CHAPTER 3

Mobility

The idea that parts of the earth have moved slowly relative to each other over distances comparable to the size of the globe belongs mostly to the twentieth century. There were some earlier suggestions of catastrophic global displacements, but it was in the twentieth century that large slow displacements of the continents were proposed and systematically advocated, and eventually their existence was decisively established.

Historically, the idea of mantle convection is closely entwined with the ideas of continental drift and plate tectonics. The idea that the earth's interior is mobile can be traced back at least to the middle of the nineteenth century, but it became the focus of sharp debate early in the twentieth century with the acquisition of seismological evidence that below the crust is a solid, rocky mantle extending about halfway to the centre of the earth. To many this seemed to make continental drift impossible.

After Holmes proposed mantle convection as a possible mechanism of continental drift around 1930, most thinking about continental drift and the emerging plate tectonics was strongly conditioned by expectations of how such convection would work. One can argue that this interaction of ideas actually held back the recognition of the pattern of movements on the earth's surface (Section 3.8).

Others have told the story of the theory of continental drift and its ultimate evolution into the theory of plate tectonics [1–6]. There are several reasons for recounting it here, rather than simply proceeding to a description of the plates and how they work. Mainly, I want to include the complementary development of ideas about the mobility of the mantle, which I think played a larger role in the story than has been appreciated. Also I want to highlight some aspects of the plate tectonics story that I think deserve more attention than they have received, and to continue the theme of

Chapter 2, showing some of the context from which important ideas emerged. Without this, it is easy to overlook the large amount of good science, done by a great many scientists, upon which such insights are usually founded. Finally, when a theory becomes as widely accepted as the theory of plate tectonics, it is easy to lose sight of why we think it is a good theory. I therefore describe some of the early and compelling evidence for the existence of the plates and their motions.

There is another story that developed in the shadow of plate tectonics: the story of mantle plumes. Plumes are a distinct component of mantle convection. They have played a significant role in the history of the continents, and possibly had a larger role early in earth history, though that is still quite uncertain. They may have had decisive effects on the history of life. The story of the idea of plumes is just as long but not as complex as the story of plates. Perhaps four names can be identified as principals: Darwin, Dana, Wilson and Morgan, with a footnote for Holmes.

This chapter is about how the idea of large, slow displacements became established. It is a long story, so my account here has to be selective, including only key evidence and key arguments. Because the stories of continental drift and plate tectonics have been told before in some detail, I do not provide a lot of detail nor many illustrations here, except where I want to emphasise particular points.

3.1 Drifting continents

In 1912 Alfred Wegener first spoke publicly and wrote of his idea that whole continents had undergone large, slow displacements. These were published in book form in 1915 under the title *Die Entstehung der Kontinente und Ozeane* (*The Origin of Continents and Oceans*), which went through several editions both before and after his death [7, 8].

The seed of Wegener's theory came from the similarity in map view of the shapes of the continental margins on either side of the Atlantic Ocean. This similarity is reflected more crudely by the coastline, and had been remarked upon previously. His ideas became more definite when he learned of similarities in fossils occurring on opposite sides of the Atlantic, and later of geological similarities. He developed his ideas into the proposal that all of the continents had been grouped into one supercontinent, and that this had fragmented and the pieces had drifted apart starting in the Mesozoic era, about 200 Ma ago.

In later editions he added more data, and also used evidence from various kinds of deposits that could be used to infer climate and thus to distinguish equatorial and polar regions, arguing that the distributions of palaeoclimate made more sense if the continents had moved.

The similarities in fossils were already being noted, and had led palaeontologists to postulate the past existence of land connections between the widely separated continents. Initially these connections were assumed to be continents that had later subsided under the ocean. Geologists had for decades held that continents episodically emerged from and subsided into the ocean, on the basis of the widespread occurrence of marine fossils in continental sediments. Little was known of the nature of the rocks of the sea floor at this time, which limited the possibility of direct geological tests of this idea.

Wegener argued that such large-scale vertical movements of continents were not viable, because gravity measurements had shown that the earth's crust is close to isostatic equilibrium. I will discuss this in more detail in the next section. The point is that it was already established that the continents stand higher than the seafloor because continental crust must be less dense, and that the continents in effect 'float' in a denser substratum. If large continental blocks subsided by several kilometres, there would be a large negative gravity anomaly created, and such large anomalies were excluded by the observations. One response to Wegener's isostasy argument was to assume that the land connections were smaller than continental scale – narrow land 'bridges'. This proposal was *ad hoc*, and has never had any evidence to support it.

Another argument Wegener made against the idea of rising and sinking continents was that if it occurred erratically in space and time, as the geological record suggested, one would expect the elevations of the earth's surface to be spread more evenly than they are between the highest and the lowest, with a preponderance of areas at intermediate elevations. Instead, the observation is that there are two preponderant levels of the earth's surface, one at the level of the deep sea floor and one at the level of the continental surfaces, just above sea level.

This bimodal distribution of elevation had been recognised for a long time as a first-order feature of the earth requiring explanation. Wegener's point is a very sound one, that the observed topography looks like a very improbable consequence of the older idea of rising and sinking continents. Its rhetorical weakness was that he did not have an explanation for the bimodal topography either. In

the absence of explicit mechanisms for either vertical or horizontal displacements of continents, it was not possible to quantify the argument at all, and so the opponents of his theory were free to make suppositions to suit their point of view, and to note, for instance, that the particular Gaussian distribution that he assumed to make his point was no more probable, a priori, than the observed distribution.

Hallam [1] points to two factors that may have facilitated Wegener's boldness of thinking. One was that he was not a geologist by training, but a meteorologist, so he had no particular commitment to prevailing ideas in the geological community. The other was that in Germany at the time, the geophysics community embraced meteorology and climatology as well as the 'solid' earth, which perhaps made their thinking more open to mobilist ideas.

Wegener was tentative about what force or forces might cause continents to drift, writing, 'The Newton of drift theory has not yet appeared . . .' and conceding that it might be a long time before this was clarified [1]. In what can be seen in retrospect as a key tactical error, he suggested a differential rotational force and tidal forces as possible causes. (Recall that Darwin also made this tactical 'error' when he made a rough estimate of the age of an erosional episode, thereby providing a target for Kelvin to snipe at.)

By the time of his third edition (1922), which became better known in the English-speaking world, Wegener's theory began to generate very strong opposition. According to Menard [6], there were two reasons in particular that might have contributed to this. One was that Wegener (perhaps like most of us) had begun with the naive idea that his theory was so obvious that it would quickly be accepted. When this did not occur, and he observed palaeontologists failing to understand his argument against land bridges, he became more of an advocate. The other reason was that he believed that geodetic measurements showed a shift of Greenland relative to Europe, and that this was a dramatic confirmation of his theory. (The drift rate implied by the data was metres per year, but he had also correlated Pleistocene glacial moraines across the Atlantic that would have required this rate.) The data later turned out to be in error, but by this time he may have become totally convinced, and adopted a more evangelical approach. When some (but by no means all) of his arguments were found wanting (such as the correlation of moraines), his credibility dropped, and the annoyance of his detractors rose.

Prominent among the opponents of continental drift was Harold Jeffreys. Jeffreys showed that Wegener's proposed driving forces were many orders of magnitude smaller than would be required to overcome the resistance from the oceanic crust through which the continents were presumed to move. Even in 1926 Jeffreys' language reveals a reaction to Wegener's fervour. He parodied Wegener by accusing him of arguing that a small force acting for a very long time could overcome a much larger force acting for the same time, and characterised this idea as 'a very dangerous one, liable to lead to serious error' ([1], [9] p. 150).

Jeffreys' dismissal of Wegener's proposed mechanisms extended to the whole idea of drifting continents. Jeffreys' language reveals that Wegener's proposed forces merely provided a convenient weakness through which to attack the larger theory. It is not clear that Jeffreys made a serious attempt to appreciate Wegener's many geological arguments. He even attacked Wegener's geophysical argument against land bridges, a subject in which he should have been expert, but in which his arguments were inconsistent with the well-known observational basis of isostasy, as you will see in the next section. Thus we see again the process of alternating over-reactions generating a heated scientific debate, just as in the nineteenth century arguments over the age of the earth, and in many subsequent topics in many areas.

A primary source of opposition to Wegener's theory was the well-established view amongst geologists and geophysicists that continents are fixed in a strong outer shell of the earth. Compressive mountain building forces were supposed to derive from cooling and contraction of the earth, which generated compressive stresses and occasional failure in this shell, or lithosphere. This theory was attracting its own opponents, because it was far from clear that it could provide for sufficient crustal shortening to account for the major compressional mountain ranges.

Superficially the model of a cooling, contracting earth seems attractive, and very compatible with Kelvin's concept, upon which his cooling age of the earth was based (Section 2.4). However, when the earth is assumed to cool from the outside by conduction, the result is that the deep interior would not have had time to cool at all. The cooling is restricted to a gradually thickening layer at the surface. Initially the surface would be put in tension, as it contracted relative to the constant-volume interior. One has to assume that the resulting tension is relieved by failure, and then this layer would be compressed as the cooling penetrated deeper. The resulting amount of compression is much less than if the whole interior were cooling.

Anyway, Wegener's theory attracted substantial opposition, and for the next several decades it was not very respectable to

advocate anything related to continental drift. It is interesting that Gutenberg, in an article written initially in the 1930s [10], reviewed an extensive literature of speculative tectonic theories, with continental drift prominent among them. Gutenberg had been an associate of Wegener's in Germany, before Gutenberg's move to the California Institute of Technology in the late 1920s. However Gutenberg gave much less space to continental drift in his later book, published in 1959 [11].

The only prominent advocates of continental drift in this period were Alex du Toit [12], in South Africa, and Sam Carey [13] at the University of Tasmania. These two geologists enjoyed two advantages. One was that some of the clearest geological evidence for past continental connections exists in the southern continents. The other was that, being located in the further reaches of the civilised western world, they could perhaps be safely ignored in the important centres of learning. du Toit, especially, contributed a great deal of evidence and elaborated Wegener's ideas significantly, and his work attracted a significant minority of followers. Carey also contributed important evidence and arguments, and is otherwise most noted for the radical idea that the earth has expanded substantially since the Palaeozoic, this being his preferred mechanism for continental drift. Carey's ideas were an important stimulus for Wilson (Section 3.4).

3.2 Creeping mantle

The idea of a deformable mantle began to have an empirical basis when it was discovered that the gravitational attraction of mountain ranges is less than would be expected from their topography. An account is given by Daly [14]. The deficit in gravitational attraction was first recorded for the Andes mountains by Bouguer, on an expedition between 1735 and 1745, and later, in 1849, by Petit near the Pyrenees. It was analogous observations arising from Everest's surveying in India that led to a quantified hypothesis.

This came from a discrepancy between two different surveying methods in India, near the Himalayan range, one based on triangulation and the other on astronomical sighting. The astronomical sightings were done relative to the local vertical, as determined by a plumb line, that is a weight hanging on a thread. In 1855 Pratt [15] proposed that the discrepancy arose because the vertical was deflected slightly by the gravitational attraction of the Himalayas. However his calculations revealed that the deflection was only about a third of what would be expected from the visible mountain range.

Explanations were offered later by Pratt [16] and by Airy [17]. Pratt noted that the discrepancy could be accounted for by a hidden excess mass to the south or by a hidden mass deficit to the north. He suggested that such differences in density might have arisen since the earth was young and liquid, but without changing the mass in any vertical column extending down from the surface, for example by differential thermal expansion. In this case different vertical columns of equal surface area would all still contain equal amounts of mass. As an illustration, he calculated that a small density deficit extending to a depth of 100 miles (160 km), and such as to make the mass column including the mountains the same as under the adjacent plains, could reduce the vertical deflection to zero.

Every geology student learns about isostasy, and about Pratt's and Airy's variations on how to distribute the density deficit under mountain ranges. What I had never appreciated until I read Daly's extensive quotation from Airy's short paper was how penetrating and far-reaching was Airy's thinking. He is famous for hypothesising what Dutton later called the condition of isostasy [18], but his thinking goes to the core of the subject of tectonic mechanism.

Rather than assuming that mass columns had remained constant through earth history, as had Pratt, Airy thought it was necessary to consider that the earth was subject to 'disturbing causes' through its history which would change both the topography and the mass within columns. He noted that the shape of the solid part of the earth closely approximates the shape of the liquid ocean surface, that there is not a concentration of land or water near the equator, and that both of these observations had been taken by physicists to indicate 'either that the interior of the earth is now fluid or that it was fluid when the mountains took their present forms'. He goes on

> This fluidity may be very imperfect; it may be mere viscidity; *it may even be little more than that degree of yielding which (as is well known to miners) shows itself by changes in the floors of subterraneous chambers at a great depth* when their width exceeds 20 or 30 feet [7 or 10 metres]; and this degree of yielding may be sufficient for my present explanation. [Emphasis added.]

Here, very clearly, is an empirically based concept of a solid, rocky, but deformable interior.

