structure provides a linkage between institutional structure, relational structure, and individual agency. Lopez and Scott's typology gives us a good introduction to the range of meanings of *social structure*, but I shall argue below (p. 119) that embodied structure is incapable of reconciling the other two approaches, whereas an emergentist approach to social structure can provide a more viable way of integrating these three facets of social structure. Ironically, Durkheim is also the key historical source of the emergentist approach to social structure, although this aspect of his work has been widely misunderstood and neglected (Sawyer, 2005, p. 100-101).

### Social entities or social relations?

A common feature of many conceptions of social structure has been to see structure as a set of relationships of some kind. There is, however, a "persistent ambiguity" (Williams, 1976, p. 253) in the meaning of *structure* that is neglected in

most accounts of social structure. As Raymond Williams explains, the word originally referred to the process of building, but:

> The word was notably developed in C17, in two main directions: (i) towards the whole product of building, as still in 'a wooden structure'; (ii) towards the manner of construction, not only in buildings but in extended and figurative applications. Most modern developments follow from (ii), but there is a persistent ambiguity in the relations between these and what are really extended and figurative applications of (i) (Williams, 1976, p. 253).[40]

> It is clear from the history of structure and structural that the words can be used with either emphasis: to include the actual construction with special reference to its mode of construction; or to isolate the mode of construction in such a way as to exclude both ends of the process – the producers… and the product, in its substantive sense (Williams, 1976, p. 257).

In other words, *structure* can refer to the whole entity that is structured by the relations between its parts, which I shall call *structure-as-whole*, or it can refer to the way that a group of things (generally the parts of a whole) is related to each other, which I shall call *structure-as-relations*.[41]

Now most accounts of social structure generally refer to *structure* as if it means structure-as-relations. Thus, for example, we have already seen Lopez and Scott describing *relational structure* as "the social relations themselves, understood as patterns of causal interconnection and interdependence among agents and their actions" (Lopez and Scott, 2000, p. 3). Their concept of *institutional structure* is less clearly a case of structure-as-relations, but it is even further removed from structure-as-whole: "Here, social structure is seen as comprising those cultural or normative patterns that … organize their enduring relations with each other" (Lopez and Scott, 2000, p. 3).

But if these structures are not structures-as-wholes, then what are they the structures of? In the extreme case (which Lopez and Scott reject), some accounts of social structure "hold that there is no whole or totality separate from the *structuring* activities and practices that are engaged in by individual actors" (Lopez and Scott, 2000, p. 5). Here structure-as-relations is held to exist without it structuring any whole at all; structure here is synonymous with a regular patterning of otherwise unrelated entities. But it seems more typical "to talk about social structure as the arrangement or pattern

---

[40] Crothers cites this extract without appearing to recognise the importance for our understanding of social structure (Crothers, 2002, p. 7).

[41] Elsewhere in this thesis I will generally rely on the context to make clear which usage is implied in each case. This seems more accessible than Collier's more rigorous suggestion that we use 'structuratum' as a synonym of what I call *structure-as-whole*, and 'structure' to mean only *structure-as-relations* (Collier, 1989, p. 85).

among the parts of a society" (Lopez and Scott, 2000, p. 10) – in other words, to see *society* as the whole that is structured by these relations.

However, *society* is an amorphous, poorly bounded and unclearly defined agglomeration that is more analogous to *nature* than to any causally effective natural entity. I do not suggest that there is no such thing as society, only that the coherence of any bounded concept of society is extremely problematic. The concept of the state, of course, is a different matter. States are organisations, usually with well defined spatial boundaries, at least in some respects. But they do not map neatly onto *societies*; there are many potentially cross-cutting social systems that follow different boundaries, or none at all (Walby, 2005). One consequence of globalisation is that less and less social entities are coterminous with states. But many – for example multinational corporations, religions, and families – have never structured themselves on this basis. Given the lack of coherence of the concept of society, it is hard to see how such a poorly defined entity could have real causal powers. Instead, I suggest, the concept of *society* is useful only as an umbrella term like *nature*, *humanity*, or *the animal kingdom* – as a label for the collection of all that is social. Most of the powers that have been attributed to societies, I suggest, belong to somewhat lower-level social entities. If we are to explain the impact of higher-level structure on human beings, then we must find some more determinate sorts of structures at an intermediate level between individual and society that can have more specific effects.[42]

There is more than a hint in the literature that specific varieties of structure-as-relations are taken to be these intermediate levels while continuing to believe that the structure-as-whole to which they correspond is nothing less than society itself. But relations as such can have no causal effect on the world. It is only when actual entities are related that the set of entities so related can have an effect; and, as I have shown in chapter three, when we claim that a set of lower-level entities and the stable substantial relations between them have a causal effect, this is synonymous with claiming that there is a higher-level entity formed from these parts and relations that is the causally effective element. Ultimately, then, the idea that structures have causal effects is incoherent if structure is taken to mean structures-as-relations and not structures-as-

---

[42] The neglect of these intermediate levels is a common problem in treatments of social structure. Mouzelis points out, for example, that "Parsons, following Durkheim, operates within a society-individual scheme that systematically ignores the complex hierarchy of actors that provides the bridge between individual role players on the micro-level, and systematic incompatibilities on the macro-level" (Mouzelis, 1991, pp. 18-19).

wholes. Those accounts of social structure that simultaneously treat structures as relations and also claim that structures are causally effective, I suggest, rely systematically on the persistent ambiguity identified by Williams to retain the appearance of coherence – they talk of structure-as-relations, while in fact the causal part of the argument relies on *structure* meaning structure-as-wholes.

This of course presents no problem for methodological individualists. Since they deny that social structure has causal effects, they can quite consistently talk of structure as relations. But it is an error to carry over such talk to any account of structure that claims social structure does have a causal effect in its own right, as the emergentist account does. I argue, therefore, that an emergentist account of social structure must be expressed in terms of the causal powers of the social entities – the structures-as-wholes – that are formed as a consequence of the structures-as-relations that are usually implied by references to social structure. But I have denied that these social entities are *societies*. What, then, are they? This is the subject of the next section.

## The beginnings of a social ontology

Society, I argue, is populated not just by human individuals but also by a range of other social entities. Realists recognise that "each discipline has its own regional ontology" (Benton and Craib, 2001, p. 5), which identifies the types of entity that are the subjects of the discipline, and may clarify some of their more general characteristics. This section, then, starts to construct a regional ontology for the social sciences by enumerating some of the kinds of entity that are the subject matter of the social sciences (I make no claim for the completeness of this list):

1) *human individuals* – people. The status of human individuals as causally effective emergent entities will be investigated in depth in chapters six and seven below.

2) *organisations*. These social structures are the main focus of the remainder of this chapter, which will argue that these structures can have emergent properties and as a result can be causally effective. Organisations are made *from*, as well as *by,* human beings, since human beings are their parts, although it may be argued that sometimes they have other types of parts as well. Organisations arise from the power of human beings to co-operate – to use each other.

3) *human artefacts*. Artefacts are made by human beings (who thus enter their causal history as morphogenetic and sometimes also as morphostatic causes) from physical materials, and may have emergent powers of their own as a consequence of

their (designed) structure – aircraft, for example, may have the power to fly, while undercarriages, cockpits, tailfins, and food trolleys generally do not. Artefacts arise from the power of human beings to construct – to use tools.

4) *symbolic entities*. Symbolic entities are created by human interaction and used by human beings to communicate meanings. Examples include words, stories, theories, and ideologies. Clearly such entities have a central role in human social behaviour, and these too may be emergent in their own right – as argued by Archer under the heading of 'cultural emergent properties'(Archer, 1995, pp. 179-183). They are implicit in a number of the arguments of this thesis, but a detailed discussion of them is beyond its scope. Indeed I expect this to be a major area for further research. They arise from the power of human beings to communicate – to use meanings.

5) There are also *hybrid types* of the above categories; *higher-level* entities that are built from various combinations of them. Thus, for example, a book would seem to be a combination of symbols and artefact, with emergent properties not possessed by either. A more important example for the purposes of this thesis, however, is provided by *institutions*. Institutions are social practices that are followed consistently as a consequence of shared beliefs – as we saw in examining Lopez and Scott's typology of structure.[43] The ontological status of shared beliefs, however, is problematic. They appear to depend upon both human individuals and symbolic entities, but I defer more detailed consideration of their ontological structure for future work. Institutions are discussed briefly later in this chapter (p. 116).

The objects of study of the social sciences, then, comprise both human individuals and a variety of types of higher-level structures that arise as a result of human activity: as a result of humans acting upon physical things, as a result of humans acting upon each other, and as a result of humans acting upon ideas. I refer to these higher-level structures in this paper as 'social entities'. Unfortunately – or perhaps fortunately – for social scientists, the interaction between these various entities is enormously complex (Byrne, 1998, p. 20). As has already been noted, this is also true of the entities studied by the natural sciences, but it will be argued below that there are some varieties of interaction that are unique to the social sciences, and that pose particular problems for a purely naturalistic account of them.

---

[43] We must be cautious in reading the literature, however, to recognise that *institutions* is often given a wider sense than this – one that includes organisations. Durkheim, for example, talks of the state and the family as social institutions (Durkheim, 1964 [1901], p. xlvi).

**Organisations**

A great many social scientists have denied that social structure can be causally effective. In the realist tradition, a prime example is provided by Varela and Harré, who regard any such belief as "the fallacy of reifying a property of a group of social actors into an entity" (Varela and Harre, 1996, p. 314). Archer, in particular, has already replied effectively to such arguments, for example in (Archer, 1982), but their persistence indicates that a more detailed analysis of the case for the causal effectiveness of social entities would still have value. This section will develop such an analysis, focussing on the case of organisations, perhaps the type of social entity whose causal efficacy it is most straightforward to explain.

*Roles and the structure of organisations*

Any organisation, I argue, is an emergent entity composed of a group of human individuals, structured by a set of relationships between them. It is common in sociology to call these relationships the *roles* of the people who occupy the particular *social positions* in the organisation (Biddle, 1986, pp. 68-9). Such roles implicitly represent rules that define how the incumbent must relate to other members of the organisation, and also how they must relate to outsiders when acting on behalf of the organisation. Lopez and Scott, for example, write

> Each social position defines a role in social life for its occupants… Roles are definitions of those things that people are expected to do in the various situations that they encounter in their lives… They specify the rights and obligations that are entailed in social positions (Lopez and Scott, 2000, p. 29).

Occupying a social position or role means (a) to be recognised as occupying it by the other relevant role incumbents, and (b) to perform the behaviours that define the role.[44]

In the terms of an emergentist ontology, roles are not entities but relations – hence they are not constituted by parts but instead are occupied or performed by actual people. They are, as Bhaskar puts it, "the 'slots', as it were, in the social structure into which active agents must slip in order to reproduce it" (Bhaskar, 1998b, p. 40). They can therefore only have causal influence in the sense that, and to the extent that, they are so occupied, or to the extent that the role incumbents 'adopt' their characteristic

---

[44] I follow the common practice of calling the occupants of a social position *role incumbents*, although strictly speaking this is inconsistent with the usage of *role* and *social position* described above.

behaviours – which is of course another way of saying the same thing. Now, when a role incumbent adopts the behaviours defined by a role (e.g. answering the phone in the call centre if they occupy the role of a call handling agent), we have a case of *downward causation*, in the sense that the behaviour of the role incumbent is influenced by (their understanding of) the institution's expectations of (rules for) a holder of that position.

This, of course, is to claim that human behaviour is caused, which may make some social scientists uncomfortable, but it is not to claim that it is ever caused exclusively by a single factor. I make no claim here that the role incumbent's behaviour is 'determined' by the social structure – in this case by the rules of the organisation – as this would be to claim that there is only one causal factor operating on the incumbent. Rather, I argue that the action of the role incumbent is co-determined (as in all cases of actual causation) by a variety of causal powers, including the causal power of the organisation as exerted through its rules, as well as the causal powers of the individual role incumbent (cf. Archer, 1995, p. 184). Thus, the organisation has a causal effect on the role incumbent, although this effect, like any causal influence, does not fully determine a necessary outcome.

### Whose causal powers?

To the extent, however, that this causal mechanism is effective, the behaviour of the role incumbent 'in the role' is part of the behaviour of the organisation,[45] and the causal effects of the organisation are the aggregate of the causal effects of its role incumbents when they do act in role. Now, a methodological individualist would argue that this reduces the behaviour of the organisation to that of the individuals and there is no need for the organisation at all in this explanation ( e.g. King, 1999b, p. 271). However, the argument made in chapter three against eliminative reductions in general is perfectly applicable to this case. The role incumbents have the effects that they do when acting in these positions only because they are organised into this organisation through their performance of these roles. If there were no organisation there would be no such positions or roles and the people would behave differently. Hence the causal effect of the organisation cannot be eliminated from the explanation of this behaviour.

Similarly, if there were no organisation, then those with whom the role incumbents interact would treat them differently. I would not hand over my money to a

---

[45] Mouzelis seems to intend something similar when he talks of "the type of action that results from the incumbency of authority positions" as a case of "macro action" (Mouzelis, 1991). See my discussion of *mega-actors* below.

person in a shop, for example, unless I believed they had, through their role incumbency, the right on behalf of the business they represent to give me the goods I expect in return. Although I am served by a person I am served by them as a role incumbent and expect them to act in a certain way – to represent the retail business that owns the shop and its stock – as a result of this role incumbency.[46] This effect on my behaviour is itself an emergent causal power of the organisation.

To give another example, the employees of an organisation only accept and follow instructions from their managers to the extent that those managers have, through their role incumbency, the right to make such a request. In this case, the organisation has a downward causal effect on the employee's behaviour that has operated through a fellow role-incumbent; but it is nevertheless an effect of the organisation because the manager too is operating as a role incumbent and only has the authority to give an instruction because she operates as a representative of the organisation.

In discussing role performance, we must distinguish between the behaviour of an individual in general and their behaviour 'in the role'. Thus the chief executive's actions are part of the organisation's actions when she is seen as representing the institution in the terms of her role (e.g. when making an announcement at a corporate event). But they are not when she is acting outside the role – in a private capacity (e.g. when going for a swim), or when she is acting in some other role (e.g. when speaking as a candidate for election under the banner of a political party).

Even when a role incumbent is acting 'in the role', however, and thus on behalf of the organisation, this does not mean that their behaviour is entirely determined by the organisation or the role specification. The causal powers of the individual and of other factors also continue to co-determine such behaviour. This helps us to explain the otherwise problematic category of 'mega-actors', introduced by Mouzelis. These are individuals "whose economic, political or culturally based social power makes the consequences of their decisions widely felt" (Mouzelis, 1991, p. 107). Such actors can be influential in one of two distinct ways. First, they may have substantial influence in

---

[46] This is the first of several occasions on which I will use companies as an illustration of the properties of organisations. This rests on the assumption that 'company' is a sub-type of 'organisation'. Clearly sub-types have specific features that distinguish them from the broader type, e.g. the 'legal personality' of the contemporary limited liability company, and particular sub-types such as companies may represent historically specific forms of a more general type. They can still be used to illustrate the properties of the more general type, however, as long as those illustrations do not rest on features of the sub-type that are not shared by the more general type, and this is the strategy adopted here.

their capacity as private individuals. Thus, for example, a wealthy and prominent private art collector who patronises a particular style of art may have a significant effect on social tastes, and through this, for example, on the art-buying behaviours of both other private individuals and of organisations like public galleries and businesses corporations.

Secondly, and more relevantly to the current argument, individuals may be immensely influential by virtue of the way in which they perform their roles in organisations. Such roles can be performed well or badly, because role specifications do not completely describe how they are to be performed. Role specifications constrain acceptable behaviour in a role, and they may provide criteria for standards of performance, but they also provide resources that are available to the role, and leave open many alternative ways of performing the role. This is one of the crucial ways in which social roles differ from role-equivalents in the structure of natural objects: they provide the opportunity for flexible behaviour within the social position by its incumbents, and such flexibility enhances the possibilities for the individual role incumbent to have an exceptional impact, whether in the form of spectacular success or dangerous failure. Hence, for example, an exceptionally capable or lucky chief executive may contribute to the establishment of a dominant business corporation with a major impact on society. And, of course, mega-actors need not become so as a consequence of their own exceptional abilities. It is enough to be in a role that gives one personal influence over a powerful organisation. The President of the United States, for example, will inevitably be a mega-actor simply because of the combination of their personal discretion in performance of their role with the immense power of the US Government. Their actions in this role, unlike those of the wealthy private art collector, will be part of the actions of the organisation to which the role belongs, yet their position enables them to affect the behaviour of this organisation in a potentially influential way.

The behaviour of the organisation, to summarise this section, is the aggregate of the behaviours of its role incumbents 'in the role'. Although the relationship between these behaviours is additive, the organisation is nevertheless emergent, because it has a non-linear effect on these behaviours as a result of the fact that the role incumbents behave differently as role incumbents than they would have done 'in isolation' (i.e. if they were not incumbents of these roles).

### *Role incumbency and organisational morphostasis*

Thus far, I have given an analysis of organisations that addresses three of the five elements required for a full analysis of a case of emergence - it has identified the components of an organisation (people), the relations that constitute them into the organisation (roles), and how this gives the organisation emergent properties not possessed by its parts. A full analysis, then, still requires an account of the morphogenesis and morphostasis of organisations. Such an account is mostly beyond the scope of the present chapter, but there is one aspect that is worth examining: the relationship between role incumbency and organisational morphostasis.

The morphostasis of an organisation requires that (a) it has incumbents for all essential roles; and (b) those incumbents act within the expectations for their role. Now, there is, in the account so far, no necessity that these requirements will continue to be met. Being a role incumbent, for example, is purely contingent. A role incumbent may choose to leave the role (in most contemporary organisations) and if they do so then any downward influence of the organisation on the former role incumbent will cease. But organisations are a type of entity that has the ability to survive the exchange of token parts, in the terms described in chapter two. They can therefore provide for their morphostasis by replacing role incumbents who leave essential roles. Similarly, role incumbents may fail to perform according to the definition of the role, which could undermine the performance and ultimately even the continuing existence of the organisation. Again, this is generally dealt with quite simply – either by removing and replacing the incumbent, or by managing their performance (e.g. by further training or by disciplinary threats) so that it does start to conform to the role's requirements. Any organisation that is unable to deal with either of these sorts of problem is likely to fail and dissolve – although there are also of course, many other problems that could lead to such a result, and a successful organisation must have morphostatic processes in place to defeat these too.

It is worth a brief digression to examine how this story compares to similar stories in the natural world. We might argue, for example, that a hydrogen atom has a certain role as part of being part of a water molecule, that the water molecule only exists and has the powers of a water molecule as long as the hydrogen atom (or another equivalent one) performs this role, that countervailing causes may interfere with its continuing in the role in which case it may leave the molecule resulting in the dissolution of the molecule, and so on. Other sorts of natural entities may mis-perform

their roles – for example, diseased cells in a biological organism – with the result that a whole of which they are part suffers performance degradation or ultimately dies or is otherwise dissolved. And natural entities may have morphostatic mechanisms (e.g. an immune system) that defend them against such threats. In other words, some aspects of the concept of role and the performance of roles are not unique to human organisations or dependent upon any specifically human property.

On the other hand, the *particular* way in which human beings perform roles is a product of the way that human beings act in a wider sense. Factors such as consciousness, reflexivity, and the concept dependence of human action mean that there are a whole variety of different factors involved in securing human role performance – in particular, humans must be *persuaded* to perform a role. Unlike the parts of most lower-level natural entities, they cannot just be held in a particular spatial relationship as a result of which role performance becomes automatic – hence the importance of social power and socialised norms and values.

### Informal organisations

We are accustomed to thinking of organisations as rather formal structures, with roles and the procedures associated with them defined by constitutions, organisation charts, rule books, operations manuals, job descriptions, and so on. However, as many authors have shown, even the most formal organisations also depend upon informal relations between their members in order to function effectively, and roles within them may be defined *de facto* as much by these informal relations as by the *de jure* documentation that purports to be authoritative. In practice, the structure of any organisation has both a formal element and an informal element.

Some organisations, however, are more formal than others. Mature governments and large business corporations, for example, tend to be rather formal and bureaucratised, whereas smaller organisations are often more dependent upon personal relationships. We can represent this by seeing formal and informal organisations as two ideal types, with actual organisations representing a varying mix of the two. In more formal organisations, more of the behaviour of the organisation and of its members acting in role can be explained by reference to the documented rules and procedures. In less formal organisations, personal influence has a more significant effect, and we may, for example, need to adopt a social network model of the organisation, weighting the contributions of individuals by a variety of measures of their influence, if we want to explain why it and its members behave as they do (see, for example, Castells, 2000).

As we move towards the ideal type of informal organisations, we find a number of varieties of social entity that continue to fall within the structural range of the entity type even though they are rather far from our typical stereotypes of it. These are still emergent structures with causal powers that emerge in the same general way as those of more formal organisations. Thus, for example, families, married couples, and communities may exercise emergent causal powers as a unit – powers that are, by implication, different from those that would be exercised by their members were they not parts of these units. The explanatory power of the analysis of organisations presented here, then, is potentially much wider than it might at first appear.

The boundaries of the type, however, are not yet well enough defined. A particularly interesting case that could arguably fit within it at the informal end of the scale is provided by Goffman's analysis of interaction situations (Goffman, 1956). Such a situation may be rather short-lived by comparison with most organisations, but it does consist of a group of human individuals (and often more than one group of individuals – e.g. customers and staff) whose behaviour is guided by roles. These roles may often be informally or even implicitly allocated, but they do follow culturally well-established patterns and can be relied on to be instantiated in essentially similar forms in many different cases of actual interaction. Goffman himself sees such social encounters as "those entities in social life that come into being whenever persons enter one another's immediate physical presence" (Goffman, 1956, p. 246) and a number of authors have suggested that there is an emergence process at work here (for example Brante, 2001, e.g. pp. 185-6; Sawyer, 2005, pp. 198, 210-14). If this is so then it would seem to have the same form as the process at work in organisations: the members of the encounter act in a different way, and are able to have different effects on others, as a consequence of being part of an encounter of a particular type. Typically these effects are on other members of the same encounter, so here we have an instance of what Bhaskar calls *intrastructuration* at work, but there is no reason why such effects can not have wider implications, as for example in the encounter between Churchill, Roosevelt and Stalin at Yalta cited by Mouzelis (Mouzelis, 1995, p. 18).

This same encounter serves as a useful reminder of the multi-layered nature of social events. This high-level social event of negotiating and signing a treaty, when seen in downwardly inclusive terms, included a whole range of smaller events. This range included the encounter between the three leaders, but also included at the very same moment the particular actions taken by those figures as human individuals in the

encounter, the biophysical movements of their bodies, the behaviour of the cells making up their bodies, and so on. As in all such high-level events, a vast range of entities with emergent causal powers interacted in its production across all of these distinct levels. There is no contradiction, therefore, in arguing that these individuals had a causal effect on the outcome, that the emergent properties of their meeting as a fleeting interactional organisation had an effect, and that the causal powers of the states they represented also had an effect, no doubt interacting with many other causal factors. Different causal powers will have had significant effects on different aspects of the outcome, and identifying the entities with causal powers involved in the overall process is only the beginning of the process of providing an explanation of any given aspect of that outcome.

These are just the same complexities, on the whole, as we are faced with in any case of actual causation. But the story told here illustrates at least one aspect of the causal powers of social structures that make causal explanation in the social world yet more complex than that in the natural world. This is the ability of humans to perform multiple roles – as Churchill, for example, could act simultaneously in the role of participant in a social encounter and in the role of prime minister of the United Kingdom. This introduces a level of complexity into the social sciences that is unknown in the natural sciences, in addition to those features like reflexivity and concept dependence that were introduced in the previous section. These uniquely human features lead to unique features of social systems, which are the subjects of the next section.

**Unique features of social systems**

Organisations have a range of properties that may be shared by other social entities but are rare or unheard of in the natural world. It is common in sociology to stress the role of meaning and culture when considering the uniqueness of the social world, and these are certainly important, but it is important to recognise that social systems may be distinctive in other respects too. This section considers a number of such respects, beginning with the exceptional degree of flexibility of form demonstrated by organisations – both structural and spatial. This flexibility, however, is only one of a number of characteristics of social systems that make them uniquely complex. Like natural entities, social entities present us with a complex field of interacting emergent entities. Unlike most natural entities, however, (a) social entities like organisations are

complex adaptive systems that are themselves in turn composed of complex adaptive systems; (b) social entities have poorly defined boundaries (not just spatially); and (c) social entities do not have exclusive parts – the same human individual may be a part of many different social entities that are not in turn parts of each other. Social entities (like at least some natural entities) are further complicated by the intersection of different dynamic properties possessed by different but interacting social entities, and even by the different dynamic properties of different aspects of the same entity. Let me briefly discuss each of these in turn.

### *The dynamism of organisations*

Social systems are not fixed in form; in the terms of chapter two, they are changeable structures, and hence can remain the same type of structure while changing their particular form as long as they stay within the structural range of the type.[47] Human beings, for example, develop over time – in our earlier years we grow, in our middle years our brains develop more and more useful connections and hence knowledge, and in our older years we develop signs of aging. Not only do we change, but our bodies also constantly rebuild and repair themselves, so that maintenance of the biological organism implies not only a constant change in the arrangement of the parts, but also a constant change in the parts themselves – token exchangeability – as the body continually replaces one cell of a certain type with another.

Organisations are also changeable structures with area ranges, and also exhibit token exchangeability, but they also go through at least one more type of change that is of extreme importance: they can change their form entirely, so that what was once one sort of entity with a particular set of causal powers that derived from its particular structure and parts can become a different sort of entity with different parts or a different arrangement of them and hence a different set of causal powers. As Buckley puts it, "persistence or continuity of an adaptive system may require, as a necessary condition, change in its structure" (Buckley, 1998, p. 86). Thus, for example, an absolutist state can change into a democratic state (and may have to if it is to survive), a colonial administration can change into an independent state, a business that starts off as

---

[47] Note that I use *system* as a synonym of *entity*. I use *system* where this is more consistent with the terminology of the existing literature, for example on complexity theory. This does not mean that I endorse functionalist usages of *social system*, in which this term is often identified with *society*, or with arbitrarily abstracted functional subsystems of societies.

a sole trader can change into a partnership and on into a private then a public limited company, and so on.

This does cause some difficulties in the use of labels to describe changing social entities. To quote Buckley again, "the complex, adaptive system as a continuing entity is not to be confused with the structure which that system may manifest at any one time" (Buckley, 1998, p. 86). Both are important. On the one hand, it is the particular structure at any one time that determines the powers of the entity at that time, whereas we must recognise that an entity that changes its structure also remains in a sense the same entity. We might label the first sort of entity, defined by its structure, as a *structural kind*, and the second, defined by its continuity, as an *adaptive kind*. Typically we label instances of these by giving structural kinds qualified names. Thus *monarchy* may be an adaptive kind, with *absolute monarchy* and *constitutional monarchy*, for example, as structural kinds representing different moments in its structural history.

Now this is a feature that is not unique to the social world, as is of course suggested by Buckley's use of the general language of complex adaptive systems. Biological species, for example, change over time, and if we suggest, for example, that 'homo sapiens' has evolved from 'homo erectus' (Diamond, 1992, p. 30), we will be labelling two structural kinds that represent different moments in the history of the adaptive kind 'human'. But social entities have some unique features that allow them to be adaptive in more dimensions than natural entities, with the result that they have a level of potential dynamism that is unparalleled in the natural world. In particular, they have the feature of *spatial flexibility*.

### *Spatial flexibility*

As a general rule, the entities studied by the natural sciences (at least insofar as they are structural kinds) assume relatively fixed spatial relationships of parts, either static arrangements or stably dynamic relationships, such as an elliptical orbit, or the physiological structure of a body. Similarly, such entities generally have clear spatial boundaries that distinguish their internal parts from their external environments.

But social entities do not depend on these sorts of spatial fixity. Some, like households, schools, and cities, usually assume a set of relations that is spatially constrained by particular physical contexts, such as a house, a school building and yard, or the physical layout of the city, but even these have considerable room for spatial diversity of inter-relationships within these contexts. Thus, for example, school children

move around within the school, they go on school trips, they may go off site for sports activities, and all while they are still within the confines of the 'pupil' role.

Other organisations may have even looser spatial relationships. In a business corporation, for example, some of its activities may be focussed on specific sites, such as office blocks or factories, but others may be much more far-flung, such as sales reps visiting clients or buyers attending trade fairs. It is of course this inherent flexibility of spatial relationships that makes possible the unlimited extension of social entities across space, and thus the phenomenon of 'globalisation'. More generally, it has been possible for social entities that developed in relatively localised spatial contexts to spread much more widely in space *without becoming new structural kinds*. Thus, social entities do not depend on structural adaptation to increase their spatial reach or alter their spatial distribution, with the result that such changes can occur relatively quickly and easily. This, of course, brings them into new situations that may in turn contribute to actual structural adaptation, but this is additional to the potential for dynamism that is already inherent in the spatial flexibility of social entities.

### *Organisations as complex adaptive systems*

The flexibility of structural form exhibited by social structures like organisations has encouraged thinkers in the systems tradition to see them as *complex adaptive systems*. The idea of a complex adaptive system is derived from general systems theory, and most thoroughly theorised in its latest incarnation, complexity theory. A complex adaptive system is not just complicated but also capable of developing its structure in response to its environment (Buckley, 1998, p. 69; Cilliers, 1998, pp. viii-ix; Waldrop, 1994, p. 11). Over the long term, the process of evolution by natural selection makes biological organisms – and thus human beings – complex adaptive systems. Over the much shorter term, however, human brains can be seen as complex adaptive systems; they may not adapt their outward structure within a human lifetime, but they constantly develop and reconfigure their internal network of neural connections in response to stimuli from the environment. This underlies the ability of human beings to learn and to adapt their behaviour.

Human beings, then, are complex adaptive systems. Such systems are harder to explain in causal terms because they are, at least in some respects, what I have called adaptive kinds: their causal powers change over time. Thus, for example, a graduate student may develop a new ability to write complex sentences on the philosophy of the social sciences, a politician may develop the ability to lie plausibly, and so on. This is

combined in human beings with the ability to monitor their own performance. The combination of adaptivity and self-monitoring means not only that the powers of such an entity can change, but also that they can have a tendency to increase or improve, as the entity tends to become better adapted to its environment.

Now social entities are also complex adaptive systems. They change their form over time, in response to changes in their environment. Furthermore, at least some social entities – i.e. organisations – have the capability to monitor their own performance, and hence a tendency to become better adapted to their environment (although this may not be obvious, since their competitors are simultaneously doing the same thing, so the *relative* performance of an organisation may not improve as its capabilities do).

Social entities, then, are complex adaptive systems composed of complex adaptive systems. Hence, not only their structure, but also their components, tend to adapt over time. Given that the causal powers of any entity depend upon both the causal powers of its parts and the way that those parts are organised to form them into the higher level whole, the causal powers of social entities are subject to a compound rate of change.

### *Organisations have poorly defined boundaries*

A further complication for the analysis of social entities derives from the spatial flexibility discussed earlier. Physical or material entities have clear boundaries, and all of the lower-level entities within those boundaries are generally parts of it. There is one class of common exceptions: entities that are 'just passing through', such as X rays passing through an object, or food passing through an organism's digestive system. These may have an effect on the entity, but they are not generally considered part of it. Social entities, however, often have much less tangible boundaries.[48] What is the physical boundary, for example, of a trade union? If a trade union is composed of its members, and its members may be located anywhere in the world, even when acting in the role of a member, the concept of a physical boundary to such a system is essentially meaningless.

The lack of a physical boundary to a system, however, creates a further problem, since the boundary of a material system provides a clear demarcation between those entities that are its parts and those that are not. It is reasonably uncontroversial to say

---

[48] Sawyer makes a similar point (Sawyer, 2005, p. 26).

that a trade union's parts are its members, but other cases are not so clear. I have assumed above that a company's parts are its employees, but in legal terms we could argue that it is really a company's shareholders that are its parts. It seems more sensible to argue that the company's shareholders are external parties who stand in a relationship of ownership to the company, but we need a more general criterion to help us sort out such ambiguous cases. One plausible candidate is the rule that a person (or other entity) is part of an organisation when they perform a role that involves acting on behalf of the organisation. This excludes people acting in roles like 'shareholder' and 'customer': while the continuing existence of the organisation may depend on its interactions with people in roles such as these, they are not parts of it (although they are parts of complementary social institutions, such as the practices of purchasing and shareholding). Rather, they are external entities whose causal powers contribute to the morphostasis of the organisation.

Despite being at first sight plausible, however, it is not clear that this rule is really watertight. If we test it with regard to universities, for example, it suggests that a university's employees are parts of it, whereas its students are not. In one sense a university's students are its customers and hence external to it; but they are likely to be subject to its rules, they may be called 'members' of the university either before or after they graduate, and at times they may be taken as representing the university – for example, when appearing in a university sports team, or on *University Challenge*. This suggests that in borderline cases 'being a part of an organisation' is not a binary property but an ideal type; that people in certain roles are on the boundary between being internal and external to the organisation. Buckley has even suggested that "when we deal with the more open system with a flexible structure, the distinction between the boundaries and the environment becomes a more and more arbitrary matter, dependent on the purpose of the observer" (Buckley, 1998, p. 36).

Luhmann points the way to an alternative criterion: "We can speak of there being a social system only to the extent that the system can be distinguished from its environment. Moreover, the agents themselves, and not merely sociological theory, must be able to recognize the boundary between system and environment"(Luhmann, 1982b, p. 139). On this basis, we might argue that a person (or other entity) is part of an organisation when they perform a role in which they believe they are acting as part of the organisation. But this still leaves question marks: is it their own belief that should count here, for example, or the beliefs of those with whom they interact in the role?

It may be, then, that there is no unambiguous general criterion for determining who is and who is not part of a given social entity; for the moment I simply conclude that although it is often obvious who is part of a particular social entity and who is not, social systems do not necessarily have well defined boundaries.

