PART 1

ORIGINS

There is a central group of ideas that underlies our understanding of the process of convection in the earth's solid mantle. These ideas are that the earth is very old, that temperatures and pressures are high in the earth's interior, and that given high temperature, high pressure and sufficient time, solid rock can flow like a fluid. As well there is the idea that the earth's crust has been repeatedly and often profoundly deformed and transmuted. This idea is a central product of several centuries' practice of the science of geology. It is the perceived deformations of the crust that ultimately have led to the development of the idea of mantle convection, as their explanation. Our subject thus connects directly to more than two centuries' development of geological thought, especially through crustal deformation, heat, time and the age of the earth.

I think we scientists should more often examine the origins of our discipline. In doing so we gain respect for our scientific forebears and we may encounter important neglected ideas. We will usually gain a perspective that will make us more effective and productive scientists. Looking at our history also helps us to understand the way science is done, which is very differently from the hoary stereotype of cold logic, objectivity, 'deduction' from observations, and inexorable progress towards 'truth'.

We may be reminded also that science has profoundly changed our view of the world and we may feel some humility regarding the place of humans in the world. The deformation processes that are the subject of this book are only very marginally a part of immediate human experience, even though they are not as exotic as, for example, quantum physics or relativity. Partly because of this, understanding of them emerged only gradually over a long period, through the efforts of a great many scientists. It is easy to take for granted the magnitude of the accumulated shifts in concepts that have resulted.

Finally, there are many people in our society who are very ignorant of the earth and its workings, or who actively resist ideas such as that the earth is billions of years old. If we are to give our society the benefit of our insights without sounding authoritarian, we must be very clear about where those ideas derive from.

For these reasons, and because there is a fascinating story to be told, Chapters 2 and 3 present a short account of the emergence of the central ideas that have engendered the theories of plate tectonics and mantle convection. Chapter 1 outlines the rationale of the book.

CHAPTER 1

Introduction

1.1 Objectives

The purpose of this book is to present the principles of convection, to show how those principles apply in the peculiar conditions of the earth's mantle, and to present the most direct and robust inferences about mantle convection that can be drawn from observations. The main arguments are presented in as simple a form as possible, with a minimum of mathematics (though more mathematical versions are also included). Where there are controversies about mantle convection I give my own assessment, but I have tried to keep these assessments separate from the presentation of principles, main observations and direct inferences. My decision to write this book arose from my judgement that the broad picture of how mantle convection works was becoming reasonably settled. There are many secondary aspects that remain to be clarified.

There are many connections between mantle convection and geology, using the term 'geology' in the broadest sense: the study of the earth's crust and interior. The connections arise because mantle convection is the source of all tectonic motions, and because it controls the thermal regime in the mantle and through it the flow of heat into the crust. Some of these connections are noted along the way, but there are three aspects that are discussed more fully. The first is in Part 1, where the historical origins of the ideas that fed into the conception of mantle convection are described. Especially in Chapter 2 those historical connections are with geology. Another major connection is through Chapter 13, in which the relationship between mantle chemistry and mantle convection is considered. The third respect arises in the last chapter, where the broad tectonic implications of hypothetical past mantle regimes are discussed.

A theory of mantle convection is a *dynamical* theory of geology, in that it describes the *forces* that give rise to the motions apparent in the deformation of the earth's crust and in earthquakes and to the magmatism and metamorphism that has repeatedly affected the crust. Such a dynamical theory is a more fundamental one than plate tectonics, which is a *kinematic* theory: it describes the *motions* of plates but not the forces that move them. Also plate tectonics does not encompass mantle plumes, which comprise a distinct mode of mantle convection. It is this fundamental dynamical theory that I wish to portray here.

This book is focused on those arguments that derive most directly from observations and the laws of physics, with a minimum of assumption and inference, and that weigh most strongly in telling us how the mantle works. These arguments are developed from a level of mathematics and physics that a first or second year undergraduate should be familiar with, and this should make them accessible not just to geophysicists, but to most others engaged in the study of geology, in the broad sense. To maximise their accessibility to all geologists, I have tried to present them in terms of simple physical concepts and in words, before moving to more mathematical versions.

For some time now there has been an imperative for geologists to become less specialised. This has been true especially since the advent of the theory of plate tectonics, which has already had a great unifying effect on geology. I hope my presentation here is sufficiently accessible that specialists in other branches of geology will be able to make their own informed judgements of the validity and implications of the main ideas.

Whether my judgement is correct, that the main ideas presented here will become and remain broadly accepted, is something that only the passage of time will reveal. Scientific consensus on major ideas only arises from a prolonged period of examination and testing. There can be no simple 'proof' of their correctness.