Airy therefore assumed an outer, non-deforming 'crust' and a denser, fluid interior. He argued first that a broad plateau could not be supported alone by the strength of the crust, demonstrating that the leverage required at its edges required a tensile strength that

was very implausible, given that the crust is known to be riven with fractures, even if the crust is 100 miles (160 km) thick. He then asked how else such plateaus might be supported, and answered himself

> I conceive there can be no other support than that arising from the downward projection of a portion of the earth's light crust into the dense [substratum]; ... the depth of its projection downwards being such that the increased power of floatation thus gained is roughly equal to the increase of weight above from the prominence of the [plateau].

He compared the crust to a raft of timber floating on water, wherein a log whose top is higher than the others will be correctly inferred to be larger and thus to project deeper into the water than the others.

Airy then showed how the downward projection of the lower-density crust (the root, as it has become known) will reduce the net gravitational attraction, and that at a distance great compared with the depth of the projection the net gravitational perturbation will approach zero. He noted that one would not expect that there would everywhere be a perfect isostatic balance, but that the strength of the crust would allow some mountains to project higher or some roots to project deeper than in the isostatic condition. Finally he noted that this would be especially true of mountains of small horizontal extent, since the leverage required to hold them up is smaller.

In 1859 Hall [19] presented evidence of slow, continuous adjustment of the earth's surface to changing loads, by demonstrating that sediments now buried deep in thick sedimentary sequences were deposited in shallow water. This observation, and many others of its kind since, went far towards justifying Airy's assumption that the interior of the earth is fluid at present, and that the isostatic condition was not just a relic from early in earth's history.

By 1889 there was accumulating evidence that the crust on broad scales is close to isostatic equilibrium, and Dutton [18] formalised the idea and proposed the name isostasy (Greek: *isos*, equal; *statikos*, stable). (Dutton actually preferred the term isobary, or equal pressure, but this was already in use in another context.)

Helmert [20] conceived in 1909 that the depth of the compensating mass deficit could be constrained by the form and magnitude of the gravity anomaly at the edge of a broad structure, and he and others used observations near continental margins to deduce that the density anomalies extended to depths of the order of 100 kilometres. This implied that the non-deformable crust must extend to such depths, in order for the density anomalies to persist.

In 1914 Barrell [21] proposed the term 'asthenosphere' (weak layer) for the deformable region below the region of strength. By this time the term 'lithosphere' was in use to describe the non-deforming layer near the surface. This had been distinguished from the low-density compositional layer, the 'crust' in modern usage, by the discovery of the Mohorovičić discontinuity in 1909 [22], which was inferred to mark the base of the crust. Barrell was willing to assume that the thickness of the asthenosphere is as great as 600 km, in order to reduce the amount of deformation required to accommodate surface uplifts. This allowed him to argue that a deformable asthenosphere was not incompatible with the solid state, as shown by its ability to propagate seismic shear waves [23].

Thus by 1914 there was a clear picture, well based on observations, of a lithosphere about 100 km thick and strong enough, on geological time scales, to support topography up to a width of the order of 100 km. Topography on broader scales was known to be approximately in isostatic balance, including the earth's first-order topography, the continent–ocean dichotomy. It was inferred that this is because the asthenosphere, below the lithosphere, behaves like a fluid on geological time scales, in spite of being in the solid state.

A different kind of observation was developed through this period which strongly supported this picture, but there were two other kinds of observation that complicated it. The supporting observation was of a protracted 'rebound' of the earth's surface in the Fennoscandian region following melting of the glaciation from the last ice age. It was argued by Jamieson in 1865 [24] that this could be explained by a viscous outflow from under the icecap, with a return flow after the icecap melted. The delayed response, by more than 10 000 years, required more than just an elastic yielding, which would rebound immediately the ice load was removed. This hypothesis was debated for a long time, but by the 1930s well-founded estimates of the viscosity of the asthenosphere had been derived by several workers. The result obtained depends substantially on the assumed thickness of the asthenosphere. For example, van Bemmelen and Berlage [25] assumed a thickness of 100 km and derived a viscosity of 1.3×10^{19} Pa s, whereas Haskell [26] in 1937 assumed an essentially unlimited thickness and obtained a viscosity of 3×10^{20} Pa s.

The first of the complicating observations was the discovery that some earthquakes occur down to depths of nearly 700 km [27]. The second was that even on the largest scale the earth is not quite in hydrostatic equilibrium. A completely hydrostatic earth should have an equatorial bulge due to rotation, but it was

found that the equator bulges by about 20 m more than this, and that the equator itself is not uniform, bulging more in some longitudes than in others. Significant stress is required to support these bulges.

Jeffreys [9], whose classic work demonstrated the existence of the bulges, argued that these and the deep earthquakes required the interior to have substantial strength, by which he meant that it could not be deformable over geological time scales. An alternative explanation of the excess bulges is that they are supported by stresses in a fluid mantle, which implies that the mantle would be in sustained internal motion. However this possibility does not seem to have been seriously advocated until it was taken up by Runcorn in 1962 [28]. In 1969 Goldreich and Toomre [29] argued further that the variations around the equator were not consistent with the previously preferred explanation that the 'equatorial bulge' was frozen in from times when the earth's rotation was faster, and they demonstrated that bulges generated by internal fluid motions would cause the earth to tilt so as to bring the largest bulges to the equator.

Jeffreys also argued that the approximate isostatic balance of mountain ranges was due to the fracturing of the crust by the tectonic forces, and subsequently by secondary gravitational (buoyancy) forces induced by the (supposed) resulting topography. He drew attention to the distinction between the strength of unfractured rock and the much lower strength of fractured rock. He supposed that it was the tectonic forces that first fractured the rock, and that the strength implied by remaining isostatic imbalance is a measure of the strength of fractured rock.

Daly ([14] p. 400) disputed Jeffreys on several grounds. He pointed out that Jeffreys' hypothesis could not account for slow isostatic adjustment away from mountain belts in response to erosion and sedimentation, nor for observed continuing adjustment to deglaciation. Daly also noted experiments by Bridgman that had shown that fractures healed quickly at high pressures. As well, we can note the internal contradiction in Jeffreys' argument that the remaining isostatic imbalance should still have reflected the strength of unfractured rock: any unfractured parts could still be out of equilibrium, and it would have been necessary to overcome the unfractured strength in order to bring them closer to balance.

Daly's ideas deserve more recognition. His thinking was wide-ranging and adventurous, and he came remarkably close to some modern concepts. The evidence at the time appeared contradictory, but I see Jeffreys' attempts to resolve the contradiction as limited and superficial in comparison with Daly's. Not all of Daly's ideas were well-based. For example he regarded the asthenosphere as being in a vitreous (glassy) rather than a crystalline state, despite his evident awareness of Airy's point that crystalline rocks were known to deform in deep mines, and despite his colleague Griggs' experiments on rock deformation [30]. He proposed to explain the large-scale bulges of the earth by supposing that below the asthenosphere is a 'mesosphere' of greater strength, though this neglects to explain how stresses maintained in a strong mesosphere would be transmitted through the asthenosphere to the surface.

Admitting that he was indulging in conjecture, Daly offered several suggestions to explain the occurrence of deep earthquakes. He proposed that the asthenosphere is heterogeneous, being strong enough to bear brittle fracture in some places. He proposed ways that this might come about, the most interesting being that blocks of lithosphere might founder and sink through the asthenosphere. Furthermore, noting that suddenly imposed stresses might induce fracture even in the deformable asthenosphere, he suggested that pressure-induced phase transformations, of the kind recently observed by Bridgman, in such sinking blocks might be a suitable trigger. This is an idea still very seriously entertained.

Daly proposed that the foundered or 'stoped' lithospheric blocks could plausibly originate during compressional mountain building:

> mountain making of the Alpine type seems necessarily accompanied by the diving of enormous masses of simatic, lithospheric rock into the asthenosphere. Thus the belt under the growing mountain chain is chilled by huge, downwardly-directed prongs of the lithosphere, as well as by down-stoped blocks. (p. 406.)

He noted that this would explain the occurrence of deep earthquakes 'under broad belts of recent, energetic orogeny'. This picture of the lithosphere, including 'prongs' projecting down under zones of compression, is remarkably close to the modern picture of a subduction zone, which we will get to later.

To summarise the evidence for a creeping mantle, gravity measurements established that mountains are close to an isostatic balance, and observations of associated sedimentary sequences showed that there are slow and continuous adjustments of the earth's surface to changing loads. Observations of post-glacial rebound of the earth's surface supported this inference and yielded quantitative estimates of the viscosity of the mantle. Observation of non-hydrostatic bulges were at first taken as evidence for a rigid interior, but were later reinterpreted as indicating a fluid interior

with a viscosity comparable to that inferred from post-glacial rebound. Deep earthquakes remained a puzzle, but Daly conjectured that the asthenosphere in which they occur is abnormal, and that the abnormalities might be associated with the active mountain belts that overlie them.

3.3 A mobile surface – re-emergence of the concept

Having set the scene for mobility in the earth's interior, I will now turn to the surface again, to describe how the surface came to be viewed as moving, the conception of moving rigid plates, and the strong evidence supporting this idea.

Although continental drift was not entirely ignored after about 1930 [10], it was certainly very unfashionable and was dismissed by many geologists, often with some passion. Against this, it was widely recognised that a really satisfactory theory of mountain building did not exist. The old idea of a contracting earth did not seem to provide for sufficient contraction to explain the observed crustal shortening, nor for zones of extension, without *ad hoc* elaborations of the theory. Expansion of the earth was proposed by a few people, and occasionally mantle convection was appealed to in contexts other than continental drift. Most geologists worked on narrower problems, and little progress was made on the question of fundamental mechanism, despite much conjecture [6].

This situation prevailed until about the mid-1950s, at which time two new kinds of evidence began to emerge that raised questions so serious they were harder to ignore. One kind of evidence was from palaeomagnetism, the other from exploration of the sea floor.

When a rock forms, it can record the direction of the local magnetic field, because any grains of magnetic minerals incorporated into the rock tend to align with the field like a compass needle. Collectively these grains then produce a small magnetic field that may be measurable in the laboratory. If a sufficiently large body of rock is magnetised in this way, the effect may be measurable in the (geological) field as a detectable perturbation of the earth's magnetic field.

Three distinct questions have been addressed through measurements of rock magnetism. First, have the rocks moved around on the earth's surface? Second, has the magnetic field changed through time? Third, can rock magnetism be used to map the sequence of formation of rocks, or to date their formation? The second and third questions will be discussed in Section 3.5.

The first question was pursued by British geophysicists in the 1950s, with a view to testing for continental drift. There were many complications to be dealt with, such as being sure that the original orientation of the rock could be reliably established and separating magnetisations acquired by the rock at different times through different microscopic mechanisms. There were also the possibilities that the magnetic field had not always been approximately aligned with the earth's spin axis, that it had not always been approximately dipolar, as at present, and that the earth had tilted relative to the spin axis.

By the late 1950s, these difficulties had been substantially overcome and strong evidence was emerging that North America and Europe had been closer together in the past [31], and that Australia had moved northward from near the south pole [32]. For those with knowledge of and confidence in the palaeomagnetic data, this was strong evidence that continental drift had occurred. However, the difficulties of the method were well-known, and it was hard for all but the minority involved in the measurements to know how much confidence to put in them. Nevertheless, these data were very influential in reïnstating continental drift as a respectable scientific topic.

The second important kind of evidence came from exploration of the sea floor, which increased greatly during and after World War II. An intimate and insightful account of this work was given by Menard [6]. One of the early and most startling discoveries was the absence of thick sediment on the sea floor. If the continents and oceans were permanent features, there should have been a continuous sedimentary record of most of earth history, but few rocks older than the Mesozoic were found on the sea floor, and those had affinities suggesting they are fragments of continents. Through the decade of the 1950s, the global extent of the 'midocean ridge' system was revealed, along with great 'fracture zones' on the sea floor. Fracture zones are narrow scars having the appearance of great faults thousands of kilometres long. Vast areas of the sea floor, where it was not covered with thin sediment, comprised monotonously rough 'abyssal hills' whose origin was unknown.

'Guyots' were found over a broad area of the central Pacific. Guyots are submarine mountains with flat tops. They were presumed to be of volcanic origin, and were and still are interpreted as former islands whose tops were eroded to sea level and which subsequently subsided below sea level. Both Hess [33, 34] and Menard [6, 35] inferred the former existence, about 100 Ma ago, of a midocean rise that has now subsided. Menard called it the

'Darwin Rise', in honour of Charles Darwin's correct explanation for the formation of coral atolls upon such drowned islands.

It was found that the heat flux conducted through the sea floor is as large or larger than on continents, despite the continental crust having a considerably higher content of radioactive heat sources.

Many of these discoveries were quite unexpected and difficult to make sense of. We must realise that the area being explored was vast, and that the picture was at first very patchy and incomplete. Nevertheless it was clear that old ideas had to be revised in major ways. The fracture zones are uniquely long and linear features, and it is hard to interpret them as anything other than strike-slip faults with large displacements, but they seem to disappear at continental slopes and have no obvious extension into the continents. The thin sediment covering on the sea floor required either that the rate of sedimentation had been very much less in the past than at present or that the sea floor is no more than about 200 Ma old. The abyssal hills topography looks chaotic, suggesting widespread tectonic disruption but the sediments overlying them on the older sea floor are largely flat-lying and undisturbed.

The relationship of fracture zones to midocean rises, if any, was unclear. In the north-east Pacific several major fracture zones connect to nothing obvious at either end. In the east they run up to the edge of the continent and appear to stop, while in the west they peter out. In the Atlantic, the rough topography and mostly east–west surveys left the picture confused, with Heezen [36] inferring that east–west troughs were part of a continuous graben on the ridge crest. Only later were they interpreted as fracture zones offsetting the ridge crest.

The origin of the midocean rise system was obscure. Where it was traced onto land in Iceland and East Africa, it was undergoing extension. This was consistent with the presence of an axial trough along much of the crest of the Mid-Atlantic Ridge, and Heezen inferred that the entire system of rises was extensional [36]. However, for some time seismic reflection data seemed to show a covering of sediment over the East Pacific Rise, and Menard inferred that it might be young and had not yet begun active rifting. Menard and Hess inferred that rises are ephemeral, and Menard proposed that the East Pacific Rise is young, that the Mid-Atlantic Ridge is mature, with active rifting, and the Darwin Rise is extinct.

Menard and Hess proposed variations on ephemeral convective upwellings to explain the existence of the rises. Heezen had traced the Mid-Atlantic Ridge around Africa and into the Indian Ocean and had inferred that it is all extensional. He reasoned from this that the earth had to be expanding, otherwise Africa would be undergoing active compression because of being squeezed from both sides by the extending ridges. The idea that the sea floor is or has been mobile was implicit in the interpretation of ridges and fracture zones. The uniformity of the sea floor and the absence of widespread evidence of deformation of sediments suggested that large areas of it were moving coherently. For example, Menard thought that the pieces between fracture zones moved independently, driven by separate convection 'cells'.

I recount these things to give some flavour of the ferment of ideas that was induced by the new kinds of observations. These were so puzzling, especially while they were incomplete, and sometimes misleading, that people were willing to appeal even to such disreputable ideas as mantle convection or earth expansion. Menard ([6] p. 132) makes the points, however, that most geologists at the time were busy with other things and unaware of or unconcerned with the sea floor, and that the oceanographers' research also was 'narrow, mostly marine geomorphology, but the areas were hemispheric and the conclusions correspondingly grand'.

That is the context in which two people proposed a third type of explanation for the midocean rises. Not earth expansion and not ephemeral convection cells, but continuous convection, coming right to the earth's surface at ridge crests and descending again at deep sea trenches. Hess wrote his paper in 1960, but it was not published until 1962 [34], while Dietz's paper was written and published in 1961 [37]. Menard argues persuasively that their work was independent ([6] Chapter 13).

Hess and Dietz accepted Heezen's arguments that the midocean ridges are extensional rifts, but they did not accept his conclusion that the earth expands. Hess had a long-standing interest in ocean trenches. Vening Meinesz [38] had measured gravity at sea in submarines, and found large negative gravity anomalies over trenches that he attributed to a down-buckling of the crust where two mantle convection currents converged. He developed a 'tectogene' theory that trenches were the early stages of geosynclines where thick sediments accumulated, later to be thrust upward in association with volcanic activity. Dietz also had an interest in geosynclines, arguing in later papers that they represent former passive continental margins that are activated by subduction. Thus both Hess and Dietz were disposed to the idea of crustal convergence and descending convection at trenches.

The central ideas that have survived from these papers are that convective upwelling of the mantle reaches the surface in a narrow rift at the crest of midocean rises and forms new sea floor. This then drifts away on both sides of the rift, ultimately to descend again into the earth at an ocean trench. A continent can be carried passively by the horizontal part of the convection flow, rather than having to plough through the sea floor, as supposed by Wegener. The youth of the sea floor and the thinness of sediments would be accounted for. A uniformly thick crust might be formed, if it is all formed by the same process at a ridge crest. Dietz recognised that the abyssal hills topography might also be a residue of rifting at the ridge. The high heat flow on ridges would be explained by the close approach of hot mantle to the surface. Dietz coined the concise term 'seafloor spreading'.

Not all of the ideas from Hess's paper have survived. For example, the composition of the oceanic crust was not definitely known at the time, and he supposed it to be serpentine (hydrated mantle peridotite), whereas Dietz more correctly assumed it to be basalt produced by melting the mantle under the ridge. Hess still thought ridges were ephemeral, being misled by the assumption that the Darwin Rise was of the same type as the modern midocean ridges. The Darwin Rise loomed large in Hess's thinking, because of his discovery of guyots. Some have viewed guyots as a key link to the idea of seafloor spreading, e.g. Cox [3], but I think they distracted him into thinking more about vertical motions than horizontal, and his thinking was still a bit confused in this paper. Hess did not think fracture zones were related to ridges. Dietz did, and he proposed that the convection proceeded at different rates on either side of a fracture zone, so the sea floor is displaced by different amounts.

Hess made another important, though somewhat separate point in his paper: that continents would be piled up by convection and also eroded down towards sea level. The consequence would be that the level of the continents would be near sea level, and this would be the result of a dynamic equilibrium between the piling up and the erosion. Thus he correctly recognised the explanation for the bimodal distribution of the elevation of the earth's surface that had been an important argument of Wegener's.

It may seem curious that Hess's and Dietz's papers became famous for proposing seafloor spreading, but not for the complementary removal of sea floor at trenches, which was an integral part of their concept. The reason is probably that the understanding of trenches and their associated mountains (island arcs or active continental margins) was in a state of confusion at the time, and as a result neither of them put much stress on what we now call subduction. Although there was a widespread concept that trenches were the sites of compression and some downward buckling (Vening Meinesz [38]) or faulting (Benioff [39]) the amount of crustal motion envisaged was usually limited. As well, attempts to determine the direction of slip in earthquakes from seismic waves were yielding confusing and inconsistent results. It was not until after a world-wide network of standardised seismographs was in place in about 1963 (to monitor underground nuclear explosions) that clear results of this type emerged. However the confusion did not hinder Wilson, as you will see in the next section.

This account of seafloor spreading has been expressed very much in terms of mantle convection, because that is how Hess and Dietz conceived it. You will see in the next section that there are advantages in looking just at the surface of the earth, without worrying about what is happening underneath. However, the question of how mantle convection relates to the surface becomes more acute as the surface picture is clarified. Already with seafloor spreading there is the novel idea that mantle convection rises, right to the earth's surface, but only in a very narrow rift zone at the crest of a midocean ridge. This is a novel form of convection. As you will see later, Holmes had proposed a picture rather similar to that of Hess and Dietz, even to the point of having a regenerating basaltic oceanic crust, but his concept was conditioned by conventional ideas about convection, and he supposed that seafloor extension occurred over a broad region.

3.4 Wilson's plates

J. Tuzo Wilson was a physicist turned geologist. He is best known for recognising a new class of faults, and for naming them 'transform faults', in a paper published in 1965 [40]. This paper is widely recognised as a key step towards the formulation of plate tectonics. It is more than that. It is in this paper that the concept of plate tectonics first appears in its complete form.

Wilson's paper is called 'A new class of faults and their bearing on continental drift'. It is worth quoting the opening of the paper.

> Many geologists [41] have maintained that movements of the earth's crust are concentrated in mobile belts, which may take the form of mountains, midocean ridges or major faults with large horizontal movements. These features and the seismic activity along them often appear to end abruptly, which is puzzling. The problem has been difficult to investigate because most terminations lie in ocean basins.
>
> This article suggests that these features are not isolated, that few come to dead ends, but that they are connected into a continuous network of mobile belts about the Earth which divide the surface into several large rigid plates [(Figure 3.1)]. . . .

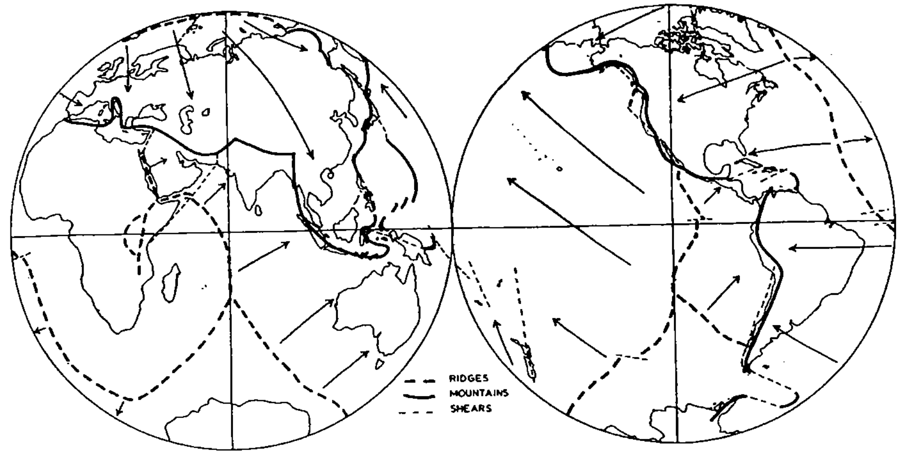

Figure 3.1. Wilson's sketch map from 1965 [40]. The original caption is as follows. 'Sketch map illustrating the present network of mobile belts around the globe. Such belts comprise the active primary mountains and island arcs in compression (solid lines), active transform faults in horizontal shear (light dashed lines), and active midocean ridges in tension (heavy dashed lines).' Reprinted from *Nature* with permission. Copyright Macmillan Magazines Ltd.

Others might have got tantalisingly close, but here, in four of the most pregnant sentences in all of geology, Wilson has defined the problem and presented its solution with simple clarity. His sketch map (Figure 3.1) gave the world its first view of the tectonic plates.

Wilson had very broad interests in geology, but he had been studying in particular large transcurrent faults. It was his recognition of the North American equivalent (the 'Cabot fault') of Scotland's Great Glen fault that first aroused his interest in continental drift [42]. He was also puzzled by the great fracture zones that were being discovered on the ocean floor, because they seemed to be transcurrent faults of large displacement, but they stopped at the continental margin, with no equivalent expression on the adjacent continent. He actually had not believed in continental drift until about 1960, but the publication of Dietz's seafloor spreading paper in 1961 convinced him that it must be right and he set about finding more evidence from the ages of oceanic islands (see Section 3.7).

Wilson's clinching insight was his recognition of the way these great faults can connect consistently with midocean ridges or with 'mountains' (meaning island arcs or subduction zones) if pieces of the crust are *moving relative to each other as rigid blocks* without having to conserve crust locally. Continuing the above quotation,

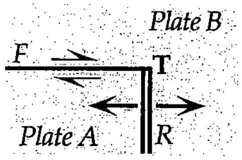

Figure 3.2. Illustration of Wilson's idea of how a transform fault (F) is transformed at the point T into a midocean ridge spreading centre (R). The concept depends on the pieces of crust on either side of F and R moving independently as rigid blocks or plates without requiring the area of each plate to be conserved locally.

Any feature at its apparent termination may be transformed into another feature of one of the other two types. For example, a fault may be transformed into a midocean ridge as illustrated in [Figure 3.2]. At the point of transformation the horizontal shear motion along the fault ends abruptly by being changed into an expanding tensional motion across the ridge or rift with a change in seismicity.

'... with a change in seismicity'? I'll return to that.

Wilson explains how his 'transform' faults may connect a ridge to a trench, or to another ridge segment, or may connect two trenches. He points out the crucial properties that transform faults may grow or shrink in length as a simple consequence of symmetric ridge spreading and asymmetric subduction, that the sense of motion on a transform fault joining two ridge segments is the reverse of the superficial appearance (Figure 3.3), and that the traces left by such faults beyond the ridge segments they connect are inactive. He does a fast tour of the world, explaining relationships between major structures, explicating what we now know as plate boundaries.

The language of the paper is terse. One senses the excitement of the rush of insights as pieces of a puzzle (literally) fall into place, and the desire to pack as much as possible into a short, crucial paper. Key information is almost lost. He forgets to spell out that it was known that the only seismically active parts of the great fracture zones cutting across the equatorial Atlantic sea floor are the parts between the ridge segments [43] (Figure 3.3), and that this was a major puzzle. That information appears only in the caption of his sketch map, and ambiguously, where he distinguishes active faults as solid lines and 'inactive traces' as dashed lines, without making clear that this had already been observed, and was not just a prediction of his theory. The cryptic 'with a change in seismicity' noted above means that the type of earthquake changes from strike-slip to normal faulting where a transform fault joins a ridge segment.

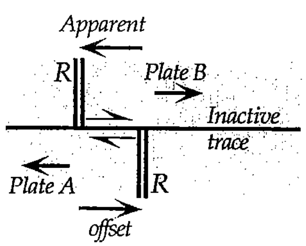

Figure 3.3. Illustration of the distinction between Wilson's 'transform fault' interpretation of ridge segments offset by a fracture zone and the 'transcurrent fault' interpretation. In Wilson's interpretation, the sense of motion across the active transform fault joining two ridge segments is right lateral in this example (looking across the fault, the other side moves to the right). In the transcurrent interpretation, the apparent offset is left lateral. Also in Wilson's interpretation the extensions of the fracture zone beyond the ridges are inactive scars within the plates, whereas in the transcurrent interpretation the extensions would also be active. Wilson noted the crucial observation that earthquakes occur mostly on the segment connecting the ridges, and only infrequently on the extensions.

Wilson was thinking as a structural geologist, and that was crucial. He envisaged rigid blocks bounded by three types of boundary that correspond to the three standard fault types: strike-slip (transform fault), normal (ridge) and reverse (subduction zone). Conceptually he narrowed the old notion of mobile belts down to sharp boundaries, and he explicitly adopted the long-standing implication of that old term, that there is little deformation outside the mobile belts, taking it conceptually to the limit of proposing that there is *no* deformation. He was explicit in the fourth sentence, quoted above, that the plates are 'rigid'. The point is explicit also within the paper: 'These proposals owe much to the ideas of S. W. Carey, but differ in that I suggest that the plates between the mobile belts are not readily deformed except at their edges.'

This was why Wilson was able to see the plates in all their simplicity. A unique and crucial feature of mantle convection, as distinct from other forms of convection, is that part of the medium behaves as a viscous fluid and part as a brittle solid, as I will explain in later chapters. This had been a source of confusion in attempts to formulate and relate ideas about continental drift, seafloor spreading and mantle convection. This can be seen for example by contrasting Holmes's concept of a new ocean in which there is broad deformation across the sea floor, reflecting the behaviour of a viscous fluid, with Hess's and Dietz's narrow spreading centres and the angular, segmented geometry of midocean ridges. By focussing on the motions that can be discerned at the surface, Wilson recognised the behaviour of a brittle solid, and successfully defined plate tectonics in those terms.

There can be no doubt that Wilson was aware of the implications of his new structural concepts for continental drift, justifying the second part of his title '... and their bearing on continental drift'. That is explicit in his explanation of the transform concept and in the last sentence of the paper where, referring to transform faults, he says 'proof of their existence would go far towards establishing the reality of continental drift and showing the nature of the displacements involved.' Perhaps too modestly, he implies here that he has not already pointed out compelling evidence, in the form distribution of earthquakes on fracture zones and implicitly in the wealth of geological and seismological evidence that had given rise to the concept of mobile belts and the complementary idea of internally stable blocks.

Contrast Wilson's paper with a little-known paper by Coode [44], also published in 1965. In this very brief note, Coode elegantly presents the conception of a ridge-ridge transform fault, along with a diagram explaining how both the ridge crest and magnetic anomalies (next section) are offset. That is all Coode does. The further implications are not developed. The paper was almost unknown until it was pointed out by Menard [6] (though this was also because it was in a journal where oceanographers were unlikely to see it).

There can be no doubt also that Wilson appreciated that he had taken a major step towards a unifying dynamic theory of the earth that would probably involve mantle convection. Two years earlier he had published several papers containing the fruits of another remarkable burst of creativity, including the seminal insight that led to the idea of mantle plumes [45, 46], and a wide-ranging article in *Scientific American* on continental drift [42]. In the latter it is clear that he has a comprehensive grasp not only of a large number of geological observations but also of the arguments from isostasy, post-glacial rebound, materials science and gravity observations over ocean trenches that the mantle is deformable and undergoing convection. His map of convection currents bears a strong resemblance to his 1965 map of the plates, and he writes of moving crustal blocks.

Reading the 1965 paper, we may see a structural geologist presenting a brilliant and novel synthesis. Reading it in conjunction with the 1963 papers, we see more: a scientist in the full pursuit of the secrets of the earth, chasing whatever kind of evidence will serve. Reading them all, I see a man move, in little more than five years, from first conversion to mobilism through to clarity of understanding of geology's major unifying concept.

In frankly championing Wilson, I do not wish to detract from the contributions of many others. I just think that his grasp of what he was doing has not been fully appreciated, perhaps because his 1965 paper is so terse, concentrates necessarily on the novel technicalities of transforms, and its title does not fully portray the unity and simplicity of his concept. I think he deserves a special place in the pantheon of geology for being the first to see the plates in complete and simple form.

I will explain further what I mean, so as to avoid unnecessary confusion. I take the essence of plate tectonics to be the concept of rigid, moving pieces of the earth's surface meeting at three kinds of boundary. I distinguish this two-dimensional concept, which can be displayed on a map, from the three-dimensional concepts of thick lithosphere and of mantle convection; these have continued to be debated and refined without detracting from the plate concept. Wilson's concept was not confined to planar geometry. Although Wilson sketched the transform concept in planar maps, that idea transfers completely to a sphere, because it involves the relationships of boundaries meeting at a point (Figure 3.2). There is no doubt that Wilson was thinking of rigid plates on a sphere. I also distinguish the *concept* from its quantitative, mathematical *description*. The idea of using Euler's theorem of rotation to describe the motions of plates on a sphere (Section 3.6) was powerful and productive, but it was a quantification of the *pre-existing* idea of moving, rigid, spherical plates.

Comparing the plate-tectonic revolution to the Copernican revolution in his preface to a collection of *Scientific American* articles [47], Wilson made the following observation.

> That the earth is the centre of the universe and that it rests on a fixed support was the obvious and early interpretation. To realise that the earth is spinning freely in space and that the sun, and not the earth, is at the focus of the solar system required a prodigious feat of imagination.... Changing the basic point of view created a new form of science with a different frame of reference. It was this change in the manner of interpreting the observations that constituted the scientific revolution.

Though others were close, both before and after, I think Wilson was the first to complete the change in point of view. Once Wilson had stood upon the far shore, it was easier for others, knowing it was there, to follow.

3.5 Strong evidence for plates in motion

3.5.1 Magnetism

About 1960 studies of palaeomagnetism began to focus strongly on the second question posed earlier (Section 3.3): has the earth's magnetic field changed through time? Specifically, has it reversed polarity? Matuyama, in 1929 [48], had studied the magnetisation through a sequence of lava flows erupted by a Japanese volcano. He found that the younger flows near the top were magnetised parallel to the present earth's magnetic field lines, but that the older flows near the bottom of the sequence were magnetised in the opposite direction. During the 1950s the question of whether this was due to reversal of the earth's field or to a peculiar response of some rocks was vigorously debated [3]. It seemed that there may have been many reversals of the earth's field, but this was difficult to demonstrate convincingly. From 1963 two groups in particular used a combination of magnetisation measurements and potassium–argon dating to try to resolve the question and to establish a chronology of reversals. These groups were at the U.S. Geological Survey in Menlo Park, California [49], and at the Australian National University in Canberra [50]. They found that the ages of normally and reversely magnetised rocks correlated around the world, which supports the idea that the earth's field had indeed reversed. By about 1969, the time sequence of reversals was established with some detail to an age of about 4.5 Ma, beyond which the K–Ar dating method did not have sufficient accuracy [51].

Meanwhile Ron Mason, of Imperial College, London and the Scripps Institute of Oceanography in California, was trying to identify magnetic reversals in oceanic sedimentary sequences. Because of this work, but still almost by chance, a magnetometer was towed behind a ship doing a detailed bathymetric survey off the west coast of the U.S. ([6], p. 72). From this magnetic survey there emerged a striking and puzzling pattern of variations in magnetic intensity: alternating strips of the sea floor had stronger and weaker magnetic field strengths [52]. The pattern was parallel to the local fabric of seafloor topography, and later was found to be offset by fracture zones, by about 1000 km in the case of the Mendocino fracture zone. It was presumed that the pattern might be explained by strips of sea floor with differing magnetisations, but its origin was obscure. In retrospect it was unfortunate that in this area the ocean spreading centre at which the sea floor formed no longer exists, and so there was no obvious association with midocean ridges.

In subsequent surveys, elsewhere, it was found that ridge crests have a positive magnetic anomaly (meaning merely that the field strength is greater than average), which some people presumed to indicate 'normal' (i.e. not reversed) magnetisation. Beyond this, on either side of the ridge crest, there was a negative (i.e. weaker than average) anomaly. In 1963 Fred Vine, then a graduate student at Cambridge University, was analysing the results of one such survey over the Carlsberg Ridge in the Indian Ocean. He noticed that the seamounts near the ridge crest were reversely magnetised. This is easier to infer for seamounts, because they are more like point sources and produce a more distinctive three-dimensional pattern of anomalies, whereas a long strip of sea floor produces a two-dimensional pattern that is more ambiguous. While his supervisor Drummond Matthews, who had collected the data, was away, Vine conceived an explanation for the magnetic stripes ([6], p. 219).

Lawrence Morley in Canada was involved in aeromagnetic surveys over Canada, and was familiar with many aspects of geomagnetism. In his seafloor spreading paper, Dietz had commented on how the magnetic stripes off the western U.S. seemed to run under the continental slope, and had suggested that they were carried under the continent by subduction and destroyed by subsequent heating. It is well known that magnetisation does not survive if rocks are heated. Conversely, it is reacquired by magnetic materials upon cooling. Morley realised that the oceanic crust could be magnetised as it formed and cooled at a spreading ridge ([6], p. 217).

What has become known as the Vine–Matthews–Morley hypothesis combines the hypotheses of seafloor spreading and magnetic field reversals. The idea is that oceanic crust becomes magnetised as it forms at a spreading centre, and a strip of sea floor accumulates that records the current magnetic field direction (Figure 3.4a). If the magnetic field then reverses and the seafloor spreading continues, a new strip will form in the middle of the old strip (Figure 3.4b), the two parts of the old strip being carried away from the ridge crest on either side. Subsequent reversals would build up a pattern of normal and reverse strips, and the pattern would be symmetric about the ridge crest (Figure 3.4c).

Vine later commented that the hypothesis required three assumptions each of which was, at the time, highly controversial: seafloor spreading, magnetic field reversals, and that the oceanic crust (the seismic 'second layer') was basalt and not consolidated sediment ([6], p. 220). Morley submitted a paper about the beginning of 1963, which was rejected by two journals in succession, the second with unflattering comments. Vine and Matthews submitted

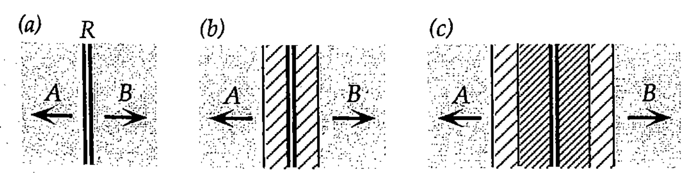

Figure 3.4. Illustration in map view of how seafloor spreading and magnetic field reversals combine to yield strips of sea floor that are alternately normally and reversely magnetised. The resulting pattern is symmetric about the crest of the ridge if the spreading itself is symmetric (meaning that equal amounts of new sea floor are added to each plate).

a paper in about July 1963, which was published in September [53]. Morley's story emerged later [3, 5, 6].

Subsequent exploration revealed extensive patterns of magnetic stripes on the sea floor, with an astonishing degree of symmetry about ridge crests (Figure 3.5), and which correlated with the field reversal chronology established on land [54, 55]. These magnetic stripes provided strong and startling evidence in favour of seafloor spreading. They also opened the prospect of assigning ages to vast areas of the sea floor on the basis of the reversal sequence, which was rapidly correlated from ocean to ocean [56]. Thus was the third question addressed through rock magnetism (Section 3.3) answered with a resounding yes: rocks can be dated using rock magnetism.

We should reflect on the magnitude of that last paragraph. Assigning ages to rocks always has been and still is a central occupation of geologists. It is painstaking work, whether the method is correlation of fossils or measurement of radioactive decay. It has taken much of this century to develop the ability to get reliable ages accurate to within a small percentage or less for many kinds of rocks. As Menard remarked ([6], p. 212)

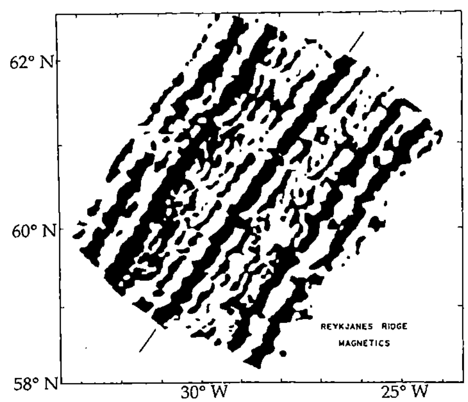

Figure 3.5. The pattern of magnetic anomalies across the Mid-Atlantic Ridge south of Iceland, where it is known as the Reykjanes Ridge. Black indicates a positive anomaly, inferred to be due to normally magnetised crust, and white indicates a negative anomaly, inferred to be due to reversely magnetised crust. The short lines mark the location of the Ridge crest, along which there is a positive anomaly. Despite the irregularities, the pattern shows a striking symmetry about the Ridge crest. From Heirzler *et al.* [57]. Copyright by Elsevier Science. Reprinted with permission.

To general astonishment, magnetic reversals provide the long-sought global stratigraphic markers that are revolutionising most of geology. At sea, as though by a miracle, magnetic anomalies give the age of the sea floor without even collecting a sample of rock.

3.5.2 Seismology

Seismology had already provided a key piece of evidence even before Wilson and Coode conceived of transform faults, as will be reiterated shortly. The Lamont (now Lamont–Doherty) Geological Observatory of Columbia University in New York state, directed by Maurice Ewing, had pioneered the exploration of the Atlantic sea floor, and then of other oceans. After Dietz's paper on seafloor spreading, Ewing turned much of the effort to testing the hypothesis. Part of this programme was to study the earthquakes in oceanic regions, and it was already known that these occur mainly on midocean ridges. By 1963 there was a better distribution of modern seismographs around the world, including the World-Wide Standardised Seismograph Network already mentioned. This permitted earthquakes in remote regions to be located with an accuracy about ten times better than previously.

Lynn Sykes, working at Lamont, found that the earthquakes are located within a very narrow zone along the crests of midocean ridges, and along the joining segments of fracture zones, where these were known or could be inferred [43, 58]. He made the explicit point that earthquakes on fracture zones occur predominantly on the segments that connect segments of ridge crest, and hardly at all on segments beyond ridge crests (Figure 3.6). This had been very puzzling when it was thought that fracture zones had offset ridges by motion along the length of the fracture zone. However it was explicitly predicted by the transform fault concept, and was noted (barely) by Wilson as evidence in its favour (Figure 3.3).

When Sykes saw the evidence of his colleagues for symmetric magnetic anomalies, he was convinced of seafloor spreading, but realised that he could make another decisive test through seismology. The elastic waves emitted by an earthquake have a distinctive four-lobed pattern. In two opposite lobes, the waves that arrive first are compressional. In the intervening two lobes the 'first arrivals' are dilatational. These waves spread through the earth's interior in all directions. With a global distribution of seismographs, it is possible to sample these waves with sufficient density to reconstruct the orientation of the lobed 'radiation pattern' and the orientation of its two 'nodal planes'. One of these planes corresponds to the fault plane, and the other is perpendicular, though you can not tell

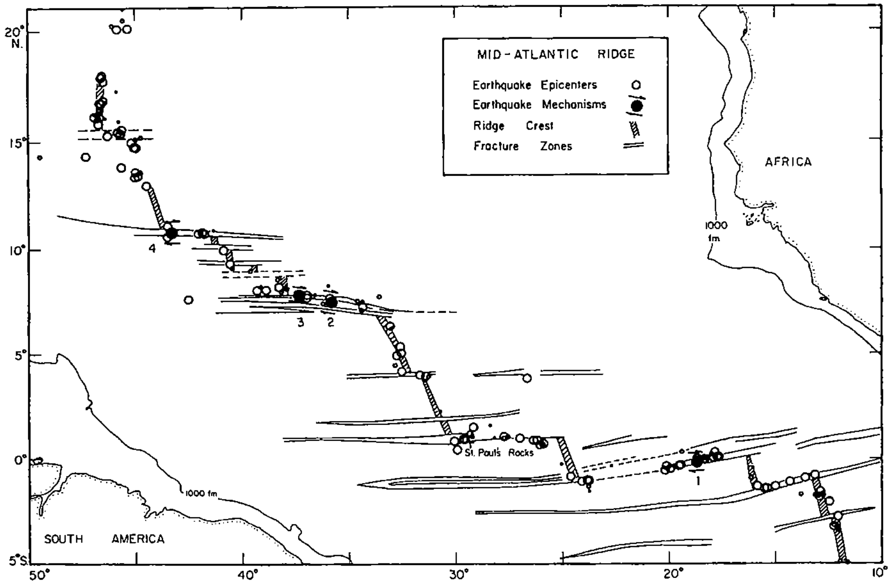

Figure 3.6. Earthquakes along the Mid-Atlantic Ridge. Open symbols show locations (epicentres) of earthquakes, and solid symbols with arrows show the sense of slip inferred from fault plane solutions. The fracture zones (oriented east–west) have earthquakes mainly between segments of the ridge crest and only rarely on the extensions beyond ridge crests. These locations and the sense of slip on the active segments are consistent with Wilson's transform fault hypothesis. From Sykes [59]. Copyright by the American Geophysical Union.

which is which just from the seismic waves. It is also possible to infer directly the orientation of stresses at the earthquake source. The result of this determination was called a 'fault plane solution'.

Sykes knew that some previous fault plane solutions on ridges were suggestive, but that he could get much more reliable results from the new global seismographic network. This he did [59]. He found that for earthquakes located on segments of ridge crest, the solutions indicated normal faulting, consistent with the ridge crest being extensional. Earthquakes located on active segments of fracture zones had one nodal plane approximately parallel to the fracture zone, consistent with strike-slip faulting (Figure 3.6). Most importantly, the sense of strike-slip motion was consistent with that predicted by the transform fault hypothesis, and opposite to that predicted by the simple transcurrent offset interpretation. This was another kind of observation strongly supportive of seafloor spreading.

3.5.3 Sediments

Ewing, during the same period, had used seismic refraction to determine the thickness of sediments in the Atlantic. If seafloor spreading were occurring, the thickness of sediments should increase with distance from the ridge crest. The results were confusing [60], partly because of the rough seafloor topography of the Atlantic, and Ewing was reluctant to come out in support of seafloor spreading.

Later a different approach became possible through a deep-sea drilling programme, which allowed the recovery of long sediment cores. An early cruise in the South Atlantic Ocean was aimed specifically at testing seafloor spreading. The results were spectacular [61]. It was found that the age of the oldest sediment, just above the basaltic basement, determined from micro fossils, increased in simple proportion to distance from the ridge crest, exactly as predicted by assuming seafloor spreading at a nearly constant rate (Figure 3.7). The results also provided an important calibration of the magnetic reversal chronology, which until then was well-calibrated only for the first few million years.

Menard and others have remarked that most scientists are converted to a new idea by observations from within their own speciality. Thus palaeomagnetic polar wandering converted a small minority of geophysicists to continental drift. Later the dramatic evidence of seafloor magnetic stripes, earthquake distributions and fault plane solutions converted a majority of geophysicists to seafloor spreading. To many of the more traditional geologists, however, such geophysical observations were still unfamiliar, and they were unsure how to regard them. However, fossil ages are a long-standing concept in geology, and something most geologists can readily relate to. Thus, although the deep-sea sediment ages were not published until 1970, they were important for spreading the word to the great majority of geologists who work on continental geology.

This completes my short survey of some of the most direct and compelling evidence that led to the acceptance of plate tectonics by a majority of geologists. There is much other evidence and there were many more players, but a knowledge of these observations suffices to place plate tectonic motions on a firm empirical footing.

3.6 Completing the picture – poles and trenches

With compelling evidence for seafloor spreading and strong evidence for the rigidity of plates from the lack of deformation of

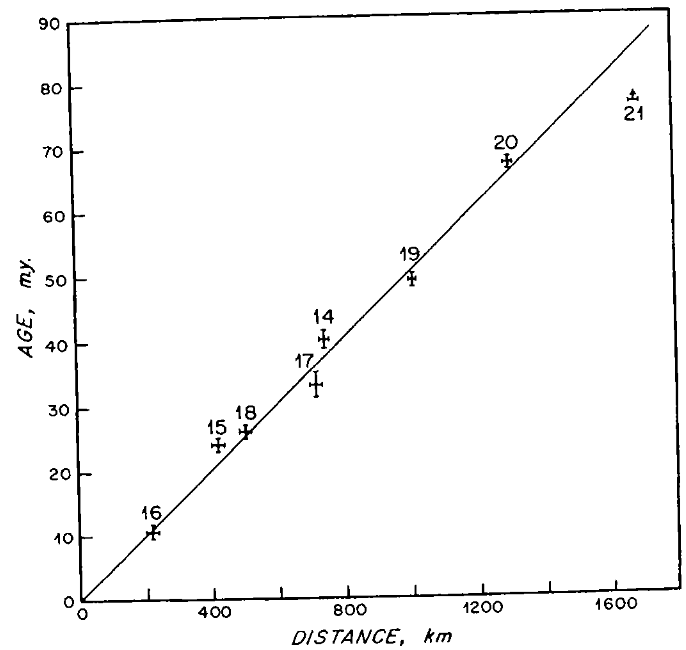

Figure 3.7. Ages of sediments immediately above the basaltic basement of the sea floor of the South Atlantic, plotted against distance from the crest of the Mid-Atlantic Ridge. The ages are inferred from micro fossils. From Maxwell *et al.* [61]. Copyright American Association for the Advancement of Science. Reprinted with permission.

sediments and magnetic anomalies over large areas of sea floor, Wilson's concept of moving rigid plates was strongly supported. However the focus had been mainly on ridges and seafloor spreading. There remained the untidiness at the other end of the conveyor belt: the ocean trenches. Also, would Wilson's tightly integrated global concept withstand further close quantitative examination? Were such quantitative tests possible? Was there, perhaps, some possibility that earth expansion had not been excluded?

3.6.1 Euler rotations

Two people independently conceived that there is a simple way to describe quantitatively the motions of rigid plates on a sphere: Jason Morgan and Dan McKenzie. The idea had been used in 1965 by E. C. (Teddy) Bullard and others at Cambridge to test

the idea that continental outlines on either side of the Atlantic matched [62], but it was originally due to the mathematician Euler. Euler's theorem is that any motion of a rigid piece of a sphere over the surface of the sphere can be described as a rotation about some axis through the centre of the sphere. The intersections of the axis with the surface are called poles of rotation. Menard coined the terms Euler pole and Euler latitude ([6], p. 324). An Euler latitude is analogous to geographic lines of latitude relative to the earth's geographic north pole. Lines of latitude are 'small circles' on a sphere (that is, the intersection of a plane with the surface of the sphere), except for the equator, which is a 'great circle' (that is, the intersection of a plane passing through the centre of the sphere with the surface of the sphere). Morgan and McKenzie each realised that in Wilson's theory plate motions should be described by rotations about Euler poles.

We must stop, at this point, and think about what plates are moving relative to. It is natural at first to think that there is some part of the earth's interior that is not moving, and to think of the plates as moving relative to that region. The problem is that we can't see into the earth very clearly, to identify such a region. In fact there is no evidence for such a region. Everything – crust, plates, mantle and core – seems to be in relative motion. It is true that the volcanic hotspots (discussed in the next section) seem to be moving only slowly relative to each other, and it has been useful to assume them to be 'fixed' for some purposes, but there is no reason to think they are not moving at all.

The better way to approach this problem is to realise that it is not necessary to think about the interior at all. The plate motions can be described entirely in terms of *relative* motions, that is of motions relative to each other, without reference to some internal, external or 'absolute' frame. The most convenient approach is to describe the relative motion of plates in adjoining pairs. This relative motion determines the nature of their interaction – whether they are pulling apart (at a ridge or 'spreading centre'), pushing together (at a trench or 'subduction zone') or sliding past each other (along a 'transform fault').

Extending this point, it is best to stop thinking about convection, or driving forces, and to realise that we are doing geometry. Strictly speaking, it is kinematics, the description of motion, without reference to causes or forces. If you like, it is the moving geometry of the earth's surface. Wilson succeeded, in his formulation of plates, also by ignoring convection and looking only at the surface.

In 1967, Menard published a short paper describing further mapping of the great fracture zones of the north-east Pacific [63]. In one illustration in that paper he used a great circle projection in order to show that the fracture zones are remarkably straight (a great circle being the nearest thing to a straight line on a sphere – it is the shortest distance between two points). Morgan was struck that the fracture zones could be more accurately represented as a series of small circles and that the radii of the small-circle segments increased from north to south in passing from one fracture zone to the next ([6], p. 285). In fact they could be thought of as sets of Euler latitudes. Combining Euler's theorem with Wilson's transform fault concept, he realised that the sets of small-circle fracture zones represented traces on a sphere of transform faults offsetting a former ridge that must have existed there. (Morgan may not have realised at this stage that a third plate, the Farallon plate, must have existed between the Pacific and North American plates.) Morgan then analysed the fracture zones along several presently active spreading centres and showed that they were consistent with Wilson's concept of transform faults between rigid plates moving on the spherical surface of the earth [64].

Meanwhile McKenzie also was struck, while reading the Bullard *et al.* paper [62], that relative plate motions could be represented in terms of rotations. His approach was to test present motions, as determined from earthquake fault plane solutions in the manner of Sykes. His colleague and co-author Robert Parker introduced the idea of using a Mercator projection oriented relative to the Euler pole. In this projection, earthquake slip directions (and fracture zones) would plot as horizontal lines (like lines of latitude). In this manner, they demonstrated that motions over a huge area, from the Gulf of California across the North Pacific and to Japan, were consistent with the motion of a single rigid Pacific plate relative to adjacent areas [65]. (In fact Japan is part of the Eurasian plate, but its motion is slow relative the North American plate, so this didn't affect the results much.)