### Social entities do not have exclusive parts

It is also worth expanding on the implications of a further property of social systems that we have already noted above: that a person is only part of a social entity when they are behaving 'in the role' – the role they occupy within that entity. At other times, they cannot be considered part of that system. But this means that at other times they may act in some other role, as part of some other social entity. Thus, the same human individual can be part of multiple social entities – I can be part of a university, part of a married couple, part of the British Sociological Association, and part of a great many other social systems (cf. Cilliers, 1998, p. 7; Parsons, 1966, p. 1; Parsons, 1969, p. 354). This is quite different from the situation of most natural entities. Because of the physical boundedness of most natural systems, any entity that is part of one natural system cannot be part of another, except yet higher level systems of which the first one is itself a part. An atom, for example, can only be part of one molecule at a time. It could, of course, via being part of the molecule, also be part of a cell, and via the cell, also be part of a human being, but the molecule can be part of only one cell, and the cell part of only one human being. A person, however, can be part of many different social institutions by virtue of playing different roles at different times. To an extent, it is even possible for human beings to play multiple roles at the same time. Thus, for example, if a manager in a business promotes a fellow member of a secret society whose principles include mutual assistance, she will, in a single action, be acting simultaneously 'in the role' of manager and society member.

The consequence of this is to introduce yet another unique dimension of complexity to the social sciences. By virtue of their overlapping membership of multiple social entities, human beings are subject to (possibly conflicting or contradictory) downward influences from a variety of different organisations and institutions in which they are expected to play a role. On occasions, their role performance in one role is likely to be affected by the influence of the others (Biddle, 1986, pp. 82-4; Goffman, 1956, ch 4). The possession of multiple roles is of course a feature of modernity; this is not to say that people had only one role each in pre-modern societies, but the quantity of different roles occupied by the same person increases

radically with the differentiation of society, as well as the potential for divergence between the behaviours expected of the same person in multiple roles, and hence this is a dimension in which complexity is increasing over time (Luhmann, 1982a).

### *Interacting dynamics*

Finally, social entities (like at least some natural entities) are further complicated by the intersection of different dynamic properties possessed by different but interacting social entities, and even by the different dynamic properties of different aspects of the same entity. Here I am thinking of the distinctions drawn by complexity theorists between "near to equilibrium" and "far from equilibrium" systems, and mirrored by Buckley in his distinction between "equilibrial", "homeostatic", and "morphogenic" systems (Buckley, 1998, p. 69; Byrne, 1998, p. 30; Harvey and Reed, 1995, p. 305). Parsonian theory notoriously saw social systems as essentially equilibrating systems, and was widely criticised for its failure to allow for dynamic change.[49] Complexity theorists, by contrast, have tended to stress the sheer unpredictability of dynamic systems in a far-from-equilibrium state. In many respects, this offers the prospect of a more plausible social systems theory (Byrne, 1998, p. 8). But, although complexity theorists have striven to contain this unpredictability in the development of a mathematics of complexity, there is also a risk of an ironic inversion of Parsons in which the unpredictability of social systems is over-stressed.

The real challenge for social theory is to explain the combination of an element of predictability with an element of unpredictability in different aspects of the same entity. In the typical business corporation, for example, it is very predictable that as long as they keep being paid, and except when they are ill, the vast majority of its employees will turn up roughly on time for work on every working day. But it is far less predictable whether it will choose to branch out into new product lines, or into new countries, or to continue its market strategy much as before. Any attempt to model or predict the future of a company, then, must either combine "near to equilibrium" and "far from equilibrium" behaviour within the same model, or it must abstract from the total picture and model one of these behaviours while assuming that the other is not significant for this specific explanatory purpose. The Parsonian fixation on equilibrium, then, must be eliminated, but equilibrating tendencies in social systems cannot be ignored. On the contrary, it is the massive potential for dynamism and instability in

social systems, arising from the kinds of complexity discussed above, that makes it important to identify how social systems do retain the level of stability that enables them to function in the first place. We need to explain both stability and change in social systems – and the "constant creative tension" (Buckley, 1998, p. 46) between the two – within the scope of a single theoretical approach.

This may mean that prediction is only likely to succeed for the more stable aspects of social systems. But it does not mean that we must abandon the search for the causal powers of social entities, since it is only by identifying such powers and showing how they are combined in the generation of actual events that we will ever be able to explain social events. While, at the theoretical level, this calls for the same elements of causal analysis in the social as in any other sciences, at the methodological level the prevalence of highly complex and dynamic open systems in the social world rules out not only the experimental method but also, much of the time, the possibility of identifying causal mechanisms through obvious empirical regularities.

The emergentist approach to social structure is thus able to accommodate some major forms of complexity arising from the dynamic properties of structures like organisations and the potential for overlapping between their parts. In the next section we will see how it can accommodate a different source of complexity – the relationship between institutions and organisations.

**Unifying the three facets of social structure**

Given the analysis of social structure presented in this chapter, we are now in a position to evaluate Lopez and Scott's account of the three facets of social structure. As we saw earlier in the chapter, they label these three facets *institutional*, *relational*, and *embodied* structure. Institutional structure is defined as "those cultural or normative patterns that define the expectations that agents hold about each other's behaviour and that organize their enduring relations with each other" (Lopez and Scott, 2000, p. 3). Relational structure is "the social relations themselves, understood as patterns of causal interconnection and interdependence among agents and their actions, as well as the positions that they occupy" (Lopez and Scott, 2000, p. 3). And embodied structure, they argue, consists of "the habits and skills that are inscribed in human bodies and minds

---

[49] See for example, (Black, 1976) and (Hacker, 1976), although Parsons offers a plausible defence in (Parsons, 1976, pp. 337-9).

and that allow them to produce, reproduce, and transform institutional structures and relational structures" (Lopez and Scott, 2000, p. 4).

### *Linking relational and institutional structure*

Although the first two of these appear in the literature primarily as competing conceptions of social structure, Lopez and Scott argue that

> The concept of social structure points to a complex articulation of the institutional and relational elements of social life… [hence] analyses of institutional structure and relational structure offer *complementary*, not alternative, frameworks of sociological analysis. Sociology will prosper only if it recognizes this (Lopez and Scott, 2000, p. 4).

In particular, as their definition of embodied structure indicates, they see embodied structure as at least part of the link between the two:

> A more comprehensive account of the social structuring of human bodies has been set out by Foucault, and the extensions to this made by Bourdieu connect his work with both Giddens and the earlier writers on institutional and relational structure. For both Foucault and Bourdieu, bodies are seen as the carriers of relational and institutional structures (Lopez and Scott, 2000, p. 98).

Despite the link through embodied structure, Lopez and Scott are cautious about attempting to explain the relationship between relational and institutional structure. I suggest that there is a crucial missing link in their attempt to reconcile these various views of structure. This is the concept of emergence, combined with a clear understanding of the various elements that are required to explain it and indeed of the differences between different kinds of social structure. This section relates each of these three types of structure to their roles in the emergence of social structure using the five pillars for the analysis of emergent systems identified in chapter two. Somewhat loosely, the three facets of social structure identified by Lopez and Scott can be related to three of the pillars of emergence, as they apply to organisations (although we will also have to consider their potential role in other sorts of structure). Relational structure corresponds to the organisation of the parts in organisations. Institutional structure is more problematic, as I shall show by discussing its relationship to institutions, but in the form of norms and values it provides a key element in the morphostasis of organisations, and also part of the mechanism through which their emergent properties arise. And embodied structure also provides at least part of the mechanism by which the behaviour of social entities emerges from the behaviour of the people that compose them. Let me discuss each of these three in turn.

*Relational structure corresponds to the relations between parts*

The first of these correspondences is clear and complete – relational structure is nothing more nor less than the organisation of the parts of a social whole, and corresponds precisely to this element in the explanation of emergence in general. This correspondence may be obscured, however, in many accounts of social structure by a failure to recognise that there *is* a whole of which any particular relational structure is the organisation. This tends to encourage a reductionist approach as a result of seeing the parts, not as parts at all, but simply as separate entities (generally human individuals).

In other accounts, the correspondence may be obscured more subtly by the common assumption that the whole to which relational social structure in general relates is *society as a whole* rather than more specific social entities. This approach breaks the link between structure and the existence (and powers) of the higher level entity. There is at best a tenuous link between society as a whole and particular effects of structure on individuals, whereas the links between, say, a government and its effects, or a value and its effects, are often much clearer. Society, as was argued above, is not so much a well-defined entity with causal powers as an umbrella term for all things social – a heap not a system. Hence to see structure as the structure of society as a whole is to eliminate the very social entities whose causal powers must feature in an emergent account of social structure.

The picture is also confused, however, by the common assumption that there is only one kind of social structure that we need to explain. This is an assumption that follows naturally if we see social structure as the structure of a single undifferentiated whole called *society*, but which is open to challenge as soon as we recognise that there are intermediate social entities between individuals and society. For example, it may be important to differentiate between organisations in all their various forms on the one hand, and institutions on the other. Relational structure, as the substantial relations between individuals who are parts of structures like organisations, plays a clear role in the emergence of the causal powers of those organisations. But it may be less significant relative to institutional structure in the case of social institutions like gift giving.

*Institutional structure and institutions*

The relation between institutional structure and the five pillars of social structures is much less direct. Whereas *relational structure* refers directly to the relations between an entity's parts, *institutional structure* is far from being a simple

synonym of any one of the five pillars. In the context of organisations, it plays a number of distinct roles, but before we can make sense of these, we must consider institutional structure in the context of institutions.

Institutions, in the sense that is relevant here, are social practices that are common within a given cultural setting. This commonality arises from the causal role played in generating instances of the practice by particular beliefs – norms and values, in the most common formulation – which are themselves shared by many or all of those people participating in the relevant cultural community. There is some ambiguity in many definitions of institutions as to whether they refer to the social practices themselves or to the norms and values that produce them; this is an ambiguity that helps to obscure the problematic ontology of institutions. Examining this ontology inevitably brings us up against questions about the nature of meaning and culture which are essentially beyond the scope of this thesis, but a small incursion into this territory is necessary to make clear the relationship between institutional structure and the conception of social structure advocated here.

Let us consider why the word *structure* should be used here at all. There are at least two possible reasons for describing institutions as structures, or as having structure. The first is that they represent – or rather they *are* – a widely reproduced patterning of social behaviour. In a sense, then, they can be said to *structure* social behaviour, but this is quite a different sense of *structure* than any of those we have come across so far. This is neither structure-as-relations nor structure-as-whole, but rather structure-as-empirical-regularity. While this is no doubt a valid use of *structure* for some purposes, it has nothing in common with the idea of social structure as a causal force. There *is* a consistent causal force at work here – the norms and values themselves, which produce the regularities of behaviour in conjunction with a variety of other factors. But it is not the institutions, the social practices themselves. These are outcomes, and it is no surprise that questions should be raised about the independent causal effectiveness of institutional structure if it is conceived of in this way.

There is, however, an alternative way of conceiving institutions, which is to argue that shared norms and values themselves represent a higher level causal power than merely the beliefs of human individuals: in other words, that the causation of institutionalised social practices can be traced to social beliefs that exist at a higher ontological level than human individuals. The argument here is that our shared norms and values are emergent from individual norms and values and have a causal power that

acts back upon individual behaviour, just as organisations are emergent from groups of individuals and have a causal power that acts back upon them. This is precisely what Durkheim seems to have intended with his *collective representations*, but this is an idea that has received a bad press from sociological thinkers ever since (Durkheim, 1964 [1901]; Durkheim, 1974 [1898]).

In the absence of a deeper analysis of the emergence of meaning and culture, which I intend to pursue in a future project, this thesis will not take a position on the viability of Durkheim's collective representations or analogous ideas of cultural structures. However, what the argument so far does imply is that *institutional structure* is a very problematic concept in the context of institutions themselves. Either it represents nothing more than an empirical regularity of social practices, in which case it has little value in the theorisation of social structure other than as a proxy for the underlying causal impact of individually-held norms and values. Or it must be justified in terms of emergent causal powers of collective cultural beliefs, which few sociologists seem prepared to endorse explicitly (although arguably many rely upon them implicitly).

Despite these ontological difficulties, however, there seems to be widespread agreement that shared norms and values, whatever their ontological status, do produce empirical regularities in social practices. *Institutional structure*, therefore, can perform a useful conceptual role as a placeholder. It may be cashed out in the future by a Durkheimian process of collective cultural causation, or by an individualist process in which the relevant causal powers are attributed to the beliefs of individuals, but in the meantime, we can use the term to mean whichever of these turns out to be the case.

### Institutional structure and organisations

Given this ontological fudge, we can move on to say that institutional structure plays a key role in stabilising and reproducing the patterns of behaviour upon which organisations depend. The observance of role specifications by their incumbents is deeply dependent upon their belief that this is the appropriate thing for them to do, and this in turn is a product of their norms and values. Hence my linking of institutional structure to morphostasis in the introduction to this section.

But as I have already pointed out, this linkage is much less direct than that of relational structure to the relations between the parts of a social structure. Even in organisations, for example, not all morphostatic causes are provided by institutional structure – people may continue to play their parts in an organisation for instrumental

reasons, for example, as well as normative reasons, so that its stability may be caused in part by its success in meeting their needs as well as by the existence of norms and values supporting it.

Furthermore, norms and values may also play other roles in the terms of my five pillars. They may play a morphogenetic as well as a morphostatic role – for example, when people's values contribute causally to the establishment of a new charitable organisation or political party. Equally, they may play a role in the mechanism by which particular powers or properties emerge in particular social entities. Thus, for example, governments have the power to collect taxes, but only because individuals send out tax returns, fill in tax returns, send out letters demanding payments, and make those payments – and each of those activities is motivated in part by the norms and values held by the individuals concerned. An explanatory reduction of the power of governments to collect taxes thus depends upon institutional structure. Hence, institutional structure appears in the explanation of the social structure of organisations in a variety of roles – as morphostatic cause, as morphogenetic cause, and as part of the mechanism by which social causal powers emerge.

### The role of embodied structure

The potential for *embodied structure* to bring together the other aspects identified by Lopez and Scott depends upon the argument that norms and values themselves are embodied. Once again, this argument seems to be neutral with respect to the ontological questions about the status of shared beliefs; whether shared beliefs are emergent or not, there are certainly beliefs at the level of the individual, and these include the norms and values that guide the individual in their behaviour, inducing them on many occasions to reproduce common social practices.

It is perhaps a little controversial to equate the holding of norms and values with embodiment. Although some of our behavioural dispositions may be stored in somatic rather than neural form, it seems unlikely that this extends to many of the norms concerning role performance upon which the notion of social structure depends. It is only really if we accept that the body includes the brain (as Lopez and Scott do in their definition of embodied structure quoted on p. 94 above) that it seems plausible to argue that institutional structure can be embodied in this way. This seems to me the only tenable way to justify the claim that our beliefs can be embodied, but it is far from clear that all the advocates of embodiment would accept it. On this argument, embodied structure, in the form of each individual's 'copy' of shared cultural beliefs, plays a role

in the emergence of higher level social structures, since institutional structure can be embodied.

The advocates of embodied structure, however, seem to suggest that relational structure can also be embodied (Lopez and Scott, 2000, pp. 90, 98). Lopez and Scott interpret Bourdieu to say "The objective relations and institutions are incorporated – taken into the corpse or body – as subjective dispositions to act" (Lopez and Scott, 2000, p. 101). But the idea that relational structure as such can be embodied seems incoherent: individuals may embody attitudes to their relations to others, and beliefs about those relations, but the relations themselves are inherently external to them. Embodied structure, I argue, can not therefore include the second essential element in any explanatory reduction – the relations between the individuals concerned. Hence any explanation of the behaviour of a social entity in terms of the structure embodied in the individuals concerned will always be incomplete.

Embodied structure, then, has a potentially useful role to play in the explanation of emergent social properties. However, because embodied structure relates only to the properties of individual bodies and inherently excludes the relations between different individuals, any attempt to see social structure as *only* embodied structure is inherently reductionist.[50] I suggest we must reject this – ultimately methodologically individualist – move by recognising that embodied structure can provide only part of an explanation of the mechanism by which social entities can exercise emergent powers.

The emergentist view of social structure, by contrast, can provide the integrating framework that enables us to make sense of the inter-relation between the relational, institutional, and embodied facets of structure that can be found in the sociological literature. When seen from this perspective, it is clear that these three facets, when analysed appropriately, are indeed complementary partial accounts of social structure, but to understand their relationship to each other we need to analyse them in terms of the five pillars of emergence and with a clear view of what social entities we are considering.

---

[50] This is a theme I will return to in my discussions of Giddens in chapter five and of Bourdieu in chapter seven.

**Conclusion**

This chapter has developed the outline of an emergentist theory of social structure, by applying the general principles of an emergentist ontology to some of the specific types of structure produced by and from human individuals. The sorts of things that we will find in a social ontology, like organisations, institutions, cultural constructs and human artefacts, are ultimately emergent from the sorts of things we will find in a natural ontology, like atoms, molecules, and cells. Thus, we can see both as parts of an ontologically unified world in which everything that exists emerges as part of a single hierarchy of structures. At the same time, however, this is a differentiated unity; the sorts of powers and properties that belong to entities at each branch of the hierarchy vary, depending upon the varying properties of their parts, and the varying ways in which these parts are organised. As a consequence of these differences in the objects of study of different sciences, the methods appropriate to their study also differ.

This chapter has not sought to give a comprehensive account of the differences between the natural and the social sciences, and in particular it has neglected the important question of the meaningfulness of social action and hence those issues raised by hermeneutics, which will be returned to in chapter seven. It has also paid less attention than I would like to the important class of symbolic or cultural entities which clearly has a vital part to play in social systems and which I intend to return to in future work. But it *has* shown how some other specific differences between the structures of some social entities and those of some natural entities make it difficult or impossible to adopt similar methods in the study of them both. Most particularly, the spatial flexibility and the lack of spatial boundedness of social entities means that they can interpenetrate, overlap, and proliferate in ways that are denied to physical entities. In addition, social entities seem unique in being complex adaptive systems composed of complex adaptive systems, and hence are subject to a higher order of structural change than most other entities.

These two factors combine to make the dynamics of social systems complex in dimensions that are unavailable to physical systems, and hence contribute to the inherent unlikelihood of us observing 'constant conjunctions' of events in the social world. Those relatively frequent conjunctions that we do observe are testimony to the presence of equilibrating forces in at least some aspects of some social structures. But we cannot usefully theorise social systems as either purely 'equilibrating' systems, or as

purely 'far from equilibrium' systems; we will have to combine elements of both dynamics if we are to explain the patterns of events we observe in society.

This chapter has also introduced some conceptions of social structure from the sociological literature. Lopez and Scott identified three facets of social structure – institutional, relational, and embodied structure. I have agreed with them that these are potentially complementary aspects of social structure, but I have argued that an emergentist approach is necessary to understand how they are related. In particular, institutional structure is a problematic concept that demands further study of the ontology of meaning and culture, and embodied structure can not unify institutional and relational structure. Instead, we must analyse these aspects of structure in terms of the five pillars of emergence for each type of entity involved.

The next chapter turns to a different typology of concepts of social structure, which classifies them in terms of differences in fundamental ontological approaches. These are the approaches associated with methodological individualism, methodological collectivism, realism (and related schools of thought), and structuration theory respectively. Chapter five develops a critique of a variety of these alternative ontological positions, based on the emergentist account of structure developed in this chapter.

# 5　　　Theories of Structure

　　This chapter will develop a constructive critique of existing theories of social structure in the light of the argument so far. These theories can be classified into four groups, representing four different ontological perspectives on social structure: methodological individualism, methodological collectivism, emergentism, and structuration theory. The objective of this chapter is to show, on the one hand, why emergentism is superior to the alternative approaches, and on the other, how the emergentist approach advocated in this thesis relates to other recent emergentist approaches.

　　It will begin by defining these four perspectives a little more thoroughly, using the typology of approaches to structure developed by Douglas Porpora as a reference point (this is the only section of the chapter which engages directly with methodological collectivism, although I will also cover it later in chapter seven). The next section will engage with the debate between the realist emergentist Margaret Archer and the methodological individualist Anthony King. This section has two objectives: both to show how the version of emergentism advocated here disposes of methodological individualist objections, but also to illuminate the complementarities and differences between my account of emergentism and Archer's. The chapter will then move on to another interesting contemporary account of emergence in sociology, that constructed by R Keith Sawyer on the basis of the concept of multiple realisability. It will argue that multiple realisability is not in itself a source of emergence; rather, the variety of emergence described by Sawyer is a sub-case of the relational variety espoused in this thesis. Finally, it will examine the much-discussed attempt by Giddens to resolve this question via his *structuration theory*. Although this section will endorse and enhance Archer's critique of Giddens' conflationism, it will also suggest that there is something to be gained from Giddens' work.

## Four concepts of social structure

　　Porpora provides a typology of approaches to social structure that focuses on a different dimension of methodological controversy from Lopez and Scott, and he

emphasises not the compatibility of the different positions but their incompatibility. He lists four different concepts of social structure:

1 Patterns of aggregate behaviour that are stable over time
2 Lawlike regularities that govern the behaviour of social facts
3 Systems of human relationships among social positions
4 Collective rules and resources that structure behaviour (Porpora, 1998, p. 339).[51]

Each of these corresponds to a particular position in the perennial debate over methodological individualism versus methodological collectivism (or sociological holism, as Porpora calls it) – the debate over the relationship between structure and agency.

The first concept – "structure as enduring patterns of behaviour" (Porpora, 2002, p. 44) – is characteristic of methodological individualism, which Porpora illustrates with the work of Collins. In this perspective, social structure is ontologically epiphenomenal – a by-product of individual behaviour that has no causal efficacy in its own right – and methodologically merely a useful convention for describing properties that really belong to groups of individuals. As Collins argues, "strictly speaking, there is no such thing as a 'state', an 'economy', a 'culture', a 'social class'. There are only collections of individual people acting in particular kinds of microsituations - collections which are characterized thus by a kind of shorthand" (Collins, 1981, pp. 987-8). Methodological individualism is a species of reductionism which, at least implicitly and sometimes explicitly,[52] denies that social entities can have emergent properties and hence causal influence in their own right. As such, it suffers from all the faults of reductionism in general, as discussed in chapter three. But it cannot, of course, be dismissed as reductionist on purely *a priori* grounds; any denial of a level-specific reductionism rests upon a demonstration that the proposed reduction would suppress genuinely emergent properties. I have supplied such a demonstration in the section on organisations in chapter four, and I will expand on my response to methodological individualism in my discussion of King below.

Porpora's second concept of social structure – "structure as law-like relations among social facts" (Porpora, 2002, p. 44) – is characteristic of methodological collectivism or sociological holism – traditionally associated with the methodological work of Durkheim, the later work of Parsons, and Althusser's structuralist Marxism. As

---

[51] See (Porpora, 2002, p. 44) for a more recent but essentially similar formulation of this list.

[52] See (King, 1999a) for an example of the explicit denial of emergent social properties.

Porpora argues, "Because this conception of social structure rigidly divorces sociology from psychology, it represents social structure as something entirely devoid of the influence of human agency. On this sociologically holist view, social structure operates mechanically and naturalistically over the heads of individual actors" (Porpora, 1998, p. 342). This is the mirror image of reductionism – a denial of the causal influence of the lower level (human individuals) from which social entities emerge – and, as argued in chapter three above, it is equally untenable as an ontological position. The rejection of this position is almost universal amongst sociologists, but it is sometimes extended inappropriately. As I shall argue below, the rejection of ontological holism does *not* entail the automatic rejection of any theory that posits a regular causal relationship between higher-level social entities (or 'social facts'). What it *does* entail is that any such relationship should ultimately be grounded in an explanation of how the interaction of human individuals underlies the higher-level regularity.

Porpora's third concept – "structure as a system of social relations among social positions" (Porpora, 2002, p. 44) – is, he suggests, "most characteristically associated with the Marxian tradition", but in more recent years has been adopted by a number of other groups, including symbolic interactionists, network theorists, and sociological realists (including Porpora himself) (Porpora, 1998, p. 343). In this view,

> social structure is a nexus of connections among [human actors], causally affecting their actions and in turn causally affected by them. The causal effects of the structure on individuals are manifested in certain structured interests, resources, powers, constraints and predicaments that are built into each position by the web of relationships. These comprise the material circumstances in which people must act and which motivate them to act in certain ways. As they do so, they alter the relationships that bind them in both intended and unintended ways (Porpora, 1998, p. 344).

This third approach allows for both human individuals and social structure to have causal efficacy in their own right, with the two interacting continuously. This is precisely the sort of relationship between individuals and social structure that is required in an emergentist account, and it is indeed the approach adopted in the most prominent explicitly emergentist theory of social structure to date – that of the critical realist Margaret Archer, which I will discuss in some depth in the next section.

Finally, Porpora's fourth concept – "structure as rules and resources" (Porpora, 2002, p. 44) – describes the approach taken in Giddens' *structuration theory*, and arguably corresponds to *embodied structure* as described by Lopez and Scott. For Giddens, as for the advocates of the third version of the concept, both social structure and individual agency are necessary elements in sociological explanation. But for

Giddens, structure has no existence other than as the subjective conceptions of rules held by individuals, and their perceptions of the resources they have access to:

> to say that structure is a 'virtual order' of transformative relations means that social systems, as reproduced social practices, do not have 'structures' but rather exhibit 'structural properties' and that structure exists, as time-space presence, only in its instantiations in such practices and as memory traces orienting the conduct of knowledgeable human agents (Giddens, 1984, p. 17).

As Archer has argued, this represents a denial of any real causal influence to structure, through the conflation of structure and agency, and hence this position, too, is incompatible with an emergentist approach to social structure. I shall return to discuss structuration theory in more detail in the last section of this chapter.

## Archer vs. King: the emergence of social structure

Margaret Archer has been the prime advocate of an emergentist approach to sociology over the last two decades, and has established a reputation as an important sociological thinker. She has identified herself with critical realism and the work of Roy Bhaskar, but she has developed the concept of emergence significantly further than Bhaskar, although always in the context of sociology. She has also been responsible for the introduction of Buckley's concepts of morphogenesis and morphostasis into the discussion of emergence. Archer, then, has made an important contribution to the development of an emergentist approach to sociology – and this contribution includes many valuable features that I will not be covering here.[53] However, there have been several challenges to her invocation of emergence to justify treating social structure as causally effective in its own right. This section will argue that these challenges exploit a flaw in Archer's argument, which leads her to offer some questionable examples of emergence, but that this problem can be overcome, and that once it is, emergentism can indeed help to resolve the ontological relationship between structure and agency.

The section begins by summarising both Archer's use of emergence in her account of structure and agency, and also some of the criticisms that have been made of it, notably by Anthony King. It then goes on to show how the alternative justification for the theory of emergence developed in this thesis refutes King's argument more effectively.

---

[53] See in particular, (Archer, 1995).

*The morphogenetic approach*

Archer's account of social structure draws on two key influences. The first is Buckley's concepts of morphostasis and morphogenesis, which I have also drawn on in chapter two above – indeed she calls her theory "the morphogenetic approach". The second key influence is her reading of Lockwood's well-known paper on social and system integration. From Lockwood, she takes in particular his argument that social and system integration, which she reads as agency and structure respectively, "are not only analytically separable, but also, because of the time element involved, factually distinguishable" (Lockwood, 1964, p. 250).

Archer's theory of social morphogenesis rests upon this separability of structure and individual action, and she relies heavily on Lockwood's temporal argument to justify the claim that they are distinct. The methodological consequence of this separability is that it enables us to examine the ongoing interplay between them while still insisting on the close inter-relationships between the two. And the examination of this interplay is premised in practice on the pre-existence of structure and the temporality of its interplay with human individuals: "Fundamentally the morphogenetic argument that structure and agency operate over different time periods is based on two simple propositions: that structure necessarily pre-dates the action(s) which transform it; and that structural elaboration necessarily post-dates those actions" (Archer, 1995, p. 76). This relationship is portrayed in a diagram, reproduced below as Figure 5.1.

Structure
$T^1$

Interaction
$T^2$       $T^3$

Structural elaboration
$T^4$

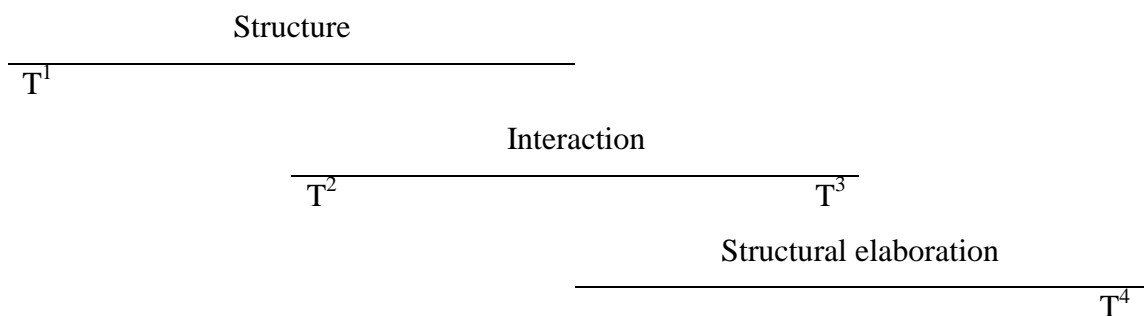**Figure 5.1 – The morphogenetic sequence (Archer, 1995, p. 76)**

Thus, on the one hand, the actions of individuals take place in the context of existing social structures, which influence those actions, while on the other, the net result of the actions of individuals is to reproduce or to transform those social structures. There is therefore a continuous cycle of interaction between social structure and

individual actions. But in the determination of any given individual action, the influencing social structures can be taken as given, as the historical residue of the sum of previous actions.

It is important to stress, however, that Archer does not argue that action *determines* structure or *vice-versa*: "it is precisely because such elaboration is co-determined by the conditional influence exerted by antecedent structures together with the autonomous causal powers of current agents, that society can develop in unpredictable ways" (Archer, 1995, p. 75). Not only agents but also culture co-determine structural change: "where any form of social elaboration is concerned, then structure, culture and agency are always involved" (Archer, 1995, p. 324). And in response to Outhwaite's suggestion that she is guilty of reification, she replies "I totally oppose any neo-positivistic notion of social hydraulics, of people reduced to inert matter which social forces mould and move" (Archer, 1990, p. 87). Her temporal account of the interaction between structure and action, then, does not constitute a social theory to the effect "structure determines action which then determines structure", but rather a methodology for analysing the interaction between the two, always in the context of other causal factors.

These causal effects of social structures (and indeed those of human individuals and of cultural entities) can be explained by their possession of emergent properties: "interaction generates emergent properties which must figure in explanatory statements" (Archer, 1982, p. 475). For Archer,

> Emergence is embedded in interaction: in the latter "we are dealing with a system of interlinked components that can only be defined in terms of the interrelations of each of them in an ongoing developmental process that generates emergent phenomena – including those we refer to as institutional structure". Emergent properties are therefore relational: they are not contained in the elements themselves, but could not exist apart from them… The highest orders of emergence are nothing more than the relations between the results of interaction. Nevertheless these "feed back" to condition subsequent interaction at lower levels (Archer, 1982, p. 475: the internal quote is from Buckley).

Emergence is thus taken to justify the central ontological claim of social realism: that social structures, although the product of human individuals, have causal powers of their own, which cannot be reduced to the powers of those individuals. As we have seen in chapters two and three, however, similar arguments apply to other emergent entities. Thus, for example, emergentists also argue that human individuals, although composed of their biological parts, have causal powers of their own, which cannot be reduced to

the powers of their biological parts. The emergentist claim that social structures have causal powers, then, in no way denies that human individuals also have such powers.

Nevertheless, this argument constitutes a direct challenge to the competing views of social ontology identified by Porpora. Let me focus first on Archer's challenge to methodological individualism.

### *Archer vs. methodological individualism*

Taking issue with J.W.N. Watkins, Archer argues that methodological individualism is only tenable if its protagonists can "show that *all* aspects of the social context, which figure in explanations… refer to nothing more than the behaviour of other people". She points out that Watkins argues that individuals have it in their power to *change* any aspect of social structure, and Archer asserts that this claim requires that all such social context can be reduced to "the effect of *contemporary* individual behaviour" (1979, p. 15). But social structures, she argues, are the residue in the present of past human behaviour, and so "the moment the individualist turns historian he also becomes a structuralist" (1979, p. 19). Here her argument appears to be that any social context carried over from the past is structural – and emergent – by virtue of its pre-existence alone, although as we have already seen, elsewhere Archer makes clear that emergence also depends on the existence of appropriate synchronic relations between a set of lower-order elements.

Archer has illustrated social emergence with a wide range of examples in the course of her work, but there are two in particular that critics have seized upon. First, she claims that "the existing demographic distribution" constitutes a social structure, arguing that "ontological status needs to be accorded to such aggregate (and emergent) social properties precisely because they are mechanisms facilitating or frustrating various policies" (1990, p. 87). The demographic distribution at any point in time is the consequence of past actions but the *current* demographic distribution cannot be altered by the *current* actions of human individuals; it must be accepted as a feature of the social context that is unalterable in the short term by human choices.

A second example of emergent social structure used prominently by Archer is the literacy rate in a model of Castro's Cuba (Archer, 1982, p. 468-9; Archer, 1995, p. 76-9). Archer takes the literacy rate to be a structure, with an independent influence on social outcomes, in this case the length of time it would take to raise literacy rates to a given target. Once again the argument is that literacy rates at any given time are the

consequence of past actions and are unalterable in the short term by human individuals, thus cannot be reduced to contemporary individual behaviour.

These two examples have been picked out by Anthony King, a methodological individualist and an advocate of the interpretive tradition in sociology, in a critique of Archer's morphogenetic approach (King, 1999a). King takes issue with Archer's apparent claim that the invocation of history is automatically a structural move:

> For the interpretive tradition the past is the meaningfully produced social relations between (now dead) individuals which have an impact on the present through the actions and interpretations of living individuals. The interpretive tradition does reduce society to people (both living and dead) but not to "these people here present" (King, 1999a, p. 205).

He illustrates the point with a response to the Castro example. Here, he says, "Archer converts the temporal priority of other people's actions into the ontological priority and autonomy of structure" (King, 1999a, p. 211). More generally, King argues that "in every case, appeals to emergentism can be reduced to the practices of other people, located at other places and times, and that, therefore, the morphogenetic approach cannot defend itself from collapsing back into an interpretive ontology" (1999a, p. 207). Thus "the concept of emergence, in fact, involves a relapse into sociological reification where society comes to exist independently of individuals, although this relapse into reification is concealed by the continual emphasis on individual practices and beliefs" (1999b, p. 270).