This point is worth elaborating a little. One often encounters the phrase 'scientifically proven'. This betrays a fundamental misconception about science. Mathematicians prove things. Scientists, on the other hand, develop models whose behaviour they compare with observations of the real world. If they do not correspond (and assuming the observations are accurate), the model is not a useful representation of the real world, and it is abandoned. If the model behaviour does correspond with observations, then we can say that it works, and we keep it and call it a theory. This does not preclude the possibility that another model will work as well or better (by corresponding with observations more accurately or in a broader context). In this case, we say that the new model is better, and usually we drop the old one.

However, the old model is not 'wrong'. It is merely less useful, but it may be simpler to use and sufficient in some situations. Thus Newton's theory of gravitation works very well in the earth's vicinity, even though Einstein's theory is better. For that matter, the old Greek two-sphere model of the universe (terrestrial and celestial) is still quite adequate for navigation (strictly, the celestial sphere works but the non-spherical shape of the earth needs to be considered). Scientists do not 'prove' things. Instead, they develop more useful models of the world. I believe the model of mantle dynamics presented here is the most useful available at present.

Mantle convection has a fundamental place in geology. There are two sources of energy that drive geological processes. The sun's energy drives the weather and ocean circulation and through them the physical and chemical weathering and transport processes that are responsible for erosion and the deposition of sediments. The sun's energy also supports life, which affects these processes.

The other energy source is the earth's internal heat. It is widely believed, and it will be so argued here, that this energy drives the dynamics of the mantle, and thus it is the fundamental energy source for all the non-surficial geological processes. In considering mantle dynamics, we are thus concerned with the fundamental mechanism of all of those geological processes. Inevitably the implications flow into many geological disciplines and the evidence for the theory that we develop is to be found widely scattered through those disciplines.

Inevitably too the present ideas connect with many ideas and great debates that have resonated through the history of our subject: the rates and mechanisms of upheavals, the ages of rocks and of the earth, the sources of heat, the means by which it escapes from the interior, the motions of continents. These connections will be related in Part 1. The historical origins of ideas are often neglected in science, but I think it is important to include them, for several reasons. First, to acknowledge the great thinkers of the past, however briefly. Second, to understand the context of ideas and theories. They do not pop out of a vacuum, but emerge from real people embedded in their own culture and history, as was portrayed so vividly by Jacob Bronowski in his television series and book *The Ascent of Man* [1]. Third, it is not uncommon for alternative possibilities to be neglected once a particular interpretation becomes established. If we returned more often to the context in which choices were made, we might be less channelled in our thinking.

1.2 Scope

The book has four parts. Part 3, Essence, presents the essential arguments that lead most directly to a broad outline of how mantle dynamics works. Part 2, Foundations, lays the foundations for Part 3, including key surface observations, the structure and physical properties of the interior, and principles and examples of viscous fluid flow and heat flow.

Parts 1 and 4 connect the core subject of mantle convection to the broader subject of geology. Part 1 looks at the origin and development of key ideas. Part 4 discusses possible implications for the chemical and thermal evolution of the mantle, the tectonic evolution and history of the continental crust. Many aspects of the latter topics are necessarily conjectural.

1.3 Audience

The book is intended for a broad geological audience as well as for more specialised audiences, including graduate students studying more general aspects of geophysics or mantle convection in particular. For the latter it should function as an introductory text and as a summary of the present state of the main arguments. I do not attempt to summarise the many types of numerical model currently being explored, nor to present the technicalities of numerical methods; these are likely to progress rapidly and it is not appropriate to try to summarise them in a book. My expectation is that the broad outlines of mantle convection given here will not change as more detailed understanding is acquired.

In order to accommodate this range of readership, the material is presented as a main narrative with more advanced or specialised items interspersed. Each point is first developed as simply as possible. Virtually all the key arguments can be appreciated through some basic physics and simple quantitative estimates. Where more advanced treatments are appropriate, they are clearly identified and separated from the main narrative. Important conclusions from the advanced sections are also included in the main narrative.

It is always preferable to understand first the qualitative arguments and simple estimates, before a more elaborate analysis or model is attempted. Otherwise a great deal of effort can be wasted on a point that turns out to be unimportant. Worse, it is sometimes true that the relevance and significance of numerical results cannot be properly evaluated because scaling behaviour and dependence on parameter values are incompletely presented. Therefore the mode of presentation used here is a model for the way theoretical models can be developed, as well as a useful way of reaching an audience with a range of levels of interest and mathematical proficiency.

1.4 Reference

1. J. Bronowski, *The Ascent of Man*, 448 pp., Little, Brown, Boston, 1973.

CHAPTER 2

Emergence

We begin with a look at some of the ‘classical’ questions about the earth: its age, its internal heat, and how rocks may deform. These questions are famous both because they are fundamental and because some great controversies raged during the course of their resolution. In looking at how the age of the earth was first inferred, we soon encounter the question of whether great contortions of the crust happened suddenly or slowly. The fact that the interior of the earth is hot is central, both to the occurrence of mantle convection and geological processes, but also historically because one estimate of the age of the earth was based on the rate at which it would lose internal heat.