McKenzie and Morgan, in these papers and in a subsequent joint publication [66], developed the ideas of rotation vectors, local velocity vectors and triple junctions. Triple junctions are points where three plates meet, and their evolution can be deduced from the behaviour of the types of plate boundary involved. This will be taken up in Chapter 9. The important point here is that the sizes and shapes of plates can evolve just by the way they are created and removed at ridges and trenches, without any changes in the motion of plates. With an understanding of this, it is possible to reconstruct the past evolution of plates.

In this way, McKenzie and Morgan were able to develop the suggestion by McKenzie and Parker that there had been another plate (the Farallon plate) between the Pacific and North American plates that has been consumed by subduction. The great fracture zones of the north-east Pacific discovered by Menard were formed at a ridge between the Pacific and Farallon plates. That ridge was subsequently replaced by the San Andreas transform fault, which represents motion between the Pacific and North American plates. Finally the mystery of these 'dangling' fracture zones was solved.

With the concept of Euler rotations, the quantitative descriptions of plate motions became possible, and it was demonstrated that indeed much of the earth's surface comprises rigid moving plates. There are some areas where large-scale rigidity does not apply, mainly where plate boundaries enter continents (for example, central Asia and western North America).

3.6.2 Subduction zones

As I mentioned in Section 3.3, the deep ocean trenches and their associated deep earthquake zones were the subjects of controversy until late in the development of plate tectonics. The trenches were found to have large negative gravity anomalies by Vening Meinesz [38], and he and others developed the idea that they are the sites of crustal compression and possibly of convective downwelling in the mantle (see Menard [6]; Chapter 10). Wadati [67] and Benioff [39] mapped the deep seismic zones. Benioff interpreted them as tracing great reverse (thrust) faults, but with limited displacements. Attempts to determine the sense of motion from fault plane solutions generated confusion because of the difficulty of obtaining consistent high-quality readings of seismograms from a sufficient global distribution of instruments. Japanese seismologists [68] detected a zone of relatively high attenuation (suggesting higher temperatures) above the deep seismic zone. This indicated some kind of spatial heterogeneity, as had been conjectured by Daly to reconcile the occurrence of deep earthquakes with strong independent evidence for an asthenosphere.

Clarity only began to emerge with the advent of a world-wide network of seismographs in the early 1960s, mentioned earlier, and with the installation by Lamont seismologists of several seismographs in the Tonga–Kermadec region north of New Zealand, this being the site of the most active deep earthquake zone. What emerged first was a higher resolution map of the Tonga–Kermadec zone, showing among other things that the seismic zone is quite thin, defining a surface dipping to the west away from the trench, and that at the northern end the trench, the volcanic line and the deep earthquakes all curve sharply to the west, suggesting an intimate association between them [69]. Second, it was found that the attenuation of seismic waves was greater both above and below the deep seismic zone [70]. This defined a tongue of mantle of anomalously low attenuation, suggesting low temperatures, about 100 km thick and including the seismic zone, which was continuous from the surface to the deepest earthquakes.

This was a spectacular confirmation of Daly's 'prongs' of lithosphere sticking down from the surface. In fact it is revealing to quote Oliver and Isacks [70] in the light of Daly's hypothesis that the asthenosphere may be laterally heterogeneous (Section 3.2).

> In retrospect, it appears quite reasonable that the zones of deep shocks should be anomalous if only because earthquakes occur there and not elsewhere. Yet most models of the earth's interior include a mantle without lateral variation, and, except in one or two cases, the models that do take into account lateral variation have not associated such variation with the entire zone of deep earthquakes. In general, in hypotheses relating to the mechanism of deep earthquakes the emphasis has been on process alone, whereas it should be on both process and the nature of the material.

Subsequent work showed that shallow earthquakes have predominantly thrust mechanisms [71], confirming the conjectures of Holmes, Vening Meinesz, Hess, Deitz, Wilson and many others that these are zones of horizontal convergence. All the findings together painted a clear picture of surface lithosphere turning and descending under the trench and island arc [71, 72].

I indulge here in a personal footnote that carries a small lesson and an indication of my own perspective. In September of 1968 I travelled from Australia to begin graduate work at the California Institute of Technology. My first publication in geophysics, in 1971 with seismologist Jim Brune, was an estimate of the rate of convergence in subduction zones using a catalogue of this century's earthquakes and magnitudes [73]. The results were of the right order (centimetres per year) predicted by plate tectonics. At the time it was clear that there were uncertainties of up to a factor of two, and subsequent work revealed that the largest earthquakes were not adequately measured by the old magnitude determinations. Nevertheless it was for some a satisfying closing of the circle that subduction was occurring at about the same rate as seafloor spreading, and no expansion of the earth was called for. To me the conclusion was no big deal: I was new to the field and it seemed quite obvious to me that plate tectonics was correct.

Had that study been done ten years earlier, it might have accelerated the formulation of plate tectonics, balancing somewhat the concentration on ridges and seafloor spreading. Had it been done twenty or thirty years earlier, it might have crystallised some of the ideas of the time into a rudimentary form of plate tectonics (see Section 3.9). I didn't know until recently that Charles Darwin had pioneered the approach 135 years earlier.

A great deal else has also been deciphered from the record of the seafloor magnetic anomalies and the rules of plate motion. It is not necessary to recount them all here. Chapter 9 explains the principles and principal consequences of plate evolution, with some representative examples. The plates and their motion are a principal boundary condition on mantle convection, and this aspect will be taken up in Part 3.

3.7 Plumes

Perhaps Charles Darwin's best-known contribution to geology is his theory of the formation of coral atolls. He proposed that the various forms of coral reefs could be arranged in a sequence (island and fringing reef, island and barrier reef, atoll), and that the sequence made sense if islands were progressively eroded and submerged, with the coral reef growing as the island shrank away. Darwin did not actually observe his proposed sequence in a contiguous set of islands. Dana, on the U.S. Exploring expedition in 1838–1842, did find Darwin's proposed sequence in place in several linear island chains, and extended the sequence to include the initial active volcano that formed the island ([6], p. 195). Dana observed the sequence in the Society, Samoan and Hawaiian islands, and correctly inferred that the Hawaiian and Society island chains age to the north-west.

When Wilson read Dietz's seafloor spreading paper, he thought of using ocean islands as probes of the sea floor, reasoning that islands should be progressively older at greater distances from spreading centres [46]. The idea was good, but the data were scattered and somewhat misleading, since some islands include fragments of continental crust, and other ages were not representative of the main phase of island formation. Even an accurate age for the main phase of building an island gives only a lower bound on the age of the sea floor upon which it is built. The sea floor will have the same age as the island if the island formed at a spreading centre, but it will be older if the island formed away from the spreading centre.

Despite the limited data available at the time, two clear ideas emerged from Wilson's work. One was that some 'lateral ridges' could be explained if they represent the traces of extra volcanism at a spreading centre. In fact, such volcanism might produce a complementary pair of ridges, one on each plate moving away from the spreading centre. Wilson cited the Rio Grande Ridge and the Walvis Ridge in the South Atlantic as an example of such a pair, the active volcanism of Tristan da Cunha being the current site of generation. A closely related idea had been proposed by Carey in 1958 [13], and Wilson has acknowledged his debt to Carey [40].

The second idea was a mechanism to explain age progressions in island chains. Wilson recognised that there is active volcanism on some islands that are located well away from spreading centres, so that some islands clearly had not formed at a spreading centre, the Hawaiian islands being an outstanding example. With the idea of seafloor moving sideways, he realised that Dana's inferred age sequence for the Hawaiian islands could be produced if there was a (relatively) stationary source of volcanism deep in the mantle that had generated the islands successively as the seafloor passed over [45]. He conjectured that this 'hotspot' source might be located near the slowly moving centre of a convection 'cell'.

In 1971 Morgan [74, 75] developed this idea by proposing that there are plumes of hot material rising from the lower mantle. His proposal actually had three parts: that the island volcanism is produced by a plume rising through the mantle, that the plume comes from the lower mantle, and that plume flow drives the plates. He also presented reconstructions of plates to argue that the volcanic centres are relatively fixed, meaning that they have low horizontal velocities relative to each other. It was a common assumption at the time that the lower mantle, or 'mesosphere', has a very high viscosity and does not partake in convection (e.g. [71]), and so presumably the assumption that plumes come from the lower mantle was a way of accounting for their slow motions. Morgan devoted most space to demonstrating hotspot 'fixity' and to arguing that plate motions are driven by plumes. Hotspot fixity has been a useful approximation that has helped to refine details of plate motions. Plumes as the primary means of driving plates received little support, and I think it is not viable, as will become clear in Part 3.

I abandon here Wilson's original meaning of the term hotspot, since his hypothesis has been superseded by Morgan's plume hypothesis. Wilson's 'hotspot' was a small, hot volume in the mantle of unspecified origin. I think it is more useful to have a term for the surface manifestation from which plumes are inferred, and that a suitable term is 'volcanic hotspot', or 'hotspot' for short. I will therefore use these terms to refer to a volcanic centre *on the earth's surface* that has the characteristics that have come to be associated with (Wilson's) hotspots and plumes: persistent volcanism in a location that is relatively independent of plate motions and moves only slowly relative to other hotspots, often with an associated topographic swell.

The reality of plumes as a source of island volcanism became commonly accepted, though not without debate. Morgan noted that volcanic hotspots often have a topographic swell associated with them, and this observation was documented more thoroughly by Crough [76]. The details of how these swells are generated has been the subject of a confused debate. This will be taken up in Chapters 11 and 12.

The concept of plumes developed only slowly as a physical and quantitative theory, with the unfortunate result that plumes came to be invoked often in very *ad hoc* ways to explain a wide range of geological observations throughout earth history and on other planets. Even when important developments in the understanding of plumes occurred, the implications were frequently overlooked.

In 1975 Whitehead and Luther [77] reported laboratory experiments that showed that the viscosity of the plume fluid has a strong effect on the form of a newly forming plume. If the plume is of higher viscosity than the surroundings, it rises as a finger. If it is of lower viscosity, it rises in a 'mushroom' or 'head and tail' form: a large spherical 'head' preceeding a narrower conduit or 'tail' up which fluid continues to flow from the source.

Morgan, in 1981 [78], pointed out that a number of 'hotspot tracks' (the volcanic chain produced on a plate as it passes over a plume) originate in flood basalt provinces. Flood basalts are the largest known volcanic eruptions in the geological record, and typically comprise basalts of the order of 1 km thick over an area up to 2000 km across. Morgan proposed that this association could be explained if the flood basalt was produced from a plume head arriving at the base of the lithosphere and the hotspot track was produced by the following plume tail.

Loper and Stacey in 1983 [79, 80] developed the quantitative theory of flow in a thermal plume tail for the case when the viscosity of the material is strongly temperature-dependent. In this case, the plume material has a low viscosity because it is hot, and the plume tail can be quite narrow, of the order of 100 km in diameter. Loper and Stacey developed the analogous theory for a hot thermal boundary layer, from which the plume was assumed to grow. Olson and Singer [81] quantified the growth and ascent of plume heads in the case where they are compositionally distinct, and some aspects of plume tail behaviour in the presence of horizontal shear flow in the surrounding fluid.

Griffiths and Campbell in 1990 [82] presented a physical theory of thermal plume heads and tails, confirmed and calibrated by laboratory experiments. They demonstrated an important distinction between compositional plume heads and thermal plume heads. In the latter, a boundary layer of adjacent material is heated by conduction, becomes buoyant, and then rises with and is entrained into the plume head. The result in the mantle can be that the plume head reaches a diameter of about 1000 km, two to three times larger than a compositional, non-entraining, plume head.

Morgan's idea that flood basalts are produced by plume heads was revived by Richards, Duncan and Courtillot in 1989 [83], with more information on hotspot track ages. Campbell and Griffiths [84] developed this hypothesis further in 1990, arguing that first-order features of flood basalts (size, temperature, composition) could be accounted for by thermal plume heads rising from the base of the mantle.

By this stage plumes were well quantified and their physics quite well understood, both for compositional and thermal plumes, and quantitative predictions were being made and tested. Many details are still debated, but a basic theory is in place and there is much observational support for the broad concept. The physical theory of plumes will be developed in some detail in Part 3.

3.8 Mantle convection

We have looked at evidence for drifting continents and for moving plates, evidence for a deformable mantle, evidence for mantle plumes, and at the development of these concepts over the past century or more. The idea of mantle convection, which arises from the convergence of these other concepts, also goes well back into the nineteenth century. Here I briefly recount the development of ideas of mantle convection that precede the conception that will be developed in Part 3.

As I have indicated, ideas about mantle convection were at times intimately linked with ideas of continental drift and the emerging idea of plate tectonics. So they should be, but the source of the convection and its relationship to surface tectonics were for a very long time unclear and puzzling. The convection was usually assumed to have a particular form, like that of the classical Benard convection [85], with steady flow, 'cells', hot upwellings and cold downwellings. It was often assumed also to occur below

the crust or lithosphere, which was assumed to be dragged around by the underlying convection. These ideas are distinct from the concept that will be presented in Part 3, which has resulted from some relatively recent conceptual shifts.

An early mention of mantle convection is by Hopkins in 1839 [86]. Fisher, in his 1881 book *Physics of the Earth's Crust* [87], proposed mantle convection as a tectonic agent, with flow rising under the oceans and descending under continents. He assumed the mantle to be relatively fluid, drawing on the concepts of isostasy being developed at that time. He envisaged that this flow would expand the oceans and compress the continents at their edges, generating mountains.

According to Hallam ([1], p. 140), the idea of a fluid mantle was more widespread in continental Europe, particularly in Germany, than in Britain and America. He cites a number of instances of this, noting that this implies a more sympathetic climate around the turn of the century within which Wegener's ideas of continental drift could develop. However, Wegener himself did not appeal to mantle convection, and his concept that continents plough through oceanic crust seems to owe little to any idea of a deformable mantle.

It was Arthur Holmes who most seriously advocated mantle convection, and he proposed it explicitly as a mechanism for continental drift, first in a talk and brief note in 1928 [88], then in a paper in 1931 [89], and finally in his book *Principles of Physical Geology*, the first edition of which appeared in 1944 [90]. Holmes's basic proposal was that convection occurs under the lithosphere and drags the continents around. His proposed flow was different from Fisher's, in that Holmes, in his initial version, reasoned that convection might rise under a continent because of the thermal blanketing effect of continental radioactivity, a subject that he was very familiar with.

Holmes then envisaged that the rising and diverging convection might rift a continent and carry the pieces apart. In his earlier version, he supposed that a piece of continent might be left over the upwelling site, because the horizontal flow would be relatively stagnant there. In his later version, he proposed instead that the crust between the diverging continental fragments might be broadly stretched and the extension accommodated by the intrusion or eruption of basaltic melts generated in the (presumed) warmer upwelling mantle.

Holmes also envisaged that a basaltic oceanic crust would be returned to the mantle. He presented the case with admirable simplicity ([90] ; see Cox [3], p. 21):

The obstruction that stands in the way of continental advance is the basaltic layer, and obviously for advance to be possible the basaltic rocks must be continuously moved out of the way. In other words, they must founder into the depths, since there can be nowhere else for them to go.

Holmes, in this later version, proposed a different driving force for his convecting system. He contrasted sialic rocks, whose density is not much affected by pressure, with basaltic compositions, which are converted by pressure first to granulites and then to eclogite, undergoing in the process a density increase from about $2.9\,Mg/m^3$ to $3.4\,Mg/m^3$. Given that it was not known then that the oceanic basaltic crust is quite thin (about 7 km), this was quite a plausible suggestion. He continues

Since this change is known to have happened to certain masses of basaltic rocks that have been involved in the stresses of mountain building, it may be safely inferred that basaltic roots would undergo a similar metamorphism into eclogite. Such roots could not, of course, exert any [positive] buoyancy, and for this reason it is impossible that tectonic mountains could ever arise from the ocean floor. On the contrary, a heavy root formed of eclogite would continue to develop downwards until it merged into and became part of the descending current, so gradually sinking out of the way, and providing room for the crust on either side to be drawn inwards by the horizontal currents beneath them.

Thus Holmes, in this later version, proposed the generation of a basaltic crust over mantle upwellings and its removal into downwellings, concluding

To sum up: during large-scale convective circulation the basaltic layer becomes a kind of endless travelling belt on the top of which a continent can be carried along, until it comes to rest (relative to the belt) when its advancing front reaches the place where the belt turns downwards and disappears into the earth.

Menard ([6], p. 157) has commented on how closely this anticipates Dietz's version of seafloor spreading, the only essential difference being that Dietz proposed that the basaltic oceanic crust is produced in the narrow rift zone at the crest of the midocean rise system, whereas Holmes assumed it would emerge over a broad extensional area.

There is another brief passage worth quoting from this section of Holmes:

The eclogite that founders into the depths will gradually be heated up as it shares in the convective circulation. By the time it reaches the bottom of the substratum it will have begun to fuse, so forming pockets of magma which, being of low density, must sooner or later rise to the top. Thus an adequate source is provided for the unprecedented plateau basalt that broke through the continents during Jurassic and Tertiary times. Most of the basaltic magma, however, would naturally rise with the ascending currents of the main convecting systems . . .

Here Holmes has proposed that the subducted eclogite might rise in two distinct ways: most of it carried up by the main circulation to form new oceanic crust, but some of it forcing its way up independently and breaking out on the surface as flood basalt. Aside from his assumption of melting at great depth, rather than as the material approaches the surface, this is broadly similar to current ideas that subducted oceanic crust is concentrated near the base of the mantle and recycled to the surface through plumes to form flood basalts and hotspot tracks [84, 91, 92].

Holmes's ideas were not entirely ignored, although they did not become part of mainstream thinking. During the 1930s, Pekeris [93] showed that convection driven by the differential thermal blanketing of continents and oceans could result in velocities of millimetres per year and stresses sufficient to maintain observed long-wavelength gravity anomalies. Hales [94] showed that plausible convection could be maintained by a mean vertical temperature gradient (above the adiabatic gradient) of as little as 0.1 K/km. Haskell's estimate of mantle viscosity from post-glacial rebound, assuming flow to penetrate deep into the mantle, appeared during this period [26]. Griggs [30] developed a number of ideas, a central one being that experimentally observed non-linearities in rock rheology could result in episodic convection. He also presented a simple laboratory realisation of the way crust might be piled into mountains over a convective downwelling. This experiment probably had a positive influence on concepts of subduction and the interpretation of the Wadati–Benioff deep earthquake zones.

I described earlier some of the ways that concepts of mantle convection entered the thinking of those who developed the ideas of seafloor spreading and plate tectonics. After the general acceptance of plate tectonics, there was a great deal of discussion of 'the driving mechanism'. I defer a detailed discussion of the ideas from this time until Chapter 12, after I have presented what I think is the most useful way to think of the relationship between plate tectonics and mantle convection. It will then be easier to discuss the limitations of some of the early plate-tectonic views. However I summarise here the nature of the required shift in thinking.

Many of the earlier discussions of mantle convection conceived it as something that happens *under* the lithosphere. One key shift in perspective is to regard the lithosphere as *part* of mantle convection. Another is to realise that the (negative) thermal buoyancy of the cold lithosphere can provide the driving force. A third is to realise that convection need not have active upwellings, but can comprise cold, negatively buoyant, *active* downwellings and complementary *passive* upwellings. A fourth is that the flow pattern is likely to be *unsteady*, especially if it is strongly affected by the changing plate configuration (Chapter 9), and in that case it is not useful to think of a 'cell' of convection.

Holmes's early idea of thermal blanketing was plausible, though the quantitative effect is rather smaller than is required. His later idea of invoking the basalt-eclogite transformation was also plausible within the uncertainties of the time about the thickness of the oceanic crust. It is a possibility still worth entertaining for some earlier stage of earth history (Chapter 14), though as an adjunct to thermal convection, not as a substitute. Daly in 1940 [14] got close to the modern concept of mobile lithosphere, with his idea that the lithosphere would be thrust downwards in compressional zones, but the mobility he envisaged was limited. He also got close by invoking gravity sliding off topographic highs, which is a form of driving by thermal buoyancy, but evidently he did not also think of the weight of his downward-projecting 'prongs' as a possible driving component.

With the advent of plate tectonics, the idea of the lithosphere being an active component was soon advanced, though there was a debate with those who thought it must still be carried passively by convection (necessarily of uncertain origin) underneath. There were also those who evidently did not think of the motions induced by an active lithosphere as convection and for whom the term mantle convection still referred to something happening under the plates [95].

The almost universal assumption that upwellings under mid-ocean ridges would necessarily be hot and active, and usually fixed as well, was a hindrance to the emergence of the concepts of seafloor spreading and plate tectonics, because in that case neither the offsets of spreading centres along transform faults nor the relative motion of different spreading centres made sense. With passive upwelling, both problems go away. Clear and direct evidence that passive upwelling under spreading centres is the norm comes from seafloor topography, as I will argue in Chapters 8, 10 and 12.

Plumes are also a form of mantle convection, as I will argue in Part 3. Furthermore, they are active upwellings. However, they do not occur universally at spreading centres but are to a substantial degree independent of the plates, and the proportion of the total

length of spreading centres affected by plumes is small. Therefore they do not contradict the point just made that upwelling under spreading centres is normally passive.

I will conclude here with a quite different kind of argument. In 1965, Tozer [96] argued that mantle convection was inevitable under very general assumptions. He observed that the viscosity of rocks is very strongly temperature-dependent, decreasing by roughly one order of magnitude for each 100 °C increase in temperature, for temperatures near that of the mantle (about 1300 °C). This has the effect of feeding back on thermal convection, so that large changes in heat transport can be accomplished by small changes in mantle temperature.

Tozer supposed that there is a certain amount of radioactive heat generation at present in the mantle. There will then be a certain mantle temperature at which the buoyancies and the viscosity are such that the convective heat loss just balances the radioactive heating. This is a stable equilibrium temperature. If the mantle were ever hotter than this, the viscosity would have been substantially less and convection very much faster, and the mantle would have rapidly cooled towards the equilibrium value. If the mantle were cooler, the viscosity would have been much higher and convective heat loss much less, and the mantle would have been warmed by the radioactivity until it approach the equilibrium temperature.

These arguments were quantified and confirmed fifteen years later [97]. The time scale of approach to the equilibrium from higher temperatures is a few hundred million years. Only if the earth started very cold does the argument fail, because it would not yet be hot enough to convect, but there is abundant evidence for a tectonically active and therefore hot earth through most of its history. An analogous argument will be given in Part 3 for the existence of thermal plumes in the mantle.

3.9 Afterthoughts

I have tried to trace the emergence of key ideas that comprise our current understanding of mantle convection. I am struck by the fact that most of them were in place well before plate tectonics was invented and mantle convection was accepted by implication. It is of course easy to be wise in retrospect, but consider that the following ideas were well established, if not widely appreciated, by about 1945.

There is a lithosphere about 100 km thick underlain by an asthenosphere with a viscosity of the order of 10^{21} Pa s [14, 26].

The major tectonic events of the Phanerozoic are recorded in long, narrow mobile belts [41].

Current tectonic activity, comprising most of the earthquakes and volcanoes, occurs in a nearly continuous network of belts [98].

By implication, there is relatively little deformation of the crust outside current and past mobile belts.

Large earthquakes involve metres of fault displacement, and recur about once a century in the most active regions. Repetition of such earthquakes would yield slip rates of centimetres per year.

Circumstantial evidence for continental drift would require drift rates of centimetres per year [7, 12].

Oceanic trenches are far from isostatic balance (having large negative gravity anomalies as well as large negative topographic anomalies), which requires a force to pull them down. It was suggested that descending convective flow would provide this force [38].

The lithosphere would be forced or required to move downward in compressional mountain belts [14].

The oceanic crust might descend under its own weight because of the transformation of basalt to dense eclogite [90].

I have not found a suggestion prior to plate tectonics that the lithosphere would sink because it is *colder* and denser, but Holmes' compositional density idea could have served to stimulate thinking about an active lithosphere.

Another point worth mentioning is that most of the key insights in the development of plate tectonics were by people once-removed from observational work. Hess, Dietz and Morley were involved in science administration. Wilson and Morgan were at institutions without big field or oceanographic programmes. Vine was a new graduate student interpreting data much of which had been obtained by his supervisor, Matthews. McKenzie had done a theoretical thesis on mantle viscosity, and his plate tectonics work was with theory and archived data.

This is not meant to disparage observational science, which is essential, and theories are only useful if their proponents attempt to connect them with observations. It does suggest, though, that often observational programmes become too dominating, and that people could take more time to think more broadly. It also clearly demonstrates that the full significance and meaning of an observation may not be evident for some time after it is made. Seafloor magnetic stripes and fracture zones are obvious examples, but there are many in my experience.

Being a theoretician, I have often encountered the attitude that theoreticians only play games, and that observational and experimental work is the 'real' science. Too many of my theoretical peers do not pay enough attention to observational constraints, so there is a grain of truth in the perception. However, the theoretical versus observational debate is sterile. Science requires both, and in my experience the most stimulating scientists are those who straddle both to some degree.

3.10 References

1. A. Hallam, *Great Geological Controversies*, 244 pp., Oxford University Press, Oxford, 1989.
2. A. Hallam, *A Revolution in the Earth Sciences*, 127 pp., Clarendon Press, Oxford, 1973.
3. A. Cox, ed., *Plate Tectonics and Geomagnetic Reversals*, 702 pp., W.H. Freeman and Company, San Francisco, 1973.
4. W. Glen, *Continental Drift and Plate Tectonics*, 188 pp., Charles E. Merrill Publishing Company, Columbus, 1975.
5. W. Glen, *The Road to Jaramillo*, Stanford University Press, Stanford, 1982.
6. H. W. Menard, *The Ocean of Truth*, 353 pp., Princeton University Press, Princeton, New Jersey, 1986.
7. A. Wegener, *Die Entstehung der Kontinente und Ozeane*, Vieweg and Son, Brunswick, 1915.
8. A. Wegener, *The Origin of Continents and Oceans*, Methuen, London, 1966.
9. H. Jeffreys, *The Earth, its Origin, History and Physical Constitution*, Cambridge University Press, 1926.
10. B. Gutenberg, Hypotheses on the development of the Earth, in: *Internal Constitution of the Earth*, B. Gutenberg, ed., Dover, New York, 178–226, 1951.
11. B. Gutenberg, *Physics of the Earth's Interior*, Academic Press, New York, 1959.
12. A. L. du Toit, *Our Wandering Continents*, Oliver and Boyd, Edinburgh, 1937.
13. S. W. Carey, The tectonic approach to continental drift, in: *Continental Drift; a Symposium*, S. W. Carey, ed., University of Tasmania, Geol. Dept., Hobart, 177–358, 1958.
14. R. A. Daly, *Strength and Structure of the Earth*, 434 pp., Prentice-Hall, New York, 1940.
15. J. H. Pratt, *Philos. Trans. R. Soc. London* **145**, 53–5, 1855.

16. J. H. Pratt, *Philos. Trans. R. Soc. London* **149**, 779, 1859.
17. G. B. Airy, *Philos. Trans. R. Soc. London* **145**, 101–4, 1855.
18. C. E. Dutton, *Bull. Phil. Soc. Washington* **11**, 51, 1889.
19. J. Hall, *Geology of New York State*, **3**, 69, 1859.
20. F. R. Helmert, *Sitzungsber. Preuss. Akad. Wiss.* 1192, 1909.
21. J. Barrell, The strength of the earth's crust, *J. Geol.* **22**, 655–83, 1914.
22. A. Mohorovičić, Das Beben vom 8.x.1909, *Jahrb. Met. Obs. Zagreb (Agram.)* **9**, 1-63, 1909.
23. R. D. Oldham, Constitution of the interior of the earth as revealed by earthquakes, *Quart. J. Geol. Soc. London* **62**, 456–75, 1906.
24. T. F. Jamieson, *Quart. J. Geol. Soc. London* **21**, 178, 1865.
25. R. van Bemmelen and P. Berlage, *Gerlands Beitr. zur. Geophysik* **43**, 19, 1934.
26. N. A. Haskell, The viscosity of the asthenoshpere, *Am. J. Sci.*, ser. 5 **33**, 22–8, 1937.
27. K. Wadati, Shallow and deep earthquakes, *Geophys. Mag. (Tokyo)* **1**, 162–202, 1928.
28. S. K. Runcorn, Paleomagnetic evidence for continental drift and its geophysical cause, in: *Continental Drift*, S. K. Runcorn, ed., Academic Press, New York and London, 1–40, 1962.
29. P. Goldreich and A. Toomre, Some remarks on polar wandering, *J. Geophys. Res.* **74**, 2555–67, 1969.
30. D. T. Griggs, A theory of mountain building, *Am. J. Sci.* **237**, 611–50, 1939.
31. S. K. Runcorn, Rock magnetism, *Science* **129**, 1002–11, 1959.
32. E. Irving, Paleomagnetic pole positions, *Geophys. J. R. Astron. Soc.* **2**, 51–77, 1959.
33. H. H. Hess, Drowned ancient islands of the Pacific Basin, *Am. J. Sci.* **244**, 772–91, 1946.
34. H. H. Hess, History of ocean basins, in: *Petrologic Studies: a Volume in Honor of A. F. Buddington*, A. E. J. Engel, H. L. James and B. F. Leonard, eds., Geol. Soc. Am., Boulder, CO, 599–620, 1962.
35. H. W. Menard, *Marine Geology of the Pacific*, McGraw-Hill, New York, 1964.
36. B. C. Heezen, The rift in the ocean floor, *Sci. Am.* **203**, 98–110, 1960.
37. R. S. Dietz, Continent and ocean evolution by spreading of the sea floor, *Nature* **190**, 854–7, 1961.
38. F. A. Vening Meinesz, *Gravity Expeditions at Sea, 1923–32. Vol.2: Interpretation of the Results*, Publ. Neth. Geod. Comm., Waltman, Delft, 1934.
39. H. Benioff, Seismic evidence for crustal structure and tectonic activity, *Geol. Soc. Amer. Spec. Paper 62*, 61–74, 1955.
40. J. T. Wilson, A new class of faults and their bearing on continental drift, *Nature* **207**, 343–7, 1965.
41. W. H. Bucher, *The Deformation of the Earth's Crust*, 518 pp., Princeton University Press, Princeton, 1933.
42. J. T. Wilson, Continental drift, *Sci. Am.*, April, 1963.

43. L. R. Sykes, Seismicity of the South Pacific Ocean, *J. Geophys. Res.* **68**, 5999–6006, 1963.
44. A. M. Coode, A note on oceanic transcurrent faults, *Can. J. Earth Sci.* **2**, 400–1, 1965.
45. J. T. Wilson, A possible origin of the Hawaiian islands, *Can. J. Phys.* **41**, 863–70, 1963.
46. J. T. Wilson, Evidence from islands on the spreading of the ocean floor, *Nature* **197**, 536–8, 1963.
47. J. T. Wilson, *Continents Adrift and Continents Aground*, 230 pp., W. H. Freeman and Company, San Francisco, 1976.
48. M. Matuyama, On the direction of magnetization of basalt in Japan, Tyosen, and Manchuria, *Proc. Japan Acad.* **5**, 203–5, 1929.
49. A. Cox, R. R. Doell and G. B. Dalrymple, Geomagnetic polarity epochs and Pleistocene geochronometry, *Nature* **198**, 1049–51, 1963.
50. I. McDougall and D. H. Tarling, Dating of polarity zones in the Hawaiian islands, *Nature* **200**, 54–6, 1963.
51. A. Cox, Geomagnetic reversals, *Science* **163**, 237–45, 1969.
52. R. G. Mason, A magnetic survey off the west coast of the United States, *Geophys. J. R. Astron. Soc.* **1**, 320–9, 1958.
53. F. J. Vine and D. H. Matthews, Magnetic anomalies over oceanic ridges, *Nature* **199**, 947–9, 1963.
54. F. J. Vine, Spreading of the ocean floor: new evidence, *Science* **154**, 1405–15, 1966.
55. W. C. Pitman III, Magnetic anomalies over the Pacific-Antarctic ridge, *Science* **154**, 1154-1171, 1966.
56. J. R. Heirtzler, G. O. Dickson, E. M. Herron, W. C. Pitman III and X. le Pichon, Marine magnetic anomalies, geomagnetic field reversals, and motions of the ocean floor and continents., *J. Geophys. Res.* **73**, 2119–36, 1968.
57. J. R. Heirzler, X. Le Pichon and J. G. Baron, Magnetic anomalies over the Reykjanes ridge, *Deep-Sea Res.* **13**, 427–43, 1966.
58. L. R. Sykes, The seismicity of the Arctic, *Bull. Seismol. Soc. Am.* **55**, 501–18, 1965.
59. L. R. Sykes, Mechanism of earthquakes and nature of faulting on the midocean ridges, *J. Geophys. Res.* **72**, 2131–53, 1967.
60. M. Ewing, J. Ewing and M. Talwani, Sediment distribution in the oceans: the Mid-Atlantic Ridge, *Bull. Geol. Soc. Am.* **75**, 17–36, 1964.
61. A. E. Maxwell, R. P. Von Herzen, K. J. Hsu, J. E. Andrews, T. Saito, S. F. Percival, E. D. Milow and R. E. Boyce, Deep sea drilling in the South Atlantic, *Science* **168**, 1047–59, 1970.
62. E. C. Bullard, J. E. Everett and A. G. Smith, The fit of the continents around the Atlantic, *Philos. Trans. R. Soc. London* **258**, 41–51, 1965.
63. H. W. Menard, Extension of northeast Pacific fracture zones, *Science* **155**, 72–4, 1967.
64. W. J. Morgan, Rises, trenches, great faults and crustal blocks, *J. Geophys. Res.* **73**, 1959–82, 1968.

65. D. P. McKenzie and R. L. Parker, The north Pacific: an example of tectonics on a sphere, *Nature* **216**, 1276–80, 1967.
66. D. P. McKenzie and W. J. Morgan, Evolution of triple junctions, *Nature* **224**, 125–33, 1969.
67. K. Wadati, Shallow and deep earthquakes, *Geophys. Mag. (Tokyo)* **4**, 231–85, 1931.
68. T. Utsu, Regional differences in absorption of seismic waves in the upper mantle as inferred from abnormal differences in seismic intensities, *J. Fac. Sci. Hokkaido Univ. Japan, Ser. VII* **2**, 359–74, 1966.
69. L. R. Sykes, The seismicity and deep structure of island arcs, *J. Geophys. Res.* **71**, 2981–3006, 1966.
70. J. Oliver and B. Isacks, Deep earthquake zones, anomalous structures in the upper mantle, and the lithosphere., *J. Geophys. Res.* **72**, 4259–75, 1967.
71. B. Isacks, J. Oliver and L. R. Sykes, Seismology and the new global tectonics, *J. Geophys. Res.* **73**, 5855–99, 1968.
72. B. Isacks and P. Molnar, Distribution of stresses in the descending lithosphere from a global survey of focal-mechanism solutions of mantle earthquakes, *Rev. Geophys. Space Phys.* **9**, 103–74, 1971.
73. G. F. Davies and J. N. Brune, Regional and global fault slip rates from seismicity, *Nature* **229**, 101–7, 1971.
74. W. J. Morgan, Convection plumes in the lower mantle, *Nature* **230**, 42–3, 1971.
75. W. J. Morgan, Plate motions and deep mantle convection, *Mem. Geol. Soc. Am.* **132**, 7–22, 1972.
76. T. S. Crough, Hotspot swells, *Annu. Rev. Earth Planet. Sci.* **11**, 165–93, 1983.
77. J. A. Whitehead and D. S. Luther, Dynamics of laboratory diapir and plume models, *J. Geophys. Res.* **80**, 705–17, 1975.
78. W. J. Morgan, Hotspot tracks and the opening of the Atlantic and Indian Oceans, in: *The Sea*, C. Emiliani, ed., Wiley, New York, 443–87, 1981.
79. D. E. Loper and F. D. Stacey, The dynamical and thermal structure of deep mantle plumes, *Phys. Earth Planet. Inter.* **33**, 304–17, 1983.
80. F. W. Stacey and D. E. Loper, The thermal boundary layer interpretation of D″ and its role as a plume source, *Phys. Earth Planet. Inter.* **33**, 45–55, 1983.
81. P. Olson and H. A. Singer, Creeping plumes, *J. Fluid Mech.* **158**, 511–31, 1985.
82. R. W. Griffiths and I. H. Campbell, Stirring and structure in mantle plumes, *Earth Planet. Sci. Lett.* **99**, 66–78, 1990.
83. M. A. Richards, R. A. Duncan and V. E. Courtillot, Flood basalts and hot-spot tracks: plume heads and tails, *Science* **246**, 103–7, 1989.
84. I. H. Campbell and R. W. Griffiths, Implications of mantle plume structure for the evolution of flood basalts, *Earth Planet. Sci. Lett.* **99**, 79–83, 1990.

85. H. Benard, Les tourbillons cellulaires dans une nappe liquide transportant de la chaleur par convection en regime permanent, *Ann. Chim. Phys.* **23**, 62–144, 1901.
86. W. Hopkins, Researches in physical geology, *Philos. Trans. R. Soc. London* **129**, 381–5, 1839.
87. O. Fisher, *Physics of the Earth's Crust*, Murray, London, 1881.
88. A. Holmes, Continental drift: a review, *Nature* **122**, 431–3, 1928.
89. A. Holmes, Radioactivity and earth movements, *Geol. Soc. Glasgow, Trans.* **18**, 559–606, 1931.
90. A. Holmes, *Principles of Physical Geology*, Thomas Nelson and Sons, 1944.
91. A. W. Hofmann and W. M. White, Mantle plumes from ancient oceanic crust, *Earth Planet. Sci. Lett.* **57**, 421–36, 1982.
92. G. F. Davies, Mantle plumes, mantle stirring and hotspot chemistry, *Earth Planet. Sci. Lett.* **99**, 94–109, 1990.
93. C. L. Pekeris, Thermal convection in the interior of the earth, *Mon. Not. R. Astron. Soc., Geophys. Suppl.* **3**, 346–67, 1935.
94. A. L. Hales, Convection currents in the earth, *Mon. Not. R. Astron. Soc., Geophys. Suppl.* **3**, 372-79, 1936.
95. D. P. McKenzie, A. B. Watts, B. Parsons and M. Roufosse, Planform of mantle convection beneath the Pacific Ocean, *Nature* **288**, 442–6, 1980.
96. D. C. Tozer, Heat transfer and convection currents, *Philos. Trans. R. Soc. London, Ser. A* **258**, 252–71, 1965.
97. G. F. Davies, Thermal histories of convective earth models and constraints on radiogenic heat production in the earth, *J. Geophys. Res.* **85**, 2517–30, 1980.
98. B. Gutenberg and C. F. Richter, *Seismicity of the Earth*, Geol. Soc. Amer. Spec. Paper 34, 1941.