I will argue below that King's general critique of emergentism is mistaken, but the specific examples cited by Archer do seem vulnerable to some of his arguments. In particular, demographic distributions and literacy rates are poor examples of emergent structure. As I shall argue, the literacy rate and demographic profile of a population are merely the sums or averages of the properties of its members. Most emergentist thinkers would consider them *resultant* and not *emergent* properties of the population, and hence reducible to the properties of human individuals.[54] A stronger defence of emergentism against King's methodological individualism, I argue, can be provided by placing more emphasis on the synchronic relational aspect of emergence.

### *Temporality, relationality, and demographic distributions*

The account of relational emergence offered in this thesis has identified two distinct but complementary, indeed mutually interdependent, aspects of emergence. On the one hand, there is a synchronic aspect of emergence – the way in which a higher

level property depends upon the presence of a particular type of parts in a particular type of relations to each other. On the other, there is a diachronic aspect – the way in which morphogenetic and morphostatic causes combine to develop and sustain the existence and characteristics of this set of parts and relations.

Archer's work focuses primarily on the latter, in her investigations of the morphogenesis of social structure, and she tends to pay less attention to the synchronic aspect of emergence. On occasion, it may even seem that she denies its significance entirely, as when she writes "Until the analytical *separability* of structure and agency was explicitly acknowledged to entail temporality *rather than* simultaneity, realists did not radically recast the form of theorising about the relations between structure and agency" (Archer, 1996b, p. 693). Similarly, temporality seems to have supplanted synchronic relations when she claims that "the moment the individualist turns historian he also becomes a structuralist" (Archer, 1979, p. 19).

A number of theorists have suggested both that Archer depends upon temporality to justify claims for the emergence of structure, and also that such an argument does not work. As we have seen, for example, in commenting on the example of literacy rates, King accuses Archer of converting "temporal priority" into the "autonomy of structure" (1999a, p. 211). Similarly, Domingues argues that Archer's treatment of emergent properties "conflates two different issues, namely the historicity of the properties… and their ontological nature – which is the really serious issue at stake here" (2000, p. 227). And Sawyer, himself an emergentist, has argued in discussing Archer's work that morphogenesis does not entail emergence (2001, p. 570; Sawyer, 2005, p. 83-4).[55] The common theme here is a rejection of the argument that the prior existence of something entails that it is emergent, and hence of the argument that historical social residues are automatically structural.

My argument so far implies that we must accept this rejection: something could exist as an unstructured aggregate of lower-level entities with purely statistical properties as a group. Such a "heap" (Laszlo, 1972, p. 28) or "taxonomic collective" (Sayer, 1992, p. 101) would possess only resultant properties, which are reducible to the properties of the individual members of the group. The overall effect of the group is merely the sum of the effects of the individuals of which it is an aggregate. Thus it could pre-exist without being emergent.

---

[54] See the discussion of resultant properties in chapter two.

[55] Healy has offered a similar argument (Healy, 1998).

However, it is much less clear that Archer can reasonably be accused of depending upon temporality in the first place. Despite her occasional neglect of the synchronic side of emergence, she does often acknowledge it. Thus for example, she writes

> Emergent properties are *relational*, arising out of combination (e.g. the division of labour from which high productivity emerges), where the latter is capable of reacting back on the former (e.g. producing monotonous work), has its own causal powers (e.g. the differential wealth of nations), which are causally irreducible to the powers of its components (individual workers). This signals the *stratified nature of social reality* where different strata possess different emergent properties and powers (Archer, 1995, p. 9).

This recognition of synchronic relationality becomes important when we consider the examples of emergent structure referred to earlier: literacy rates and demographic distributions. As the issues involved in both cases are similar, let me focus on just one of them, the claim that demographic distributions may be causally effective in their own right because they possess emergent properties. As we have seen, Archer argues that "the existing demographic distribution" constitutes a social structure, writing that "ontological status needs to be accorded to such aggregate (and emergent) social properties precisely because they are mechanisms facilitating or frustrating various policies" (Archer, 1990, p. 87). Elsewhere she writes that "a demographic structure is often treated as a mere aggregate of so many people of such and such ages, yet this structure itself can and does modify the powers of people to change it, that is, it affects the powers of its constituents – by defining the size of the relevant group of child bearing couples whose reproductive behaviour could transform the structure and thus restricting their influence upon it, however prolific or non-prolific they may be" (Archer, 1995, p. 174).

Let us analyse this example. The "demographic structure" refers to the proportions of a population in different age bands. The argument presented isolates one mechanism that contributes to the morphogenesis of this distribution – childbirth – and one factor that affects the operation of this mechanism – the number of potentially child bearing couples (abbreviated to "couples" hereafter). If the number of couples at $T_n$ increases, then, other things being equal, the number of babies at $T_{n+1}$ will increase. If we wish to explain the impact of this on the demographic structure, we will express the effects in proportional terms: a higher proportion of couples produces a higher proportion of babies in the demographic structure in the next time period.

So far, this is uncontroversial. But the point of the example is the claim that the proportional relation between the numbers of couples and the population as a whole affects the proportion of babies at $T_{n+1}$ in its own right, as an emergent property of the population, rather than merely as a summative measure of the causal effects of the individual couples concerned. But does the demographic structure really have a causal effect in its own right? I argue not, as the proportion of couples at time $T_n$ affects the proportion of babies at time $T_{n+1}$ only because the *number* of couples at time $T_n$ affects the *number* of babies at time $T_{n+1}$.

Here, the "proportion" has no effect distinct from the summed effects of the individual couples, because those couples themselves are unaffected by the size of this proportion. The proportional relation is merely a formal relation, with no substantial effect on the relata (see Sayer, 1992, p. 88 on formal vs. substantial relations). The number of babies produced by the group of couples is affected by the number of couples in the group, but not by the number of other people in the population, and hence there is no causal power exercised by the *proportion* of couples in the population that is distinct from the impact of the *number* of couples. I conclude that there is no evidence here to support the claim that the proportional relation of the group of child bearing couples to the population as a whole is an emergent property. Thus, the effect of a demographic distribution on, say, birth rates or pension costs, is merely the resultant sum of the effects of the individuals of which it is an aggregate.

This is why this example (and the similar example of literacy rates) is vulnerable to King's argument that the effects of a distribution can be reduced to the effects of the individuals composing it. The problem here is not that Archer is relying on a temporal argument for emergence; on the contrary, her claim that demographic distributions are mechanisms implies that she is making a synchronic relational claim. The problem, rather, is that in these particular cases (but not in many others she cites), this relational claim is wrong, and the supposed higher-level structure is not emergent at all. In this particular case, Archer in my view mistakenly claims emergent structural powers for an unstructured aggregate of human individuals and hence *is* assigning structural status to what is, in ontological terms, nothing more than a collection of individuals.

But the non-existence of emergent powers in these particular cases is merely a distraction from the more important argument: that there are many other cases where social structures do have such powers. The next section will show how the account of

relational emergence given in chapters two and three can be used to justify such a claim in response to King's critique of emergentism.

### Relational emergence vs. methodological individualism

Let us turn to an example of a social structure cited by Archer that definitely *does* have emergent properties: the division of labour. In Adam Smith's example of pin production, a group of workers organised according to the principles of the division of labour is able to produce substantially more than the same group of workers, each producing pins individually (Archer, 1995, p. 51; Archer, 1996b, p. 686). Hence a group organised on these principles has an emergent property that is not possessed by the same workers when they are not so organised. This capability to produce more is therefore an emergent causal power of the organised group, and not causally attributable to the individual workers.

Now King recognises that this capability cannot be ascribed to the individuals alone:

> The interpretivist tradition is no way arguing that this new division of labour can be understood through dis-aggregating the division of labour back to its molecular constituents – the individual craftsmen or individual readers. This approach fully recognizes the qualitative novelty of this situation but that newness resides precisely in the new relations between individuals (King, 1999a, p. 213).

Thus he recognises the same facts of the case as Archer and myself (Archer writes "the power of the … emergent property, mass production, did exceed those of everyone involved, because it was no aggregate of their individual productivity but the relational resultant of their *combined productive activity*" (Archer, 2000b, p. 467)). Where we differ is on the question of whether these facts entail that the group as such has causal powers in its own right. King denies that the combination of people plus relations, or people plus interaction, constitutes a higher level entity with causal effects of its own (King, 1999b, p.272).

But my discussion of organisations as emergent entities in chapter four has already shown why such combinations *do* constitute higher level entities with emergent causal powers. When people combine in this way, they form wholes that have powers those same individuals would not have if they were not organised into such a whole. These powers are therefore powers of the organisation and not of the individuals or even of the group of individuals, since they only exist when these individuals (or others like them) are related to each other in the particular ways that are necessary for an organisation of this type to exist. This is exactly analogous to the argument that human

individuals themselves have causal powers that are theirs and not those of the cells that make them up, since these causal powers only exist when those cells (or others like them) are related to each other in the particular ways that are necessary for such a person to exist.

Given that the higher level powers of any organisation depend upon the presence of both the individual members of the organisation and the particular sorts of relations that constitute them into an organisation of this type, the redescription principle deployed in chapter three also applies. We cannot say that it is not the organisation that has these causal powers, but rather its members and their relations, because the presence of these members and relations is *synonymous* with the existence of the organisation. In such cases, to say that 'these individuals and the relations between them' have a certain causal effect *is* to say that the organisation has such an effect.

This is the positive case for asserting that organisations have causal effects in their own right, but there is also a negative case, which rests on the fact that the position expressed here by King is ultimately incoherent. This is because King's argument has an entirely general form and therefore produces the same sort of reduction *ad absurdum* identified in chapter three. If it is true, say, in the case of the division of labour, that people plus a particular set of relations between them do not constitute a higher level entity with causal powers of its own, then the same argument would appear to apply to any other part-whole relation. King offers no reason why this should be true of people and their relations, but not of other sorts of entity. The argument therefore entails, for example, that when a dog barks, this is *not* because the dog has the causal power to bark, but rather because its lungs, windpipe, vocal chords, mouth, etc, and the relations between them have the power to bark. And if this is not absurd enough, we can pursue the same logic indefinitely to produce the full reduction *ad absurdum* – because on the same logic we must also deny that it is the lungs, windpipe, etc, that are doing the barking, but rather the "cells plus relations" that make those organs up, and then we must consider it to be the "molecules plus relations" instead of the cells, the "atoms plus relations" instead of the molecules, and so on to levels where science has so far failed to go.

The full irony of King's position appears when we apply it to human individuals. The same argument he applies to deny causal effectiveness to social structures could be applied equally well to the human individuals whose causal powers he seeks to privilege. Just as social structures are nothing more than the people in them

and the relations between those people, human individuals are nothing more than the cells in them and the relations between those cells. But it would be just as invalid to eliminatively reduce the causal powers of people to those of their cells as it is to eliminatively reduce the causal powers of social structures to those of people (cf. Durkheim, 1974, p. 28-9).

King goes so far as to say "For hermeneutics, social networks have their own distinctive properties which are irreducible to isolated individuals, extracted from these networks, but that does not mean that these networks are more than the individuals in them" (King, 1999b, p. 275). The first part of this reads like an acceptance of emergentism, but he seems to believe that by rejecting the claim that "networks are more than the individuals in them" he turns this into a denial of emergence. There is arguably a sense in which emergentism sees an emergent whole as more than its parts, because a specific set of relations between the parts is *also* required to constitute it into this type of whole. But as we have seen, King too accepts that relations as well as parts are required to produce the causal effect of the whole. Perhaps by "more than" he is referring to emergentism's claim that wholes have properties that are not possessed by their parts – but the first part of his statement in this quote seems to confirm that he accepts the same point. There is no other "more than" to be found in emergentism than these, and so it is hard to see what objection there is here to a synchronic relational form of emergentism.

A related confusion seems implicit in King's criticism which I cited in an earlier section:

> despite its apparent coherence, the concept of emergence, in fact, involves a relapse into sociological reification where society comes to exist independently of individuals, although this relapse into reification is concealed by the continual emphasis on individual practices and beliefs (King, 1999b, p. 270).

The usage of the word "independently" in this context, however, is extremely ambiguous. Relational emergentism certainly does not imply that society can exist independently of individuals, since social structures are always synchronically composed (at least in part) of people. It *does* imply, however, that wholes like social structures may have *distinct* causal powers from those of their parts, or in other words *causal powers in their own right*. On occasion, emergentists may have called these "independent" causal powers, and it is easy to see why they might do so, but these causal powers of the whole remain dependent on the presence and properties of the

parts. It would be desirable to avoid this usage of "independent", but when used in this way it does not entail reification of social structure.

With these clarifications, I believe the argument of this section shows that King's critique of emergentism can be successfully refuted – and the whole reductionist ontology of methodological individualism falls with it. Once we recognise that social structures are synchronically constituted by the relations between people as well as by the people themselves, it becomes impossible to eliminatively reduce the powers of those structures to the powers of the individuals composing them. Social structures like organisations have causal powers in their own right, which arise from the combination of individuals and relations that constitute them, but which are different from the causal powers that would be possessed by these same individuals if they were not organised into these social structures. They are therefore causal powers of the structures and not of the individuals.

In defending this emergentist account of social structure this section has also refined Archer's approach by emphasising the synchronic relational aspect of emergence. In particular, it has used the redescription principle to show that relational explanations of emergent properties are entirely consistent with the denial of eliminative reductionisms like methodological individualism, and it has distinguished more carefully between emergent and resultant properties in the analysis of social structure. The next section continues this engagement with existing accounts of social emergence by looking at the work of Sawyer.

**Sawyer: functional emergence in sociology**

A different tradition of anit-reductionist thought has been introduced into sociology by R Keith Sawyer, in a series of recent articles and in his book *Social Emergence* (Sawyer, 2001; Sawyer, 2002a; Sawyer, 2002b; Sawyer, 2003; Sawyer, 2005). Sawyer argues for what he calls *nonreductive individualism* in the social sciences by analogy with the functionalist argument that has been made by Fodor for *nonreductive physicalism* or *nonreductive materialism* in the philosophy of mind.[56] Nonreductive materialism argues on the one hand for

> *ontological materialism*, the belief that all that exists is matter, thus rejecting various forms of Cartesian dualism and vitalism. However, nonreductive materialism argues

---

[56] Despite sharing the same label, there is little else in common between this functionalist tradition in the philosophy of mind and the functionalist tradition in sociology.

that mental properties and states are irreducible to physical properties and states and that the science of the mind is autonomous from the science of neurons" (Sawyer, 2002a, p. 539).

Thus a denial of the reality of mind is accompanied by an insistence on the irreducibility of mental properties and states to physical ones. This stance is far from universally accepted in the philosophy of mind, but Sawyer takes its validity in that field as read and seeks to develop a similar logic for the philosophy of the social sciences, and I shall concentrate my analysis on the latter.

Nonreductive individualism, then, is the analogous combination of *ontological individualism* – the claim that only individuals, and not social entities, are real in the social world – with the denial of *methodological individualism*, on the grounds that social properties are irreducible to properties of individuals and hence the social sciences can be independent of psychology and lower-level biological sciences. Sawyer believes that social laws can be defined that are not reducible to laws about individual behaviour because the reduction to individual causes may be radically different in each case of the social event to be explained. Thus, "social properties may be irreducible to individual properties, even though social entities consist of nothing more than individuals" (Sawyer, 2002a, p. 541). Sawyer justifies this claim by invoking Fodor's account of multiple realisability and wild disjuncture.

### Multiple realisability and wild disjuncture

In an extraordinarily elegant paper, Fodor argued that the concepts and laws of higher level sciences can not be reduced to those of physics because the "natural kind predicates" of the higher level sciences – the entities or properties that appear in their laws –do not correspond to natural kind predicates in physics. In particular, the truth of law-like generalisations in the higher-level sciences was not necessarily dependent on a consistency of physical composition amongst the various instances or tokens of the higher-level predicates to which the generalisation applied.

Fodor illustrated the argument with Gresham's Law – the old adage that 'bad money drives out good'. The truth of Gresham's Law, Fodor pointed out, did not depend on any physical similarity between the different types of money to which it could be applied. Gresham's Law could apply equally well to many different kinds of money, even though their physical forms were very different. The higher level predicate 'money' could be realised in many different physical forms, which seemed to have little or nothing in common with each other as far as their physical characteristics were concerned. If there was some causal explanation for this regularity, it surely could not

be a physical explanation because it applied to what Fodor called a *multiply realisable* or *wildly disjunctive* variety of physical forms.[57] Thus, there was no regularity of physical form corresponding to the regularity of behaviour at the economic level. Predicates like money, then, are not physical but higher-level types.

Fodor's wild disjuncture converges to some extent with Cohen and Stewart's more recent concept of *complicity*: "The hallmark of complicity is the occurrence of the same feature or features in systems whose rules are either known to be very different, or are expected to be very different if only we could find out what they are" (Cohen and Stewart, 1995, p. 417). They use as an example the independent evolution of wings in birds, bats, insects, and pterodactyls. There are certain functional properties that appear to be held in common between a broad variety of wings, most obviously the ability to fly that some wings confer on their owners, despite the radically different details of their physical construction.

Although Fodor's paper did not link functionalism explicitly to emergence, Sawyer has argued that Fodor's concepts of multiple realisability and wild disjuncture justify taking an emergentist view of social structure (Sawyer, 2001). Sawyer argues that:

> Multiple realizability alone does not necessarily imply irreducibility; if there are only a few realizing states, or if those states display some common features, the reduction may not be problematic. However, reduction would be difficult if the neurobiological equivalent of a psychological term were an otherwise unrelated combination of many neurobiological concepts and terms… Fodor termed such a realization *wildly disjunctive...* Fodor argued that a true scientific law cannot have wildly disjunctive components and that wild disjunction thus implied that there could be lawful relations among events, described in psychological language, that would not be lawful relations in the language of physics (Sawyer, 2001, p. 557).

Sawyer argues that "most social properties of interest to sociologists seem to have wildly disjunctive individual-level descriptions" (Sawyer, 2001, pp. 558). For example "an individual-level description of the social-level natural kind term 'competitive team sport' is likely to be wildly disjunctive" (Sawyer, 2002a, p. 549). Any causal law predicated on the properties of such a social-level natural kind will therefore be irreducible to a law expressed in terms of properties of its lower level parts, since the lower level properties involved in each case (or group of cases) will be different. He argues that this is enough to defeat a reductionist account of such

---

[57] Wild disjuncture refers to cases of multiple realisability in which the different physical realisations are radically diverse, although the dividing line between instances of the two concepts is unclear.

properties, and suggests that this provides us with a "nonreductionist account of emergentism" (Sawyer, 2001, p. 580).

### *A critique of Sawyer*

I will argue below that multiply realisable properties are indeed a case of emergence, but that Sawyer's explanation of why they are so is wrong. In explaining why, however, it will be necessary to develop an important elaboration of the theory of emergence presented so far in this thesis. Before I do so, let me consider a couple of more obvious responses to Sawyer.

First of all, opinions differ on how effective Fodor's argument is against reductionism. Kim argues that each distinct type of lower-level instantiation of the higher-level predicate could be explained separately, and so functionalism does not constitute a viable alternative to reductionism (Kim, 1998, p. 116). Where there is a large number of radically divergent lower-level realisations, it may be impractical to provide a reductive set of explanations for a single higher-level law, but it nevertheless remains true that the behaviour of each instance of the higher-level predicate may be reductively explainable in principle. Multiple realisability therefore seems to provide an epistemological rather than an ontological argument against reductionism – which is perhaps why Sawyer seeks to combine his opposition to methodological individualism with support for ontological individualism.

Secondly, the argument for functionalism seems at first sight to have nothing in common with the concept of emergence, other than representing an anti-reductionist position. Sawyer is quite careful about constructing the anti-reductionist argument, but extremely vague about the relationship between multiple realisability and emergence itself. Let us return to the basic definition of emergence: *Emergence* is the idea that a *whole* can have *properties* (or powers) that are not possessed by its *parts*. The functionalist argument could only be related to this definition if we considered the whole in this definition to be, not a particular instance, but a general type of whole (a 'natural kind'), and if we considered the parts to be not the parts of one particular instance, but rather the many varying types of parts possessed by different realisations of the whole. But this is clearly not the intention behind the basic definition; it is about the properties of a particular type of whole and would seem to apply to any instance or token of that type. Here, then, Kim's response to functionalism must be correct: the functionalist argument has nothing to say about the relation of a *particular* whole to its parts, and therefore can not provide an argument against reduction for any given

instance of an emergent property. It might appear that it is an argument against the reduction of functional *laws* rather than against the reduction of emergent properties, and thus an anti-reductionist strategy that is tied to a covering law view of causality.

However, in this case first appearances may be misleading. I will argue below that multiply realisable properties *are* emergent properties, but they are generally emergent properties of much the same kind as those I have been discussing earlier in this thesis. Sawyer's argument for emergence therefore does not constitute a separate argument from the relational argument I have been advancing, but instead is a special case (or from a different perspective, a generalisation) of that argument.

To show why this is so we must return to the general theory of emergence. Any given emergent property of a whole, I have argued, is the product of a mechanism that can be explained in terms of the properties of the parts and the relations between them. However, I would like to suggest that the mechanism underpinning the higher level property need not depend upon *all* the properties of the parts and *all* the relations between them. Some higher level properties may depend upon a subset of the properties of the parts and of the relations between those parts. Let me call this subset the *underpinning* properties and relations for this property. Where a higher level property is emergent from an underpinning subset of properties and relations it may be realised in a variety of different types of entity which share that underpinning subset, but have parts and relations between them that vary widely in other respects.

Thus, for example, the aerodynamic properties of wings may depend upon those underpinning properties and relations that determine the shape and rigidity of the whole, but not at all on a wide variety of other properties of the wholes concerned. Wings made in many different ways from many different sorts of materials may therefore share the same aerodynamic properties. Those aerodynamic properties are thus relationally emergent in much the same sense as, say, the properties of a water molecule. In other words, the analysis of chapters two and three above applies to them. Indeed, this is part of the significance of the concept of structural ranges discussed in chapter two. The current argument, however, brings out an important supplementary point: structural ranges apply to *properties* as well as to *entities*, and the structural range of a property may be wider than that of any particular entity. This will happen whenever its underpinning properties and relations are found in a variety of different entity types, whether or not those entity types are structurally continuous in respect of their composition.

The consequence is that these so-called functional properties are indeed relationally emergent in each token case, and relationally emergent on the basis of the same causal mechanism in each token case. The only difference between these functional properties and the simpler cases of relationally emergent properties discussed earlier in this thesis is that the underpinning properties and relations may be a subset of the properties of the parts of the wholes possessing the property and/or the relations between those parts. It is entirely possible that many of the emergent properties already discussed also work in this way. Thus, for example, the emergent properties of a married couple as such may be entirely independent of a broad range of properties of the particular people who are its parts, such as their colour, physical shape, and the knowledge they possess about a wide range of subjects. A similar argument can be applied to the property of 'being a competitive team sport' cited as an example by Sawyer. This property is emergent from a certain subset of the relations between the participants in any such sport, and that subset of underpinning relations falls within a certain range for *all* competitive team sports, irrespective of the other properties of the sport concerned. This argument is synonymous with the claim that these properties are multiply realisable.[58]

As Sawyer points out, a wide range of social properties are multiply realisable in this sense, and inevitably so. Given that, as was argued in chapter four, both human individuals and many social entities are complex adaptive systems, every instance of either will differ from every other instance. Different humans have different DNA, different biological phenotypes as a result of their different DNA, different life histories which are reflected in different neuronal configurations in their brains, and different causal powers as a result of these other differences. Given that social entities are composed primarily of human beings, then this alone implies that all social entities will be multiply realisable. However, as social entities themselves are also complex adaptive systems that can be constructed in different ways and can develop over time, there is a further reason for them to be multiply realisable. Quite where the line is to be drawn between multiple realisability and wild disjuncture is unclear, but the variations between different instances (tokens) of the same type of social entity may often be

---

[58] The case of Gresham's law cited by Fodor, however, may be more complex, and there may be other kinds of functional properties that raise further issues, such as the case of keys discussed by Teller (Teller, 1992. pp. 144-6). Teller, incidentally, discusses keys under the heading of 'relational emergent properties', but means something different than I do by this term. The

substantial. This is in marked contrast to lower levels of the ontological hierarchy, where different instances of the same natural kind are generally identical in both composition and structure.

Thus it is true both that multiply realisable properties are emergent, and also that a wide range of social properties may be considered multiply realisable properties. However, in other respects this argument differs considerably from Sawyer's. Above all, it is clear from this argument that such properties are not emergent *because* they are multiply realisable, but *despite* being so. It is the commonality of their underpinning properties and relations that underlies the similarity of their higher level properties, not the divergences between the other properties of and relations between their parts. Multiply realisable properties are a fascinating and important subset of relationally emergent properties, but they are indeed a subset and not some other independent foundation for the existence of emergent properties or laws.[59]

A further implication is that, again despite first appearances, multiply realisable properties are irreducible for ontological and not epistemological reasons. They can not be eliminatively reduced for precisely the same ontological reasons as any other relationally emergent property, and entities possessing them are causally effective in just the same sense as any other causally effective entity. Hence they provide no warrant for being ontologically individualist: it is the higher level entity that possesses the emergent functional property, and not the lower level parts. Sawyer's arguments for ontological individualism thus pose no threat to the argument that relational emergence justifies the claim for social entities to have causal powers (and ontological reality) in their own right

To summarise this section: Sawyer's argument for a nonreductive emergentist sociology premised upon multiple realisability and wild disjuncture is very interesting, not least because social entities in general *are* multiply realisable. However, he fails to see that this functional argument for emergence is in fact a special case of the more general and more fundamental relational argument for emergence. He has helped to raise awareness of the possibilities of an emergentist account of sociology, and has made some valuable contributions to such an account, but the incompleteness of his

---

argument of this section, therefore, does not constitute a complete reply to Fodor, but I do not believe this undermines its critique of Sawyer's use of Fodor's work.

[59] Logically, there may also be cases of properties that are shared by different types of entity even though those entities do not provide a common set of underpinning properties and relations. The cases cited by Sawyer, however, are not of this type.

underlying general theory of emergence leaves him with an inadequate account of the emergence of social structure.

## Giddens: structuration theory

Many thinkers have highlighted the need to recognise the causal roles of *both* social structure and human agency in the social world. Mouzelis, for example, argues that we must

> view any social system (a group, an organisation, a society and so on) from both a social and a system integration point of view: to view it, in other words, as both a set of interacting actors… and as a configuration of social positions or institutions … Any attempt to eliminate agency leads unavoidably to various forms of reification; whereas any attempt to eliminate institutional structure leads to ultra-voluntaristic theories of social persistence and change (Mouzelis, 1991, p. 117).

There is, however, more than one way to integrate agency and structure. Emergentism offers one such way, but this section will examine an explicitly anti-emergentist alternative: Anthony Giddens' *structuration theory*. It will begin by outlining Giddens' ontological position, then I will consider Archer's critique of this and the defence offered by Rob Stones, before outlining my own perspective on structuration theory.

### *Structuration theory*

Giddens tells us that "as conceptualized in structuration theory, 'structure' means something different from its usual usage in the social sciences" (Giddens, 1984, p. xxvii). He argues that "a structure is not a 'group', 'collectivity' or 'organization': these *have* structures" (Giddens, 1993, p. 121), and so seems to identify structure as what I have called *structure-as-relations* rather than *structure-as-whole*. He labels collectivities and the like as *systems* rather than structures, but his discussions of structure rarely refer back to these systems; more often he seems to see structure as the structure of *practices,* to which he assigns central importance in sociology (Giddens, 1984, p. 2) or of *institutions*, which he defines as "the more enduring features of social life" (Giddens, 1984, p. 24).

The content of structure, for Giddens, is *rules* and *resources*:

> In structuration theory 'structure' is regarded as rules and resources recursively implicated in social reproduction; institutionalized features of social systems have structural properties in the sense that relationships are stabilized across time and space. 'Structure' can be conceptualized abstractly as two aspects of rules – normative elements and codes of signification. Resources are also of two kinds: authoritative resources, which derive from the co-ordination of the activity of human

agents, and allocative resources, which stem from control of material products or of aspects of the material world (Giddens, 1984, p. xxxi).

These rules and resources stabilise social practices and play a key role in their reproduction, in particular because, for Giddens, structure has no existence outside the minds of the human agents involved in its reproduction:

> Structure thus refers, in social analysis, to the structuring properties allowing the 'binding' of time-space in social systems, the properties which make it possible for discernibly similar social practices to exist across varying spans of time and space and which lend them 'systemic' form. To say that structure is a 'virtual order' of transformative relations means that social systems, as reproduced social practices, do not have 'structures' but rather exhibit 'structural properties' and that structure exists, as time-space presence, only in its instantiations in such practices and as memory traces orienting the conduct of knowledgeable human agents (Giddens, 1984, p. 17)

Not only rules, but also resources, in the sense in which Giddens uses the term, exist only internally to individuals. Even allocative resources, for Giddens, are not material products, but rather "stem from control of" such products, and exist only as memory traces (Giddens, 1984, p. 33). This enables him to reconcile structure and agency, not as the distinct causal powers of inter-related types of entity, but as different aspects of the same thing:

> The constitution of agents and structures are not two independently given sets of phenomena, a dualism, but represent a duality. According to the notion of duality of structure, the structural properties of social systems are both medium and outcome of the practices they recursively organize. Structure is not 'external' to individuals: as memory traces, and as instantiated in social practices, it is in a certain sense more 'internal' than exterior to their activities in a Durkheimian sense (Giddens, 1984, p. 25).

Giddens' strategy for reconciling structure and agency, then, seems to allow some sort of causal effect to structure – or at least to *institutional structure* in the sense defined by Lopez and Scott. But at the same time he appears to deny a distinct ontological status to structure by seeing it as 'virtual' except in those moments when it appears as a property of human individuals – as *embodied structure*. He explicitly rejects Durkheim's argument that structure can be seen as emergent and therefore as exercising a causal influence in its own right (Giddens, 1979, pp. 50-51). His position therefore seems somewhat analogous to Sawyer's attempt to combine ontological individualism with a denial of methodological individualism.

Giddens' counterposition of duality to dualism is widely interpreted to mean that no ontological separation is possible between structure and agency, since both can only

ever be effective as aspects of human individuals. This is the heart of the critique of Giddens' ontology offered by Archer.

### *Archer's critique of Giddens*

Archer recognises that both emergentism (which she tends to refer to as 'the morphogenetic approach') and structuration theory start from the presumption that both structure and agency are causally significant. As we have seen, Archer's own argument is explicitly emergentist, and leads to the conclusion that structure and agency each have their own existence and causal influence – and thus to an *analytical dualism* of structure and agency in social explanations. By contrast, Archer (along with many other critics of Giddens, e.g Craib, 1992) argues that Giddens' *duality of structure* conflates structure and agency as two sides of the same coin rather than two separate but interacting elements. In consequence, each can only be analysed by seeing how it appears through the other:

> To treat 'structure' and 'agency' as inseparable is central to the notion of 'duality'. This method of transcending dualism then produces an ontology of 'social practices' which are held to be the ultimate constituents of social reality. There is a decentring of the subject here because human beings only become people, as opposed to organisms, through drawing upon structural properties to generate social practices. There is an equivalent demotion of structure, which only becomes real, as opposed to virtual when instantiated by agency. These ontological assumptions have direct implications for practical social theorizing, for they enjoin that social theory should concern itself exclusively with 'social practices'. These alone are the subject matter of the social sciences. If this is the case then its corollary is central conflation, for the implication is that neither 'structure' nor 'agency' have independent or autonomous or anterior features, but only those properties which are manifested in and reproduced or transformed through 'social practices' (Archer, 1995, p. 101).

Thus, "by enjoining the examination of a single process in the present tense, issues surrounding the relative independence, causal influence and temporal precedence of the components have been eliminated at a stroke" (Archer, 1995, pp. 93-4). Archer labels this as *elisionism* or *central conflation* of structure and agency, and argues that it "deprives *both* elements of their relative autonomy, not through *reducing* one to the other, but by *compacting* the two together inseparably" (Archer, 1995, p. 101).

Archer's theory of social morphogenesis, by contrast, rests upon the analytical separability of structure and individual action, as we saw earlier in this chapter (p. 127). By contrast with structuration,

> 'Morphogenesis' is also a process, referring to the complex interchanges that produce change in a system's given form, structure or state… *but it has an end-product*, structural elaboration, which is quite different from Giddens's social system as merely a 'visible pattern'. This to him can best be analysed as recurrent social

practices, whereas to general systems theorists, the elaborated structure has properties which cannot be reduced to practices alone, although these are what generated both it and them (Archer, 1982, p. 458).

To put this into my terms, Archer argues that structuration theory sees the causal powers of structure and agency as inseparable because each enters into the constitution of the other, and neither has a distinct existence in its own right. This can be contrasted with an emergentist ontology, which insists that both structure and agency have causal powers in their own right. A number of authors have disputed this argument, and in the next section I will discuss one such response: the defence of Giddens offered by Rob Stones.

### Stones' defence

The primary thrust of Stones' argument is to dispute the extent of the underlying differences between structuration theory and emergentism, and to argue instead that social theorists should be working towards a productive synthesis of these two essentially compatible approaches, each of which has something useful to contribute to the study of structure and agency (Stones, 2001, p. 177; Stones, 2002, pp. 223-4). This argument rests in particular upon a denial of the claim that Giddens conflates structure and agency. As Stones sees it, Giddens argues "that structures enter into people and that the agency within people draws on those internalized structures" but "would not deny that there are social structures outside the person that are separate from that person – that would be absurd – only that these can't be *drawn upon* by agents as structures unless they enter into the phenomenological 'horizon' of agents" (Stones, 2002, p. 223).

This idea that structures are separate from people but also enter into people is clarified in another article in which he tells us that Giddens does preserve an ontological distinction between structure and agency:

It is a different notion of dualism that Giddens rejects, the kind of dualism that sees structure as always entirely external to agency, in which structure is conceptualized as akin to the walls of a room and agency as akin to the space to move within the room. This kind of dualism is rejected because structuration theory conceptualizes structure as being partly within the agent as knowledgeability or memory traces. So the structure enters into the person (or corporate agent) such that we can say *both* that agency is a part of the person and that social structure is a part of the person. Structure, for Giddens, is something that is conceptualized as inhabiting people in the sense that it enters into the constitution of the reflexive and prereflexive motivations, knowledgeability and practices of people (Stones, 2001, p. 184).