Much of my limited knowledge of the history of geology prior to this century comes from Hallam’s very readable short book *Great Geological Controversies* [1]. I make this general acknowledgement here to save undue interruption of the narrative through this chapter. My interpretations are my own responsibility.

2.1 Time

The idea that continents shift slowly about the face of the earth becomes differentiated from fantasy only with an appreciation of time. One of the most profound shifts in the history of human thought began about 200 years ago, when geologists first began to glimpse the expanse of time recorded in the earth’s crust. This revolution has been less remarked upon than some others, perhaps because it occurred gradually and with much argument, and because the sources of evidence for it are less accessible to common observation than, for example, the stars and planets that measure the size of the local universe, or the living things that are the products of natural selection.

During the time since the formulation of the theory of plate tectonics, my home in Australia has moved about 1.8 m closer to the equator. Within the same period, that displacement has become accessible to direct scientific observation, but not to unaided human perception. Mostly the landscape is static to human perception. It is not an uncommon experience to see the aftermath of a landslide or rockfall, and it is occasionally possible to see a fresh fault scarp after an earthquake. Students of geology now take for granted that these are irreversible events that are part of the processes of erosion and tectonic deformation. However, the relationship of these observations to the form of the land surface and to folded and faulted rock strata is not at all immediately obvious. Indeed it is only 200 years since this connection began to be made seriously and systematically, and less than 100 years since earthquakes, fault scarps and sudden slip on buried faults were coherently related through the ideas of accumulated elastic stress and frictional fault surfaces.

One person's dawning comprehension of the expanse of geological time is recorded in the account (quoted by Hallam [1], p. 33) by the mathematician John Playfair of his visit in 1788 to Siccar Point in Britain, in the company of the geologists Hutton and Hall, to observe a famous unconformity where subhorizontal Devonian sandstones rest on near-vertical Silurian slates (which he called schistus).

> We felt ourselves necessarily carried back to the time when the schistus on which we stood was yet at the bottom of the sea, and when the sandstone before us was only beginning to be deposited, in the shape of sand and mud, from the waters of a superincumbent ocean. An epocha still more remote presented itself, when even the most ancient of these rocks, instead of standing upright in vertical beds, lay in horizontal planes at the bottom of the sea, and was not yet disturbed by that immeasurable force which has burst asunder the solid pavement of the globe. Revolutions still more remote appeared in the distance of this extraordinary perspective. The mind seemed to grow giddy by looking so far into the abyss of time . . .

Playfair and Hutton did not have a clear quantitative measure of the time intervals they were contemplating, but they knew they were dealing with periods vastly greater than the thousands of years commonly believed at the time. Hutton especially must have appreciated this, because he is perhaps most famous for expounding the idea of indefinite time in a famous statement from that same year [2] '. . . we find no vestige of a beginning, no prospect of an end.' (I said 'indefinite time' rather than 'infinite time' here because Hutton's words do not necessarily imply the latter. In modern parlance, we could say that Hutton was proposing that the earth was in a steady state, and it is characteristic of steady-state processes that information about their initial conditions has been lost.)

The work and approach of Lyell in the first half of the nineteenth century provided a basis for quantitative estimates of the elapse of time recorded in the crust. Lyell is famous for expounding and applying systematically the idea that geological structures might be explained solely by the slow action of presently observable processes. He and many others subsequently made use of observations that could be related to historical records, of erosion rates and deposition rates, and of stratigraphic relationships, to demonstrate that a great expanse of time was required. Though still rather qualitative, an eloquent example comes from an address by Lyell in 1850 (Hallam [1], p. 58; [3]).

> The imagination may well recoil from the vain effort of conceiving a succession of years sufficiently vast to allow of the accomplishment of contortions and inversions of stratified masses like those of the higher Alps; but its powers are equally incapable of comprehending the time required for grinding down the pebbles of a conglomerate 8000 feet [2650 metres] in thickness. In this case, however, there is no mode of evading the obvious conclusion, since every pebble tells its own tale. Stupendous as is the aggregate result, there is no escape from the necessity of assuming a lapse of time sufficiently enormous to allow of so tedious an operation.

According to Hallam (p. 106), it was Charles Darwin who made one of the first quantitative estimates of the lapse of geological time, in the first edition of his *Origin of Species* [4]. This was an estimate for the time to erode a particular formation in England, and Darwin's estimate, not intended to be anything more than an illustration, was 300 million years. Though it might have been only rough, Darwin's estimate conveys the idea that the time spans involved in geology, that can be characterised qualitatively only by vague terms such as 'vast', are not 300 000 years and not 300 billion years, for example.

During the middle and later years of the nineteenth century, a great debate raged amongst geologists and between geologists and physicists, particularly Lord Kelvin, about the age of the earth (which I will discuss in Section 2.4). What impresses me is not so much the magnitudes of the differences being argued as the general level of agreement and correctness, especially amongst geologists' estimates. We must realise that initially they knew only that the number must be orders of magnitude greater than the 10^4 years or so inferred from scriptures, a number that was then still commonly