1

Historical Background

1.1 Introduction

The origin of mountain ranges, volcanoes, earthquakes, the ocean basins, the continents, and the very nature of the Earth's interior are questions as old as science itself. In the development of modern science, virtually every famous Natural Philosopher has conjectured on the state of the deep interior and its relationship to the Earth's surface. And every one of these thinkers has come to essentially the same general conclusion: despite the obvious solidity of the Earth beneath our feet, the interior must have flowed, in order to create the complex surface geology we see today. Although we can trace this idea as far back as written scientific record permits, it nevertheless remained a strictly qualitative hypothesis until the early part of the twentieth century. Then several timely developments in physics, fluid mechanics, geophysics, and geology finally established a true physical paradigm for the Earth's interior, mantle convection.

In reviewing the development of the concept of mantle convection, we find it is impossible to identify one particular time or event, or one particular individual, as being decisive in either its construction or its acceptance. Instead, the subject's progress has followed a meandering course, assisted along by the contributions of many. Still, there are a few scientific pioneers whose insights were crucial at certain times. These insights deserve special recognition and, when put together, provide some historical context with which future progress can be measured.

The idea of flow in the Earth's interior was popular among the early Natural Philosophers, as it was commonly assumed that only the outermost portion of the Earth was solid. Descartes imagined the Earth to consist of essentially sedimentary rocks (the crust) lying over a shell of denser rocks (the mantle) with a metallic center (the core). Leibniz proposed that the Earth cooled from an initially molten state and that the deep interior remained molten, a relic of its formation. Edmond Halley argued that the flow of liquids in a network of subsurface channels would explain his discovery of the secular variation of the geomagnetic field. Both Newton and Laplace interpreted the equatorial bulge of the Earth to be a consequence of a fluid-like response to its rotation. The idea that the Earth's interior included fluid channels and extensive molten regions, as expressed in the artist's drawing in Figure 1.1, was the dominant one until the late nineteenth century, when developments in the theory of elasticity and G. H. Darwin's (1898) investigation of the tides indicated that the Earth was not only solid to great depths, but also "more rigid than steel."

In the eighteenth and early nineteenth centuries, resolution of an old controversy led geologists to accept the idea of a hot and mobile Earth interior. This controversy centered on

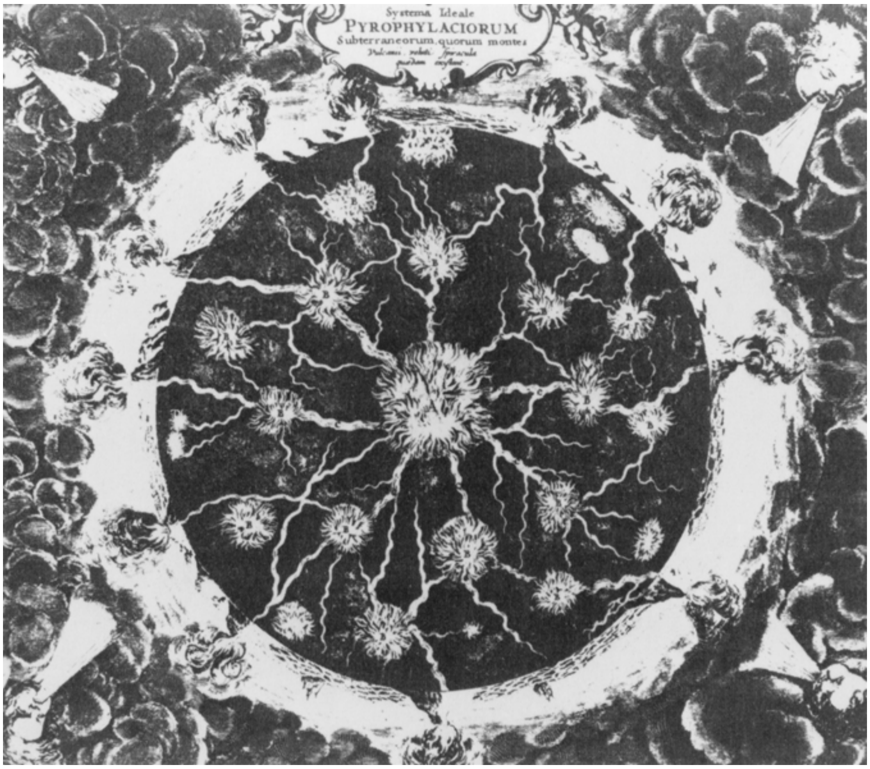

Figure 1.1. An early depiction of the Earth's interior, showing channels of fluid connected to a molten central core. This was the prevailing view prior to the nineteenth century.

the origin of rocks and pitted the so-called Neptunists, led by Abraham Werner, who thought all rocks were derived by precipitation from a primeval ocean, against others who held that igneous rocks crystallized from melts and were to be distinguished from sedimentary rocks formed by surficial processes. Their most influential member was an amateur scientist, James Hutton (Figure 1.2), who advanced the concept of uniformitarianism, that the processes evident today were those that shaped the Earth in the past. He also held to the idea of a molten, flowing interior, exerting forces on the solid crust to form mountain ranges, close to the modern view based on mantle convection. Ultimately Hutton's view prevailed, and with it, an emphasis on the idea that the fundamental physical process behind all major geological events is heat transfer from the deep interior to the surface. Thus, geologists were receptive to the idea of a hot, mobile Earth interior. However, most of the geological and geophysical evidence obtained from the continental crust seemed to demand vertical motions, rather than horizontal motions. For example, in the mid-nineteenth century it was discovered that mountain ranges did not have the expected positive gravity anomaly. This was explained by low-density continental roots, which floated on the denser mantle like blocks of wood in water, according to the principle of hydrostatic equilibrium. This implied, in turn, that the mantle behaved like a fluid, allowing vertical adjustment. However, the notion that the crust experiences far larger horizontal displacements was less well supported by evidence and was not widely held.

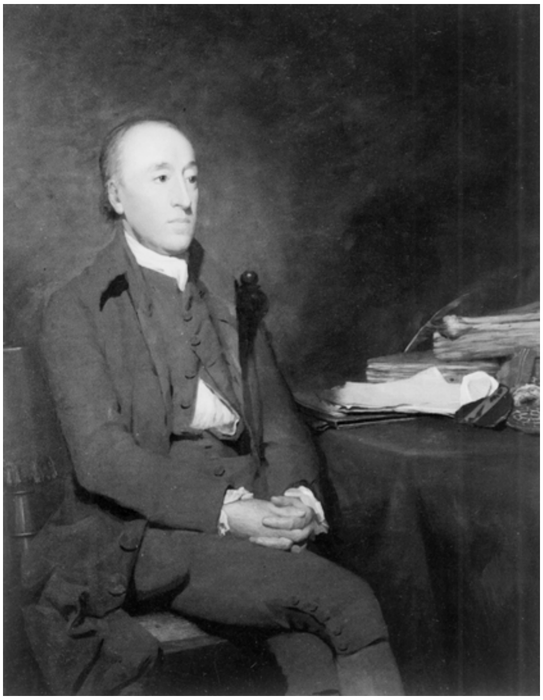

Figure 1.2. James Hutton (1726–1797), the father of Geology and proponent of internal heat as the driving force for Earth's evolution.

The idea that flow in the Earth's interior is a form of thermal convection developed slowly. Recognition of the significance of thermal convection as a primary fluid mechanical phenomenon in nature came from physicists. Count Rumford is usually given credit for recognizing the phenomenon around 1797 (Brown, 1957), although the term convection (derived from convectio, to carry) was first used by Prout (1834) to distinguish it from the other known heat transfer mechanisms, conduction and radiation. Subcrustal convection in the Earth was first suggested by W. Hopkins in 1839 and the first interpretations of geological observations using convection were made by Osmond Fisher (1881). Both of these presumed a fluid interior, so when the solidity of the mantle was established, these ideas fell out of favor.

The earliest experiments on convection in a layer of fluid heated from below and cooled from above were reported by J. Thompson (1882), who observed a "tesselated structure" in the liquid when its excess temperature, compared to that of the overlying air, was sufficiently large. But the name most closely associated with convection is Henri Bénard (Figure 1.3). Bénard (1900, 1901) reported the first quantitative experiments on the onset of convection, including the role of viscosity, the cellular planform, and the relationship between cell size and fluid layer depth. Bénard produced striking photographs of the convective planform in thin layers of viscous fluids heated from below (Figure 1.4). The regular, periodic, hexagonal cells in his photographs are still referred to as Bénard cells. Since Bénard used fluid layers in contact with air, surface tension effects were surely present in his experiments, as he himself recognized. It has since been shown that Bénard's cells were driven as much by surface tension gradients as by gradients in buoyancy. Still, he correctly identified the essentials of thermal convection, and in doing so, opened a whole new field of fluid mechanics. Motivated by the "interesting results obtained by Bénard's careful and skillful experiments," Lord Rayleigh (1916; Figure 1.5) developed the linear stability theory for the onset of convection

Figure 1.3. Henri Bénard (1880–1939) (on the left) made the first quantitative experiments on cellular convection in viscous liquids. The picture was taken in Paris about 1920 with Reabouchansky on the right.

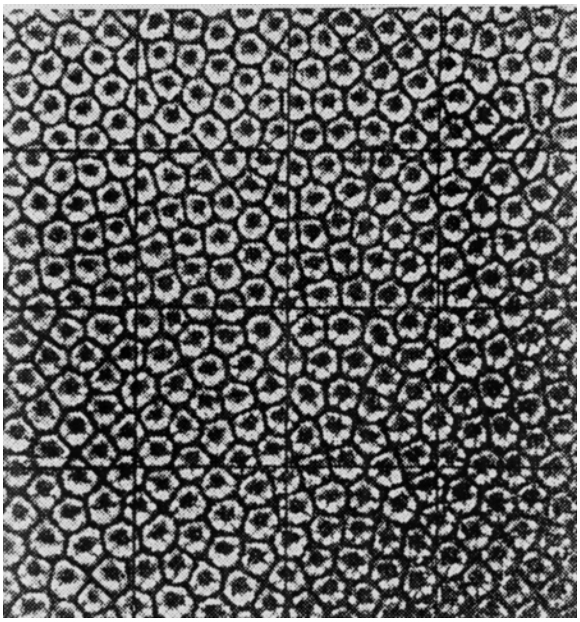

Figure 1.4. Photograph of hexagonal convection cells in a viscous fluid layer heated from below, taken by Bénard (1901).

in a horizontally infinite fluid layer between parallel surfaces heated uniformly from below and cooled uniformly from above, and isolated the governing dimensionless parameter that now bears his name. It was unfortunate that these developments in fluid mechanics were not followed more widely in Earth Science, for they might have removed a stumbling block to acceptance of the milestone concept of continental drift.

Figure 1.5. Lord Rayleigh (1842–1919) developed the theory of convective instability in fluids heated from below.

1.2 Continental Drift

The earliest arguments for continental drift were largely based on the fit of the continents. Ever since the first reliable maps were available, the remarkable fit between the east coast of South America and the west coast of Africa has been noted (e.g., Carey, 1955). Indeed, the fit was pointed out as early as 1620 by Francis Bacon (Bacon, 1620). North America, Greenland, and Europe also fit as illustrated in Figure 1.6 (Bullard et al., 1965).

Geological mapping in the southern hemisphere during the nineteenth century revealed that the fit between these continents extends beyond coastline geometry. Mountain belts in South America match mountain belts in Africa; similar rock types, rock ages, and fossil species are found on the two sides of the Atlantic Ocean. Thus the southern hemisphere geologists were generally more receptive to the idea of continental drift than their northern hemisphere colleagues, where the geologic evidence was far less conclusive.

Further evidence for continental drift came from studies of ancient climates. Geologists recognized that tropical climates had existed in polar regions at the same times that arctic climates had existed in equatorial regions. Also, the evolution and dispersion of plant and animal species was best explained in terms of ancient land bridges, suggesting direct connections between now widely separated continents.

As previously indicated, most geologists and geophysicists in the early twentieth century assumed that relative motions on the Earth's surface, including motions of the continents relative to the oceans, were mainly vertical and generally quite small – a few kilometers in extreme cases. The first serious advocates for large horizontal displacements were two visionaries, F. B. Taylor and Alfred Wegener (Figure 1.7). Continental drift was not widely discussed until the publication of Wegener's famous book (Wegener, 1915; see also Wegener, 1924), but Taylor deserves to share the credit for his independent and somewhat earlier account (Taylor, 1910). Wegener's book includes his highly original picture of the breakup

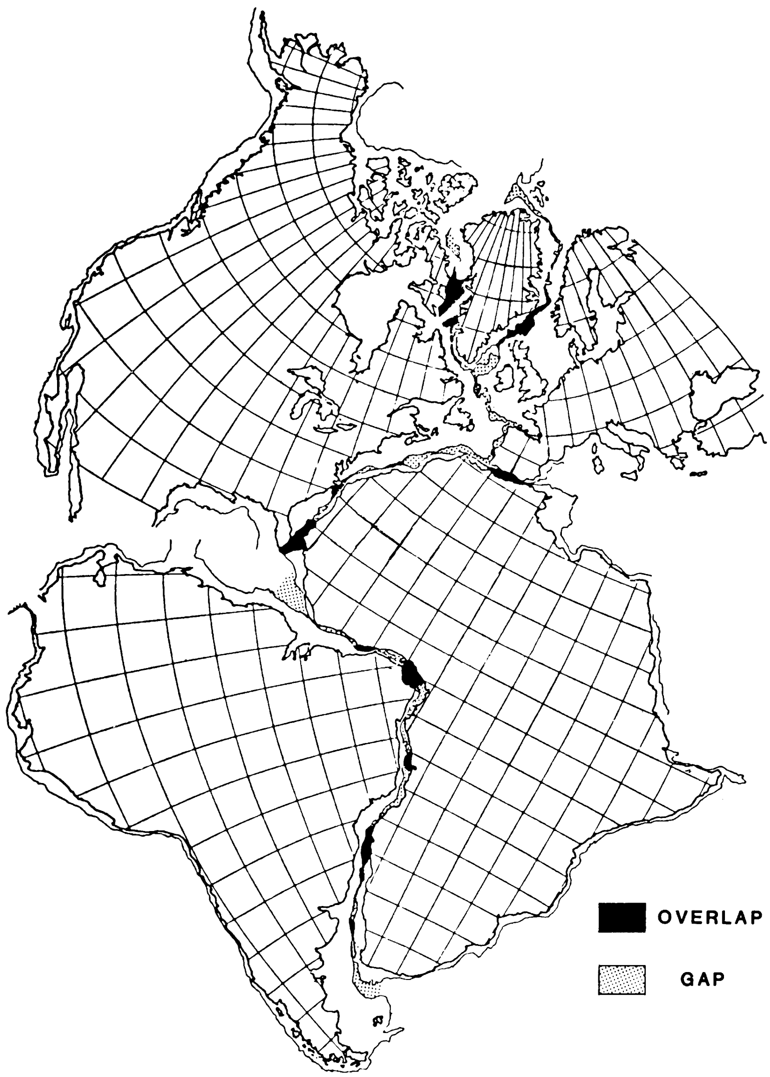

Figure 1.6. The remarkable “fit” between the continental margins of North and South America and Greenland, Europe, and Africa (from Bullard et al., 1965). This fit was one of the primary early arguments for continental drift.

and subsequent drift of the continents, and his recognition of the supercontinent Pangaea (all Earth). (Later it was argued (du Toit, 1937) that there had formerly been a northern continent, Laurasia, and a southern continent, Gondwanaland, separated by the Tethys ocean.) Wegener assembled a formidable array of facts and conjecture to support his case, including much that was subsequently discredited. This partially explains the hostile reception his book initially received. However, the most damaging criticisms came from prominent geophysicists such as H. Jeffreys in England and W. Bowie in the U.S., who dismissed the idea because the driving forces for continental drift proposed by Taylor and Wegener (tidal and differential centrifugal forces, respectively) were physically inadequate. (Wegener was a meteorologist and recognized that the Earth’s rotation dominated atmospheric flows. He proposed that these

Figure 1.7. Alfred Wegener (1880–1930), the father of continental drift.

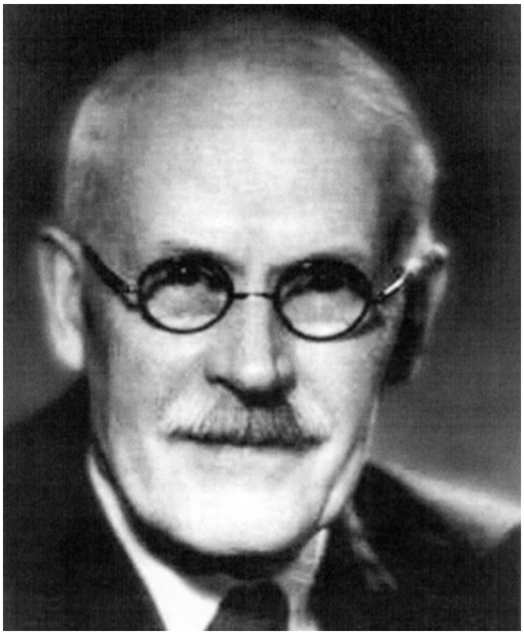

Figure 1.8. Harold Jeffreys (1891–1989), the most influential theorist in the early debate over continental drift and mantle convection.

rotational forces were also responsible for driving the mantle flows resulting in continental drift.) At the same time, seismologists were exploring the Earth's deep interior, and were impressed by the high elastic rigidity of the mantle. In his influential book, *The Earth* (Jeffreys, 1929), Sir Harold Jeffreys (Figure 1.8) referred to the mantle as the "shell," arguing that this term better characterized its elastic strength. Paradoxically, Jeffreys was at the same time making fundamental contributions to the theory of convection in fluids. For example, he showed (Jeffreys, 1930) that convection in a compressible fluid involved the difference between the actual temperature gradient and the adiabatic temperature gradient. This result would later figure prominently in the development of the theory of whole mantle convection. But throughout his illustrious career, Jeffreys maintained that the idea of thermal convection in the highly rigid mantle was implausible on mechanical grounds. The realization that a solid could exhibit both elastic and viscous properties simultaneously was just emerging from the study of materials, and evidently had not yet come fully into the minds of geophysicists.

The failure of rotational and tidal forces meant that some other mechanism had to be found to drive the motion of the continents with sufficient power to account for the observed deformation of the continental crust, seismicity, and volcanism. In addition, such a mechanism had to operate in the solid, crystalline mantle.

Question 1.1: *What is the source of energy for the tectonics and volcanism of the solid Earth?*

Question 1.2: *How is this energy converted into the tectonic and volcanic phenomena we are familiar with?*

The mechanism is thermal convection in the solid mantle, also referred to as subsolidus mantle convection. A fluid layer heated from below and cooled from above will convect in a gravitational field due to thermal expansion and contraction. The hot fluid at the base of the layer is less dense than the cold fluid at the top of the layer; this results in gravitational instability. The light fluid at the base of the layer ascends and the dense fluid at the top of the layer descends. The resulting motion, called thermal convection, is the fundamental process in the Earth's tectonics and volcanism and is the subject of this book. We will see that the energy to drive subsolidus convection in the mantle and its attendant geological consequences (plate tectonics, mountain building, volcanic eruptions, earthquakes) derives from both the secular cooling of the Earth's hot interior and the heat produced by the decay of radioactive elements in the rocks of the mantle.

The original proposal for subsolidus convection in the mantle is somewhat obscure. Bull (1921, 1931) suggested that convection in the solid mantle was responsible for continental drift, but he did not provide quantitative arguments in support of his contention. About the same time, Wegener came to realize that his own proposed mechanism was inadequate for continental drift. He apparently considered the possibility of mantle convection, and made passing reference to it as a plausible driving force in the final edition of his book (Wegener, 1929). It was during this era that the importance of convection was first being recognized in his own field of meteorology. But Wegener chose not to promote it as the cause of continental drift, and the idea languished once again.

1.3 The Concept of Subsolidus Mantle Convection

Arthur Holmes (1931, 1933; Figure 1.9) was the first to establish quantitatively that thermal convection was a viable mechanism for flow in the solid mantle, capable of driving continental drift. Holmes made order of magnitude estimates of the conditions necessary for convection, the energetics of the flow, and the stresses generated by the motion. He concluded that the available estimates of mantle viscosity were several orders of magnitude less

Figure 1.9. Arthur Holmes (1890–1965), the first prominent advocate for subsolidus mantle convection.

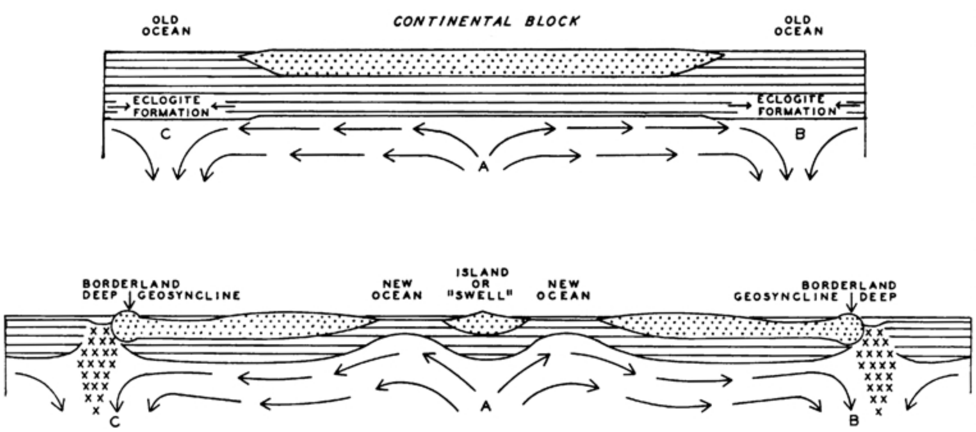

Figure 1.10. Arthur Holmes' (1931) depiction of mantle convection as the cause of continental drift, thirty years prior to the discovery of seafloor spreading.

than that required for the onset of convection. He also outlined a general relation between the ascending and descending limbs of mantle convection cells and geological processes, illustrated in Figure 1.10. Holmes argued that radioactive heat generation in the continents acted as a thermal blanket inducing ascending thermal convection beneath the continents. Holmes was one of the most prominent geologists of the time, and in his prestigious textbook *Principles of Physical Geology* (Holmes, 1945), he articulated the major problems of mantle convection much as we view them today.

The creep viscosity of the solid mantle was first determined quantitatively by Haskell (1937). Recognition of elevated beach terraces in Scandinavia showed that the Earth's surface is still rebounding from the load of ice during the last ice age. By treating the mantle as a viscous fluid, Haskell was able to explain the present uplift of Scandinavia if the mantle has a viscosity of about 10^{20} Pa s. Remarkably, this value of mantle viscosity is still accepted today. Although an immense number (water has a viscosity of 10^{-3} Pa s), it predicts vigorous mantle convection on geologic time scales.

The viscous fluid-like behavior of the solid mantle on long time scales required an explanation. How could horizontal displacements of thousands of kilometers be accommodated in solid mantle rock?

Question 1.3: *Why does solid mantle rock behave like a fluid?*

The answer was provided in the 1950s, when theoretical studies identified several mechanisms for the very slow creep of crystalline materials thereby establishing a mechanical basis for the mantle's fluid behavior. Gordon (1965) showed that solid-state creep quantitatively explained the viscosity determined from observations of postglacial rebound. At temperatures that are a substantial fraction of the melt temperature, thermally activated creep processes allow hot mantle rock to flow at low stress levels on time scales greater than 10^4 years. In hindsight, the flow of the crystalline mantle should not have been a surprise for geophysicists since the flow of crystalline ice in glaciers had long been recognized and accepted.

In the 1930s a small group of independent-minded geophysicists including Pekeris (1935), Hales (1936), and Griggs (1939) attempted to build quantitative models of mantle convection. Figure 1.11 shows an ingenious apparatus built by Griggs to demonstrate the effects of mantle convection on the continental crust. Griggs modeled the crust with sand–oil mixtures, the mantle with viscous fluids, and substituted mechanically driven rotating cylinders for the thermal buoyancy in natural convection. His apparatus produced crustal roots and near-surface thrusting at the convergence between the rotating cylinders; when only one cylinder was rotated, an asymmetric root formed with similarities to a convergent plate margin, including a model deep sea trench. The early work of Pekeris and Hales were attempts at finite amplitude theories of mantle convection. They included explanations for dynamic surface topography, heat flow variations, and the geoid based on mantle convection that are essentially correct according to our present understanding.

In retrospect, these papers were far ahead of their time, but unfortunately their impact was much less than it could have been. In spite of all the attention given to continental drift, the solid foundation of convection theory and experiments, and far-sighted contributions of a few to create a framework for convection in the mantle, general acceptance of the idea came slowly. The vast majority of the Earth Sciences community remained unconvinced about the significance of mantle convection. We can identify several reasons why the Earth Science community was reluctant to embrace the concept, but one stands out far above the others: the best evidence for mantle convection comes from the seafloor, and until the middle of the twentieth century the seafloor was virtually unknown. The situation began to change in the 1950s, when two independent lines of evidence confirmed continental drift and established the relationship between the continents, the oceans, and mantle convection. These were paleomagnetic pole paths and the discovery of seafloor spreading. We will consider each of these in turn.