This, he believes, constitutes a denial of conflation, with the consequence that Archer ought to be able to accept structuration theory:[60]

> I see no reason why the morphogenetic approach could not accept this notion, of what I will call *duality within agents (or people)*, as long as structuration theory reassures realism that it does hold significant parts of the context of structures (i.e. the parts external to the memory traces and knowledgeability of the agents in focus) to really exist *outside* the agents, and that memory traces and knowledgeability also *really exist as potentials* even when they are not drawn upon or instantiated. Indeed, I should say that a reliance upon *duality within agents* already runs right through the morphogenetic approach, for example … within the very idea of structural conditioning (Stones, 2001, p. 184).

But it is highly questionable whether Stones' argument rebuts the charge of conflationism at all. In seeing structure as part of the human agent and the agent also as part of structure he presents an ontological argument that, as written, is simply incoherent on any plausible understanding of part-whole relations (see Stones, 2005, p. 21 for an even clearer statement of the mutual constitution of structure and agency). For A to be part of B entails that A is less than the whole of B (i.e. that there is at least one other part of B that is not part of A). Since it is impossible for A to be less than B and also for B to be less than A, it is impossible for A to be part of B and B also to be part of A. What *is* logically possible is for *part of A* to be part of B, and *part of B* to be part of A – ontological interpenetration – and although this is not what Giddens or Stones actually say it does seem to be what they really mean. However, for this to be a plausible ontological claim they would need to be much more precise about *which* parts of *which* wholes interpenetrate.

Giddens and Stones see no need for such an analysis because of another flaw in their argument: they assume a false identity between causal interdependence and ontological composition. This operates most subtly in the frequent and systematic misuse of the word 'constitution' and its cognates to mean both or either of these things interchangeably, as for example in: "Structure … enters into the constitution of … motivations [and] knowledgeability" (Stones, 2001, p. 184 - see above for the full quote). Here we see a transmutation, via the semantic duality of 'constitution', of the thoroughly reasonable argument that structure has a causal effect on knowledge into the thoroughly unreasonable argument that structure thereby becomes a part of knowledge.

Both Giddens and Stones consistently assume that whenever a structure contributes causally to a belief it thereby becomes part of that belief. Thus, for example,

---

[60] And indeed that structuration theorists ought to be able to accept emergence (Stones, 2001, pp. 194-5).

the claim that structural conditioning rests on duality (made by Stones in a quote cited above) and thus on ontological interpenetration conflates knowledge that arises causally from the agent's experience of a structure with the structure itself. Emergentists are perfectly happy with the arguments that social structures causally affect our knowledge, and that our knowledge causally affects our action.[61] What we reject is that this entails that structures *become* part of our knowledge.

But perhaps the greatest problem with both Giddens' ontology and Stones' defence of it is a persistent failure to distinguish between different kinds of structure. Addressing this failure is the key to resolving the debates over structuration theory, and it is this that I turn to next.

### *One size does not fit any: institutions*

It is common for social theorists to write of social structure as if it was homogeneous, in the sense that they assume that all social structures can be theorised in the same way – that 'one size fits all' when it comes to the analysis of social structure. In this section I will argue that there are important differences between different classes of social structure, and that the neglect of these differences has been a persistent problem in the debate over structuration theory. In particular, we must distinguish, as I have done in chapter four, between the ontological structure of organisations and that of institutions (irrespective of whether we equate institutions to collective representations or to individual beliefs). By conflating the two, many debates on structure have simply confused the picture. In order to assess the debate between structuration theory and emergentism effectively we need to consider it from both possible angles – as an analysis of organisations, and as an analysis of institutions.

Let us begin with structuration theory as an analysis of institutions, as I have defined them in chapter four: social practices that are common within a given cultural setting. As I argued there, such commonality arises from the causal role played by particular beliefs (norms and values) which are shared by members of the cultural community concerned. However, there are two distinct and different possible ways of theorising this causal role. The first argues that these norms and values are only causally effective as items of knowledge or belief held by individual human agents. The second, however, argues that it is not individual norms and values but collective norms and

---

[61] Stones seems to think that Archer would reject the last point (Stones, 2005, p. 55). This misunderstanding, I believe, arises from a corresponding failure to distinguish between composition and causation in the emergentist account of social ontology.

values that are causally effective here: that individual-level norms and values, related to each other in the sense of being shared over a certain community, form the parts (or underpinning properties) of an emergent whole (or an emergent property of a higher level entity) which we may call a *collective representation*, to use Durkheim's phrase, and that it is this collective representation that is causally effective. One could argue, for example, that the *commonality* of social practices cannot be explained by the causal effects of individual norms and values, but only by the commonality of those norms and values across the community, and hence (by the redescription argument) that it is the collective norm or value that is causally effective and not the individual one.

Giddens' duality of structure seems designed to allow him to hold both of these theoretical positions simultaneously while also denying them both. This is precisely the intention of the passage already quoted:

> Structure thus refers, in social analysis, to the structuring properties allowing the 'binding' of time-space in social systems, the properties which make it possible for discernibly similar social practices to exist across varying spans of time and space and which lend them 'systemic' form. To say that structure is a 'virtual order' of transformative relations means that social systems, as reproduced social practices, do not have 'structures' but rather exhibit 'structural properties' and that structure exists, as time-space presence, only in its instantiations in such practices and as memory traces orienting the conduct of knowledgeable human agents (Giddens, 1984, p. 17)

Thus he simultaneously claims that structure "makes it possible for discernibly similar social practices to exist" but that it "exists, as time-space presence, only in its instantiations in such practices and as memory traces". Thus, on the one hand, he claims that structure makes possible the commonality of practices, which would appear to require a quasi-Durkheimian notion of structure as something that is wider than the beliefs of individual human beings. This is the view that also seems implicit when he defines structure as "rules and resources, organised as properties of *social* systems" (Giddens, 1979, p. 66, emphasis added). Yet on the other hand he insists that structure exists only as instantiations in the practices and minds of individual human beings, and thus denies the Durkheimian notion. But it seems that such structure could only produce a commonality of social practices if there is a commonality to its instantiations in many human beings, and if this is case surely it is the latter commonality – a relational property of the group, and not a property of any individual or even of an unrelated group of individuals – that produces the former.

Reading Giddens' structuration theory as an account of the structure of institutions, then, seems to leave us with a contradiction. He wants norms and values

simultaneously to be more widely binding than their individual instantiations because of their social character, but also nothing more than their individual instantiations in ontological terms. The former claim depends upon accepting the claim that 'collective representations' are causally effective, while the latter depends upon denying it.

I tentatively suggest that there is a more plausible way of understanding the ontology of institutions: to recognise that collective norms are emergent properties of cultural groups, with the individual norms of members of the group as their underpinning properties. This suggests that collective norms *are* causally effective, although further ontological work is required to clarify their underpinning properties and causal mechanisms, and that they affect individual norms, which are *also* causally effective. In this picture, it is individual norms that appear to directly bring about individual instantiations of social practices, but their relationship with the cultural group holding the norms collectively is in some way analogous to that between role incumbents and organisations. Thus, for example, individuals holding norms may operate on behalf of the cultural group, in the same way that causally effective organisations work through their role incumbents. Alternatively, (or in addition) individual norms may be causally influenced by collective representations.

This is an area that requires more theorisation, and I have already indicated I intend to address this in my future work,[62] but it already seems a much more attractive solution than the ontology of structuration theory, for a number of reasons. First, there is no need to equivocate about the causal status of collective representations, unlike structuration theory's 'now you see it, now you don't' treatment, which depends upon the causal effectiveness of collective representations but simultaneously denies it. Second, there is no need to confuse the compositional relationship between institutional structure and individual agency. Individual norms may be underpinning properties of collective representations, and the individuals holding the norms may be the parts of the corresponding cultural groups, but the converse is not true, hence we escape the ontological conflationism of structuration theory. And third, in doing so we escape the need to conflate composition and causality: we can accept that collective representations causally affect individual norms without thereby arguing that they become parts of them.

---

[62] Archer has already made a significant contribution to theorising this area in (Archer, 1996a).

*One size does not fit any: organisations*

We can deal more briefly with the relationship between structuration and social structures that take the form of organisations, as defined in chapter four above. Here we must relate Giddens' account of structure to what he calls *collectivities* and *systems*. As Mouzelis and Giddens himself have pointed out, the latter concepts provide at least a partial defence against the charge of conflationism, as both are varieties of higher-level social entities which Giddens treats as having some ontological significance in their own right (Mouzelis, 2000, p. 760). In defending himself against charges of conflationism, Giddens has argued that his rejection of dualism in favour of duality "does not mean denying that there are social systems and forms of collectivity which have their own distinct structural properties" (Giddens, 1993, Introduction to second edition, pp. 4-5).

Giddens seems to have a somewhat functionalist conception of *system* (Giddens, 1979, p. 74) which he describes as a "structured totality" (Giddens, 1979, p. 64). *Collectivities* are perhaps closer to the concept of organisations I have emphasised in chapter four, although perhaps he means the two terms to be synonymous: he argues that "collectivities 'consist of' interactions between members but structures do not; any system of interaction, however, from a casual encounter up to a complex social organisation may be analysed structurally" (Giddens, 1976, p. 118). Returning to his response to accusations of conflationism, he says

> In structuration theory, the concept of 'structure' presumes that of 'system': it is only social systems or collectivities which have structural properties. Structure derives above all from regularized practices and is hence closely tied to institutionalization; structure gives form to totalizing influences in social life (Giddens, 1993, Introduction to second edition, p. 7).

In analysing organisations, then, Giddens seems to argue that they are structured by the "regularized practices" which are, or are produced by, institutions. Giddens himself, however, does not make the ontological distinction between collectivities and institutions so clear, because he does not treat institutions as entities or properties at all, but rather as processes.[63] It therefore seems in his exposition that there is only one sort of structured entity, corresponding to my concept of organisations, while institutionalization is nothing more than a part of the process by which their structuring occurs. This is entirely consistent, of course, with his nominal rejection of collective

---

[63] For an interesting and important discussion of the process oriented nature of Giddens' ontology, see (Sawyer, 2002b).

representations as entities or emergent properties in their own right in his treatment of institutions.

As I argued in chapter four, institutions do indeed play an important role in the emergence of organisations, contributing to their morphostasis, their morphogenesis, and indeed to the causal mechanisms underpinning some of their properties. But we cannot understand the ontological status of organisations in terms of institutional structure alone. We also need to understand their relational structure: to recognise that they are made up of people, but that although their parts are people, organisations have powers distinct from those of the people who make them up by virtue of the ways in which those people are related to each other in the organisation. Those relationships can be expressed as roles, and roles themselves can only be effective because the people involved have shared understandings of them. In other words, the relations that constitute organisations depend causally upon institutions. But the relations are not the same things as the institutions; we must avoid the error of confusing causation and composition. Giddens' account of the ontological basis of organisations or collectivities, then, is radically incomplete by comparison with the emergentist account presented in chapter four.

### Reconciliation?

Unlike Stones, then, I see no hope for a reconciliation between emergentism and structuration theory that is premised on accepting a structurationist ontology. Structuration's ontology is not only conflationist but fundamentally confused and incoherent, resting as it does on (a) an inherently contradictory attitude to collective representations; and (b) the systematic conflation of composition and causation. If there is to be a reconciliation, then, it must rest on accepting an emergentist ontology of structure and agency. But what opportunity does such a conclusion offer for a reconciliation?

The opportunity for reconciliation, I believe, arises from the potential for disarticulating structuration theory and its ontology. If we read Giddens as presenting a causal story about the relationships between agency and various forms of structure, and disregard his conflationist ontology, it becomes possible to extract some value from his work as a *theoretical* account. An opportunity may arise from the consideration of organisations above. Institutions and organisations are *different* kinds of structure and we cannot understand either unless we recognise this and theorise them differently, while recognising that both are emergent structures with causal powers in their own

right. Nevertheless, we do need to recognise that they are *interdependent* kinds of structure, especially in the sense that institutions are implicated in the structuring of organisations. Structuration theory is an attempt to theorise this interdependence as well as an attempt to develop a social ontology, and there may be significant value in Giddens' theory even if we reject his ontology. If structuration theorists are prepared to abandon their ontological claims there is room for dialogue about the processes by which structure and agency interact.

**Conclusion**

This chapter has compared and contrasted a number of the existing approaches to social structure with the account developed in chapter four. It has considered two explicitly anti-emergentist approaches – King's defence of methodological individualism, and Giddens' structuration theory – and two different attempts to develop an emergentist approach – those of Archer and Sawyer.

King's defence of methodological individualism fails in the face of the redescription principle, since he claims for individuals causal powers that can only exist when they are organised into higher-level social entities. He has nevertheless exposed a flaw in Archer's account of emergentism, which is her use of some examples that are really aggregates with no distinct causal powers in their own right as though they were cases of emergence. This flaw can be corrected, however, and once we have done so it is clear that Archer's account is broadly compatible with that developed in this thesis. Her emphasis is on morphogenesis, whereas this thesis pays more attention to the synchronic relational side of emergence, but the two accounts are essentially complementary.

Like Archer, Sawyer has done much valuable work in raising awareness of emergence amongst sociologists, but I have argued in this chapter that his account of emergence is flawed. He borrows Fodor's argument that higher level laws can be irreducible by virtue of multiple realisability or wild disjuncture – in other words, that there can be higher level functional properties that are built upon many differing lower level structures. However, this chapter has shown that multiply realisable properties are usually a special case of relational emergence, and that in such cases their irreducibility is a product of relational emergence and not of multiple realisability as such. Such properties provide an interesting and important extension to the account of emergence

presented in this thesis, but not a viable alternative basis for a general theory of emergence.

Finally, I have agreed with many other critics that Giddens' structuration theory fails to conceptualise adequately the relation of structure and individual action as a result of conflating the two and in effect denying the emergence of social entities. In terms of institutions, structuration theory's duality of structure seems like an attempt to evade the problematic ontological nature of collective representations, but it is an attempt that fails; he relies on the existence of collective representations while simultaneously denying it. In terms of organisations, he ignores the relational aspect of their structure and thus misses the very thing that makes them emergent wholes. Some of Giddens' work may be useful in helping us to understand the processes by which agency and certain sorts of social structure interact, but as an ontology of social structure it must be discarded.

I conclude, first, that the emergentist account of social structure developed in this thesis is decisively superior to the non-emergentist alternatives that have been reviewed here, and second, that it is both complementary to, and an enhancement of, improved versions of the emergentist alternatives offered by Archer and Sawyer.

The last two chapters, then, have justified the claim that social structures can be causally effective in their own right as a result of being emergent entities. Traditionally, however, such claims have been seen as antithetical to human agency. The next two chapters will show that this is not so, by developing an emergentist account of agency and showing how agency interacts with social structure.

# 6        Agency

As human beings, we are inclined to believe that the conscious exercise of our own capacities makes us the source of what happens in society. This chapter will examine this belief as it relates to the behaviour of the human individual. In what sense, and to what extent, can we human beings ourselves be regarded as agents with a causal impact on our social world?

There are a number of different definitions of agency, and we must distinguish at the outset between two groups of these – the concepts of *political agency* and *individual agency*. Political agency is the possession of "the power to bring about effective change in collective life" (Coole, 2005). Political agency, however, may potentially be exercised by other things than individual human beings – in Archer's account, for example, it is exercised by groups (Archer, 1995, pp. 257-8). This chapter is concerned, however, with the specific powers of human individuals, and therefore with agency in its second sense: individual agency. Individual agency can be defined as "the power of actors to operate independently of the determining constraints of social structure. The term is intended to convey the volitional, purposive nature of human activity as opposed to its constrained, determined aspects" (Jary and Jary, 2000,p. 9).

This chapter will argue that we human individuals do indeed have causal powers of our own, and that those causal powers are emergent from the organised set of our parts in accordance with the same general logic (outlined in chapters two and three above) that applies to social structures, and indeed to the structures of the natural world. Just as it is for any other emergent entity, it is the fact that a human individual is a particular organisation of particular sorts of parts that makes it possible for that individual to possess causal powers in his or her own right. Thus, in explaining the powers of human individuals – their capability of demonstrating agency – we must be prepared to consider how the five pillars of emergence apply: What are the parts, how related, that constitute human individuals, how does this sort of structure lead to the powers that they possess, and how is this sort of structure brought about and sustained? Clearly there are many levels at which we could pursue these questions, and as a work of social theory this thesis will neglect much of the biological explanation of the workings of human beings. But I do not believe we can neglect the biological basis of

human capacities entirely if we are to construct an adequate understanding of human social functioning.

Nor, however, can we neglect the fact that human behaviour is causally influenced by external factors. The idea of agency is often claimed to entail the freedom of human action from the external constraint of social structure (Jary and Jary, 2000, p.9; Loyal and Barnes, 2001, p.507). I will certainly accept that human action is not fully determined by social causes, but I will also argue that social entities do have a causal impact on our actions. The critical realist account of the co-determination of actual events by a multiplicity of causes, potentially from a variety of levels in the hierarchy of emergence, provides the framework needed to reconcile the claim for agency with the recognition of the causal impact of external factors on human action (both natural and social).

The first two parts of the chapter will address the questions: how do the causal powers of humans emerge, and how are these causal powers combined with external causal influences in the determination of human behaviour? They offer a positive theory of agency – of human action – based on the emergent properties of human individuals, as part of a hierarchy of entities with emergent powers, including both the biological parts of human beings and the social entities composed (at least in part) of human beings. In bringing together the biological, the individual, and the social contributors to the determination of human action, this chapter will draw on a variety of disciplinary approaches, including for example neuroscience, phenomenology, and the philosophy of action. It is one of the strengths of emergentism that it allows us to recognise the complementarities between these approaches and reconcile what often seem to be competing arguments about the determination of human action.

The remainder of the chapter will relate this emergentist theory of action to two important existing approaches. Pierre Bourdieu and Margaret Archer have advanced what seem at first sight to be incompatible theories of human agency. While Archer places heavy stress on conscious reflexive deliberation and the consequent choices of identity and projects that individuals make, Bourdieu's concept of *habitus* places equally heavy stress on the role of social conditioning in determining our behaviour, and largely ignores the contribution of human reflexivity. Despite this, I argue that these two approaches, with some modification, can both be accommodated within the emergentist theory of human action which is sketched out in this chapter. This is not to suggest that Archer and Bourdieu can be completely reconciled; their divergences are

multi-layered, and the resolution of the theoretical conflict depends in particular in rejecting the apparent conflationism of Bourdieu's ontology.

Overall, this chapter will focus on the positive construction of an emergentist ontology of individual human agency and a complementary emergentist theory of human action. However, this approach directly contradicts the ontological perspectives of many existing explanatory accounts of agency, including reductionist accounts from three different perspectives – biological, individualistic, and structuralist – as well as conflationist accounts. It also conflicts with those that explicitly or implicitly deny that we can provide a causal account of human action, including approaches inspired by the doctrine of free will and by hermeneutics. These various challenges to an emergentist view of agency will be discussed and rejected in chapter seven.

**Brains, minds, and agency**

Human individuals, I argue, are emergent systems, with properties and powers that arise from their components and the way in which these are organised to form a human being. For the purpose of explaining how societies work, the most important of these powers is our generic power to act, including our power to act communicatively. The next two sections will therefore be concerned with the emergent roots of our power to act, and with an analysis of the sorts of factors that affect how we can and do realise that power. These two sections are structured as an analysis of Davidson's influential claim that there are causally effective mental phenomena such as beliefs and desires, and that these can compose 'reasons' which can be causes of our actions (Davidson, 2001).

The current section will consider the ontological status of mental phenomena, and in particular the question of how they may be emergent from the networks of neurons that make up our brains, although it will engage only very briefly with the underlying neuroscience.[64] This section, then, constitutes an analysis of what parts of human individuals, and what organisation of them, underpin the emergence of our power to make decisions. The following section will consider the relationship between such decision making and the causation of human behaviour; a relationship, I will argue, that is less obvious and straightforward than is generally assumed. It will then go on to fit this in to a wider account of how the causal powers of a variety of different

---

[64] Nor will it attempt to explain or explore the nature of our experience of the mental, which is emphasised in phenomenological approaches, or in philosophical analyses of *qualia.*

entities can co-determine human action. Taken together, these two sections offer an explanatory reduction of human action in terms that allow appropriate roles to both mental phenomena and the 'hardware' of our brains.[65]

### *What are mental phenomena?*

We must begin by asking what mental phenomena are. While there is some agreement on what sorts of things are mental phenomena – sensations, beliefs, desires, intentions, concepts, reasons, and decisions, for example – the criteria that circumscribe the mental are more controversial. I shall accept Searle's approach, which implies that mental phenomena are thoughts of any type of which we can be conscious. Thus, for something to be mental, we must be able to think with it. This does not mean, of course, that we are conscious of, or thinking with, it all the time; we always hold a great many concepts and beliefs that we are not conscious of at that particular moment (Searle, 1992, p. 172).

This immediately leads us on, however, to another question: how can mental phenomena exist? What is it that connects the 'mind' – the array of mental phenomena we experience, if experience is the right word – to our bodies? The relatively recent development of the sciences of the brain – the neurosciences – is starting to provide answers to questions like this, although those answers are still highly incomplete. What they do seem to show is that mental phenomena, both when we are conscious of them and when we are not, are somehow produced by the networks of neurons that make up a large part of our brains.

John Searle, for example, argues that consciousness itself must have neurophysiological causes: "the *mental* state of consciousness is just an ordinary biological, that is, *physical,* feature of the brain" (Searle, 1992, p. 13). Searle sees this as an emergence relation:

> The brain causes certain 'mental' phenomena, such as conscious mental states, and these conscious states are simply higher-level features of the brain. Consciousness is a higher-level or emergent property of the brain in the utterly harmless sense of 'higher-level' or 'emergent' in which solidity is a higher-level emergent property of $H_2O$ molecules when they are in a lattice structure (ice)… Consciousness is a mental, and therefore physical, property of the brain in the sense in which liquidity is a property of systems of molecules (Searle, 1992, p. 14).[66]

---

[65] See (Searle, 1992, pp. 234-7) for more on 'hardware' processes in our brains.

[66] Also see (Searle, 1997, pp. 17-18).

I would agree: mental states (like concepts and beliefs) are emergent from inter-related groups of physical entities (like neurons and synapses). Conceived in downwardly inclusive terms, they are therefore both mental and physical; conceived in level-abstracted terms, we can see mental states as having distinct causal powers of their own (if they are entities) or as *being* causal powers in their own right (if they are properties). Beliefs, for example, may have causal effects on us that neurons cannot have when they are not structured through synaptic connections into a particular sort of higher level entity. For many purposes, we can neglect the emergence base and talk in terms of the causal impact of mental states, but for others we may have to take account of the neurophysiological underpinnings of the mental.

### The emergence of the mental

There are a variety of open questions here, as a consequence of the relatively underdeveloped state of the brain sciences. For example, it is not clear whether mental phenomena should be regarded as entities in their own right or as emergent properties. I tend to the view that mental phenomena are emergent properties, but this raises a second open question, for if mental states are properties of an entity then it is not immediately obvious what that entity is. If any given belief, for example, follows from the state of an identifiable set of neurons and the relations between them (let me call this a neural sub-network), then it would be a property of that particular sub-network. However, if each belief (or mental state) depends upon the state of all or most of the neurons in the brain, as some neuroscientists believe (Freeman, 2000, p. 148), and if these beliefs are interdependent with other beliefs which are also emergent from the same network, it may be more appropriate to see mental states as properties of a larger whole. This larger whole could be a particular subset of the brain – perhaps a subsystem for processing mental phenomena – or it could be the whole brain, as implied by Freeman's argument, or it could even be the whole human body. The latter is suggested by the recognition that some mental phenomena, such as perception, depend upon an interaction between parts of the brain and other parts of the body, and also because the nervous system, which for some purposes may behave like a part of the brain, extends throughout the body. This sort of consideration tends to encourage the idea of mental states, dispositions, and the like being 'embodied' but it still seems to depend upon the logic of neural networks, and thus on emergence from an identifiable subset of the body: the nervous system. It would be pointless to speculate further on the resolution of these questions, or indeed on many of the details of the mechanisms that seem to relate neural

networks to mental phenomena; these questions cannot be resolved until the neurosciences provide us with greater insight into these aspects of the brain.

But there are features of the mechanism that seem reasonably well established. In particular, the networks of neurons and synaptic connections between them that make up much of our brain are conditioned or configured by our experience. The mechanisms by which such networks can be conditioned to store knowledge, beliefs, and the like are relatively well understood at one level as a result of computer simulations of neural networks. Our mental states, such as beliefs, seem to be underpinned at the neuronal level by connections of varying strengths (frequencies) between neurons and groups of neurons. These connections tend to be strengthened when we have experiences that appear to confirm the mental state, and weakened when we have experiences that appear to undermine it. These neural connections, then, do not represent individual experiences, one at a time, but a kind of weighted summary of them. Our experiences, then, are a morphogenetic cause of the particular configurations of neurons and synapses that are the emergence base of our mental states.

It is the effects on our neural network, and therefore on our beliefs, of our experiences that 'conditions' us to possess certain mental states. Such conditioning need not be conscious – if we experience a particular pattern of stimuli repeatedly then our brain will learn from it without any necessary conscious intervention, as for example in the phenomenon of subliminal learning (Freeman, 2000, p. 191). On the other hand, our conscious thinking may itself provide inputs to the learning process. This conditioning mechanism provides the route by which 'socialisation' or analogous processes may play a significant role in establishing our beliefs and dispositions.

On the basis of this account, we can loosely describe the five pillars of the emergence relation by which mental entities or properties emerge from neural networks. In this emergence relation (i) the parts are neurons; (ii) these are related by synaptic connections which connect some pairs of neurons and not others, and in which the synaptic connections can have various strengths (firing frequencies); (iii) this network can produce a meaningful mental state because the varying strengths of these connections underpin our ability to think of concepts as having certain types of relationships with other concepts; (iv) the network connections underpinning a given mental state are created as a result of our experiences, and may be modified, weakened, or strengthened by further experiences; (v) the network representing a given concept is sustained over time by physiological processes that need not concern us, except to say

that such networks may fade over time and be forgotten or partially forgotten, but repeated exposure to or exercise of particular mental states will tend to lead to renewal of their strength.

This picture of mental phenomena and their emergence base is admittedly incomplete, and it is no doubt over-simplified in many respects. However, I believe it is sufficiently accurate to serve its purpose here: to make clear that a full account of human action must recognise and seek to theorise the biological basis of that action and its relationship to the higher-level influences on that action, and to show that this relationship can potentially be theorised as an emergence relation of mental phenomena from our physical brains and perhaps bodies.

With this picture of mental entities and their emergence base, we can now turn to the question of how they contribute to the determination of human action.


## Decisions, actions, and behaviour

This is an appropriate moment to return to Davidson's influential account of the relationship between mental entities and the determination of human action (Davidson, 2001). Davidson argues that mental phenomena, specifically *reasons*, can be causes of our actions. Primary reasons for an action arise from the combination of a belief (e.g. doing *x* will have the result *y*) and an attitude (e.g. I desire *y*), where beliefs and attitudes are mental phenomena. And, he says "A primary reason for an action is its cause" (Davidson, 2001, p. 12). This section will evaluate Davidson's argument.

Davidson himself recognises that it is necessary to deal with certain criticisms of such a view. For my purposes, the most relevant of these is the argument that causes imply causal laws, but there seems to be no regularity in the relation between the reasons people hold and their actions. He has two related responses to this, the first being to suggest that in any given case of action, there will be a mixture of reasons contributing to its causation (Davidson, 2001, p. 16), and the second being that the causal regularity underlying this process may exist at a lower level – "neurological, chemical, or physical" (Davidson, 2001, p. 17) – upon which the mental level supervenes.

I accept the suggestion that multiple different factors interact to cause a particular action, which is of course in keeping with the account of cause given in chapter 3 above. One might also point out that causation, as Bhaskar in particular has described it, does not require exceptionless empirical regularities at *any* level, so there is

no need to invoke lower levels in response to this particular criticism. We *do* need to consider the lower levels, however, if we are to provide a plausible explanation of *how* reasons could operate as causes.[67] My main concern in this section, therefore, will be with the examination of the brain processes required to make Davidson's mentalistic account work. I will argue that Davidson's account is incomplete but that we can produce a modified emergentist version of it that is consistent with what we know about the brain.

### The indirect link from reasons to actions

We must begin by asking what is meant by *reason* in this context. There are at least three alternatives. The first is an after-the-event description of what we now believe our motivation was for the action in question. We could call this a *post-event reason* or a rationalisation. But the verbalisation of such after-the-event descriptions is a separate action from the one we are attempting to explain. Such verbalisations may misrepresent our thinking at the time of the action, and since they occur after the event to be explained they cannot be its causes. At best, they are useful but fallible evidence about our motivations at the time of the original action. The second alternative is what I will call a *conscious reason*. This version of the concept implies that we acted because of a decision we made, through a process of conscious consideration of the reason in question. The third alternative we may call an *unconscious reason*. This would count as a cause of an action if there were beliefs, desires, and hence reasons implicit in our neural networks at the moment immediately preceding the action – as mental entities that we were not conscious of at the time – and these combined to generate our action without us being conscious of the fact. Davidson appears to intend the second of these alternatives, but I will argue below that to construct a viable version of the argument requires that we explain human behaviour in terms that combine explanations of both the second and the third type.

The argument is most easily approached by considering how decisions and behaviour are related to each other over time. Experimenters have shown that to take a conscious decision and implement it takes a minimum of a quarter of a second; yet top tennis players, for example, can react to a serve in a tenth of a second (Dennett, 2003, p. 238). How can this be? Dennett argues that "the tennis player commits to a simple plan

---

[67] I do not, however, accept Davidson's belief that supervenience provides an appropriate model of how to conceive of this relationship with the lower levels (see for example, Charles, 1992, p. 275; Savellos and Yalçin, 1995, pp. 9-10).

and then lets 'reflexes' execute her intentional act" (Dennett, 2003, p. 238). The 'simple plan' here consists of a set of consciously-chosen strategies, the precise strategy to be adopted being conditional on what type of serve is received, and the 'reflexes' consist of the ability of our brain and body not only to execute pre-determined strategies but also, when they have already been suitably trained by previous experience, to determine *how* to execute them (e.g. just how high and how wide to swing the racquet head) independently of any further conscious decision-making. Thus, the conscious decision takes place at one time, and the execution of that decision is done non-consciously at a later moment.[68] There is a decision *before* the other player serves, but there is no decision *between* the serve and the return, only an implementation of that previous decision. Furthermore, the conscious decision only partially describes the behaviour to be undertaken, leaving other details to be 'filled in' non-consciously.

Phenomenological investigation leads to a similar conclusion. Freeman, for example, writes that Merleau-Ponty "proposed that actions are not controlled by consciousness, because experience has already created an understanding of the present, from which action flows without need for reflection. Awareness is not essential for intentional coping, because many of our daily actions emerge without reflection" (Freeman, 2000, p. 171).[69]

Yet it is also true that our brains at least sometimes offer us the opportunity to consciously review and alter our behaviour when we are on the point of implementing it, as suggested by Freeman: "Brain activity preceding the initiation of an intentional act starts before the onset of awareness of an intent to engage in that action. The subjects also report that, after becoming aware that they are about to act, they can abort the action" (Freeman, 2000, p. 170). In cases like this, it seems that the brain activity preceding the initiation of the action represents the beginning of an action implementation process, which may be driven to some extent by past decisions, but which is potentially modifiable by a 'last minute' conscious review.

I suggest that *all* decision making works like this: that we do make conscious decisions but these decisions are only the *indirect* and *partial* causes of our behaviour,

---

[68] I follow Searle here in using *nonconscious* to refer to brain entities and events of a type which we can never be conscious of, and *unconscious* to refer to those that we are not conscious of at the time but could be conscious of at some other time (i.e. *mental* states of which we are not currently conscious) (Searle, 1992, p. 155).

[69] Merleau-Ponty sometimes seems to come close to expressing an emergentist view (e.g. Merleau-Ponty, 1963, pp. 150, 184) but there is also a strong hint of conflationism in his work which may have influenced both Giddens and Bourdieu.

in that (a) they occur a variable length of time before the action concerned; and (b) they are always incomplete regarding the details of the action to be taken.

Let us imagine, for example, the case in which I decide 'I'll have lunch when I've finished this paragraph'. Clearly I could represent this as a decision based on a conscious reasoned balancing of a number of beliefs and desires, and argue that these reasons caused my subsequent action of ceasing to write, getting up from my chair, walking into the kitchen, and preparing my lunch. However, it is clear that this is an incomplete account of the causation of this behaviour. First, some explanation is required of how my decision at one point in time becomes activated at another, say ten minutes later, when I come to finish the paragraph. Note that this is far from an automatic process. I may, for example, become engrossed in what I am writing and go on for several more paragraphs before I remember my intention. Or I may find the current paragraph impossibly difficult and decide to give up and have lunch before I finish it. Or I may finish the paragraph, start getting up for lunch, but alter my decision at the last minute because something else now seems more important. Yet, if I do have lunch at the end of the paragraph, my earlier decision to do so surely contributed causally to that outcome. This, of course, is entirely consistent with Bhaskar's account of real causal powers and multiple determination (see chapter three): the earlier decision has a tendency to produce a certain effect, but that effect may be frustrated by the effects of other causal factors.

Secondly, this decision is incomplete as a determination of my action because it says nothing about *how* I will implement that decision. It is quite likely, for example, that when I get up out of my chair I will walk through to the kitchen without paying the slightest conscious attention to how I move my legs in order to achieve this – there is no conscious decision at all involved in this part of my behaviour. As Freeman says, "we perform most daily activities that are clearly intentional and meaningful without being explicitly aware of them" (Freeman, 2000, p. 23). Thus, some parts of the behaviour I have decided upon are not themselves decided upon. Other parts may be decided upon, but as a result of some other decision at some other time. Take the question of how I sit when I eat – another part of implementing this decision to have lunch. I may have decided years ago to sit up straight at the table when eating, and go on to do so without re-making this decision.

Decisions, then, may have variable size or scope, in the sense that, say, a decision to drive to work has greater scope than the decision to turn left at a particular

junction on the way. This in turn has greater scope than the decision to turn the steering wheel a bit further to get round this corner successfully (although, of course, experienced drivers often do not make conscious decisions about how far to turn the steering wheel; they delegate this to a non-conscious skill established by previous training). Thus any single case of human behaviour may represent the (full or partial) realisation of a series of nested decisions of various sizes or scope.

### *Decisions produce dispositions*

It seems difficult to avoid the conclusion that our decisions, and with them the conscious reasons that motivate them, are merely inputs, among others, into the determination of our behaviour. Furthermore they are inputs with variable degrees of effect. As Barnes and Loyal have suggested, "Might it not be that all actions are chosen but that there is a range of chosen actions from those readily modified to those carried out with implacable will and determination?" (Loyal and Barnes, 2001, p. 523). Decisions, then, do not seem to produce behaviour directly, but rather produce dispositions to behave in a given way in the future in certain circumstances. Indeed, we may be able to define a decision as an event in which an episode of conscious reflection (a process) leads to changes in our dispositions (our tendencies to behave in particular ways). These dispositions then seem to be held with varying degrees of commitment, through being implemented in the brain as neural networks, in an analogous way to beliefs which we hold with varying degrees of confidence.

This suggests a model of the determination of human behaviour which fits neatly with the Bhaskarian conception of actual causation as the outcome of the interplay of a variety of causal powers. Let me represent this analytically as a series of steps:

(i) belief formation: we develop beliefs as a result of our experience, which are implemented at the neural level as neural networks;

(ii) decision making: we possess the power to think consciously about our plans, and make decisions, which are co-determined causally by our thinking powers and the network of beliefs that they work upon;

(iii) decision storage: having made decisions, these are stored in our neural networks as new or modified dispositions (note that there may be multiple loops back to step (ii) before an action actually occurs, including the 'last minute' conscious review of some of our decisions);

(iv) action implementation: our actions are determined directly and immediately by non-conscious brain processes which use our beliefs, previous decisions, and skills as inputs.

This same story can be told in two apparently contradictory ways. We can tell it with our conscious thinking 'in charge', on the grounds that we do consciously make decisions about what we are going to do – thus emphasising reflexivity. Or we can tell it with our nonconscious behaviour-determination processes 'in charge', on the grounds that those decisions are merely inputs to the real determination, that they can be overridden, and that they only ever relate to part of the determination of what we do in any single action – thus emphasising dispositions. While each of these stories may have its merits for the purpose of answering different questions, the most balanced story is one in which our conscious decision-making and our non-conscious behaviour-determination appear as complementary and mutually-necessary components in the causation of our actions.

Where does this leave Davidson's account of reasons as causes? Whether we read it in the second or the third of the senses I suggested earlier, as a conscious or a nonconscious account of the role of reasons, my account suggests that reasons *can* be causes of our actions, but they are only ever partial and contingent causes. Reasons co-determine our decisions, and decisions are stored in our brains as neural configurations – dispositions – which in turn co-determine our actions. But other factors are also involved, and these other factors can lead to some of our decisions not being realised. There are therefore good reasons why there is no exceptionless empirical regularity connecting reasons and actions: like any other causal power, the causal powers of reasons to motivate actions are contingent on the operation or non-operation of other causal powers with the capacity to co-determine our decisions and our subsequent behaviour.[70]

Reasons, then, can indeed be causes, but they are only co-determining causes and always operate in conjunction with a complex of other factors in determining our actual behaviour. Despite our intuition that our actions are determined immediately and

---

[70] Technically, there are yet further opportunities for other causal powers to co-determine our actions. So far we have only accounted for what we might call the determination of our motor movements. But in a sense our physical behaviour itself is co-determined by further external factors. Thus, for example, walking is a movement that is co-determined by (a) our motor movements; and (b) gravity – if there was no gravity operating, then these same motor movements would not result in 'walking'. But in practice, and certainly for most sociological purposes, we can usually abstract from this sort of physical co-determination of our behaviour

directly by our conscious decisions, the process by which our behaviour is determined (including the 'filling in' of details beyond our conscious decisions) is at least partially non-conscious.

Thus, the theory of action briefly outlined here shows how it might be possible that our actions are directly and non-consciously determined by our current dispositions, while allowing that those dispositions are themselves the outcome of a series of past events. Those events include (i) very recent reflections that we tend to see as directly causally effective 'decisions'; (ii) older reflections that shaped our dispositions consciously at the time but which we may now have forgotten; and (iii) experiences that affected our dispositions (for example in the subliminal acquisition of a habit or skill) without us ever consciously deciding how. The next section will look further into the question of non-conscious influences on our behaviour, and relate the argument to Bourdieu's concept of the *habitus*.

## Habits, skills, and habitus

What are these non-conscious causal influences on our behaviour? There are at least two types, which we can label *habits* and *skills*. But Bourdieu has developed a more ambitious account of human action which rests on the much broader concept of *habitus*. This section will begin by looking briefly at habits and skills, then examine Bourdieu's approach in more detail. It will argue that habitus provides a valuable component of the theory of action, but one that is seriously misleading unless it is placed in the wider context of an emergentist ontology, and the theory of action developed earlier in this chapter.

### *Habits and skills*

*Habits* are simply particular ways of doing things that we tend to repeat without thinking about them at the time. So, for example, I may roll out of bed the same way every morning without really thinking about how I'm doing it, or take the same cereal box out of the same cupboard when I arrive in the kitchen for breakfast without really thinking about what to eat, or where to find it. Habits like these may or may not be the product of earlier conscious decisions. I may never, for example, have thought consciously about how I get out of bed in the morning – I may just have developed a particular technique by a process of unthinking trial and error. Whereas I may have made a definite decision about what sort of cereal to eat in the mornings, perhaps years

ago, for carefully considered conscious reasons, and indeed a conscious decision about where to keep the box. Thus, my cereal-box habit may be unconscious at the time I do it but at least partly caused by conscious decisions in the dim and distant past, although other aspects of the way I perform it, such as the way I bend down to get the box, may have been acquired as non-consciously as the way I get out of bed.

*Skills* are ways of doing things which give rise to particular causal powers – learned techniques enabling us to do things we could not otherwise do, which generally require some effort to acquire. Like habits, skills may be consciously learned or merely internalised as a result of past experiences of observation, imitation, trial, and error. Thus, when I lift the spoon to my mouth to eat my morning cereal, I am executing a mundane but nevertheless skilled performance of spoon control. No doubt once upon a time my mother told me 'hold your spoon like this' and 'no, like this', but it is equally likely that I learned the finer points of spoon control by avoiding those movements that left my food on the table or the floor rather than in my mouth, and/or by watching and copying others. Perhaps a more subtle but significant example is provided by the acquisition of social skills. An example is the question of how close one stands to other people. There are no formal rules about how close one should stand to other people. No doubt there are books that will advise you on this question, but most of us learn what is appropriate by observation, imitation, trial, and error. We come to learn what is considered an appropriate distance in various contexts in our own particular culture by doing what other people do, and by tacitly noticing when others react to what they consider inappropriate distancing on our part.

Now, although there may be moments of both conscious awareness and conscious decision-making involved in the acquisition of skills and habits, we do not think consciously about them on each occasion that we enact them. Until I started writing this section, for example, I doubt whether I had thought about how I use my spoon in the last decade. For all practical purposes, we can regard these as non-conscious inputs to the determination of our behaviour.

### Bourdieu's concept of the habitus

A number of social theorists and philosophers have recognised the significance of habits and/or skills.[71] In an important extension of this argument, Bourdieu has

---

[71] Margaret Archer, for example, talks of *embodied knowledge*: "a 'knowing how' when doing, rather than a 'knowing that' in thought" (Archer, 2000a, p. 162), although she argues that its

centred his influential account of human social behaviour on the concept of *habitus*. Habitus, for Bourdieu, is the set of dispositions inculcated in each of us by the conditioning that follows from our social environment.

> The conditionings associated with a particular class of conditions of existence produce *habitus*, systems of durable, transposable dispositions, structured structures predisposed to function as structuring structures, that is, as principles which generate and organize practices and representations that can be objectively adapted to their outcomes without presupposing a conscious aiming at ends or an express mastery of the operations necessary in order to attain them (Bourdieu, 1990b, p. 53).

The conditioning that follows automatically from the opportunities and necessities inherent in our social position, he argues, tends to "generate dispositions objectively compatible with these conditions and in a sense pre-adapted to their demands" (Bourdieu, 1990b, p. 54). This is an effect that is particularly powerful in early life, generating a durable attitude to the world that motivates us to see the world in the terms dictated to us by our early social position, and to behave in the ways more or less mandated to us by that position (Bourdieu, 1990b, p. 53). Since all those who share a given social position are exposed to the same opportunities and necessities, they tend to develop a similar habitus, hence their social practices tend "to be objectively harmonized without any calculation or conscious reference to a norm and mutually adjusted in the absence of any direct interaction or, *a fortiori*, explicit co-ordination" (Bourdieu, 1990b, pp. 58-9).

Thus, the habitus, produced by social conditioning, tends to encourage us to behave in ways that reproduce the existing practices and hence structure of society. This conditioning is so effective that the dispositions it generates are below consciousness, and in some cases embedded in the most physical ways in which we use our body, becoming "embodied history, internalized as a second nature and so forgotten as history" (Bourdieu, 1990b, p. 56).[72] Not only do they *exist* below consciousness, though, Bourdieu also clearly suggests that they *operate* below consciousness. In *The Logic of Practice* he tells us "The *habitus* is a spontaneity without consciousness or will" (Bourdieu, 1990b, p. 56). And in *Distinction*:

> The schemes of the habitus, the primary forms of classification, owe their specific efficacy to the fact that they function below the level of consciousness and language, beyond the reach of introspective scrutiny or control by the will. Orienting practices practically, they embed what some would mistakenly call *values* in the most

---

role is confined to our dealings with the natural world. Archer, Bourdieu, and Giddens may all have been influenced by Merleau-Ponty in their treatment of embodiment.

[72] There is a fascinating echo here of Durkheim (Durkheim, 1964 [1901], p. 6).

automatic gestures or the apparently most insignificant techniques of the body – ways of walking or blowing one's nose, ways of eating or talking (Bourdieu, 1984, p. 466).

This is an enormously powerful account of the way our past moulds certain aspects of our behaviour. Thompson illustrates this with Bourdieu's explanation of accents:

> The linguistic habitus is also inscribed in the body... A particular accent, for instance, is the product of a certain way of moving the tongue, the lips, etc.: it is an aspect of what Bourdieu calls, following Pierre Guiraud, an 'articulatory style'. The fact that different groups and classes have different accents, intonations and ways of speaking is a manifestation, at the level of language, of the socially structured character of the habitus (Thompson, 1992, p. 17).

This disposition to form our mouths into certain shapes, and thus produce a certain accent, when we speak is one that is generally neither consciously learned nor consciously considered when we speak yet it tends to reflect our social origins. This is an example of what Freeman calls "classical conditioning of behaviour, by which we can learn without being aware of the process and the outcome or being able to recall them" (Freeman, 2000, p. 191).

The primary difficulty with Bourdieu's habitus appears when it is seen as an explanation of all, or a large proportion, of human behaviour. Many of our dispositions, for example, seem to be *learned* quite consciously via explicit verbal instruction, rather than being absorbed and embodied sub-consciously. Bourdieu, however, could presumably accept this modification of the argument while still maintaining that such learning becomes embodied, internalized, and in a sense forgotten – as happens when we learn a new sport, for example. This would still leave us with a habitus of dispositions that derived largely from the opportunities and necessities inherent in our social position, and able to operate sub-consciously on our subsequent behaviour.

A similar but more serious objection can be made to the suggestion that the *operation* of habitus is sub-conscious. A number of authors have criticised Bourdieu for his apparent denial of conscious or deliberative or strategic decision making in the determination of human behaviour, in marked contrast to most theorists of agency. In their view, habitus becomes nothing more than a conveyor belt for the determination of human behaviour by social forces. King lists no less than eight authors who have interpreted Bourdieu in this way (King, 2000, p. 418) and Wacquant lists another three (Wacquant, 1993, p. 238). However, as Wacquant points out, there are many authors who see another side to habitus (Wacquant, 1993, p. 238). Brubaker and Bouveresse both suggest that Bourdieu positions habitus as the explanation for a certain class of

actions, rather than as the single principle of all actions, and thus as operating alongside other principles, such as rational calculation or conscious norm-observance, which explain other classes of actions (Bouveresse, 1999, p. 15; Brubaker, 1993, p. 214).

Furthermore, Bourdieu stresses that the habitus itself does not operate in as deterministic a way as some authors suppose. In particular, he argues that what the habitus produces is not automatically determined actions, but what has been called a "capacity for constant improvisation" (Postone et al., 1993, p. 4). He strenuously resists the labelling of his work as undermining this capacity:

> It is easy to see how absurd is the cataloguing which leads people to subsume under structuralism, which destroys the subject, a body of work which has been guided by the desire to reintroduce the agent's practice, his or her capacity for invention and improvisation ... I wanted to emphasise that this 'creative', active, inventive capacity was not that of a transcendental subject in the idealist tradition, but that of an acting agent (Bourdieu, 1990a, p. 13)

Part of the reason for this divergence of opinions among commentators, however, seems to be that "it is not clear *how* dispositions produce practices" (Jenkins, 2002, p. 79) and thus "it is difficult to know where to place conscious deliberation and awareness in Bourdieu's scheme of things" (Jenkins, 2002, p. 77). In the absence of a clear explanation of how dispositions produce practices, it is understandable that there is confusion about the apparent conflict between Bourdieu's stress on the subconscious operation of habitus and his insistence that it operates through active, creative, invention and improvisation.[73]

## Archer vs. Bourdieu

Bourdieu, then, despite his insistence on the inventive and creative capacity of the habitus, seems less than decisively committed to allowing a role for conscious deliberation in the determination of action. By contrast, Archer's account of human action places conscious reflexive deliberation at its heart.

### Archer and reflexivity

For Archer, reflexivity is a power which human beings possess: it is the ability to monitor ourselves in relation to our circumstances (Archer, 2003, pp. 9, 14). It is exercised through a process of conscious *reflexive deliberations* during which we

---

[73] Thus the divergent views on habitus are *not* just a product of the inappropriate imposition of categories derived from Anglo-American sociology on Bourdieu's work, as Wacquant suggests (Wacquant, 1993).

conduct internal conversations with ourselves about ourselves (Archer, 2003, p. 25) – our situation, our behaviour, our values, our aspirations. The inner conversation "is a ceaseless discussion about the satisfaction of our ultimate concerns and a monitoring of the self and its commitments" (Archer, 2000a, p. 195).

Such reflexivity, she argues, is a "mature ability" and a precursor to the development of a *personal identity* and a *social identity*. These senses of who we are depend upon us delineating what we care about (thus defining one's personal identity) and relating it to our social context to develop projects that are based upon our ultimate concerns and which we use to guide the conduct of our lives (thus defining our social identity) (Archer, 2000a, pp. 9-10, 219). And for Archer reflexivity is specifically a *causal* power (Archer, 2003, p. 9). Thus in our reflexive deliberations we come to conclusions that affect our behaviour in the social world.

There is a strongly humanistic element to Archer's stress on the conscious nature of our reflexive deliberations and the opportunity that they present us to make decisions for ourselves about how we will conduct our lives. This is not, however, at the expense of social influences on human behaviour; as she says, "we do not make our personal identities under the circumstances of our own choosing. Our placement in society rebounds upon us, affecting the persons we become, but also and more forcefully influencing the social identities which we can achieve" (Archer, 2000a, p. 10). And indeed Archer has devoted two volumes to showing that social structures and cultural systems have causal powers in their own right (Archer, 1995; Archer, 1996a). At the same time, she rejects the implication that one's social position fully determines one's subjectivity or behaviour, pointing out (contra Bourdieu) that these develop in very diverse ways amongst people with the same social background (Archer, 2003, p. 348).

What is critical for Archer, as for me, in these relationships is that we continue to recognise that human beings, social structures, and cultural entities each have their own distinct existences and influences on social outcomes. None of these types of entity can be eliminated from the explanation of social events, nor conflated with each other in such explanations. In accordance with this ontology, Archer rejects views of human action that deny causal power to individual humans and their reflexivity. Thus she criticises those who argue that human action can be explained without recognition of the causal powers of human beings as such – whether because they substitute the powers of our biological parts for the powers of the whole human being (e.g. neural reductionists)

or because they substitute social forces for them (e.g. accounts of human action as socially-determined discourse) (Archer, 2003, pp. 10-14).[74] And she criticises the view that human agency and social structure can be conflated, which she perceives most clearly in Giddens' structuration theory (Archer, 1982) but also in the work of Bourdieu himself (Archer, 2003, pp. 11-12).

Archer and Bourdieu are therefore opposed at two distinct levels: in terms of both their theoretical and their ontological views of human agency. At the theoretical level, the conflict turns on the extent to which human beings influence their own destiny. While Archer rejects "contemporary social theory that seeks to diminish human properties and powers" (Archer, 2000a, back cover) Bourdieu sees human action as driven by a socially-derived habitus that provides "a spontaneity without consciousness or will" (Bourdieu, 1990b, p. 56). At the ontological level, the question turns on whether social structure can be seen as distinct from human beings or whether the two are mutually constitutive.

### Ontological conflation in Bourdieu

Archer discusses the ontological differences using the example of how Bourdieu might see one of her research subjects ('Graham') – perhaps making some conscious choices, but, "largely unaware that his horizons have been socially reduced" as a consequence of social conditioning (Archer, 2003, p. 11). For Archer, the problem with this position is that

> there never comes a point at which it is possible to disentangle Graham's personal caution (a subjective property of a person) from the characteristics of his context (objective properties of society)… All that is certain is that he does not have the last word about himself, his intentions or actions. Therefore, it becomes impossible that Graham can deliberate upon his circumstances as subject to object, because these are now inseparable for 'Graham' (Archer, 2003, p. 12).

This is an example of the more general ontological error of conflationism, which "rests upon conceptualising 'structures' and 'agents' as ontologically inseparable because each enters into the other's constitution" (Archer, 2003, p. 1).

Thus Archer sees the divergence between Bourdieu and herself as primarily ontological, mirroring precisely her critique of Giddens' structuration theory (see chapter five above, and Archer, 1982; Archer, 1995, ch. 4). By comparison with both Archer and Giddens, however, Bourdieu is rather vaguer about the ontological relationship between structure and agency. Like both, he clearly rejects both

---

[74] The next chapter discusses these forms of reductionism in more detail.

methodological individualism (in the form of Sartre's subjectivism) and methodological collectivism (in the structuralism of Levi-Strauss and Althusser), and seeks to find a middle way that can accommodate some features of both (Bourdieu, 1990a, pp. 9-13). But does he take the conflationist or the emergentist route between these two? Strong support for the accusation of conflationism can be found in Bourdieu's description of habitus as "systems of durable, transposable dispositions, structured structures predisposed to function as structuring structures, that is, as principles which generate and organize practices and representations" (Bourdieu, 1990b, p. 53). Dispositions are features of human individuals, so here he seems to be equating structure with internal human properties in much the same way that Giddens equates structure with rules (Giddens, 1984, pp. 17-25).

This seems to fit Archer's characterisation of conflationism (cited above), with agency and structures each entering into the constitution of the other. On the one hand, agents and their knowledge are constitutive of structures:

> To speak of habitus is to include in the object the knowledge which the agents, who are part of the object, have of the object, and the contribution this knowledge makes to the reality of the object. But it is not only a matter of putting back into the real world that one is endeavouring to know, a knowledge of the real world that contributes to its reality (and also to the force it exerts). It means conferring on this knowledge a genuinely constitutive power, the very power it is denied when, in the name of an objectivist conception of objectivity, one makes common knowledge or theoretical knowledge a mere reflection of the real world (Bourdieu, 1984, p. 467).

And on the other, structures are constitutive of agents:

> Overriding the spurious opposition between the forces inscribed in an earlier state of the system, outside the body, and internal forces arising instantaneously as motivations springing from free will, the internal dispositions – the internalization of externality – enable the external forces to exert themselves, but in accordance with the specific logic of the organisms in which they are incorporated (Bourdieu, 1990b, pp. 54-5).

If both of these claims are maintained, then it is difficult to see how agents can be distinguished from structure and vice-versa. However, I suggest, Bourdieu's position can be made compatible with an emergentist ontology with some relatively subtle changes that leave his *theoretical* agenda intact.

To begin with, we need not alter the claim that agents are constitutive of structures. Indeed, the emergence relationship is concerned precisely with the question of how parts interact to generate wholes with emergent properties. Thus it is perfectly compatible with an emergentist ontology to argue that structures ('the object') are made up of agents, thereby inherently including in the structure the knowledge that agents

have of the structure by virtue of including the agents as its parts and thus their knowledge as properties of these parts. This knowledge has a central role to play in the interplay of structure and agency which perpetuates that structure. It is therefore perfectly compatible with the emergentist position to see this knowledge as constitutive of structure.

The second claim, however, brings us to the heart of the ontological disagreement, with the phrase "the internalization of externality".

### *The internalization of externality*

On a metaphorical reading of *internalization*, the second claim is entirely compatible with an emergentist ontology; on a literal reading, it is entirely incompatible. Let me begin with the metaphorical reading. In this sense, when we 'internalize' something, our beliefs about the world are affected by our experience in such a way that we accept a belief about that thing as a fact. Thus, for example, we may internalize a sense of inferiority as a result of being persistently treated as though we are inferior by people around us. Metaphorically, we may say that we have internalized our inferiority, but literally, what we mean is that we have acquired the *belief* that we are inferior. Now in this sense of *internalization*, Bourdieu's passage above means that our beliefs about the world, or our dispositions towards acting in it, are affected by our experiences of social structures, and as a consequence those social structures have an effect on our behaviour. These beliefs and dispositions are not to be equated with social structure, nor to substitute for the notion of a distinct social structure, but to be seen as features of the human beings who are parts of the structure. This does indeed overcome a "spurious opposition between the forces inscribed in an earlier state of the system, outside the body, and internal forces arising instantaneously as motivations springing from free will" since it helps to make clear the mechanism through which the external forces causally affect the internal ones. Here, the 'external forces' do not disappear into the body but their effectiveness derives in part from a process that depends upon their effects on the body.

Unfortunately the literal sense of *internalization* leads to a very different interpretation of Bourdieu's argument, and it is this sense that is encouraged by the description of habitus as "structured structures predisposed to function as structuring structures". In this sense, when we *internalize* something it becomes literally part of us. In this sense, habitus is not merely a set of dispositions that has been causally influenced by our experiences of social structure. Instead habitus literally *is* structure,

internalized into our bodies – a view that closely reflects Giddens' conception of structure, which I criticised in chapter five. And on this reading, Bourdieu is not simply rejecting a spurious opposition between external and internal forces, but also denying the real distinction between external and internal forces. Now, beliefs and dispositions are no longer properties of human beings who are distinct from social structures; rather they represent an ontological penetration of the individual by the social structure. On this reading, structures really *are* parts of people. If this is what Bourdieu intends, then his position is indeed conflationist. Such a view, however, is not only incompatible with an emergentist ontology; it is also a clear ontological error, in that it fails to distinguish between a thing and its causal consequences. To be more specific, it fails to distinguish between a social structure and the consequences it has for our mental states. This is the same species of error as the claim that a child leaving a zoo has animals in their head, rather than thoughts or beliefs *about* the animals they have seen.

Distinguishing which of these readings Bourdieu really intends is not easy. He does not seem to have considered emergence at all, and he pays little attention to the ontological niceties required to distinguish an emergentist from a conflationist perspective, and so his account is open to a variety of ontological interpretations. At the ontological level, then, I suggest there is scope for reconciling Archer and Bourdieu through an emergentist reading of Bourdieu's ontology. And as I have argued both above and in chapter five, the conflationist alternative is untenable. Hence Bourdieu's ontology is only viable if we give it an emergentist interpretation, whatever his own intentions were.

### The theoretical tension between Archer and Bourdieu

Ontology, however, is not entirely independent of theory; this strategy will therefore only work if Bourdieu's *theoretical* position is compatible with such a reading. And of course we must still consider the second apparent conflict between Archer and Bourdieu: their differing perspectives on the theoretical relationship of human causal powers to human action. The theory of action outlined earlier in this chapter provides the basis for resolving both of these questions.

Before examining my proposed resolution, we must briefly consider the most obvious way of resolving the theoretical conflict – the argument that some actions are reflexively determined and others are determined by the habitus, so that both Archer's and Bourdieu's theories are right, but about different actions. Thus, for example, I might

exercise my reflexivity in deciding how to vote, but be driven by my habitus in the degree of deference I display towards the officials in the polling station.

In a sense, both authors allow space for just such a reading of their argument. Bourdieu, for example, writes "if one fails to recognize any form of action other than rational action or mechanical reaction, it is impossible to understand the logic of all the actions that are reasonable without being the product of a reasoned design" (Bourdieu, 1990b, p. 50), which seems to suggest that he accepts that *some* actions are indeed the product of reasoned design.[75] And Archer suggests that personal identity, which seems to be a corequisite of reflexive deliberation "comes only at maturity but it is not attained by all" (Archer, 2000a, p. 10). Hence at any one time some people will not yet have become reflexive, and others will never do so – leaving them, it would seem, in the grip of their habitus. On this reading Bourdieu's insistence on the role of the habitus, and Archer's insistence on the role of reflexivity can be seen as logically compatible, with their different emphases reflecting either a desire to stress the importance of their own theoretical perspective; or an implicit argument about what proportion of our actions fits into each category.

This chapter will argue, however, that there is a stronger way to reconcile these two theoretical perspectives. The heart of the argument will be that many and perhaps most of our actions are co-determined by *both* our habitus and our reflexive deliberations; and that despite the apparently conflicting implications of these two perspectives for our sense of our ability to choose our actions, they in fact represent two complementary moments of one and the same process. Before we turn to this theoretical disagreement, however, it will be useful to address the internal tensions in Bourdieu's view of the habitus.

**Resolving the tensions in Bourdieu's *habitus***

The primary tension I have identified within Bourdieu's account is that between his stress on the subconscious operation of the habitus and his insistence that the habitus operates through active, creative, invention and improvisation. The route to resolving this tension is provided by the theory of action developed earlier in this chapter, and the emergence relation it describes between the mental and the neural.

---

[75] Bouveresse reads him in this way: (Bouveresse, 1999, p. 49). And Bohman argues the case for explaining some social phenomena in terms of habitus but not others (Bohman, 1999, p. 132-3).

*Finding a place for consciousness*

In particular, my argument that our actions are caused by the dispositions stored in our neural networks as a result of past decisions and experiences maps closely onto the claim that our practices are caused non-consciously by our habitus, although Bourdieu sees this in somatic rather than specifically neural terms. At the same time, the role I allow to decision-making in amending this set of dispositions provides the mechanism by which the operation of our dispositions can be inventive and improvisational – not directly, in the action-implementation phase, but indirectly, in the decision-taking phase of the process, which can be invoked up to the very last moment, perhaps for example when the set of existing dispositions does not provide decisive guidance to the brain on how to implement a given action. This process of interaction between an emergent mental layer invoked in the process of decision-making and the underlying neural layer which translates our dispositions into actual behaviour provides a clear account of how dispositions can indeed produce practices while leaving space for conscious decision making.

Just as importantly, this account shows how it is possible that some parts of our actions can be determined more or less unconsciously while others are determined as a consequence of conscious, and perhaps rational, decision making. Where the translation into behaviour of a disposition that has been embedded in our neural network is unproblematic – such as the usual way in which we shape our mouth to speak, and thus of course the accent that we produce – then the process of action implementation can proceed with no reference to the conscious level. Thus, we can often "react appropriately to situations instantly, without reflection" (Sayer, 2005, p. 951). But where this translation *is* problematic – say, when we need to decide which way to turn *en route* to a place we have never visited before – then our consciousness must be invoked to provide a conscious decision which will complete the set of dispositions required to determine the action to be implemented.

It is, however, typically different aspects of the same behaviour that need to be explained in these two different ways, as opposed to entirely distinct actions. Say, for example, I need to reply to a difficult question. In doing so, I may reproduce an accent by shaping my mouth in ways that I implement entirely without conscious thought, but in the very same speech act I may express an idea which I must carefully think through in a conscious decision-making process. Even the implementation of a conscious

decision into the form of a socially competent performance is thus achieved as a matter of routine.

We can relate this back to the voting example introduced earlier: it is not that the act of voting is consciously reflexive while the act of speaking to the polling official is driven by habitus. Rather, some aspects of *both* actions are driven by conscious decisions taken in the very recent past, whereas other aspects of the same actions are driven unconsciously from our accumulated set of dispositions – our habitus.

Although this account of action is therefore consistent with many aspects of Bourdieu's habitus, it provides an explicit role for conscious input to our dispositions that Bourdieu largely neglects. As I have argued above, there is an important role for *conscious learning* in the construction of our habitus. To be told "that's not for the likes of us" (Bourdieu, 1984, p. 471), for example, may deeply affect our habitus but it does so through a conscious process. And *conscious decision-making*, too, plays a key role in my account because it alters our set of dispositions.

Of course, that decision-making itself is always heavily influenced by our existing set of dispositions. As Thompson has put it, "To view action as the outcome of conscious calculation… is to neglect the fact that, by virtue of the habitus, individuals are *already predisposed* to act in certain ways, pursue certain goals, avow certain tastes, and so on" (Thompson, 1992, pp. 16-17). Thus decision-making is never independent of the habitus, of our existing set of dispositions. But it does provide a mechanism for the amendment of our dispositions, most obviously in response to new situations which are not congruent with our previous experience. For example, when we adopt a new role, we may have to think carefully about *how* to perform it, and this may be guided not only by the dispositions arising from our previous social positions, but also by consciously absorbed new information, such as instruction from a supervisor, or information from a book. Thus, we cannot account fully for our dispositions without taking into account the role of both conscious learning and conscious decision making in their determination.

On the one hand, then, we need to recognise the role of decision making in contributing to our dispositions; and on the other, we need to recognise the essential role played by unconscious dispositions in the implementation of decision making and also as influential inputs to the making of those decisions. Thus we have what Bourdieu has called "a permanent dialectic between an organizing consciousness and automatic behaviours" (Bourdieu, 1990b, p. 80). Although I have placed more emphasis on the

role of the conscious in this dialectic than Bourdieu, I suggest that my account here is essentially consistent with his apparent intentions.

### The durability of the habitus

Nevertheless, this account of human action is still in tension with Bourdieu's on at least one more count: his claim that the habitus is durable, "ingrained in the body in such a way that they endure through the life history of the individual" (Thompson, 1992, p. 13). On my account, our dispositions can clearly develop substantially over time. Even the most deeply embodied of our dispositions can be altered. We may consciously choose to alter our accent, for example, if we move into new circumstances where people find us hard to understand, or where we feel out of place.[76] The tension here is between the idea that the habitus consists of all our dispositions, and the idea that the habitus endures unchanged throughout an individual's life. Only one of these ideas can be definitional to the habitus, since we are constantly altering our dispositions. The habitus can *either* be defined as the set of all our dispositions, in which case it is not entirely durable; *or* it can be defined as the set of our durable dispositions, in which case it is clearly only going to be a subset of our dispositions. It would seem more consistent, with Bourdieu's intention and with the theory of action outlined here, to regard the habitus as *all* of our dispositions.

Where does this leave Bourdieu's claim that the habitus is durable? I suggest that we can still maintain a version of this claim, not as definitional to the habitus, but as an empirical claim about certain aspects of the habitus in certain situations. Thus, while accepting that certain of our dispositions are changeable, and others are developed as short-term responses to particular situations, it is possible to argue that the social conditions that frame our dispositions are such that an important subset of our dispositions are determined relatively durably by our early experiences arising from our social position. This also leaves open the possibility that the habitus of certain groups of people in certain socio-historical situations are more stable and more durable than those of others. It is tempting to argue, for example, that habitus were far more stable in the feudal period than they are in the richest countries today given the vastly different rates of social change (and thus of changes in the social positions of individuals over their lifetimes) between these two types of society (Bourdieu, 1990a, pp. 73-4).

---

[76] Thus Archer is correct to criticise Bourdieu for portraying bodily skills as 'beyond the grasp of consciousness' (Archer, 2000a, p. 166).

The tensions within Bourdieu's account of habitus, then, with some relatively subtle amendments and qualifications to that account, would seem to be resolved by integrating that account with the theory of action developed in this chapter. By showing how action can be the product of our dispositions, while also providing a place for creative conscious decision-making in the determination of these dispositions, I believe the account above resolves the apparent contradiction between elements of voluntarism and determinism that many thinkers have wrestled with in Bourdieu's thought. Ultimately, this resolution rests upon the recognition – and theorisation – of the emergent relation between our mental states and processes on the one hand, and our neural states and processes on the other. By placing Bourdieu's theory within the emergentist framework, therefore, it is possible to increase both its internal coherence and its potential as a theory. [77] And the converse is also true: the demonstration that the theory of action developed in this chapter is consistent with a plausible and coherent interpretation of Bourdieu's habitus both lends credibility to the theory and also fleshes it out with Bourdieu's extremely powerful set of existing theoretical tools.

## Reconciling Archer and Bourdieu

It should already be clear that in resolving the tensions in Bourdieu's thought, the previous section has also opened the way to reconciling his theory of the habitus with Archer's account of reflexivity. It is precisely by showing how reflexive deliberation can enter into the same process of action determination as the habitus that the previous section reconciled conflicting interpretations of the habitus. The same argument allows us to position both Bourdieu's work and Archer's within a fuller account of human action. The purpose of this section is to show how Archer's contribution can be brought into the synthesis.

### *The effect of reflexivity on dispositions*

The starting point here is the recognition that we are constantly presented with opportunities for reflexive review of our beliefs and intentions, which have an effect on

---

[77] In an alternative attempt to reconcile the habitus with emergentism. Lau has argued that habitus is itself emergent (Lau, 2004, p. 370). I believe this is an error. A habitus is a collection of dispositions, each of which has a causal impact of its own (whether as an emergent property or an emergent entity), but I do not believe that the habitus has any causal powers of its own beyond those of the dispositions themselves. Hence it is a heap (or perhaps just a somewhat arbitrary collection of properties) and not an emergent entity, although it serves a useful theoretical purpose as an abstraction.

our actions via altering our dispositions. Conscious reflexive deliberation therefore plays a role in influencing the dispositions that in turn largely determine our actions.

In practice this means that when we act, some aspects of our actions may be determined with little or no conscious input – such as the way our mouth movements form our accent when we speak – while others are strongly influenced by recent reflection. The extent to which reflection affects our actions is, however, left open by this theory. It seems likely that this extent is highly variable, across a number of dimensions. Let us consider three of these. First, the same individual may be highly reflexive with regard to some aspects of their behaviour, but strongly driven by their social conditioning with regard to others. Consider, for example, the radical male political activist who is highly reflexive in his response to globalization, war, or capitalism, yet uncritically reproduces the attitudes and behaviours towards women acquired from the culture of his upbringing. Second, individuals from different backgrounds may display a different balance of reflexive and unreflexive action – which is, of course, a key part of Bourdieu's argument – such that *on average* individuals from an intellectual background, for example, may be more questioning of their dispositions than those from a working class background.[78] And thirdly, as noted above, different societies in different historical periods may show marked differences in the degree of reflexivity demonstrated by their members; thus, for example feudal societies probably discouraged any sort of challenging of the habitus, whereas contemporary post-industrial societies positively demand it, with their constantly changing environments constantly disrupting the assumptions of the habitus, and with education systems that must increasingly prepare children to be flexible in later life (cf. Bourdieu, 1990a, pp. 73-4).

If these speculations are valid, the contribution of reflexivity to the causation of human action varies by individual, by social class, and by historical context. Hence we need to theorise the ways in which reflexivity develops and operates, as well as theorising the less reflexive aspects of the development and operation of the habitus. We need a theory of reflexivity to complement Bourdieu's theory of the habitus, and Archer's *Being Human* and the fascinating empirical work in *Structure, Agency, and the Internal Conversation* offer a substantial contribution to just such a theory (Archer, 2000a; Archer, 2003).

---

[78] Even if this is true as a generalisation, of course, there are many individuals who are exceptions.

*The effect of the social on dispositions*

Once again, however, some reinterpretation of the argument will be required. Archer's analysis of the acquisition of personal and social identity is a compelling story about the development of reflexivity, but it is a story that neglects the role of the habitus. Archer certainly argues that social structure does affect human action, but she does not see its effects being channelled through our dispositions. Thus, for example, she argues that "we do not make our personal identities under the circumstances of our own choosing. Our placement in society rebounds upon us, affecting the persons we become, but also and more forcefully influencing the social identities which we can achieve" (Archer, 2000a, p. 10). Thus both our choices of primary concerns, and our choices of roles and projects through which we can pursue them, are constrained by our social context. However, Archer tends to stress the externality of social forces, as when she says that the individual is right to believe:

> that he lives in a social world that has different properties and powers from his own – ones which constrain (and enable) his actions. These are *temporally prior* to his conceiving of a course of action, *relatively autonomous* from how he takes them to be, but can *causally influence* the achievement of his plans by frustrating them or advancing them (Archer, 2003, p. 14, also see pp. 134-5).

Structures are thus seen as having an influence on the outcome of our plans rather than on our subjectivity itself. The reason appears to be her desire to retain the human individual as an independent actor in their own right:

> Our reflexive deliberations are held to be the processes through which we agents selectively mediate structural and cultural properties and also creatively contribute to their transformation. Therefore to rob agency of its first-person powers, by accrediting them to third parties, is to cut back on the causal powers which make each and every agent an active contributor to social reproduction or transformation. Agency needs to be granted autonomous properties in order to play this role. To eliminate their first-person perspective on themselves deprives them of this autonomy by discrediting their powers and explaining them away as the results of childhood influences, society's discourse or brain states (Archer, 2003, pp. 38-9).

Like Archer, I strongly believe that we cannot eliminate the first-person perspective, nor the causal powers of human individuals, from the explanation of human action. But I believe we can retain these without denying the impact of the social world on human subjectivity, and without denying the role of our biological parts in underpinning our behaviour. I argue, in short, that we can explain the powers of human individuals without explaining them away.

Thus we can accept that some day we may be able to explain the neurological underpinnings of human behaviour without this entailing neurological reductionism. To

say that we might be able to identify the neural network that has a particular disposition as an emergent property, for example, is not to say that our behaviour is determined by our neurons. We cannot explain human behaviour *purely* in terms of the causal powers of our biological parts. Those parts can not produce our human causal powers unless they are combined in the particular set of structural relations that constitutes them into a human being. It is as whole human beings that we have the capacity to decide, to act, and to affect the social world.

Similarly, to say that our social background and experiences influence our dispositions is not to cede all causal power to the social level at the expense of the individual. Our dispositions may sometimes be heavily and unconsciously affected by social factors, but none of us is ever completely at the mercy of our habitus. We all have the capability of reflecting critically upon our beliefs, though that capability may be more strongly developed in some people than others. Our habitus at any one time is not the unmediated product of social structures, but the result of a lifetime of critical reflection upon our experiences, including our experiences of those structures. To accept that social conditioning affects our beliefs, then, is quite different from believing that social structures determine our behaviour. Of course, they have an influence on our behaviour, and this influence operates in part through the causal impact they have on our beliefs, but they do not *determine* those beliefs. Thus the human individual remains the prime mover of human action, even if we accept that social conditioning plays a crucial part in forming our dispositions.

Once this is accepted, then Archer's account of the development of personal identity and social identity can be seen as an argument about the extent to which we are able to modify our habitus. As we develop a personal identity, we become more able to evaluate our concerns – to become reflexive – and to modify our dispositions as a consequence. Indeed, developing projects is precisely this – an example of the process of modifying our dispositions for future action. Reflexivity thus becomes a critical attitude towards the dispositions we have acquired from our past, as well as towards the contemporary social situation that we face.

With these re-interpretations, then, Archer's account of reflexivity can be integrated with the theory of action outlined in this paper and thus with a similarly reinterpreted version of Bourdieu's account of the habitus. The resulting synthesis, I argue, provides us with a powerful and coherent account of human action.

**Conclusion**

This chapter has done more than placing a reinterpreted habitus within the emergentist ontology; it has also placed it within a specific theory of action and thus stepped beyond the realm of social ontology and into that of social theory. But, as I argued in chapter one, this is entirely consistent with the ontological thrust of this thesis. As metatheory, a generic emergentist ontology can only take us so far in its own right towards understanding the relations between levels; to go beyond a certain point, level-specific theory is also required. But the two (the emergentist framework and the level-specific theory) play complementary roles, and it is impossible to show how the emergentist framework can apply to particular levels without developing at least the skeleton of some level-specific theory. Thus the development – and/or appropriation – of a certain amount of level-specific theory is inescapable if this thesis is to achieve its objective of showing how emergence applies to social structure and agency.

From Archer, this theory takes both her ontological insistence on the distinct existence of uniquely human causal powers and her theoretical insistence on the need to take account of conscious reflexive deliberation in the explanation of human action. It is thus able to draw on her account of the development of personal and social identity to expand and consolidate its account of reflexivity. But her account must also be modified, most particularly to allow for the role of acquired dispositions in causing our behaviour, and the effect of social context on those dispositions.

From Bourdieu, the theory takes his penetrating examination of the construction and operation of the habitus, and his recognition that our socially-influenced beliefs contribute to our reproduction of social structure. But his account too must be revised. Ontologically, it must be clarified by recognising that social structures are not literally internalized by individuals, but only metaphorically, through the influence they have on our subjectivity. Theoretically, it must be modified to show how we, as reflexive beings, are sometimes able to critically evaluate and thus modify our dispositions in the light of our experience, our reasoning capacities, and our value commitments.

The outcome is a view of human action as "a permanent dialectic between an organizing consciousness and automatic behaviours" (Bourdieu, 1990b, p. 80). While this chapter has only been able to present a brief sketch of the resulting theory of action, I suggest that this synthesis overcomes the imbalances and tensions that can be found in earlier accounts. This arises, I believe, from the adoption of an emergentist perspective in which social structures, cultural systems, human individuals and indeed our

biological parts are all recognised as possessing relevant causal powers, and examining how those causal powers interact in practice, rather than seeking to deny the causal influence of any of these, or to conflate multiple levels into one.

In the case of agency, the emergentist approach enables us to recognise that human individuals themselves are emergent entities with causal powers of their own as a consequence of the inter-relations between their biological parts. Those powers depend upon the structure and properties of our brains (and indeed our bodies as a whole) but they cannot be reduced to an explanation in neural terms. It is only because our neurons are organised into the complex networks that constitute our brains that they can provide the foundation for mental properties like beliefs and desires that found our ability to make decisions. Thus, neurons cannot take decisions – only human beings can.

But we must also balance this claim for the autonomy of the human individual with a full recognition of the interplay of biology and society with our causal powers in the determination of human action. The theory of action developed here, and its assimilation of Bourdieu's habitus, shows how the interplay of our social context with our biological powers to form and store dispositions and to translate them into behaviour plays a fundamental role in the causation of our behaviour. Our reasons, our dispositions, our beliefs, are all emergent properties of the human being as a whole, but they are emergent from a neuronal base, and social causes play a central part in their morphogenetic and morphostatic histories.

The possession of agency, then, means that human beings can have a causal impact on the world in their own right, but this does not mean that theirs are the only causal powers that can influence the social world; and it does not mean that the actions of human beings are not themselves caused. Social events are the outcome of complex interactions between the causal powers of individuals, organisations, institutions, natural objects, human artefacts, and symbolic structures.

Many explanatory accounts of agency, however, deny key elements of the emergentist account, whether by denying the causal significance of the biological level, the individual level, the social level, or some mixture of these. Others deny the distinctions between these levels and conflate them. And others deny that human behaviour can be causally explained at all. These latter accounts of agency seem to imply a new dualism, a humanistic dualism that echoes vitalism and Cartesianism in

seeing human behaviour as causally effective, yet somehow itself exempt from causal explanation. The next chapter will criticise these competing theories of agency.

# 7        Agency and Cause

Agency, I argue, is nothing more nor less than the possession and exercise by human individuals of emergent causal powers of their own – the ability to have a causal impact on the world in their own right. Thus, for example, most humans have the power to act, to speak, to communicate, to walk, to eat, and to perceive their surroundings. But the assertion of human agency does not imply that these human causal powers are themselves uncaused, or entirely independent of the lower levels from which they emerge. Those powers emerge from the inter-relations between our biological parts, including the neurons that make up our brain. And those parts and their configuration (notably that of the neurons that provide the emergence base for our mental properties) are not fixed biologically but develop during our lifetimes as a consequence of our interactions with the social and natural world. Thus, when we exercise our agency there is always a biological emergence base underlying it and a social causal history behind it. Human agency, then, arises from an ongoing series of interactions between our biological parts, our individual causal powers, and our social context.

Yet almost all previous approaches to agency have denied the significance or even validity of some part of this picture. Some approaches deny that human agency is caused at all, and many of these seem to be motivated by an attachment to the libertarian version of the philosophical doctrine of *free will* – the belief that we have free will and that this is incompatible with our will itself being causally determined. As Diana Coole has pointed out, the conception of free will, though often tacit, seems to permeate the debate on agency (Coole, 2005).  Some thinkers in the interpretative tradition also suggest that the inherent meaningfulness of human action makes it impossible to explain it in causal terms. And many others, while accepting that human action may be caused, deny the causal significance of one or more of the social, individual, or biological levels. This chapter is primarily intended as a defence of the emergentist theory of agency against these various alternative views.

The first part of the chapter will address various varieties of the denial of causal influence to different levels. These are summarised in Figure 7.1 below, which identifies each of these competing ontologies of human action in terms of the levels of entity to which it ascribes causal efficacy. This part will begin by considering

methodological collectivism, then methodological individualism, and finally biological reductionism. Although conflationism also has a bearing on agency, this has already been addressed in the discussions of Giddens in chapter five and Bourdieu in chapter six, so nothing more need be added here.
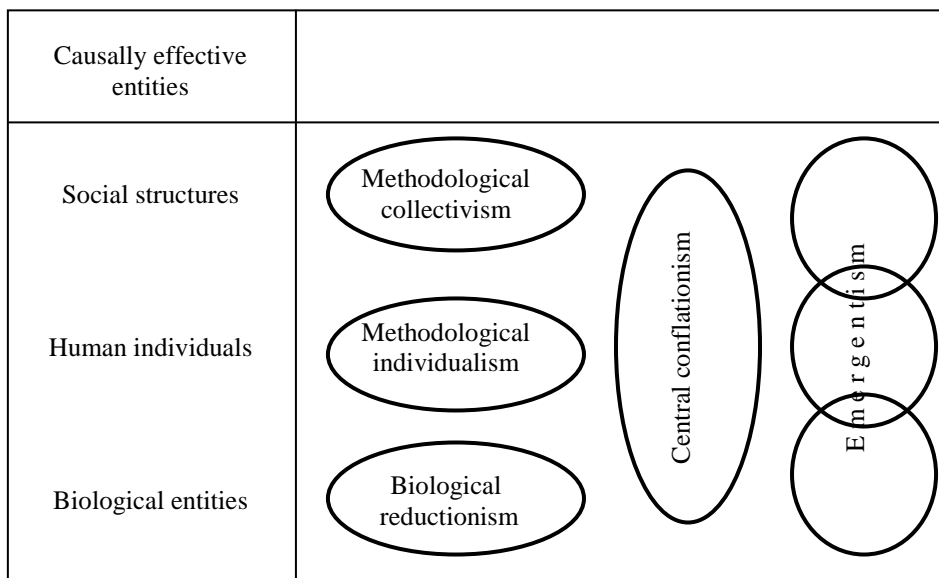


**Figure 7.1 – Competing ontologies of human action**

The chapter then moves on to two groups of approaches which deny that agency is itself caused. The second main part of the chapter criticises libertarian approaches to free will in theories of agency. The third responds to varieties of the interpretative tradition in social thought that deny the viability of a causal approach to human action.

## Holisms and reductionisms

### *Methodological collectivism*

Methodological collectivism is represented most simply as Durkheim's view that social facts are to be explained by other social facts (Durkheim, 1964 [1901], p. 110). Although Durkheim justified this using an emergentist argument (see chapter four above), his version of emergentism saw no need for lower level explanations of the causal mechanisms involved. This implies that neither human individuals nor biological entities have any causal impact in their own right on the causation of social events. The advocates of methodological collectivism need not deny the existence of lower level entities such as human individuals and material objects, but in effect they argue that they have no causal impact in their own right on the determination of social facts. The

rate of suicide, for example, may be determined by social facts even though individual suicides are not.

This can be seen as a claim for extreme causal isolation between the levels of the emergence hierarchy. In this view, the entities at each layer may be causally efficacious in the determination of events in that same layer, but unable to affect entities in other layers. At first sight, it might seem viable to argue that social facts determine social facts without denying the existence of human causal powers or that they determine the behaviour of individual human beings. But once we recognise that social facts are composed of individual human facts, this becomes problematic, since the explanation of any given social fact must be consistent with the explanation of the individual human facts of which it is composed. If only the social facts are causally effective in the determination of other social facts, it would seem that this consistency can only be sustained if the social facts also determine the individual facts, and hence at least some aspects of individual human behaviour must be determined 'from above'. The portrayal of methodological collectivism as extreme causal isolation, then, cannot be sustained; it inescapably implies that at least some events at the individual level are 'socially determined' by facts at the social level.

To say that an event is 'determined' by a particular class of causes in this sense is to claim that it is not just co-determined by social causes alongside others from different levels, as implied in the Bhaskarian model of actual causation, but rather is completely and exclusively determined by the causal powers of social structures to the exclusion of all other factors. As a general ontological claim about social events it is easy to show that social determinism is false. Even neglecting the influence of human individuals, it is clear, for example, that social facts are co-determined by natural physical causes. As I write this, for example, the volume of tourist activity in Thailand has suffered seriously as a consequence of the tsunami in December 2004. Part of the effect has been indirect, as a result of changing perceptions of how attractive Thailand is as a tourist destination; but part has been very direct indeed, through the physical destruction of tourist facilities. The influence of mega-actors on social events shows equally clearly that social events can be co-determined by individual agency (Mouzelis, 1991, pp. 77-8, 107). A case in point is the major role played by George Soros in forcing sterling out of the European ERM in 1992, and the consequent impact on sterling's exchange rate, import and export prices, and knock-on effects on the health of the British economy.

Although generalised *ontological* social determinism is thus clearly false, there is a weaker *theoretical* version of this argument: that in a given class of cases social factors may effectively determine events because the actual impact of other co-determining factors is negligible. Thus there is room within an emergentist ontology for theoretical arguments asserting that the determination of a particular class of events is causally dominated by a particular type of causal power, despite the rejection of true social determinism.

The implication is that some social theory that makes what appear to be socially deterministic claims may still be valid as social theory once it is repositioned within an emergentist ontology – whether or not the original formulation of the theory explicitly endorsed methodological collectivism. Of course, much socially determinist theory may be false, as a result of neglecting the causal impact of lower level entities, but in some cases it may be true that the causal powers of social structures are the dominant factors in particular classes of causal situation. To the extent that this is so, it may be valid to abstract from the lower level causal factors in these explanations. This does not necessarily imply that individual behaviour is determined by social factors, simply that it can be abstracted from in theorising the higher level. At least some social facts, and even their composing individual human facts, may be 'determined' (at least in the sense of 'predominantly caused') by higher-level factors.

More commonly, perhaps, the *normal* behaviour of a class of people, or at least some aspects of it, may be predominantly caused by social structures, even though there is always the possibility that in some cases other factors will come into operation and prevent the realisation of these causal influences. This is, of course, the heart of Bourdieu's argument on habitus, and it is also how role performance, and thus the morphostasis of social institutions, works. It is in part the causal powers of Birkbeck College, for example, that lead me to turn up every Thursday evening in the autumn term to teach Political Sociology, although it is always possible that on any given Thursday some other factor will intervene and lead me to do something else instead.

In contrast with strict social determinism, then, emergentism implies, first, that although social structures do have causal powers that can systematically affect individual behaviour, these factors can always be overridden by other co-determining causes. And second, that human beings are causally effective entities, whose behaviour on the one hand can influence social facts, and on the other is itself caused by a variety of interacting factors, including the influence of social structures, but is not *determined*

monolithically by social structure. The causal effectiveness of emergent social structures, then, in no way undermines the argument for the causal effectiveness of human beings. Indeed, it rests upon it, since the emergence of social structure depends upon its parts being causally effective in their own right.

### *Methodological individualism*

Methodological individualism denies the causal efficacy of social structures and insists that all social causation can be reduced to the causal impacts of individual human beings. Chapter four discussed and dismissed this denial of downward influence from the social as it appears in the work of King, and there is no need to revisit this argument here. Just as the causal effectiveness of emergent social structures does not undermine the argument for the causal effectiveness of human beings, the causal effectiveness of human beings does not undermine the argument for the causal effectiveness of social structures.

However, looking down the emergence hierarchy rather than up it, methodological individualism can also be taken to imply a denial of the causal efficacy of the biological parts of human individuals in contributing to the determination of human behaviour. When we turn to this side of the ontology of human individuals, there are at least two possible versions of methodological individualism.

On the one hand, there could be an emergentist version of methodological individualism, which regarded human individuals as causally effective entities in their own right while recognising that they stand in an emergent relation with their parts and allowing for the explanatory reduction of human properties and powers. Emergentist methodological individualism would imply that all social properties are resultant rather than emergent – that all social structures are like statistical distributions rather than having unique causal powers as a consequence of the way they are organised. While I argue that such a claim is empirically untenable, it is nevertheless ontologically coherent. However, in accepting an emergentist approach to the foundations of human behaviour, this version of methodological individualism can no longer deny the *ontological possibility* of causally effective emergent structures at a higher level. It is difficult to see how anyone accepting this possibility could go on to deny the causal effectiveness in their own right of structures like states and business corporations, and I am not aware of any actual emergentist methodological individualists, so we may move on quickly to the other variety.

On the other hand, there is an anthropocentric or dualistic version of methodological individualism, which implies that human individuals are causally independent of their lower level parts and denies the possibility of an explanatory reduction of human powers in terms of the biological level. As Barnes puts it,

> The full implications of speaking of human beings naturalistically as 'creatures' are no more welcome in many parts of sociology and social theory than those of speaking of them as 'social'. Much is made in these fields of an allegedly irreducible distinction, between human beings and the natural order in which they are set. As sources of intentional actions, human beings are regarded as exempt from the normal run of naturalistic explanation and accounting applied everywhere around them. And this is often justified by an explicit dualism: a fundamental distinction is alleged to exist between natural objects and events, linked by relations of cause and effect, and human beings, whose independently inspired voluntary actions are set completely apart from the causal nexus. This particular form of dualism is completely rejected here (Barnes, 2000, p. x).

I share Barnes' rejection of such a dualism. I also believe he is right in linking it to the question of free will, and I will return to this connection in a later section. This section, however, will consider the strictly causal implications of this dualistic stance. This stance is in effect a modern form of Cartesian dualism, in that it claims that human behaviour can be explained in terms that completely deny any causal impact of the body, including the brain, on that behaviour. Our mental 'selves' are somehow able to direct that behaviour without any causal connection in the opposite direction. It is undeniable that if our mental life is to be of any significance at all, then it must affect our physical behaviour, but the dualistic stance implies that the causation is all one way. Just as Descartes, having asserted the independence of the soul from the body, had to invent a connective function via the pineal gland in order to provide a channel for the soul to influence the body, these modern dualists also need a mechanism through which the mind can affect the body while remaining unaffected by it (Lokhorst, 2005).

This one-way view of the causation of human behaviour, however, is easy to disprove. It is very clear, for example, that aspects of our behaviour are strongly influenced by our emotions, and that there is a substantial chemical or hormonal impact on the determination of our emotions. If proof of this is necessary, it is provided by the fact that medication which interferes with or counteracts these chemical influences is effective in altering people's behaviour (Freeman, 2000, p. 160). It is simply not plausible, then, to deny in principle that biological factors can affect human behaviour; by implication, then, we must see human behaviour, like other sorts of events, as co-determined by causal powers from a variety of levels of the emergence hierarchy.

On the other hand, just as the ontological refutation of social determinism does not rule out cases in which social facts are the predominant causal factors in determining other causal facts, the ontological refutation of dualistic individualism does not rule out cases in which mental facts are the predominant causal factors in determining human behaviour. It is entirely consistent with emergentism for whole classes of human behaviour to be predominantly caused in this way – by the causal powers adhering to human individuals in their own right.

Of course, it may become possible – at some time in the future – to provide comprehensive explanatory reductions of these causal powers in terms of our biological parts and their relations to each other. But a complete explanation of human functioning in terms of our parts would not mean that human agency had been eliminated; just as explanatory reduction does not entail eliminative reduction in the natural sciences, nor does it in the human sciences. The powers that humans have by virtue of the way they are built from their parts will never become powers of their parts alone, however thoroughly we are able to explain them.

This still allows us to construct explanations of human behaviour in terms of intentions, reasons, social learning, shared meanings, norms, values, institutions, and the like. And it allows for the possibility that the causal explanation of human behaviour in these terms is extremely complex, with the consequence that it may in some respects seem unpredictable. The point that is essential here is that human beings are causally effective entities, whose behaviour is itself in turn caused by a variety of interacting factors, but it is not *determined* entirely by factors from any single level of the emergence hierarchy, including social structure or indeed the biological structures of their bodies.

Our biology, then, cannot be denied, but neither can human behaviour can be explained *entirely* in terms of biology. The claim that it can be is the subject of the next section.

### Biological reductionism

Biological reductionism is the claim that human behaviour is caused entirely by the properties of our biological parts. Dennett, for example, has been accused of "viewing agency simply as a biological phenomenon" (Malik, 2003, p. 49). Biological reductionism exists in a number of varieties, depending upon which sorts of biological entities are assigned causal effectiveness, but the implication in each case is a denial of our ability to act in ways that are not directly determined by those biological entities.

On some interpretations, sociobiology and more recently evolutionary psychology can be seen as examples of biological reductionism, in which human behaviour is to be explained entirely by our genes. Similarly, there are advocates of neurophysiological reductionism, in which human behaviour is to be explained entirely by our neurons. In this section I will discuss each of these briefly, and argue that neither provides a viable account of human behaviour, for just the same reasons that all other attempts to eliminatively reduce emergent properties must fail.

Evolutionary psychologists have argued that many aspects of contemporary human behaviour can be explained as the phenotypical manifestations of genetic adaptations which were naturally selected in response to the environments in which human beings evolved. Since complex social organisation is a very recent innovation in terms of the rate at which genetic change is presumed to occur, they argue that there has not been enough time for the genetic basis of our behaviour to adapt to the conditions of such social organisation. The consequence, they claim, is that much of our behaviour is determined by genetic adaptations to the sort of environment faced by Stone Age human beings (as argued, for example, in Tooby and Cosmides, 1992).

There has been a widespread critical response to this argument (see, for example, Dupre, 2001; Edwards, 2003; Rose and Rose, 2001). Perhaps the core element of this response is that the properties of our bodies and particularly those of our brains are not determined entirely or even predominantly by our genes. Dupré, for example, argues that our intelligence and the development of our brains depend at least as much on nourishment and the impact of our contemporary environment as they do on our genes (Dupre, 2001, p. 30). Whatever our genetic inheritance, the bodies we end up with depend upon a process of embryological development that is influenced by our physical environment, and more importantly, the brains we end up with depend upon a process of neurological development that is dominated by the inputs provided by our experience – in a word, by learning. Of course our genes also have a substantial effect on the development of our biological and indeed our mental structures, and hence indirectly affect our behaviour, and it is possible that this genetic effect has changed very little since the Stone Age. But because our brains today develop – learn – in very different conditions from those of the Stone Age, "there is no reason to suppose that the outcome of that development was even approximately the same then as now" (Dupre, 2001, p. 31). The sheer plasticity of our brains allows our environment, and in particular our social environment, to play a substantial causal role in moulding our beliefs, desires,

and other mental properties, as well as our skills and capabilities – our dispositions and our causal powers.

Genetic reductionism, however, is not the only form of biological reductionism. There is also the variety that suggests our behaviour is determined by our neurons. But we cannot explain human behaviour *purely* in terms of the causal powers of our biological parts. As I have argued earlier, these parts are not enough to produce our human causal powers unless they are combined in the particular set of structural relations that constitutes them into a human being. The person we are, the character we have, the sorts of projects we want to pursue, flow from the combination of all of these parts into a single biological and social human individual with a body and a brain. That person will have been shaped over time by both genetic and neurological effects, but ultimately they possess powers that can only exist when all the various parts of the human being are brought together into a unique whole with emergent properties of its own. One merit of the emergentist approach is therefore that it enables us to connect the human individual back up to the whole person including the non-mental aspects of the body – its emotions, physical needs, health and disease, and the use and constraint of that body in time and space.[79]

To explain the biological constitution of such an individual is not to eliminate the causal powers that person possesses in their own right as a structured whole. Thus it is only ever the whole person who makes choices and acts, even though (a) the person doing the choosing has the capabilities and inclinations they do entirely as a result of their causal history; and (b) it may some day be possible to make an explanatory reduction of the process of choosing that shows how it is made up of lower-level events and the relations between them. Despite, and indeed because, of (a) and (b), we are emergent entities with the emergent power to make decisions about our actions, in a sense that neurons, and even neuronal networks unattached to bodies, could never be.

Some authors, however, find a tension between this causal account of human action and the claim that we do indeed make choices. This brings us to the question of free will, which runs as a persistent subtext through much of the debate on agency.

---

[79] See, for example, Sayer's discussion of the role of emotions in the judgements that affect our dispositions (Sayer, 2005, p. 950).

**Free will and agency**

The dualistic individualists who argue for the primacy of agency over structure in sociology often seem to be motivated by a *libertarian* view of free will.[80] There is a huge philosophical literature on free will and determinism, most of which I must ignore here for reasons of space and relevance. However, the central role of libertarianism in the dualistic individualist approach means that no alternative account of agency can be convincing unless it has a response to this treatment of free will. This section, therefore, will briefly outline the main positions in the debate on free will, then analyse how this affects the question of agency. It will argue that the adherence of many individualists to libertarianism confuses a humanistic (indeed an excessively humanistic) moral argument with a causal one. Their account of agency as free will purports to have explanatory content but is in fact tangential to explanatory discourse. By separating out these non-explanatory considerations I will show that we can address valid humanist concerns without somehow excluding human beings from the influence of causality in our explanatory account of the world. I will therefore argue that an emergentist conception of agency is entirely compatible with free will of a sort – but not the sort the libertarians believe in.

*Defining free will and determinism*

Let me begin with some brief definitions, drawn from the literature [81]:

(i) *determinism* is the idea that all events are fully caused by the preceding state of the universe (or the preceding events in the universe). In other words, given the previous state of the universe, it was inevitable that certain events would follow, and nothing could have happened differently than it did.[82] Since the advent of quantum physics, strict determinism has largely been replaced by the doctrine of *near-determinism*, which is the idea that most events are fully caused by the preceding state of the universe, but that certain causes are probabilistic within a certain range. Hence

---

[80] Brief definitions of *libertarian* and the other main concepts in the debate on free will are given below.

[81] The definitions in the following paragraphs are particularly influenced by (Honderich, 1993) and (Watson, 1982).

[82] This is therefore a different sense of *determinism* than that implicit in the earlier discussion of *social determinism*.

there is an element of randomness in the causation of quantum events, but this randomness is strictly limited in its range and scope.[83]

(ii) *free will* is the ability of humans to choose their actions. It is commonly understood to imply that the individual concerned could choose to act otherwise, and that as a consequence they can be held morally responsible for their actions. This, however, gives rise to at least two varieties of the concept, depending upon what is meant by "could choose to act otherwise". In one sense, it means that the individual has the *opportunity* to act otherwise. In this sense, a galley slave with a brutal overseer has very little free will, since being chained to their oar they have no opportunity to leave it, and being subject to flogging they have no viable opportunity to stop rowing – though they could in principle choose to be flogged instead. Let me call free will in this sense *opportunity free will*.[84] In the second sense, it means that the individual's choices are effectively free of causal determination, since if the individual's choices were caused by the previous state of the universe (e.g. by brain states or events) then, given the previous state of the universe, they could not have chosen otherwise than they were caused to do. Let me call free will in this sense *metaphysical free will*.[85]

(iii) *compatibilism* is the argument that determinism and free will are compatible. Broadly speaking, compatibilism rests on defining free will as *opportunity free will* rather than as *metaphysical free will* (and implicitly denying the possibility of the latter form). Thus, compatibilists assert that determinism is compatible with opportunity free will, and that this is the only kind of free will that is worth considering. Typically they go on to assert both that determinism (or more usually, near-determinism) is true, and also that we *can* have free will in their chosen sense of it – although of course the degree to which our opportunities are constrained (and hence our degree of free will, or of moral freedom) is a contingent, practical, and political question. Opportunity free will is sometimes held to justify holding individuals

---

[83] Since the advent of chaos theory, some have suggested that randomness at the quantum level could sometimes be amplified to have macro level effects, with the consequence that the effects of quantum indeterminacy may be greater than previously believed. However, even in this scenario, once a quantum event has occurred, its further consequences follow in a deterministic fashion.

[84] Honderich calls actions which are freely chosen in this sense *voluntary* (Honderich, 1993, p. 86). This word is sometimes used with other connotations, however, so I have not followed his usage here. Opportunity free will also seems synonymous with Berlin's negative liberty (Berlin, 1969).

[85] Honderich describes free will in this sense as the *origination* of the decision (Honderich, 1993, pp. 39-43).

responsible for their actions (i.e. liable to be praised or punished for them) on the grounds that – and to the extent that – they had the opportunity to act otherwise, irrespective of the other causal factors at work.

(iv) *incompatibilism* is the argument that determinism and free will are incompatible. It rests on (and follows from) defining free will as *metaphysical free will*. There are two varieties of incompatibilism, depending upon whether determinism or free will is taken to be true:

(v) On the one hand, *hard determinists* argue that determinism (or near determinism) is true and hence we have no free will.

(vi) On the other, *libertarians* argue that metaphysical free will is true, which is to say that human behaviour is not fully determined by prior causes. This represents a stronger denial of determinism than is implied by near-determinism. Near-determinism still implies that human choices are causally determined: it just introduces an element of randomness into the quantum-level determination of them – and this does not seem to justify holding the individual morally responsible for their choices any more than if they were fully determined.[86] For libertarians, human behaviour is caused by human choices, but human choices are not themselves caused at all – 'we' just make them, free of any prior causal determination, and hence 'we' can be held responsible for them. This entails denial of any causal account of human choices and hence also the denial of any possibility of the explanation of mental entities in terms of brain activities, or of causal influences of social entities on decisions through such a process.

### Cause and responsibility

Libertarians, therefore, cannot be emergentists about human beings; they must be dualistic individualists (see p. 194). Both positions entail the same image of the human being as causally autonomous, not just in the emergentist sense of having causal powers of their own, but in the full-blown dualist sense of being free from causal co-determination by the causal powers of their parts (and indeed the causal powers of other entities). Yet it seems increasingly clear that human actions *are* caused, at least in part; that our choices are only one part in the causal chains leading to our actions and not even the final part; that our choices themselves are  emergent from brain processes, caused by the configurations of neural networks that constitutes them into the emergent beliefs and desires we have at the moment of choice; and that a wide variety of factors

can affect those choices, whether through experiences that affect the beliefs and desires that we have, or by altering our brain chemistry, e.g. by good/bad nutrition, or through behaviour-altering drugs, whether Prozac or Ecstasy.

The more we learn about how human choices are caused, the clearer it becomes that libertarianism is wrong.[87] Yet it represents a doctrine that has played an important social role, since the argument that we, and not other entities, are causally responsible for our own actions has played a central role in legitimating what Barnes calls "the institution of responsible action" which underlies the concepts of law and crime, and the practice of punishment (Barnes, 2000, p. 6). As Barnes says, "Explicitly, the entire edifice of our law is built upon belief in free will", but the actual practice of law and punishment "raise the question of whether concepts like 'responsibility', 'choice', 'free will', 'agency' and so forth might not be secondary features of the institution of responsible action, mere rationalising accompaniments of procedures moved by pragmatic expediency" (Barnes, 2000, pp. 12, 14).

I suggest that the role of these concepts within the discourses of responsible action is indeed to legitimate a regime of enforcement of social norms, although I would hesitate to call this 'pragmatic expediency'; rather, such enforcement is fundamental to the establishment and maintenance of society. We hold people responsible for their actions because this is the very foundation of sociality, and to sustain this institution of responsible action, societies must sustain the belief that it is justified. The argument that people are causally responsible for their actions has played a key role in sustaining this belief, and this is an argument that has seemed strong because it conforms with the subjective sense we have of being in control of our own actions – of making decisions that are subsequently implemented in our actions.

As long as science was completely unable to explain human choices in causal terms, it was perfectly viable for this conception of free will to coexist with what Barnes calls "the institution of causal connection" (Barnes, 2000, p. 11) – the set of discourses in which events are explained in terms of their causes. Yet within the institution of causal connection itself – within science – more and more progress is being made in explaining human decisions causally. There are still, of course, those who deny that such explanations will ever be complete (to the extent that a complete explanation is

---

[86] For a penetrating explanation of why near-determinism does not help the libertarian case, see (Richards, 2000, pp. 139-40).

possible). Dupré, for example, asserts that causal completeness is unprovable, and a matter of faith, with the implication that human behaviour may never be shown to be causally determined (Dupre, 2001, pp. 157-8). Dupré may or may not be correct about causal completeness, but it is already demonstrable that human behaviour is at least in part causally determined (e.g. by the effects of drugs). The idea of humans as uncaused causal movers, as causal originators, is becoming increasingly untenable; and as it declines in plausibility its effectiveness in reconciling the institutions of responsible action and causal connection is evaporating. Because we do not possess metaphysical free will, we can no longer justify excluding human decisions from the institution of causal connection, and the complex of ideas upon which our institution of responsible action rests can now be seen to be internally inconsistent.

To digress for a moment: we therefore need to reconstruct the case for the institution of responsible action – for law, crime, and punishment – perhaps in terms that rest, not upon the concept of metaphysical free will, but on that of opportunity free will. In other words, perhaps we should be held responsible for our actions because, and to the extent that, we had opportunities to do something different and we chose not to, irrespective of the fact that our choice was caused. The brutally overseen galley slave can hardly be held responsible for rowing, because there is no viable opportunity to do otherwise. But if, for example, a politician in a democratic society chooses to take a bribe to alter a political decision, she is responsible for her decision, because she has a viable opportunity to take a different one. Freedom, then, in the sense required to provide a compatibilist legitimation of the institution of responsible action, is a matter of the opportunities and constraints we are faced with, not a matter of how the choice we make between our opportunities is caused.

Ironically, if our decisions were uncaused, in the sense of being totally unaffected by external factors, then it would become pointless to assign responsibility for them. The very point of responsibility is to legitimate a regime of punishment and reward which would be totally irrelevant if it did not exercise some causal influence on our future behaviour. Indeed, it is *because* such a regime can have a causal effect on human behaviour that it can contribute to maintaining functioning societies. The compatibilist account of freedom and responsibility is therefore capable of being reflexive in a way that is denied to the libertarian account: because it sees human

---

[87] Richards also provides an interesting philosophical argument for the impossibility of metaphysical free will (Richards, 2000, pp. 145-6).

behaviour as causally influenced, it can show why the institution of responsible action is necessary and how it works, as well as supplying part of its legitimation.

Let me round up the implications of this argument for the debate on agency between emergentists and dualistic individualists. Individualism often seems to be motivated by a desire to deny that human choices are caused because such causation seems to entail both an assault on the unique status of human beings and the destruction of the institution of moral responsibility. But I have argued in this section that human choices *are* caused: the human actor is a deciding individual, exercising causal powers of his or her own, but subject like the rest of nature to the process of causation. Nevertheless, the causation of human choice is entirely compatible with the institutions of moral responsibility, as long as we recognise that we need to alter the discourses that have been used to legitimate those institutions. Furthermore, emergentism maintains the unique status of human beings, as a class of entities with a range of causal powers that is unmatched in the known universe, while rejecting the anti-scientific denial that human choices are caused like any other natural event.

I therefore suggest that the emergentist account of agency offered here is entirely consistent with a reasonable humanistic desire to respect the uniqueness of human beings and sustain the institutions of moral responsibility. What it is not consistent with is an over-humanistic desire to introduce an anti-scientific and mystificatory dualism into the explanation of human action.

## The interpretative tradition and the denial of cause

There is another important tradition in the philosophy of the social sciences that denies the possibility of a causal approach to the explanation of human action, and thus conflicts with the theory of agency presented in this and the previous chapter. This is the interpretative tradition, and this section will engage with it by discussing the canonical works of Peter Winch and to a lesser extent Charles Taylor (Taylor, 1994; Winch, 1958).

These thinkers proceed from the belief that any human action is inherently meaningful; that it has the character of action rather than of mere physical movement because of the meanings that co-constitute it. Thus, for example, the behaviour of marking a cross on a piece of paper constitutes the action of voting only because of the meanings infusing it (Winch, 1958, p. 49). Winch argues, for example, that "people cannot be said to be 'voting' unless they have some conception of what they are doing.

This remains true even if the government which comes into power does so in fact as a result of the 'votes' cast" (Winch, 1958, p. 51). These meanings themselves are logically dependent upon the prior existence of a language community and of certain social institutions (Winch, 1958, p. 44). A similar argument applies to people's motives for their actions. It is argued that we cannot ascribe a motive to an actor unless this motive is something that the actor understood (Winch, 1958, pp. 45-6), and this remains true even if we are claiming (correctly) that the voter was not consciously aware of why they acted. Thus in Winch's view, they could have acted for reason $x$ even if they did not consciously decide to do so, but they could not have acted for reason $x$ if they had no grasp of the meaning of $x$. All our actions, then, are suffused with meaning, and if we are to give a coherent account of human actions we must bring out these meanings.

The consequence, for interpretative thinkers, is that actions have some of the character of texts: they can only be explained by attempting to develop a coherent rendering of their meaning (Taylor, 1994, p. 185). The production of such a rendering generally depends upon the interpreter sharing an understanding of the language, culture, and/or social institutions within which the action was produced (Taylor, 1994, p. 183, 188).

### The denial of cause

Now this argument is commonly interpreted by these thinkers as implying that causal explanations of human action are inappropriate or impossible, and must be replaced in the social sciences with an interpretative approach to understanding human action (Fairclough et al., 2002, p. 2; Martin, 1994, p. 263). However, it is not immediately apparent why the discussion so far should have this implication. Winch offers a number of arguments, which are formulated as a response to Mill's claim that the explanation of human social behaviour is just a more complicated version of the explanation of natural events. Let me discuss two of these briefly before turning to the most significant argument in more detail.

First, he rejects any possibility of a physiological account of human motives:

> To discover the motives of a puzzling action *is* to increase our understanding of that action; that is what 'understanding' means as applied to human behaviour. But this is something we in fact discover without any significant knowledge about people's physiological states; therefore our accounts of their motives can have nothing to do with their physiological states (Winch, 1958, p. 78).

But 'understanding' has multiple meanings when applied to human behaviour. The fact that sometimes we seek to understand human behaviour in terms of motives in

no way implies that we cannot also seek to understand it in terms of physiological mechanisms, and scientists have made great progress in understanding aspects of human behaviour in these terms. Furthermore, the fact that we can learn about people's motives without examining their physiological states does not mean that their motives are unconnected to those states.[88] It would be entirely consistent with this fact for people's motives to be mere epiphenomena of their physiological states, or, as I argue, for them to be emergent from those states. Winch's argument here is an example of the dualistic individualism dismissed earlier in this chapter, and I have nothing more to add to that earlier discussion.

Secondly, he is critical of the notion that there are regularities in social activity that are similar to the physical regularities implicit in causal laws (Winch, 1958, p. 88). Clearly it is true that social activity does not exhibit the degree of empirical regularity that some physical phenomena do, such as the movement of the stars and planets. However, it is an error to regard *any* physical laws as producing exceptionless empirical regularities. As Bhaskar shows, even in the natural world causation is always an interaction between a variety of causal powers, whose result is contingent upon just which powers are relevantly present for any particular event (Bhaskar, 1978, pp. 109-11). Many natural events are just as unpredictable as social events, but in both cases this is entirely consistent with them being the product of causation, and with the possibility of providing a (fallible, as always) causal explanation. Once we have dismissed the Humean connection between causal explanation and exceptionless regularities, the lack of the latter is no barrier to the former.

I believe, then, that these first two arguments can be dismissed by reference back to earlier discussions in this thesis. Let me now turn to what seems to be his main argument. Winch writes:

> But the issue is not an empirical one at all: it is *conceptual*. It is not a question of what empirical research may show to be the case, but of what philosophical analysis reveals about *what it makes sense to say*. I want to show that the notion of a human society involves a scheme of concepts which is logically incompatible with the kinds of explanation offered in the natural sciences (Winch, 1958, pp. 71-2).

Human reactions, he argues, are not just more complex than those of other animals; this difference in complexity has become a difference in kind, because of the role of meaning and understanding in human behaviour, and their roots in a social

---

[88] Here we have an epistemological argument masquerading as an ontological argument – what Bhaskar calls the epistemic fallacy (Bhaskar, 1978, p. 27).

context (Winch, 1958, pp. 72-4). Therefore the inherently meaningful nature of human action makes it qualitatively different from the objects of the natural sciences. This, of course, is true. However, it is equally true, as he also points out here, that the writhing around of an injured cat is qualitatively different from the movement of a chopped-down tree – yet Winch himself seems happy to accept that the behaviour of the cat can be explained causally. The existence of a qualitative difference, it seems, is not enough in itself to establish the argument against cause.

What is it, then, that makes human action, but not the behaviour of the cat, exempt from causal explanation? The heart of Winch's argument seems to be that because the objects of social science are inherently meaningful, they are only knowable through a fallible and potentially contested process of interpretation. As a consequence, he argues, the process by which social scientists must attempt to understand their objects of study is not like the process by which natural scientists do so, but rather it is like the process by which natural scientists understand each other (Winch, 1958, pp. 86-9).

This is not unreasonable as a description of a significant difference between the natural and the social sciences, but it does not entail that human actions are uncaused. Rather, I suggest, it places an epistemological obstacle in the way of us understanding them. This obstacle is the need to interpret the meaning of any behaviour before that behaviour can be treated as an action to be explained, and to interpret the meaning of an actor's statements before those statements can be considered descriptive of motives, reasons, beliefs, desires, or the like. But this is not an insuperable obstacle: both Winch and Taylor seem to accept that, as long as the investigator understands the language, culture, and social institutions within which an action is set, the investigator may be able to interpret the action accurately (Taylor, 1994, pp. 183, 196-7; Winch, 1958, p. 86). Clearly such interpretations are fallible, but then so are our observations of natural events, particularly given that such observations may be theory-laden (Martin, 1994, p. 266). At worst, there is an extra layer of fallibility involved in the social case. But there is also an extra layer of linguistic and social mechanisms underpinning effective communication which we use constantly to ameliorate the risk of misunderstanding each other. In both cases, it is very often possible for us to establish the facts with a level of confidence that is adequate for practical purposes. This epistemological challenge, then, does not prevent us making causal explanations of human action, and as

a purely epistemological issue, it has no bearing whatsoever on the question of whether human action is causally determined.

Although it is not brought out explicitly by either Winch or Taylor in the pieces I have been considering, there is another possible interpretation of the hermeneutic argument that *would* provide an ontological challenge to the causal determination of human action. This is the suggestion that actions and motives are radically inseparable: that the status and content of an action as an action is logically rather than causally dependent upon its motive, so that a motive is part of an action and not a cause of it. Thus, for example, the action of voting is not caused by my intention to vote, but ontologically constituted as the act of voting by this intention. Let me call this the *inseparability challenge* to the causal explanation of human action. The inseparability challenge provides an ontological objection to causal explanations in the social sciences: that our actions are not caused by our motives, even though our motives are the only factors that Winch allows are relevant to their explanation, hence only an interpretive and not a causal account of our actions is possible. Although this argument is not stated explicitly by Winch or Taylor in these two pieces, it is the only interpretation I can see of their argument that constitutes a new challenge to the causal explanation of human social action, and I suspect that at least some advocates of the interpretative approach found their denial of cause on some version of the inseparability challenge.

### *Actions and motives* **can** *be separated*

It will be useful in responding to this argument to consider, as an example of a human social action, the act of voting by raising one's hand in a meeting. I suggest that we cannot understand the issues raised by the inseparability challenge without analysing this event into three potentially distinct elements, representing the parts played in it by individuals occupying three different roles with regard to the action. We have, first, the decision of the actor to vote and the implementation of this by the raising of their hand. Let us call this *action enactment*. Secondly, there is the recognition by an observer that this hand raising represents the action of voting. I shall call this *action ascription*. Thirdly, there is the decision of an authorised person, let us say the chairperson of the meeting, to accept, register, or count that hand-raising as a properly-executed vote. I shall call this *action registration*.

Let us begin with action enactment. Now, it is clear from my argument in the previous chapter that I regard a decision, say to vote, as a mental event that is distinct

from the physical implementation of that decision, in this case by raising one's hand. I have argued that such a decision leads to a mental state (an intention to vote, which is a sort of disposition) that is emergent from a neural configuration. In the context of the brain's processes for implementing behaviour, this mental state interacts with other factors to co-determine the subsequent raising of the hand (if indeed such behaviour follows; it may not if other causal factors intervene). I have also argued that such a decision will have a causal history of its own, arising in part, for example, from our knowledge of voting procedures, from our beliefs about the likely consequences of the various options that are the object of the vote, and from how those consequences relate to our various desires. Thus, the actor's decision is on the one hand caused, and on the other contributes causally to the subsequent implementation of the corresponding action.

The question that arises now is whether there is any inconsistency between this causal account and the inherent meaningfulness of human social actions. I suggest there is not. The fact that raising their hand in this context will have the meaning of voting enters into the causal explanation of the behaviour through the actor's knowledge of that fact. They raise their hand for a directly causal reason: because (a) they want to vote; and (b) they know that raising their hand is likely to have the (causal) effect of being registered as voting. Both of these factors are mental states, emergent from neural states, and the actor's brain works in such a way that these mental states are able to affect their physical behaviour in a process of actual causal interaction. For the actor's decision making, the meaningfulness of social action is a directly causal story – first, because the relevant meaning of the action is the actor's belief about the kind of causal impact the behaviour will have; and second, because this belief itself has a causal effect on the subsequent behaviour.

Let us turn, then, to the second moment of the event: action ascription. Any given action may be recognised as an action by any number of other people who (i) are in a position to perceive the physical behaviour;[89] and (ii) have beliefs about the meaning of that physical behaviour. Once again, it is straightforward to construct a causal account of action-ascription that is consistent with the role of meaning. The observer's sensory organs perceive the actor's physical behaviour and send the corresponding neural signals to the brain; the brain processes these signals, and whether

---

[89] I include speech acts in the category of physical behaviour, although the voting example simplifies the case by neglecting whether further issues arise in the interpretation of speech acts.

consciously or subconsciously, combines these with its existing knowledge about the sorts of behaviour that count as different types of action in this sort of situation. In this process the end result is a belief (in the observer) about what action the actor has just performed, and this belief is caused by a brain process which is co-determined by (a) the actor's behaviour; and (b) the beliefs the observer has about what counts as this sort of action. Once again, there is a clear causal story, and in this causal story the individual draws upon their knowledge of social meanings to arrive at the result.

The significance of that result, however, will sometimes depend upon whether the interactor is endowed with the authority to register an action of the kind perceived. If an observer interprets the actor's raising of their hand as waving, it has this meaning for them irrespective of their social role, and any social significance attached to the action of waving inherently follows. But some actions are meaningful as such only within the context of a social institution that requires organisational recognition of the action. In such cases, there is a third moment in the event: action registration.

In these cases, it matters a great deal whether the observer making an action-ascription is endowed with the authority to count that action as voting or not. In the case of our voting example, the actor's vote will only be registered as the action of voting if the chairperson of the meeting so ascribes it. It does not matter at all whether other observers recognise the actor's behaviour as voting, for the purpose of action registration. Of course, if the actor or another observer realises that their action has been registered in a different way from that which was intended, they may call out to correct this interpretation, and succeed in overturning it. But this merely confirms the point, since now the chairperson has changed her interpretation, and it remains true that if the chairperson interprets it as a vote, it *is* a vote.[90]

When considered at the level of the authorised observer as a human individual, the causal process of action registration is just the same as that of action ascription (although it may be followed by further actions, such as writing down the number of votes cast, and declaring a result). It is because the observer performing the action registration is acting in a role which authorises them to register actions on behalf of the organisation concerned that action registration is different from action ascription. It therefore follows from the argument of chapter four above that it is the organisation

---

[90] In his treatment of voting, Winch comes to the opposite conclusion (Winch, 1958, p. 51). But his conclusion can only be sustained if action is defined entirely from the actor's point of view; it arises from neglecting the fact that different individuals may ascribe different meanings to the same behaviour and that it is not just the meaning intended by the actor that matters.

which has the causal power to register votes rather than the individual who is authorised to perform this task as part of their role within the organisation. Since the meaning recognition implicit in action registration is performed by the individual in just the same way as it is in action ascription, no further possibilities arise of incompatibility between this causal account and the meaningfulness of social action.

Now, although there are effective mechanisms in place in all functioning societies that ensure that most social actions are ascribed and registered as intended by the actor, there is no *necessity* for the actor's meaning to coincide with the observer's. Some human behaviour is intended as an action and not recognised as such; and other behaviour is interpreted as an action although it was not intended as such. To characterise action as meaningful behaviour, then, without asking whose meaning makes it such, appears to conflate the meanings of the different individuals involved. Once we look at the picture in more detail, it is clear that there are three different senses in which we might consider a behaviour to be an action – it may be implemented, ascribed, or registered as such. To represent action as abstractly meaningful behaviour, rather than behaviour with specific meanings for specific individuals, is to conflate these different senses, and to obscure the vital fact that the same behaviour may have different meanings for different individuals, all of which are socially significant in their own right.

On the basis of this analysis of the role of meaning in action, I suggest that the inseparability challenge can be dismissed. Its denial of a distinction between action and motive is unsustainable, for two main reasons.

First, it is clear that action ascription and action registration are distinct from action enactment, not just analytically but because they are performed by different people. The concept of human social action as we generally understand it implies both that an action has been implemented by the actor and that an action has been ascribed and (sometimes) registered by observers. The meaning of an action to its observers, then, is just as important to the theory of action as its meaning to the actor, and its meaning to its observers is independent of its motive. It is only if we falsely conflate the meanings of the action as they appear to both actor and observers that we can consider the action inseparable from the motive of the actor.

Secondly, it seems to be possible to construct a plausible (if incomplete) causal account of human action. In these accounts full recognition can be made of the meaningful nature of action, and indeed they imply a more thorough recognition of how

meaning operates in the process of human action than non-causal accounts. The insistence of interpretative thinkers on the important role of meaning in human action can therefore be sustained without denying the (ontological) causation of human action or the (epistemological) possibility of explaining it causally. This allows us to develop a social science which recognises the value of *both* interpretive and causal methodologies, and indeed their inter-dependence.

The interpretative tradition, then, identifies some complications that must be allowed for in the theory of human action, but there is no reason to suppose that it undermines a causal or an emergentist account of agency.

## Conclusion

This chapter has argued that the emergentist account of human agency provides the basis for an effective critique of over-socialised accounts of the human individual, but also of purely individualistic accounts, and of accounts that either overstate or deny the impact of biology on the causation of human behaviour. The argument for emergence implies the existence of complementary and interacting powers at each of these levels, and not the denial or conflation of any of them.

The approach to agency advocated here is somewhat naturalistic, in the sense that it sees human agency as simply one of many levels of causal power, each with their own unique properties, which interact in the causation of actual events. By contrast, many accounts of agency seem to imply a new dualism, a humanistic dualism that echoes vitalism and Cartesianism in seeing human behaviour as causally effective, yet somehow itself exempt from causal explanation. This chapter has discussed and dismissed two types of anti-causal theories of agency: those arising from a libertarian attitude to the question of free will, and those arising from the interpretative tradition. Both are founded on valid humanistic concerns: on the desire to sustain the institution of responsible action, and on the desire to recognise the inherent meaningfulness of human action, respectively. But as I believe this chapter has shown, both of these concerns can be met by the emergentist account of agency, and indeed can be met more effectively by the emergentist account than by anthropocentric dualisms that deny the causation of human behaviour.

Emergence, then, provides a framework within which we can recognise the respective roles of human agency, social structure, human freedom, and the meaningfulness of social action. As John Dupre has argued,

the only hope for serious illumination of [the nature and causes of human behaviour] is a pluralistic one, an approach that draws both on the empirical knowledge derivable from the (various) sciences, and on the wisdom and insight into human nature that can be derived from more humanistic studies (Dupre, 2001, p. 4).

The challenge, of course, is how to reconcile these various elements. To some extent the various sciences and humanistic studies address different purposes and hence do not conflict. But they intersect in the question of human agency, and they have long seemed in irreconcilable conflict here. I believe the emergentist account of agency outlined in the last two chapters provides an effective framework for their reconciliation.

# 8    Conclusion

This thesis has sought to develop a version of the theory of emergence that can provide a clear basis for understanding causal relations, and to show that such a theory can improve our understanding of society. In addressing the latter objective, it has attempted not only to clarify the emergent properties and entities that constitute social structure and human individuals, but also to validate the general theory of emergence developed here by testing its power to explain the ontology of a specific domain: the social world.

This concluding chapter will draw together the threads of the argument and review how they achieve these objectives, what is original about the argument, and why these results are important for sociological theory. It will reprise in turn the key features of my accounts of emergence, of social structure, and of agency and tie up the threads of the sociological argument by summarising some of the broad variety of ways in which structure and agency can interact. But it will also recognise that there are still a number of loose threads that cannot be tied up without substantial further work, and indicate what some of those threads might be. Finally, it will return to the implications of the argument for the metatheoretical questions introduced in chapter one.

**Emergence**

From the point of view developed in this thesis, the world in which we live is made up of entities with emergent properties (which may also be called causal powers). Each entity is a hierarchical structure composed of other such entities. At one level down from the entity we find a set of its parts, but each of these is also an entity composed of parts, and this hierarchical structure continues all the way down to the most fundamental components of our world, whatever they may be. The concept of emergence is essential to our understanding of these structures and their causal powers because it enables us to see how the entities at each level can have causal powers of their own despite being in a sense 'nothing more' than a collection of lower level parts.

Many versions of the concept of emergence can be found in the various literatures on the subject. Not all of them, however, are capable of explaining the existence of causal powers at each level. It is useful to distinguish between three broad

varieties of the concept, which we may call *temporal emergence*, *strong emergence*, and *relational emergence*. *Temporal emergence* focuses on something like the lay sense of 'emergence' – the sense of emergence as the first appearance of something new – but *anything* that exists, whether it is an entity with emergent properties or not, would appear to be emergent in this temporal sense. A purely temporal concept of emergence, therefore, is unable to distinguish emergent properties from resultant properties, or entities from heaps – collections of intrinsically unrelated entities that do not possess emergent properties as a group. Heaps, unless they are taken to have always existed, must presumably have first appeared at some moment in the past. Hence they are temporally emergent without possessing causal powers in their own right, and so temporal emergence does not provide us with a viable argument for the existence of such powers. The thesis has demonstrated this point by examining Archer's attempt to justify an emergentist perspective by using examples of structure that on reflection are heaps and not cases of emergence.

*Strong emergence* is the idea that an entity can possess emergent properties that are in principle impossible to explain in terms of the properties of its parts and the relations between those parts. Strong emergence has so far failed to provide a viable argument for the existence of higher level causal powers, most obviously because no-one has plausibly demonstrated that any actual property is emergent in this sense. There are, of course, many higher level properties that we can not yet fully explain in terms of their parts and the relations between them, and in principle some of these may be strongly emergent, though it seems likely that explanations will eventually be developed for many of them. Any claim that a particular property is strongly emergent, however, will always be open to refutation by the production of an explanation of its causal mechanism, so all such claims must be treated as provisional. More significantly, perhaps, even if some properties really are strongly emergent, strong emergence fails to provide an explanation of how the rest – all the properties we are currently or potentially able to explain – could be causally effective in their own right.

This thesis has therefore advocated a variety of *relational emergence*. Relational emergence ascribes the possession of higher level properties to the particular way in which the parts of the entity possessing the property are organised, and accepts that it may be possible to produce an explanation of how those parts, organised in that way, produce such a property at the higher level. This, however, does not mean that relational emergence is compatible with the eliminative reduction of these higher level properties,

as it might seem to be. This is because of what I have called the *redescription principle* – the most important innovation made in this thesis with regard to the general theory of emergence.

The redescription principle applies to any explanation of a causal power in terms of (a) the parts of a higher-level entity; and (b) the relations between these parts that occur only when they are organised into such an entity. The principle states that such an explanation has not eliminated the higher-level entity from its terms, but instead has merely redescribed it. Since an emergent entity is nothing more than its parts and their organisation, any explanation that depends upon the properties of its parts and on the characteristic way that they are related within this type of higher level entity is in fact an explanation in terms of the higher level entity. Thus, because emergent properties depend upon the existence of particular sets of relations between the parts of the entity possessing the property (unlike resultant properties), the higher level entity cannot be eliminated by any reductionist strategy from causal accounts. Even though it may be possible to explain relationally emergent powers, we cannot *explain them away*.

The relational concept of emergence, then, has the twin benefits that it provides a justification for treating the emergent properties of higher level entities as causally effective in their own right, while at the same time allowing us to explore the ways in which these properties are produced as a consequence of the properties of the parts and the way in which they are organised to form this particular sort of higher level entity.

The properties that emerge in this way are, I have argued, identical with the *real causal powers* described by Bhaskar in his theory of causation, and this thesis has accepted his argument that each particular case of *actual causation* is the outcome of the interplay of a variety of such real causal powers. Actual events, then, are co-determined by the causal powers or emergent properties of the entities that are significantly involved in the production of that event. And as was argued in chapter three above, because the causal effect of any given entity is a consequence of the level of organisation that it represents, it is entirely possible for entities at different levels of the ontological hierarchy to interact in the production of actual events.

The arguments of chapters two and three therefore provide a general ontological framework which can be applied to entities and their properties at any level of the emergence hierarchy. It is this framework that underpins the argument that social structures and human individuals are entities with emergent properties, which can interact to co-determine social events. These arguments also provide an analytical or

methodological framework for examining the relationship between a whole and its parts, which I have drawn on in considering the relationship between social structure and its human parts and that between human beings and their biological parts. This latter framework can be summarised as what I have called the *five pillars* of emergence: if we are to explain the emergent properties of any entity, we must consider (i) its parts; (ii) the relations between those parts that are characteristic of this particular type of entity; (iii) the set of morphogenetic causes that have produced the entity in its current form; (iv) the set of morphostatic causes that stabilise the entity and ensure its continued survival; and (v) the mechanisms by which its parts and relations produce the specific properties of the entity. The next sections will review the application of this framework to social structure and to human agency.

## Structure

The term *social structure* has been used by sociologists in many different ways. For my purposes here, we may analyse this variety into two different dimensions of variability. In what we may call the analytic dimension, structure is taken sometimes to refer to higher-level social entities, sometimes to the relations between their parts, and sometimes to a property of the parts themselves (as in Giddens' treatment of rules). This thesis has sought to clarify the relationship between these different senses of structure, applying the general theory of emergence to show that it is the higher level social entities that are causally effective, by virtue of possessing properties that depend upon both the properties of their parts and the relations between them. Thus each of the three senses of structure identified here refers to a useful and necessary part of the explanation of social structure, but none of them can be substituted for the others.

In the second dimension, which we might call the typological dimension, there is a broad range of different types of social entity that could be seen as structures, such as socioeconomic classes, capital, markets, patriarchy, and social practices such as common courtesy. If we are to place social theory on a sound ontological footing, we will need to investigate the whole range of possible social structures, identify where the boundaries lie between the different types, establish which really do have emergent causal powers, and apply the five pillars approach to understanding how those causal powers emerge. But there is an essential precursor to such an exercise: to demonstrate that the concept of emergence can apply to any social structures at all. That is what this

thesis has attempted, and to achieve this we need only show that at least *one* type of social structure is indeed emergent. The thesis has focussed on the case of organisations.

Chapter four sought to show that organisations are entities with emergent properties. Their parts are (at least primarily) human individuals, related to each other by a set of practices that are specific to the particular roles that each individual adopts within the organisation. In following these practices, the individuals that make up an organisation behave differently than they would do if they were not parts of it, and when they do follow these practices they have different effects on the world than they could have as individuals outside the organisation. The effect of this behaviour must then be seen as a causal effect of the organisation itself, and not of the individual *qua* individual. Methodological individualists may seek to give an account of such behaviour in terms of the behaviour of the individuals concerned and their relations with other individuals, but where these relations are those that derive from their role in the organisation, then the redescription principle applies: if we attribute the causal effects here to the individuals concerned *plus* the set of relations in which they stand as a concomitant of being parts of an organisation, then in fact we are attributing those causal effects to the organisation. This ontological justification for a belief in the causal efficacy of organisations is the first of two original and important consequences of this thesis's emergentist approach to the explanation of social structure.

Organisations, then, *are* causally effective in their own right, and a relational emergentist account shows us how this can be. However, this should give no succour to methodological collectivists, because the emergentist account developed here also recognises that social structures do not entirely *determine* human behaviour, but only *co-determine* it in conjunction with the causal powers of a great many other entities from a variety of levels of the emergence hierarchy, including human individuals themselves. Furthermore, this relational emergentist account encourages us to analyse *how* the emergent properties of organisations can be explained as an outcome of lower-level interactions. It is the recognition that such explanations are entirely compatible with the rejection of methodological individualism that is the second distinctive and important outcome of this approach to the emergence of social structure.

## Agency

Chapters six and seven have argued that human individuals themselves, like organisations, are entities with emergent properties. As in all cases of emergence, those

properties arise from the unique way in which our parts are organised to form us as human beings. An emergentist theory of agency, then, recognises that our causal powers arise from our biological constitution. However, this does not mean that human actions are biologically determined. Like any other events, they are co-determined by a variety of interacting causes from a variety of ontological levels. The original theory of action developed in this thesis has sought to show that the mechanisms through which human action is determined provide opportunities for our action to be influenced both by the social structures that are so central to our environment, but also by our own uniquely human powers of conscious reflexive thinking.

Human beings, then, are entities with emergent causal powers. These are not, however, the god-like causal powers to uniquely and totally determine subsequent effects that sometimes seem implicit in discussions of the human causal role. Rather, we have only the ability to exercise a causal influence on the world around us, alongside many other factors, and the events that result are always the outcome of many interacting factors, of which our input is only one. In providing an explanation of *how* human individuals can be causally effective (rather than merely taking this for granted), the argument of this thesis therefore also shows how this causal effectiveness can be reconciled with the causal effectiveness of *other* entities that affect social behaviour.

Furthermore, this recognition of human powers should give no succour to those anthropocentric methodological individualists who dismiss on principle all attempts to explain the biological basis of human behaviour. Human actions are caused and can be explained (though in practice they can only be explained to the extent that we understand the particular mechanisms involved). Hence, as chapter seven made clear, libertarian conceptions of free will must be rejected. In another close parallel with its analysis of social structure, the relational account of emergence encourages us to consider how the emergent properties of human individuals can be explained as an outcome of the behaviour and inter-relationships of our biological parts. Once again, the recognition that such explanations are entirely compatible with the denial of reductionism – here, biological reductionism – is a critical contribution of the approach to emergence developed in this thesis.

## The interplay of structure and agency

To put it another way, the value of the emergentist approach to structure and agency advanced in this thesis is that it shows how a broad range of explanations of

different facets of them both can be fitted together in a coherent and mutually-reinforcing whole. The example that has been developed most thoroughly in the body of the thesis has been the argument that Bourdieu and Archer's at first sight radically incompatible accounts of human agency both have a great deal to offer and can be fitted together productively in an emergentist theory of human action. But this is only a particular case of a more general benefit of the emergentist approach. While methodological individualists deny the value of theories that allow any explanatory role to social structures, and methodological collectivists deny the value of theories that explain social facts in terms of human individuals, realist emergentists can see the value of both and integrate the two. While conflationists like Giddens and Bourdieu argue that structure is effective through becoming a part of individuals, and Archer tends to see the effects of structure on individuals as entirely external, the approach to emergence advocated here allows us to accept that structure does work at least partly through its metaphorical internalisation, but also that structure still retains causal power in its own right. The consequence is that the emergentist *ontological* approach developed in this thesis allows us to adopt *theoretical* arguments advanced from all sides of the debate where they cast useful light on the relations of structure and agency, and to integrate arguments from a variety of these perspectives.

Structure and agency, then, interact in a great many ways. Using the five pillars of emergence as the principle for organising the analysis, this section will summarise the sorts of interaction that this thesis has discussed.

Let me begin with agency. Human individuals are composed of their biological *parts*, organised through the anatomical *relations* that are characteristic of human beings. It might seem at first sight that these parts and relations are independent of social causes, although the increasing capabilities of human biotechnology such as genetic engineering are making us increasingly aware that even our biological composition is subject to social influence. More significant in the present, though, is the key role played in human beings by our brains. Critically, our brains are composed of neurological parts arranged by relations that are remarkably plastic: they change as a result of our experience, and hence it is possible for the social world, through its effect on our experience, to affect the (neurological) relations between our parts.

The combination of these parts in these relations results in a number of *mechanisms* that provide human beings with the full range of our emergent capabilities. These include the ability to breathe, to move, to eat, and so on, but in social theory we

are most interested in the generic ability of human beings to act intentionally in a social context. As with any other mechanism, this one requires a theoretical explanation – an explanatory reduction of the mechanism – and chapter six advanced a theory of action that showed one way in which it is possible that this ability to act emerges from the organisation of our parts and the sorts of processes, both conscious and non-conscious, that arise from that organisation.

The particular actions that follow from this mechanism in any particular case depend upon the prior shaping of our human totality through a process of altering our anatomical, and most particularly our neurological relations – the *morphogenetic* process we call learning, which produces the beliefs and dispositions which in turn co-determine our actions. In this process a variety of different causes may interact, and the theory of action developed here enables us to see how the social environment may causally affect our habitus, as argued by Bourdieu, while still retaining space for the reflexive deliberations of the human agent, as stressed by Archer. These reflexive deliberations create the opportunity for the agent to take conscious account of the social structures that form their context and hence a second way in which the social environment can affect our beliefs and dispositions. Thus conditioning and decision-making provide the processes by which our neurological relations are rearranged to produce new emergent beliefs and dispositions – the process by which social causes can affect our human structure and thus our emergent properties.

Of course, such influences can also have a *morphostatic* effect; as Bourdieu has stressed, our experiences often tend to reinforce our existing dispositions. Thus the learning process need not change our human structure; it may sometimes stabilise it. Meanwhile, there are also a great many other physiological processes that maintain the existence of the human form, which we can often take for granted in the social sciences.

By considering the five pillars in the emergence of the human capability to act, then, we can see how that capability is simultaneously underpinned by our biological parts and structure and causally influenced (co-determined but not determined) by social structures.

Next we must consider the other side of the story – how agency influences structure – by considering the five pillars of the emergence of organisations. The *parts* of organisations are primarily human beings organised through their acceptance of roles

which define the *relations* required to form the organisation.[91] The *mechanism* that produces the emergent properties of the organisation is provided by the enactment of these roles by their incumbents according to a set of (often loosely defined and/or implicit) rules. Thus there is an important place in the understanding of structure's dependence on agency for the rules stressed by Giddens, or perhaps of the dispositional rule-equivalents offered by Searle in place of rules – the "skills and abilities that are, so to speak, functionally equivalent to the system of rules" (Searle, 1995, p. 142). Role enactment is itself always a human action, and hence a product of human agency, operating through the mechanisms discussed above. Here then we have the first aspect of the influence of human agency on social structure – the acts of social entities depend upon an emergence mechanism that itself depends upon human agency.

This mechanism, however, also leaves open many opportunities for the exercise of individual discretion in the details of *how* a role is to be enacted. Rules are always indeterminate in boundary cases (Wittgenstein, 1953); and they only ever specify some aspects of how a role is to be enacted. Indeed, many roles are at least partly *strategying* roles, in Mouzelis' terms (Mouzelis, 2000), and in these cases the enactment of the role is inherently a matter of exercising individual judgement and discretion.

Role enactment thus provides opportunities both for role incumbents to act in routine ways that reproduce the organisation – and thus as *morphostatic* causal influences on the organisation as a whole – and for role incumbents to act in strategying ways that may lead to changes in the organisation – and thus as *morphogenetic* causal influences on the organisation as a whole.

Applying the 'five pillars' analysis to human individuals and organisations, then, enables us to see how a broad variety of interrelationships between structure and agency fit together to produce social action.

## Outstanding issues

Significant gaps, however, remain in the account of these interrelationships that has been presented in this thesis. Most obviously, in the space and time available for a work of this kind it has only been possible to address a subset – though a significant one – of the various kinds of social structure. Ultimately, we need to apply the emergentist

---

[91] It is open to debate whether other things might also be parts of organisations, e.g. computers performing roles previously performed by human beings. If they were, this would complicate the story told here, but it would not invalidate it as far as it goes.

perspective and methodology developed here to the analysis of a broad variety of putative social structures if we are to place the study of social systems onto an ontologically rigorous footing.

This thesis has begun to consider one further major class of social structures: institutions, which I have defined as social practices that are followed in a particular community. Institutions are dependent upon shared beliefs, and upon shared meanings – the agreement amongst a community that certain sorts of physical behaviour count as certain sorts of practices. I have suggested that shared beliefs may be seen as properties of groups of people rather than entities in their own right, but a great deal more work is required if we are to be able to give a thorough ontological analysis of institutions. We must also remember that institutionalised practices play a significant role in organisations, so that a more complete theorisation of organisations themselves also depends upon further progress in this area.

The question of institutions also opens up a much broader area that requires deeper ontological investigation: the structure of meaning and culture as emergent systems. Meaning and culture are themselves key objects of study for sociology in their own right, but it is also clear that structure and agency are strongly mutual interdependent with meaning and culture. As was pointed out in chapter one, for example, social conditions have a strong causal influence on the development of culture; human action is inherently meaningful; and the functioning of social organisations and institutions rests to some extent upon our beliefs about them.[92] These are interdependences that have been recognised in this thesis but the further development of the theory initiated here depends upon the complementary theorisation of meaning and culture, and of their inter-relations with structure and agency.

A comprehensive emergentist approach to meaning and culture would seem to require the integration of at least four distinct analyses. These are: the roots of meaning in the neural architecture that underlies the human mind; the role of interaction with other people, and with the natural and social world more generally, in forming our meanings; the relations between signs, signifieds, and referents that are analysed by semiotics; and the development by human communities of shared meanings and thus of culture as a communal as opposed to a purely individual possession. Just as emergence

---

[92] John Searle, for example, has argued that "language is essentially constitutive of institutional reality" – that money, for example, can only function as such because of the linguistic meanings we associate with it (Searle, 1995, p. 59).

provides an ontological framework that enables us to reconcile social structures and human agency in a way that allows causal influence to both, we need to investigate whether and how the same sort of framework can enable a similar reconciliation of these various perspectives on meaning and culture, and thus underpin a clearer account of what sort of causal effects we can attribute to them.

Although I have focussed here on the outstanding issues in the theory of structure and agency, there are also many ways in which the general theory of emergence could be developed further. In this thesis, for example, I have set aside the question of supervenience, the concept of internal relations, process-based versus configurational varieties of emergence, and a critical examination of the possible ways in which emergentism could be false, incomplete, or only partially valid.

Although I believe that these omissions do not undermine the main arguments presented in this thesis, they do mean that it is something less than a complete general theory of emergence, and equally something less than a complete account of structure and agency. Further work remains to be done in both areas, and I expect to address issues in both areas in a continuing research programme.

**The implications for metatheory**

Despite these omissions, I believe that this thesis has presented an account of structure and agency that is richer and more powerful than those implicit in more restrictive ontological perspectives. It is richer and more powerful, I argue, because the emergentist ontological framework allows for the full range of interactions between causally effective entities at both levels, while at the same time making clear the relationship between the different elements of these interactions. Here we have a case of that mutually supporting relation between domain theory and metatheory discussed in chapter one: the emergentist ontological framework allows us to theorise the inter-relationships of structure and agency more fully than its competitors and this improvement at the domain-theoretical level validates the corresponding innovation at the metatheoretical level.

In validating an emergentist ontology, this analysis also implicitly demonstrates the value of ontological work in the social sciences, and in particular the value of being more rigorous about the ontological status of the concepts that are employed in it. Sociological concepts are often pressed into service with loose contextual definitions and no attempt to establish what their real referents are or whether it is valid to treat

them as having causal significance. In the natural sciences, it would be unthinkable to employ concepts without attempting to understand what range of entities they apply to, what those entities are made of, how their parts must be related to form such an entity, what properties and powers flow from that structure, how these entities come into existence, and how that existence is maintained. But social scientists often seem happy to employ entity concepts while ignoring some or all of these questions about them (the examples of 'discourse', 'culture', 'values', and of course 'agency' spring to mind). There is frequently a presumption that we can usefully analyse the social role of such concepts while utterly disregarding their ontological basis.

I believe this is a major source of error and particularly of conceptual confusion in the social sciences; my project in this thesis has in part been to offer a method of resolving such confusion and an illustration of the application of that method. One benefit of emergentism is that it provides a framework for the pursuit of this sort of domain-ontological rigour. On the one hand the general theory of emergence provides a general ontological foundation for the development of domain-specific conceptual understanding; and on the other the *five pillars* provide a useful analytical framework to help us clarify our thinking on any particular entity.

By implication this is a naturalistic approach to domain ontology. The general theory of emergence describes a world in which the emergence of social entities depends on the same sort of structural relations as the emergence of natural entities, and indeed they contain natural entities as their lower-level parts. Thus, as advertised in chapter one, this is a naturalistic ontology. However, we have now gone a step further, since the 'five pillars' approach is also a naturalistic *methodology*: it applies equally well to the social as to the natural world.

Nevertheless, the emergentist approach advocated here also retains what is in a sense an anti-naturalistic element. In recognising that entities at different ontological levels have different sorts of properties that must be studied in different sorts of ways, it also implies that at least *some* parts of the methodology of the social sciences will differ from the methodologies of other sciences. Thus for example we need an interpretative element in the social sciences that we do not need in, say, physics. This particular difference seems to distinguish sharply between the natural and the social sciences, and thus to have *anti-naturalistic* implications.

However, other methodological differences between the sciences do not neatly follow the boundary between the natural and the social sciences. The natural sciences

are by no means homogeneous in their methods, for just the same reason as has been identified above: the different natural sciences differ in their objects of study, and these objects differ in their structures and properties (see chapter one). In a sense, then, the methodology of the natural sciences themselves assumes that methods of study will differ between different disciplines, with the consequence that it is internally contradictory to regard the need for different methods in the social sciences as either anti-naturalistic or naturalistic.

These methodological arguments reflect the ontological structure of the world as it is perceived in emergentist theories. In this view of the world, we no longer have the undifferentiated monism of reductionist ontologies which see all causal powers belonging to the most fundamental physical particles and everything else as just aggregates of such particles. Nor do we have the extreme dualisms of views that insist that life is explanatorily unreachable from the inorganic level, or mind from the physical, or the implicit dualism in the belief of many social scientists that human behaviour is explanatorily unreachable from the biological level or indeed any causal level at all. In place of both we have a *differentiated monism,* which could equally well be described as an *integrated pluralism.* Under either label, this is a recognition that everything is ultimately composed of the same lower level types of components but that they really do have distinct existences and properties by virtue of being organised into higher level structures. And corresponding to this differentiated monism or integrated pluralism at the ontological level, we have a differentiated monism or integrated pluralism at the methodological level. Both of these terms describe the methodological consequences of emergentism far better than either side of the false antinomy of naturalism vs. anti-naturalism.


**Conclusion**

I claim that this thesis has made important and original contributions, both to the general theory of emergence, and to its application to the question of structure and agency. With regard to the first of these, I believe it has shown that the concept of emergence can indeed provide a firm ontological foundation for the generic rejection of eliminative reductionism, when conceived in the terms developed in the early chapters of this thesis – but also that other existing accounts of emergence fail to provide such a foundation. With regard to the second, it has shown that this ontological argument can be applied to the social realm – to individual human agency and to at least some cases

of social structure – and that doing so produces a stronger defence against the ontological errors of methodological individualism, methodological collectivism, and central conflationism than those that have been advanced hitherto.

It thus grounds an improved understanding of structure and agency; an understanding that does not replace existing theoretical approaches to these issues, but does show which of these add value, how they do so, and how they can be reconciled. It has also taken some small steps towards the development of a substantive theory of human action as part of a process of filling in the gaps that this ontological approach reveals in the existing accounts. Although these contributions are original, they do of course draw heavily upon the existing literature. One of their strengths is that they show how some of the contributions in the existing literature can be integrated in a richer and more complete synthesis than those licensed by previous domain ontologies of structure and agency. Most significantly, the more synchronic account of structure and agency developed here is complementary in many respects to the more morphogenetically-oriented account developed by Margaret Archer; but I have also sought to integrate insights from Giddens, Bourdieu, and Sawyer while criticising other aspects of their work.

The theory of emergence elaborated in the early part of this thesis has not only provided the ontological principles upon which this analysis of structure and agency has been built. It has also provided the outline of a methodology for such work: the 'five pillars' approach to analysing the ontology of any given type of object. The analysis of structure and agency has been constructed largely using this method, although I have gone beyond it where this has seemed to help to validate the ontological analysis. This thesis, then, does not only make important claims in its own right. It also initiates an approach to analysing the social world that can be applied much more widely, an approach that has the potential to cast new light on a variety of other problems in the social sciences.

# Bibliography

Archer, M. (1979). *Social origins of educational systems*. London: Sage.

Archer, M. (1982). 'Morphogenesis versus Structuration'. *British Journal of Sociology*, 33, 455-83.

Archer, M. (1990). 'Archer replies to Outhwaite'. In J. Clark, C. Modgil & S. Modgil (Eds.), *Anthony Giddens: Consensus and Controversy* (pp. 86-88). London: Falmer.

Archer, M. (1995). *Realist Social Theory: The Morphogenetic Approach*. Cambridge: Cambridge UP.

Archer, M. (1996a). *Culture and Agency*. Cambridge: Cambridge UP.

Archer, M. (1996b). 'Social integration and system integration: developing the distinction'. *Sociology*, 30, 679-699.

Archer, M. (1998). 'Realism and morphogenesis'. In M. Archer, R. Bhaskar, A. Collier, T. Lawson & A. Norrie (Eds.), *Critical Realism: Essential Readings* (pp. 356-381). London: Routledge.

Archer, M. (2000a). *Being Human: The Problem of Agency*. Cambridge: Cambridge UP.

Archer, M. (2000b). 'For structure: its reality, properties and powers: A reply to Anthony King'. *Sociological Review*, 48, 464-472.

Archer, M. (2003). *Structure, agency, and the internal conversation*. Cambridge: Cambridge UP.

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

Barnes, B. (2000). *Understanding Agency*. London: Sage.

Benton, T. (1985). 'Realism and Social Science'. In R. Edgley & R. Osborne (Eds.), *Radical Philosophy Reader*. London: Verso.

Benton, T., & Craib, I. (2001). *Philosophy of Social Science*. Basingstoke: Palgrave.

Berlin, I.S. (1969). *Four essays on liberty*. Oxford: Oxford University Press.

Bertalanffy, L.v. (1971). *General system theory*: Allen Lane.

Bhaskar, R. (1978). *A realist theory of science*. Hassocks: Harvester Press.

Bhaskar, R. (1986). *Scientific realism and human emancipation*. London: Verso.

Bhaskar, R. (1993). *Dialectic : the pulse of freedom*. London ; New York: Verso.

Bhaskar, R. (1994). *Plato etc*. London: Verso.

Bhaskar, R. (1998a). 'Critical realism and dialectic'. In M. Archer, R. Bhaskar, A. Collier, T. Lawson & A. Norrie (Eds.), *Critical Realism: Essential Readings* (pp. 575-640). London: Routledge.

Bhaskar, R. (1998b). *The possibility of naturalism*. London ; New York: Routledge.

Biddle, B.J. (1986). 'Recent Developments in Role Theory'. *Annual Review of Sociology*, 12, 67-92.

Black, M. (1976). 'Some Questions About Parsons' Theories'. In M. Black (Ed.), *The social theories of Talcott Parsons : a critical examination* (pp. 268-288). Carbondale: Southern Illinois UP.

Blitz, D. (1992). *Emergent evolution*. Dordrecht ; London: Kluwer Academic Publishers.

Bohman, J. (1999). 'Practical Reason and Cultural Constraint: Agency in Bourdieu's Theory of Practice'. In R. Shusterman (Ed.), *Bourdieu: a critical reader* (pp. 129-152). Oxford: Blackwell.

Boogerd, F.C., Bruggeman, F.J., Richardson, R.C., Stephan, A., & Westerhoff, H.V. (2005). 'Emergence and Its Place in Nature: A Case Study of Biochemical Networks'. *Synthese*, 145, 131-164.

Bourdieu, P. (1984). *Distinction : a social critique of the judgement of taste*. London: Routledge & Kegan Paul.

Bourdieu, P. (1990a). *In other words*. Cambridge: Polity.

Bourdieu, P. (1990b). *The logic of practice*. Cambridge: Polity.

Bouveresse, J. (1999). 'Rules, Dispositions, and the *Habitus*'. In R. Shusterman (Ed.), *Bourdieu: a critical reader* (pp. 45-63). Oxford: Blackwell.

Brante, T. (2001). 'Consequences of realism for sociological theory-building'. *Journal for the Theory of Social Behaviour*, 31, 167-95.

Broad, C.D. (1925). *The Mind and its Place in Nature*. London: Kegan Paul & Co.

Brubaker, R. (1993). 'Social Theory as Habitus'. In C. J. Calhoun, M. Postone & E. LiPuma (Eds.), *Bourdieu: critical perspectives* (pp. 212-234). London: Polity Press.

Buckley, W. (1967). *Sociology and modern systems theory*. Englewood Cliffs, N.J.: Prentice-Hall.

Buckley, W. (1998). *Society - a complex adaptive system*. Australia: Gordon and Breach.

Bunge, M. (1996). *Finding philosophy in social science*. New Haven ; London: Yale University Press.

Bunge, M. (2003). *Emergence and convergence*. Toronto ; London: University of Toronto Press.

Byrne, D. (1998). *Complexity Theory and the Social Sciences*: Routledge.

Campbell, D.T. (1974). ''Downward Causation' in Hierarchically Organised Biological Systems'. In F. J. Ayala & T. Dobzhansky (Eds.), *Studies in the Philosophy of Biology* (pp. 179-186). London: Macmillan.

Cartwright, N. (1983). *How the law of physics lie*. Oxford ; New York: Clarendon Press : Oxford University Press.

Cartwright, N. (1999). *The dappled world : a study of the boundaries of science*. Cambridge, UK ; New York, NY: Cambridge University Press.

Castells, M. (2000). *The Rise of the Network Society*. Oxford: Blackwell.

Charles, D. (1992). 'Supervenience, Composition, and Physicalism'. In D. Charles & K. Lennon (Eds.), *Reduction, Explanation, and Realism* (pp. 265-296). Oxford: Oxford UP.

Cilliers, P. (1998). *Complexity and Postmodernism*. London: Routledge.

Cohen, J., & Stewart, I. (1995). *The collapse of chaos*. New York ; London: Penguin.

Collier, A. (1989). *Scientific realism and socialist thought*. Hemel Hempstead: Harvester Wheatsheaf.

Collier, A. (1994). *Critical Realism*. London: Verso.

Collins, R. (1981). 'On the Microfoundations of Macrosociology'. *American Journal of Sociology*, 86, 984-1014.

Coole, D. (2005). 'Rethinking Agency: A Phenomenological Approach to Embodiment and Agentic Capacities'. *Political Studies*, 53, 124-142.

Corning, P.A. (2002). 'The Emergence of "Emergence": Now What?' *Emergence*, 4, 54-71.

Craib, I. (1992). *Anthony Giddens*. London: Routledge.

Crothers, C. (2002). 'History of Social Structural Analysis'. In S. C. Chew & J. D. Knottnerus (Eds.), *Structure, Culture, and History*. Lanham, MD: Rowman & Littlefield.

Davidson, D. (2001). 'Actions, reasons, and causes', *Essays on Actions and Events* (pp. 3-19). Oxford: Clarendon.

Dean, K., Joseph, J., & Norrie, A. (2005). 'Editorial: New Essays in Critical Realism'. *New Formations*, 7-26.

Dennett, D. (2003). *Freedom Evolves*: Allen Lane.

Diamond, J.M. (1992). *The rise and fall of the third chimpanzee*. London: Vintage.

Domingues, J.M. (2000). 'Social Integration, System Integration and Collective Subjectivity'. *Sociology*, 34, 225-241.

Dupre, J. (2001). *Human Nature and the Limits of Science*: Clarendon Press.

Durkheim, E. (1964 [1901]). *The Rules of Sociological Method*. New York: Free Press.

Durkheim, E. (1974 [1898]). 'Individual and Collective Representations', *Sociology and Philosophy*. New York: The Free Press.

Durkheim, É. (1974). *Sociology and philosophy*. New York: Free Press.

Edwards, J. (2003). 'Evolutionary psychology and politics'. *Economy and Society*, 32, 280-298.

Elder-Vass, D. (2005a). 'Emergence and the Realist Account of Cause'. *Journal of Critical Realism*, 4, 315-38.

Elder-Vass, D. (2005b). 'The emergence of social structure and the question of naturalism', *BSA Annual Conference*. York.

Elder-Vass, D. (2005c). 'Towards an inter-disciplinary theory of emergence', *Complexity Science and Society Conference*. Liverpool.

Elder-Vass, D. (forthcoming, 2006). 'Re-examining Bhaskar's three ontological domains: the lessons from emergence'. In C. Lawson, J. Latsis & N. Martins (Eds.), *Contributions to Social Ontology*. London: Routledge.

Emmeche, C., Koppe, S., & Stjernfelt, F. (1997). 'Explaining Emergence: Towards an Ontology of Levels'. *Journal for General Philosophy of Science*, 28, 83-119.

Fairclough, N., Jessop, B., & Sayer, A. (2002). 'Critical Realism and Semiosis'. *Journal of Critical Realism*, 5, 2-10.

Fleetwood, S. (2001). 'Causal Laws, Functional Relations and Tendencies'. *Review of Political Economy*, 13, 201-220.

Fodor, J.A. (1974). 'Special Sciences (Or: The Disunity of Science as a Working Hypothesis)'. *Synthese*, 28, 97-115.

Freeman, W.J. (2000). *How brains make up their minds*. London: Phoenix.

Gell-Mann, M. (1995). *The quark and the jaguar*. London: Abacus.

Giddens, A. (1976). *New Rules of Sociological Method*. London: Hutchinson.

Giddens, A. (1979). *Central problems in social theory*. Basingstoke: Macmillan Press.

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

Giddens, A. (1993). *New Rules of Sociological Method*. Cambridge: Polity.

Goffman, E. (1956). *The presentation of self in everyday life*. Edinburgh: University of Edinburgh, Social Sciences Research Centre.

Gribbin, J.R., & Gribbin, M. (1999). *Almost everyone's guide to science*. London: Phoenix.

Hacker, A. (1976). 'Sociology and Ideology'. In M. Black (Ed.), *The social theories of Talcott Parsons : a critical examination* (pp. 289-310). Carbondale: Southern Illinois UP.

Harré, R., & Madden, E.H. (1975). *Causal powers : a theory of natural necessity*. Oxford: Blackwell.

Harvey, D.L., & Reed, M. (1995). 'Social Science as the Study of Complex Systems'. In L. D. Kiel & E. W. Elliott (Eds.), *Chaos theory in the social sciences : foundations and applications* (pp. 295-325). Ann Arbor, MI: University of Michigan Press.

Healy, K. (1998). 'Conceptualising Constraint: Mouzelis, Archer and the Concept of Social Structure'. *Sociology*, 32, 509-522.

Holland, J.H. (1998). *Emergence: from chaos to order*. Oxford: Oxford University Press.

Honderich, T. (1993). *How Free Are You?* Oxford: Oxford UP.

Honderich, T. (1995). *The Oxford companion to philosophy*. Oxford ; New York: Oxford University Press.

Horgan, T. (2002). 'From Supervenience to Superdupervenience'. In J. Kim (Ed.), *Supervenience* (pp. 113-144). Aldershot: Ashgate/Dartmouth.

Humphreys, P. (1997). 'How properties emerge'. *Philosophy of Science*, 64, 1-17.

James, S. (1984). *The Content of Social Explanation*. Cambridge: Cambridge UP.

Jary, D., & Jary, J. (2000). *Collins Dictionary of Sociology*. Glasgow: Harper Collins.

Jenkins, R. (2002). *Pierre Bourdieu*. London: Routledge.

Kauffman, S.A. (1995). *At home in the universe : the search for laws of self-organization and complexity*. London: Viking.

Kim, J. (1992). ''Downward Causation' in Emergentism and Nonreductive Physicalism'. In A. Beckermann, H. Flohr & J. Kim (Eds.), *Emergence or Reduction?* (pp. 119-138). Berlin: de Gruyter.

Kim, J. (1993). 'The Non-Reductivist's Troubles with Mental Causation'. In J. Heil & A. R. Mele (Eds.), *Mental causation* (pp. 189-210). Oxford: Clarendon 1993.

Kim, J. (1998). *Mind in a physical world : an essay on the mind-body problem and mental causation*. Cambridge, Mass. ; London: The MIT Press.

Kim, J. (1999). 'Making Sense of Emergence'. *Philosophical Studies*, 95, 3-36.

Kim, J. (2002). 'Concepts of Supervenience'. In J. Kim (Ed.), *Supervenience* (pp. 37-62). Aldershot: Ashgate/Dartmouth.

King, A. (1999a). 'Against structure: a critique of morphogenetic social theory'. *Sociological Review*, 47, 199-227.

King, A. (1999b). 'The impossibility of naturalism: The antinomies of Bhaskar's realism'. *Journal for the Theory of Social Behaviour*, 29, 267-288.

King, A. (2000). 'Thinking with Bourdieu against Bourdieu: A 'practical' critique of the habitus'. *Sociological Theory*, 18, 417-433.

Kitcher, P. (1998). '1953 and All That: A Tale of Two Sciences'. In M. Curd & J. A. Cover (Eds.), *Philosophy of Science: The Central Issues* (pp. 971-1003). New York: W W Norton.

Klee, R.L. (1984). 'Micro-Determinism and Concepts of Emergence'. *Philosophy of Science*, 51, 44-63.

Laszlo, E. (1972). *The systems view of the world*. Oxford: Blackwell.

Lau, R.W.K. (2004). 'Habitus and the Practical Logic of Practice: An Interpretation'. *Sociology*, 38, 369-388.

Lawson, T. (1997). *Economics and reality*. London: Routledge.

Lewes, G.H. (1874-9). *Problems of life and mind*. London: Trubner.

Lloyd Morgan, C. (1923). *Emergent evolution*. London: Williams and Norgate.

Lockwood, D. (1964). 'Social Integration and Systems Integration'. In G. K. Zollschan, W. Hirsch & D. Martindale (Eds.), *Explorations in social change* (pp. 244-257). Boston: Houghton Mifflin.

Lokhorst, G.-J. (2005). 'Descartes and the Pineal Gland', in E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2005 Edition)*.

Lopez, J., & Scott, J. (2000). *Social Structure*. Buckingham: Open University Press.

Loyal, S., & Barnes, B. (2001). ''Agency' as a Red Herring in Social Theory'. *Philosophy of the Social Sciences*, 31, 507-524.

Luhmann, N. (1982a). 'The differentiation of society', *The differentiation of society* (pp. 229-254). New York: Columbia University Press.

Luhmann, N. (1982b). 'Politics as a Social System', *The differentiation of society* (pp. 138-165). New York: Columbia University Press.

Machlup, F. (1994). 'Are the social sciences really inferior?' In M. Martin & L. C. McIntyre (Eds.), *Readings in the philosophy of social science* (pp. 5-19). Cambridge, MA: MIT Press c1994.

Malik, K. (2003). 'Science has stolen our soul', *New Statesman* (pp. 48-49).

Martin, M. (1994). 'Taylor on Interpretation and the Sciences of Man'. In M. Martin & L. C. McIntyre (Eds.), *Readings in the Philosophy of Social Science* (pp. 259-279). Cambridge, MA: MIT Press.

McLaughlin, B.P. (1992). 'The Rise and Fall of British Emergentism'. In A. Beckermann, H. Flohr & J. Kim (Eds.), *Emergence or Reduction?* (pp. 49-93). Berlin: de Gruyter.

Merleau-Ponty, M. (1963). *The structure of behaviour*. Boston: Beacon Press.

Merton, R.K. (1968). *Social theory and social structure*. New York: Free Press : Collier-Macmillan.

Meyering, T.C. (2000). 'Physicalism and downward causation in psychology and the special science'. *Inquiry-an Interdisciplinary Journal of Philosophy*, 43, 181-202.

Mihata, K. (1997). 'The Persistence of 'Emergence''. In R. A. Eve, S. Horsfall & M. E. Lee (Eds.), *Chaos, Complexity, and Sociology* (pp. 30-38). Thousand Oaks, Calif.: Sage.

Mill, J.S. (1900). *A system of logic*. London: Longmans.

Mills, C.W. (1970). *The Sociological Imagination*. Harmondsworth: Penguin.

Mouzelis, N. (2000). 'The subjectivist-objectivist divide: Against transcendence'. *Sociology-the Journal of the British Sociological Association*, 34, 741-762.

Mouzelis, N.P. (1991). *Back to sociological theory*. Basingstoke: Macmillan.

Mouzelis, N.P. (1995). *Sociological theory: what went wrong?* London ; New York: Routledge.

Nagel, E. (1998). 'Issues in the Logic of Reductive Explanations'. In M. Curd & J. A. Cover (Eds.), *Philosophy of Science: The Central Issues* (pp. 905-921). New York: W W Norton.

Newman, D.V. (1996). 'Emergence and Strange Attractors'. *Philosophy of Science*, 63, 245-261.

Ollman, B. (2003). *Dance of the dialectic : steps in Marx's method*. Urbana, Ill.: University of Illinois Press.

Parsons, T. (1966). *Societies : evolutionary and comparative perspectives*. Englewood Cliffs, N.J.: Prentice-Hall.

Parsons, T. (1969). *Politics and social structure*. New York: The Free Press ; London : Collier Macmillan.

Parsons, T. (1976). 'The Point of View of the Author'. In M. Black (Ed.), *The social theories of Talcott Parsons : a critical examination* (pp. 311-363). Carbondale: Southern Illinois UP.

Porpora, D. (1998). 'Four concepts of social structure'. In M. Archer, R. Bhaskar, A. Collier, T. Lawson & A. Norrie (Eds.), *Critical Realism: Essential Readings* (pp. 339-355). London: Routledge.

Porpora, D. (2002). 'Social Structure: The Future of a Concept'. In S. C. Chew & J. D. Knottnerus (Eds.), *Structure, Culture, and History*. Lanham, MD: Rowman & Littlefield.

Postone, M., LiPuma, E., & Calhoun, C.J. (1993). 'Introduction'. In C. J. Calhoun, M. Postone & E. LiPuma (Eds.), *Bourdieu: critical perspectives* (pp. 1-13). London: Polity.

Prigogine, I., & Stengers, I. (1984). *Order out of chaos : man's new dialogue with nature*. London: Heinemann.

Richards, J.R. (2000). *Human nature after Darwin : a philosophical introduction*. London: Routledge.

Rose, H., & Rose, S. (2001). *Alas Poor Darwin: Arguments against evolutionary psychology*. London: Vintage.

Savellos, E.E., & Yalçin, Ü.D. (1995). 'Introduction'. In E. E. Savellos & Ü. D. Yalçin (Eds.), *Supervenience : new essays* (pp. 1-15). Cambridge [England] ; New York: Cambridge University Press.

Sawyer, R.K. (2001). 'Emergence in sociology'. *American Journal of Sociology*, 107, 551-585.

Sawyer, R.K. (2002a). 'Nonreductive individualism Part I - Supervenience and wild disjunction'. *Philosophy of the Social Sciences*, 32, 537-559.

Sawyer, R.K. (2002b). 'Unresolved tensions in sociocultural theory: Analogies with contemporary sociological debates'. *Culture & Psychology*, 8, 283-305.

Sawyer, R.K. (2003). 'Nonreductive individualism - Part II - Social causation'. *Philosophy of the Social Sciences*, 33, 203-224.

Sawyer, R.K. (2005). *Social Emergence*. Cambridge: Cambridge UP.

Sayer, A. (1992). *Method in Social Science*. London: Routledge.

Sayer, A. (1997). 'Critical Realism and the Limits to Critical Social Science'. *Journal for the Theory of Social Behaviour*, 27, 473-488.

Sayer, A. (1999). *Realism and Social Science*. London: Sage.

Sayer, A. (2005). 'Class, Moral Worth and Recognition'. *Sociology*, 39, 947-964.

Searle, J.R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.

Searle, J.R. (1995). *The construction of social reality*. London: Allen Lane, Penguin Press.

Searle, J.R. (1997). *The mystery of consciousness*. London: Granta.

Smith, T.S. (1997). 'Nonlinear Dynamics and the Micro-Macro Bridge'. In R. A. Eve, S. Horsfall & M. E. Lee (Eds.), *Chaos, Complexity, and Sociology* (pp. 52-63). Thousand Oaks, Calif.: Sage.

Sperry, R.W. (1969). 'A modified concept of consciousness'. *Psychological Review*, 76, 532-36.

Sperry, R.W. (1986). 'Macro- versus Micro-Determinism'. *Philosophy of Science*, 53, 265-270.

Stephan, A. (1992). 'Emergence - A Systematic View on its Historical Facets'. In A. Beckermann, H. Flohr & J. Kim (Eds.), *Emergence or Reduction?* Berlin: de Gruyter.

Stephan, A. (2002). 'Emergentism, Irreducibility, and Downward Causation'. *Grazer Philosophische Studien*, 77-94.

Stones, R. (2001). 'Refusing the Realism-Structuration Divide'. *European Journal of Social Theory*, 4, 177-198.

Stones, R. (2002). 'Structuration'. *Sociology-the Journal of the British Sociological Association*, 36, 222-224.

Stones, R. (2005). *Structuration theory*. Basingstoke: Palgrave Macmillan.

Taylor, C. (1994). 'Interpretation and the Sciences of Man'. In M. Martin & L. C. McIntyre (Eds.), *Readings in the Philosophy of Social Science* (pp. 181-211). Cambridge, MA: MIT Press.

Teller, P. (1992). 'A Contemporary Look at Emergence'. In A. Beckermann, H. Flohr & J. Kim (Eds.), *Emergence or Reduction?* (pp. 139-153). Berlin: de Gruyter.

Thompson, J.B. (1992). 'Editor's Introduction'. In P. Bourdieu (Ed.), *Language and symbolic power* (pp. 1-31). Cambridge: Polity.

Tooby, J., & Cosmides, L. (1992). 'The Psychological Foundations of Culture'. In L. Cosmides, J. Tooby & J. H. Barkow (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford: Oxford UP.

Varela, C.R., & Harre, R. (1996). 'Conflicting Varieties of Realism'. *Journal for the Theory of Social Behaviour*, 26, 313-25.

Wacquant, L.J.D. (1993). 'Bourdieu in America'. In C. J. Calhoun, M. Postone & E. LiPuma (Eds.), *Bourdieu: critical perspectives* (pp. 235-262). London: Polity.

Walby, S. (2005). 'No one Polity Saturates the Political Space in a Given Territory'. *Sociology*, 38, 1035-1042.

Waldrop, M.M. (1994). *Complexity : the emerging science at the edge of order and chaos*. London: Penguin Books.

Watson, G. (1982). 'Introduction', *Free Will* (pp. 1-14). Oxford: Oxford UP.

Williams, R. (1976). *Keywords : a vocabulary of culture and society*. [London]: Fontana/Croom Helm.

Wimsatt, W.C. (2000). 'Emergence as Non-Aggregativity and the Biases of Reductionisms'. *Foundations of Science*, 5, 269-297.

Winch, P. (1958). *The Idea of a Social Science*. London: Routledge & Kegan Paul.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell.