The Rejection of Continental Drift

THEORY AND METHOD IN AMERICAN EARTH SCIENCE

Naomi Oreskes

The Rejection of
CONTINENTAL DRIFT

The Rejection of
CONTINENTAL DRIFT

Theory and Method in
American Earth Science

NAOMI ORESKES

New York Oxford
Oxford University Press
1999

Oxford University Press

Oxford New York
Athens Auckland Bangkok Bogotá Buenos Aires Calcutta
Cape Town Chennai Dar es Salaam Delhi Florence Hong Kong Istanbul
Karachi Kuala Lumpur Madrid Melbourne Mexico City Mumbai
Nairobi Paris São Paulo Singapore Taipei Tokyo Toronto Warsaw

and associated companies in
Berlin Ibadan

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

Oxford is a registered trademark of Oxford University Press

Library of Congress Cataloging-in-Publication Data
Oreskes, Naomi.
The rejection of continental drift : theory and method
in American earth science / Naomi Oreskes.
p. cm.
Includes bibliographical references and index.
ISBN 0-19-511732-8; ISBN 0-19-511733-6 (pbk)
1. Continental drift.
2. Geology—United States—History—20th century.
I. Title.
QE511.5.074 1999
551.1'36—dc21 98-4161

1 3 5 7 9 8 6 4 2

Printed in the United States of America
on acid-free paper

To Shlomo Flötzgebirge, my faithful companion,
and to K. B., who sees connections that no one else notices
and always keeps me on track.

PREFACE

This book began in 1978 when I first studied geology at Imperial College in London. I had completed two years as a geology major at a leading U.S. university and counted myself lucky to have chosen a field of science heady in the wake of revolutionary upheaval: geologists around the globe were reinterpreting old data and long-standing problems in the new light of plate tectonics. It seemed a good time to be an aspiring young earth scientist. Imagine my surprise—and dismay—to discover in England that the radically new idea of plate tectonics had been proposed more than half a century before by a German geophysicist, Alfred Wegener, and widely promoted in the United Kingdom by the leading British geologist of his era, Arthur Holmes. The revolution that had been described by my professors in the United States as the radical revelation of a dramatically new vision of the earth was viewed by many of my professors in England as the pleasing confirmation of a long-suspected notion. Whereas my textbooks in the United States had proclaimed the explanatory power of the new ruling theory, Dorothy Rayner, the doyenne of British Stratigraphy, dryly instructed in her text, *Stratigraphy of the British Isles*, "our stratigraphy and history has certainly been illuminated by the current hypothesis, but so far the light shed is somewhat uneven." And although I had only just learned of these ideas two years before, my English flatmate could pull out the dog-eared copy of Arthur Holmes's 1945 textbook she had read in elementary school.

The seeds of an intellectual inquiry were sown. For some years they lay dormant while I lived the life of a field geologist, although the ground in which they lay was being heavily fertilized. Working as a professional geologist in Australia, I learned—often from mildly indignant colleagues—not only that many Australian geologists knew about and believed in the idea of continental drift in the 1940s and 1950s but also that in several instances they were ridiculed at international meetings or on visits to the United States by rude and arrogant Americans. I also learned that other theories of crustal mobility, including the expanding earth hypothesis, had been advocated and in some cases were still being advocated by Australian geologists. One Australian who was receptive to the idea of an expanding earth was my employer, the director of exploration of Australia's third-largest mining company, who periodically circulated papers on this topic among his scientific staff. It was evident that the recent history of earth science was much more complex, much more nationalistic, and much more *interesting* than my professors and textbooks—or my readings in the philosophy of science—had ever suggested.

In the early 1980s, I returned to the United States to pursue graduate studies in geology and again encountered a conundrum. My English training and Australian experience had inculcated in me an inductive methodology, in which scientific problems originated in the observation of geological phenomena in the field, but many of my American professors disdained inductive science and what they pejoratively dismissed as "outcrop" geology. They encouraged me to pursue a deductive strategy and to rely primarily on the tools of laboratory analysis. This was particularly true of younger professors and those who had achieved a high level of professional recognition. The issue was not one of theoretical belief but of methodological commitment. My American and British professors promoted contradictory and ultimately incompatible views about the right way to generate scientific knowledge. Two strands began to merge: divergent visions of the recent history of earth science and divergent methodological commitments. Was there some relation between the two? So began the active portion of the inquiry represented by this book.

My debts are thus spread over several continents. My research advisors at Stanford University, Peter Galison and Marco T. Einaudi, encouraged me to pursue the questions raised in this book while still engaging in scientific research. Neither of these men has ever allowed his thinking to be constrained by the historically contingent boundaries of academic disciplines, and for this I am deeply grateful. Among the Stanford faculty, past and present, I am also indebted to Nancy Cartwright, who profoundly influenced my thinking; to Dennis Bird, John Bredehoeft, John Dupree, Jane Maienschein, James O'Neil, Tjeerd van Andel, and Norton Wise; and to fellow graduate students David Magnus, Carey Peabody, Barbara Bekken, Lisa Echevarria Benatar, Peter Mitchell, Hilary and Jon Olson, Nicolas Rose, and Allan Rubin. In Australia I am indebted to Roy Woodall, formerly Director of Exploration of Western Mining Company, and to colleagues at Western Mining and BP Minerals who shared historical anecdotes; at Imperial College, London, to John Knill, Paul Garrod, Angus Moore, Mike Rosenbaum, Ernie Rutter, Richard Sibson, and the late Janet Watson, all of whom inspired me in important ways; and at Dartmouth College to Claudia Henrion, Richard Kremer, and the late, great Chuck Drake, who I dearly wish had lived to read this book.

Throughout this project I have benefited from the intellectual generosity and moral support of many colleagues: Duncan Agnew, Ron Doel, Robert Dott Jr., Mott Greene, David Kaiser, Homer LeGrand, Phil Pauley, Ron Rainger, Martin Rudwick, and Kenneth Taylor commented on the manuscript; Allan Allwardt, Richard Creath, Henry Frankel, Eli Gerson, Carl-Henry Geschwind, Bruce Hevly, Rachel Laudan, Leo Laporte, Chandra Mukerji, Robert Smith, and Don White provided feedback and information; Helen Wright Greuter and Finley Wright gave me access to their father's letters; Michele Aldrich sent historical tips. Among librarians and archivists, I am indebted to Deborah Day of the Scripps Institution of Oceanography, Charlotte Dirksen and Henry

Lowood at Stanford, Barbara DeFelice and the late Susan George at Dartmouth College, Ronald Brashear at the Huntington Library, Susan Vasquez and Shaun Hardy at the Carnegie Institution of Washington, and the staff of the Yale University Libraries. In Ireland, I thank Patrick Wyse Jackson and the staff of the Trinity College Archives; in South Africa, Leonie Twentyman-Jones and Jill Gribble at the University of Cape Town, the family of Alexander du Toit in Johannesburg, and Mr. Gerry Levin of the Geological Society of South Africa, all of whom provided extremely generous help. Thanks are also due to my research assistants Mary Ann Marcinciewicz, Drew Tenenholz, Jennifer Wehner, Virginia Terry, and Laurie Norris; to Karen Endicott, who proofread the entire text; to the staff of the Dartmouth College Day Care Center and the University Plaza Nursery School; and to Joyce Berry and Cynthia Garver at Oxford University Press.

I have had generous financial support from the National Science Foundation (SBE-9222597 and EAR-9357888), the National Endowment for the Humanities (Fellowships for University Teachers), the Mellon Foundation New Directions Fellowship (Stanford University), the William and Mary Ritter Memorial Fellowship of Scripps Institution of Oceanography, and the Burke Research Initiation Award Program at Dartmouth College. For continuous intellectual support and stimulation, I am grateful to my parents, Susan and Irwin Oreskes; my siblings Michael, Daniel, and Rebecca Oreskes; my husband, Kenneth Belitz, and, above all, my children, Hannah and Clara Belitz, who I hope will never stop asking "why?"

New York N. O.
March 1998

CONTENTS

The Rejection of
CONTINENTAL DRIFT

Evidence is hard to come by, it is largely
circumstantial, and there is never enough of it.
—*J. Hoover Mackin*

We only know causes by their effects.
—*William Whewell*

Introduction

The Instability of Scientific Truth

Scientists are interested in truth. They want to know how the world really is, and they want to use that knowledge to do things in the world. In the earth sciences, this has meant developing methods of observation to determine the shape, structure, and history of the earth and designing instruments to measure, record, predict, and interpret the earth's physical and chemical processes and properties. The resulting knowledge may be used to find mineral deposits, energy resources, or underground water; to delineate areas of earthquake and volcanic hazard; to isolate radioactive and toxic wastes; or to make inferences and predictions about the earth's past and future climate. The past century has produced a prodigious amount of factual knowledge about the earth, and prodigious demands are now being placed on that knowledge.

The history of science demonstrates, however, that the scientific truths of yesterday are often viewed as misconceptions, and, conversely, that ideas rejected in the past may now be considered true. History is littered with the discarded beliefs of yesteryear, and the present is populated by epistemic resurrections. This realization leads to the central problem of the history and philosophy of science: How are we to evaluate contemporary science's claims to truth given the perishability of past scientific knowledge? This question is of considerable philosophic interest and of practical import as well. If the truths of today are the falsehoods of tomorrow, what does this say about the nature of scientific truth? And if our knowledge is perishable and incomplete, how can we warrant its use in sensitive social and political decision-making?[1]

For many, the success of science is its own best defense. From jet flight to the smallpox vaccine, from CD players to desktop scanners, contemporary life is permeated by technology enabled by scientific insight. We benefit daily from the liberating effects of petroleum found with the aid of geological knowledge, microchips manufactured with the aid of physical knowledge, materials synthesized with the aid of chemical knowledge. Our view of life—and death—is conditioned by the results of scientific research and the capabilities of technological control. Our moral and political judgments are colored by what science tells us is natural. Nearly every aspect of our lives has been affected by scientific knowledge or technical innovation facilitated by that knowledge. But science and technology have also brought us to the edge of an environmental abyss with its familiar litany of silent crises: global deforestation, ozone depletion, greenhouse effects, groundwater contamination, nuclear annihilation. Most of us stand as mute witnesses to this leviathan: we know surprisingly little about the process of scientific research that has brought us this confusing largesse.

Why do we know so little about how scientific knowledge is generated? Throughout the first half of this century, philosophers and historians of science were concerned primarily with the progression of scientific theories and rarely with the processes by which those theories came to be.[2] The mechanics of scientific research—the choice of questions and methods, the construction and application of instruments, the resolution of contradictions between overlapping data sets—traditionally have received scant attention. The reason was deliberate: analytic philosophers were interested in the logic of scientific theories and their demonstration, not in their genealogy. Viewing science as a rational enterprise, philosophers such as Hans Reichenbach explicitly denied the relevance of the methods by which scientific ideas came to fruition. According to Riechenbach, the philosopher of science was "not much interested in the . . . processes which lead to scientific discoveries. . . . He is interested not in the context of discovery, but in the context of justification."[3]

Reichenbach's ideal has been radically challenged in the last several decades. Most contemporary historians and sociologists of science now believe that the social context of science—the context of discovery—is *more* important that the context of justification.[4] Any theory can be rationally reconstructed to sound logical, even inevitable, in retrospect, but this tells us little about how scientific knowledge actually develops.[5] Scientists are not free agents, historians and sociologists have argued, and the social context of their work not only delimits their options but may even determine the content of their knowledge.[6] And if all knowledge is socially constructed, then objectivity is a chimera. This radical claim strikes at the heart of scientists' beliefs about their enterprise.

Not surprisingly, scientists (and many others) resist such a view and argue that these historians have missed the point. Scientists are interested in truth, and since scientific knowledge works, it must be more or less correct. Human factors sometimes "get in the way" of objective knowledge, but the point of science is to resist such influences. If scientists permit social pressures to distort their perceptions of the physical world, then they are failing to live up to the scientific ideal. Of course, scientists do fail, because they are human, but in most cases it scarcely matters because bad science gets weeded out by the collective scrutiny of the scientific community and—imperfect or not—the surviving knowledge *works*.[7] To attack science simply because it exists within human culture is pettiness, at best. At worst, some believe, it is to attack the very foundations of rationalism.

Historians would argue that it is scientists who have missed the point. *All* science is socially structured, both the good and the bad, and so is the peer review system that adjudicates between them. But can historians prove this point? Historical case studies can illustrate how the development of a particular idea—including our best science—

reflects the constraints of historical situations, and in recent years historians have produced many such studies.[8] But in many such contextualized histories of science, social context is a kind of miasma that pervades scientific thinking in an intangible and ultimately inexplicable manner.[9] The evidence for the role of social forces in the production of scientific knowledge is almost always circumstantial. If scientists are consciously struggling to generate knowledge independent of time and place and historians are claiming that this is an illusory or even meretricious goal, then either scientists are self-deluded or historians are disingenuous.[10] Can these positions be reconciled? Can knowledge be both contingent and transcendent?

Part of the answer may lie in the realm of scientific methods and the process of scientific research. Until recently, few historians, philosophers, or scientists focused their concern on the epistemic status of scientific process.[11] The scientific method, always in the singular, has been taken as monolithic and unproblematic—a textbook cliché, the one sure thing we all know about science. But is there such a thing as *the* scientific method? The answer is clearly no. From the past two decades of historical scholarship, one insight has emerged unequivocally: the methods of science are complex, variegated, and often local. Throughout history scientists have drawn on a wide variety of epistemic commitments and beliefs, linguistic and conceptual metaphors, and material and cognitive resources, all of which have changed with time and varied in space. At different times or in different social contexts, scientists have preferred either inductive or deductive modes of reasoning and argumentation, experimental or theoretical approaches to problem-solving, laboratory- or field-based methodologies. Various attempts have been made to prove the superiority and permanently establish the hegemony of particular methodological approaches, but most of these attempts have failed. The diversity of scientific methodology persists. And whatever methods scientists have chosen, it is through these choices that representations of the natural world have been forged from the infinite sea of sensory perceptions. Scientific practices are the tools with which scientists link the phenomenological world and their representations of that world. Perhaps they provide the link between science and the human world as well.

The Problem of Continental Drift

The story of continental drift illustrates how choices about methods constrained possibilities for scientific truth. In the early part of the twentieth century, a number of scientists—most notably Alfred Wegener—proposed that the relative positions of the continents of the earth were not fixed. Among the leading North American scientists of the 1920s and 1930s—members of the U.S. National Academy of Sciences, presidents of scientific societies, professors at distinguished universities, and men for whom we have medals named today—continental drift was widely discussed and almost uniformly rejected, not merely as unproved, but as wrong, incorrect, physically impossible, even pernicious. American scientists were much more hostile to the idea than their European counterparts; some even labeled the theory *unscientific*.[12] Yet forty years later, the basic central idea of continental drift—that the continents are not fixed but move horizontally over the face of the earth—was established as scientific fact. Why did distinguished scientists adamantly reject as false a claim now universally accepted as true?[13]

The standard answer, to be found in most geology textbooks and many works on the history of science, is that continental drift was rejected for lack of an adequate causal mechanism. Because scientists could not explain *why* continents moved, they concluded that they *could not move.* But this is an example of a rational reconstruction in the history of science. Because moving continents are now accepted as fact, scientists have tended to assume that if the idea was rejected by earlier scientists, then those scientists must have had a good reason to reject it. Perhaps there was not enough evidence. Perhaps the arguments were inadequately articulated. Perhaps the people who proposed the ideas were not well-known, or published their ideas in obscure places. When scientists and historians in the 1960s and 1970s looked back at Alfred Wegener's work, they found an obvious deficit in that the mechanism he proposed was not the same as that accepted today. Indeed, it was patently implausible by today's standards. The implausibility of Wegener's mechanism was taken as the obvious explanation of why his theory was rejected.

But this explanation is anachronistic. What matters in historical argument is not what seems plausible to us, looking backwards, but what was said and done at the time the events took place. What matters is what was plausible to *them.* Viewed this way, the standard account is demonstrably false. Part I of this book shows how Wegener's mechanism was not considered hopelessly implausible at the time it was first proposed; it drew on a large body of work in geodesy and physical geology that pointed to flow in a plastic substrate beneath the crust. Wegener argued that this flow, widely accepted in the 1920s as a demonstrated phenomenon, could help to account for continental drift. Broadly construed, Wegener's argument holds today. But more important, Wegener's proposed mechanism was just the opening round in a series of discussions and debates about the mechanism of continental drift. By the early 1930s, a number of mechanisms had been proposed for continental drift, including the one that is generally accepted as the driving force of plate tectonics: convection currents in the earth's asthenosphere.

So why was continental drift rejected? The thesis of this book is that American earth scientists rejected the theory of continental drift not because there was no evidence to support it (there was ample), nor because the scientists who supported it were cranks (they were not), but because the theory, as widely interpreted, violated deeply held methodological beliefs and valued forms of scientific practice. The idea of the motion of continents, the empirical evidence for it, and the mechanical explanation of it developed by Arthur Holmes have all been corroborated by contemporary earth science.[14] But to accept these ideas in the 1920s or early 1930s would have forced American geologists to abandon many fundamental aspects of the *way they did science.* This they were not willing to do.

The conclusion of this book, therefore, is that *science is not about belief; it is about how belief gets formulated.* At any given moment, only a finite set of knowledge satisfies the reigning criteria for the formulation of scientific belief, and only *this* knowledge is eligible as truth. But the discriminating criteria are historically contingent; over time and across communities, they shift, they evolve, they are overthrown, they transmute. The changing criteria for the formulation of belief provide the pathways through which cultural context delimits the boundaries of scientific knowledge.

I

NOT THE MECHANISM

Joly and Holmes have a beautiful theory,
and I believe it will be epoch-making.
—*Chester Longwell, January 1926*

This page intentionally left blank

1

Two Visions of the Earth

Plate tectonics is the unifying theory of modern geology. This theory, which holds that the major features of the earth's surface are created by horizontal motions of the continents, has been hailed as the geological equivalent of the "theory of the Bohr atom in its simplicity, its elegance, and its ability to explain a wide range of observation," in the words of A. Cox.[1] Developed in the mid-1960s, plate tectonics rapidly took hold, so that by 1971, Gass, Smith, and Wilson could say in their introductory textbook in geology:

> During the last decade, there has been a revolution in earth sciences . . . which has led to the wide acceptance that continents drift about the face of the earth and that the sea-floor spreads, continually being created and destroyed. Finally in the last two to three years, it has culminated in an all-embracing theory known as "plate tectonics." The success of plate tectonics theory is not only that it explains the geophysical evidence, but that it also presents a framework within which geological data, painstakingly accumulated by land-bound geologists over the past two centuries, can be fitted. Furthermore, it has taken the earth sciences to the stage where they can not only explain what has happened in the past, and is happening at the present time, but can also predict what will happen in the future.[2]

Today moving continents are a scientific fact. But some forty years before the advent of the theory of plate tectonics, a very similar theory, initially known as the "displacement hypothesis," was proposed and rejected by the geological fraternity. In 1912, a German meteorologist and geophysicist, Alfred Wegener, proposed that the continents of the earth were mobile; in the decade that followed he developed this idea into a full-fledged theory of tectonics that was widely discussed and debated and came to be known as the theory of continental drift. To a modern geologist, raised in the school of plate tectonics, Wegener's book, *The Origin of Continents and Oceans*, appears an impressive and prescient document that contains many of the essential features of plate tectonic theory.[3] His ideas were substantiated by an impressive volume of data culled from diverse branches of earth sciences. Yet Wegener's theory was rejected by most of his contemporaries. In the British Commonwealth and in continental Europe, he gained a minority following; in the United States his ideas were resoundingly rejected and in some quarters ridiculed. Thirty years after the last edition of Wegener's book was published, however, a major revolution in earth sciences occurred that incorporated many of the essential features of Wegener's theory. In light of the considerable overlap between Wegener's concepts and a modern theory that commands virtually universal support, why was the theory of continental drift rejected?[4] Why did American geologists in particular react so negatively to the idea of moving continents? To answer these questions, we must first ask where the theory of continental drift came from and what it sought to replace.

Tectonics in the Nineteenth Century

One of the outstanding problems of early-twentieth-century geology was the origin of mountains. Throughout the eighteenth and nineteenth centuries, geologists had focused interest on mountain ranges, for their economic importance as physiographic boundaries and sources of mineral resources, for their scientific importance as repositories of large-scale exposures of the earth's internal structure, and for the psychic impression made by their size and beauty. In the nineteenth century, causal theories of mountain-building became a primary focus of geological concern in both the United States and Europe.[5] Theorists such as Eduard Suess in Austria and James Dana in the United States proposed that mountains formed through compressive stresses generated by a gradual thermal contraction of the whole earth. But these two men—and the colleagues they influenced—held very different views of contraction.

Suess's Collapsing Earth

Eduard Suess (1831–1914) was a field geologist and professor of geology at the University of Vienna (figure 1.1) who dedicated his career to unraveling the structure of the Swiss Alps. From 1883 to 1904, Suess published his magnum opus, a four-volume treatise entitled *Das Antlitz der Erde,* in which he articulated the theory of thermal contraction as the earth's mechanism of mountain-building. (These volumes were published in English between 1904 and 1909 as *The Face of the Earth,* translated by the English geologist Hertha Sollas.) Suess likened the process of terrestrial contraction to the wrinkling of the skin of a desiccating apple and suggested that mountains resulted from a wrinkling of the earth's crust to accommodate diminishing surface area. The earth's contraction, he proclaimed, allowed us to witness "the breaking up of the terrestrial globe."[6]

According to Suess's theory, the features of the earth were explained in terms of vertical motions of its crust. Initially, the entire earth was covered by a continuous continental crust. As the earth cooled and began to shrink, portions of the outer crust collapsed and began to form the ocean floor, causing the earth to become differentiated into continents and oceans. With continued cooling, the remaining uplifted portions were undermined and became unstable, and they collapsed to form the next generation of ocean basins. What had formerly been ocean became dry land. With

Figure 1.1. Eduard Suess. (From Zittel 1901, p. 513.)

each additional increment of cooling, what had been continent became ocean basin and what had been ocean became continent.

The continual transformation of oceans into continents and vice versa explained a number of puzzling geological phenomena, such as the presence of marine deposits on dry land and the numerous alternating episodes of marine and terrestrial conditions recorded in stratigraphic successions. Perhaps most important, the theory explained the well-known similarity of fossil assemblages in widely separate continents: animals and plants had migrated and been dispersed across now sunken continents. Darwin had explained the divergence of plant and animal species as the result of natural selection in isolated communities under contrasting environmental conditions, but paleontologists had found the *same* fossil forms in widely separated continents with radically different biological and climatic environments. Paleontologists concluded that these areas must once have been contiguous and must later have broken apart. Suess's world of gradually decreasing environmental continuity could account for gradually increasing faunal diversity.

This evidence from historical geology was a primary motivation for most paleontologists and stratigraphers to accept the concept of "Gondwanaland"—named after the Gondwana system of India—the term used by Suess to describe the giant supercontinent that had once united much or all of the globe.[7] Sunken portions were referred to as "land bridges" across which ancient species had traveled (although the broad swatches of sunken continents were more like platforms than bridges). But Suess's theory also had consequences for structural geologists: it suggested that mountains would occur across the globe—on all continents and all parts of continents—a prediction borne out by the widespread occurrence of mountains throughout Europe and nearby parts of Asia and North Africa. The collapse of continental blocks was also linked to the origin of igneous intrusions and volcanism, as molten rock escaped from the earth's interior along radial cracks formed during periods of collapse.

Contraction theory was a unifying account of global progression that explained both the history of life and the history of the planet on which life evolved. Writing in

1901, Karl Zittel, professor of geology at Munich and president of the Bavarian Royal Academy of Sciences, triumphantly (and to his mind generously) declared:

> In the hands of one of the most accomplished of foreign geologists and one of the strictest logicians of any age, crust-tectonics [has] been elevated into a new inductive philosophy of earth-configuration. . . . A work like that of Suess, so cosmopolitan in its standpoint, reminds all workers of their community of aim, rouses each one from the particular to the general, and brings him back with renewed vigor and mental insight to the particular. The time was ripe for an effort to establish systematic clearness in the acquired abundance of detail and to seek for comprehensive laws and principles.

Quoting Marcel Bertrand, pioneer in determinative mineralogy and petrology, Zittel continued, "The creation of a science, like that of a world, demands more than a single day; but when our successors write the history of our science, I am convinced that they will say that the work of Suess marks the end of the first day, *when there was light.*"[8]

The European Tradition of Secular Cooling

Suess's work was the culmination of a long-established tradition of interpreting earth history in terms of secular cooling, an idea promulgated in the early nineteenth century by geologists such as Léonce Elie de Beaumont and Henry De la Beche. In 1834, De la Beche had written:

> If we suppose with M. Élie de Beaumont, that the state of our globe is such that, in a given time, the temperature of the interior is lowered by a much greater quantity than on its surface, the solid crust would break up to accommodate itself to the internal mass; almost imperceptibly when time and the mass of the earth are taken into account, but by considerable dislocations according to our general ideas on such subjects.[9]

The first director of the British Geological Survey and founder of the Royal School of Mines, Henry De la Beche (1796–1855) was a principal proponent of the thermal contraction theory in Great Britain, which stood in opposition to Charles Lyell's uniformitarianism, which subsequently has been better known. Lyell argued that past processes must be interpreted in terms of presently observable and incremental causes, even when the empirical evidence on the face of it might suggest otherwise. Thus large gaps in the rock record—unconformities—could and should be explained by slow uplift and erosion, rather than by a sudden event that stripped away the missing strata. In contrast, De la Beche argued that there was strong empirical evidence and good theoretical justification for episodic geological upheaval, progressively decreasing in intensity through geological time.

De la Beche found support for his progressionist views among many geologists who have been labeled in retrospect "catastrophists," such as Adam Sedgwick, who objected to the *a priori* and rigid character of Lyell's philosophy, and William Conybeare, who found steady-state uniformitarianism inadequate to account for many aspects of the geological record, including the obvious power of past orogenic activity displayed in mountain belts.[10] In 1831, De la Beche arranged for publication of part of Elie de Beaumont's theory in the British *Philosophical Magazine,* and his own *Manual of Geology* (1831) and *Researches in Theoretical Geology* (1834) drew heavily on his French colleague's work. Paralleling the French geologist's arguments, De la

Beche proposed a series of periodic, moderate upheavals in earth history to account for the earth's distinct mountain chains.

De la Beche's argument for greater intensity of force at some intervals during geological history was subsequently seconded by William Whewell in his famous account of "The Two Antagonist Doctrines of Geology."[11] Using an empiricist argument against Lyell's "pseudo-empirical" uniformitarianism, Whewell wrote:

> It must be granted at once, to the advocate of . . . geological uniformity, that we are not arbitrarily to assume the existence of catastrophes. The degree of uniformity and continuity with which terramotive forces have acted, must be collected, not from any gratuitous hypothesis, but from the facts of the case. We must suppose the causes which have produced geological phenomena, to have been as similar to existing causes, and as dissimilar, as the effects teach us. . . . But when Mr. Lyell goes further, and considers it a merit in a course of geological speculation that it *rejects* any difference between the intensity of existing and past causes, we conceive that he errs no less than those he censures. . . . We are in danger of error, if we seek for slow causes and shun violent agencies further than the facts naturally direct us, no less than if we were parsimonious of time and prodigal of violence. *Time*, inexhaustible and ever accumulating his efficacy, can undoubtedly do much for the theorist in geology; but *Force*, whose limits we cannot measure, and whose nature we cannot fathom, is also a power never to be slighted: and to call in the one to protect us from the other, is equally presumptuous, to whichever of the two our superstition leans.[12]

Despite Whewell's warnings about the excesses of Lyellianism, De la Beche was overpowered by the more eloquent and more adamant Lyell (although thermal contraction would be taken up again later in the century by British geophysicists Osmond Fisher and George Darwin).[13] In contrast, Léonce Elie de Beaumont (1798–1874) became one of the most influential figures in French science of the mid-nineteenth century. A founder of the Société Géologique de France, Elie de Beaumont was inspired by his teacher at L'Ecole des Mines, René Just Haüy, to look for mathematical regularity in geological phenomena. Professor of physics and mineralogy, Haüy pioneered the science of crystallography and is credited with formulating the mathematical law of rational intercepts to explain crystal morphology and symmetry.[14] As a field geologist with the Corps des Mines, responsible for producing the first geological map of France and delineating the contained mineral resources, Elie de Beaumont sought to find in macroscopic geological phenomena the geometric regularity his teacher had found within microscopic phenomena.

The geometric analysis of geological strata was well established in mines, where tracing of dipping and intersecting strata was essential for exploiting ore horizons and coal seams. But Elie de Beaumont wanted to extend this form of analysis to larger scale geological features, including the largest of all—mountain belts.[15] One key empirical observation was that the orientations of folded structures within mountain belts were not random: deformed structures of similar age tended to be oriented along similar trends, and the observed regularities commonly extended over large geographical areas. Systems of mineral veins within mountain chains characteristically followed a preferred orientation as well. Elie de Beaumont concluded that the orientation of a mountain chain was related to the age of its formation and that the forces involved were global in scale: all of the world's mountains could be connected along a series of great circles that defined a pentagonal network, or *réseau pentagonal*, along which global contraction occurred. Building on the ideas of his predecessor, Georges Cuvier (1769–1832), who had proposed periodic organic upheaval to explain the extinction of fossil species, Elie de Beaumont proposed periodic physical upheaval, caused by gradual secular cooling and thermal contraction of the earth.[16]

Elie de Beaumont's notion of secular cooling was drawn from a consensus forged by several of the leading scientists of seventeenth- and eighteenth-century Europe. Descartes, Leibniz, and Buffon had all supposed that the earth had formed by cooling from a molten mass. Using this assumption, Buffon (1707–1788) produced one of the earliest estimates of the age of the earth—approximately 70,000 years.[17] In 1796, the origin of the solar system by condensation from a hot gaseous cloud was developed theoretically by Laplace, in what came to be known as the Kant–Laplace nebular hypothesis, and the theory was widely accepted by European scientists in the nineteenth century.[18]

If the earth had formed by consolidation from a primordial nebular gas, this had obvious consequences for its structure and subsequent history. One consequence was that the earth was probably still cooling. This seemed to be supported by the available empirical evidence: it was well known that mines were hot, and the temperatures in mines greatly increased with depth. Volcanic phenomena also pointed to a hot, if not necessarily molten, interior. In 1837, Joseph Fourier applied his new method of mathematical analysis to the theory of secular cooling and concluded that the shape of the earth and its thermal gradient proved that it must have formed at high temperature and have been still in the process of cooling from this original hot state.[19] By the mid-nineteenth century, the hot origin of the earth and its subsequent conductive cooling was a well-established scientific principle, one that could be drawn upon in developing theories of the earth's geological history and evolution. Cuvier's and Elie de Beaumont's *révolutions* were part of the normal evolution of the earth's crust. Periodic upheaval was the consequence of continuous secular cooling and resulted in lateral pressure along zones of weakness within the earth. The geological consequence was the formation of mountain belts.[20]

Suess's *Das Antlitz der Erde* was thus the culmination both of one man's life's work and of a long-standing and distinguished tradition in continental European science bound to a dynamic, changing earth. But while contraction theory in Europe was viewed in opposition to Lyell's steady-state uniformitarianism, the situation in the United States was different. There, James Dana had developed an entirely different version of contraction in which oceans and continents were permanent features of the earth's crust.

Dana's Stable Continents and Permanent Oceans

James Dwight Dana (1813–1895) was a professor of geology at Yale University in the latter half of the nineteenth century, the son-in-law of Benjamin Silliman, America's first professor of geology and founder of the *American Journal of Science* (figure 1.2).[21] As a student under Silliman at Yale, Dana focused his early work on the chemistry and properties of minerals. In 1837, at the age of 24, he published his *System of Mineralogy*, a book that remains the basis of modern mineralogy texts.[22] The following year, Dana joined the Wilkes expedition—America's first federally financed scientific expedition. After his return, he married his mentor's daughter and published exten-

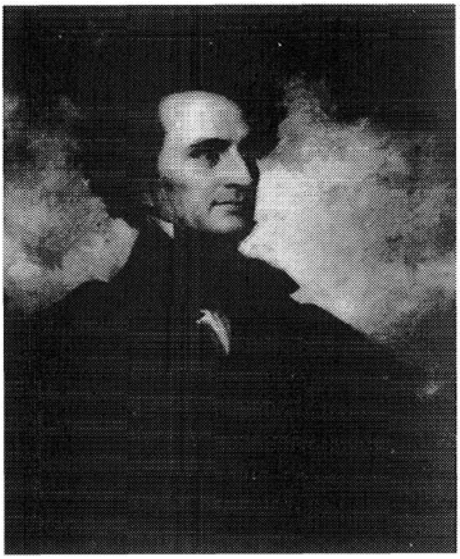


Figure 1.2. James Dwight Dana. (Reproduced with permission of Yale University Art Gallery, Bequest of Edward Dana Salisbury, B.A., 1870.)

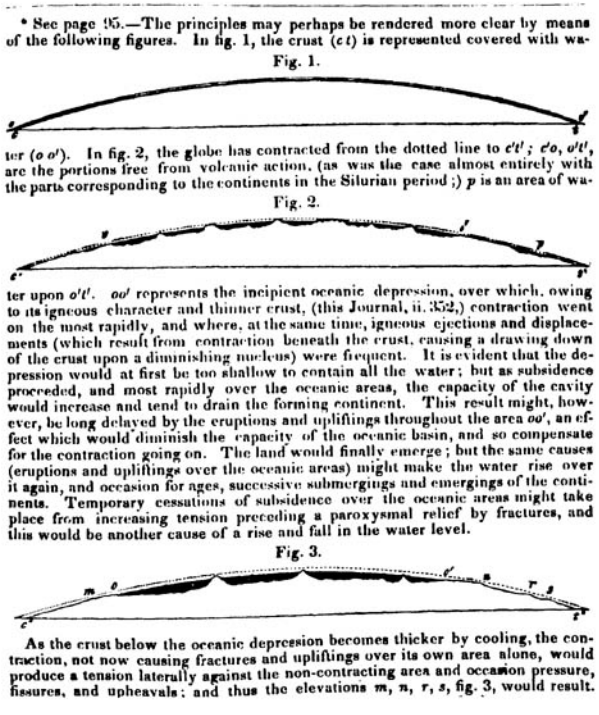
sively on the results of his observations of geology, volcanology, and paleontology in the Pacific region. In 1850, Dana succeeded his father-in-law as Silliman Professor of Geology at Yale, where he continued his work on the results of the Wilkes expedition. Following in his father-in-law's footsteps, Dana served as editor of the *American Journal of Science* from 1846 to 1890, and he passed the position on to his own son, Edward S. Dana.

In addition to his work on the Wilkes expedition, Dana published some 250 scientific articles on a wide range of subjects, including mineralogy, petrology, and dynamical geology, and a number of popular and pedagogical texts. His *Textbook of Geology* (1863) was widely used in burgeoning geology courses around the United States; this book, along with his position as editor of the *American Journal of Science*, helped to establish Dana as one of the most prominent academic geologists of late-nineteenth-century America.[23]

Dana's interests were both systematic and synthetic, and in the mid-1840s he began work on a theory of the origins of the earth's major physiographic features. His theory rested on three legs: his detailed knowledge of mineralogy, his experience and observations on the Wilkes expedition, and his reading about the properties of the lunar surface. Astronomical observations indicated that the lunar surface was heterogeneous, cratered and mountainous, and perhaps not unlike the surface of the earth.[24] Presuming that lunar craters were similar to the volcanic craters he had seen in the Pacific Islands, Dana theorized that they had formed from similar materials and by similar processes. Dana connected this supposition to the mineralogical observation that different minerals had different melting temperatures and proposed that the lunar topography had formed when the surface of the moon first became solid in an inhomogeneous manner, just as "a melted globe of lead or iron . . . when cooling unequally, becomes depressed by contraction on the side which cools last."[25] The lunar craters were the areas of lowest melting temperature; the lunar highlands, the areas of highest temperatures. By analogy, Dana suggested that the continents and oceans of the earth were primordial features formed when the earth solidified from a molten globe.

> The areas of the surface constituting the continents were first free from eruptive fires. These portions cooled first, and consequently the contraction in progress affected most the other parts. The great depressions occupied by the oceans thus began; and for a long period afterward, continued deepening by slow, though it may have been unequal, progress.[26]

In Dana's view, the oceans were permanent, primordial features that continued to be the site of thermal contraction for a long period of geological history; Charles Darwin's study of the subsidence of coral islands in the Pacific Ocean suggested that oceanic sinking still continued. Dana explained the presence of marine deposits on the continents as the product of a former era when "the depth [of former oceans] would be too shallow to contain the seas; and consequently the whole land would be under water."[27] The occurrence of mountain ranges on the edges of continents Dana attributed to a "lateral pressure" induced by "the greater subsidence of the oceanic parts relative to the continental."[28] As the whole earth contracted, its rocks would be squeezed, and the squeezing would be greatest on the continental margins, which were the sites of greatest differential stress.[29] Local vertical uplift might occur in regions subjected to the expansive force of ascendant magma, but the net direction of crustal movement was down (figure 1.3).



* See page 95.—The principles may perhaps be rendered more clear by means of the following figures. In fig. 1, the crust (*c t*) is represented covered with water (*o o'*). In fig. 2, the globe has contracted from the dotted line to *c't'*; *c'o*, *o't'*, are the portions free from volcanic action. (as was the case almost entirely with the parts corresponding to the continents in the Silurian period;) *p* is an area of water upon *o't'*. *oo'* represents the incipient oceanic depression, over which, owing to its igneous character and thinner crust, (this Journal, ii. 352,) contraction went on the most rapidly, and where, at the same time, igneous ejections and displacements (which result from contraction beneath the crust, causing a drawing down of the crust upon a diminishing nucleus) were frequent. It is evident that the depression would at first be too shallow to contain all the water; but as subsidence proceeded, and most rapidly over the oceanic areas, the capacity of the cavity would increase and tend to drain the forming continent. This result might, however, be long delayed by the eruptions and upliftings throughout the area *oo'*, an effect which would diminish the capacity of the oceanic basin, and so compensate for the contraction going on. The land would finally emerge; but the same causes (eruptions and upliftings over the oceanic areas) might make the water rise over it again, and occasion for ages, successive submergings and emergings of the continents. Temporary cessations of subsidence over the oceanic areas might take place from increasing tension preceding a paroxysmal relief by fractures, and this would be another cause of a rise and fall in the water level.

As the crust below the oceanic depression becomes thicker by cooling, the contraction, not now causing fractures and upliftings over its own area alone, would produce a tension laterally against the non-contracting area and occasion pressure, fissures, and upheavals; and thus the elevations *m*, *n*, *r*, *s*, fig. 3, would result.

Figure 1.3. Dana's representation of contraction theory. In Dana's model of the earth, the oceans and continents formed by differential thermal contraction early in earth history. (From Dana 1847c.)

In contrast to Suess's work, a consequence of Dana's theory was that mountains would only form on the edges of continents, and this was borne out by the Appalachians, the principal mountain chain of American experience. Dana's theory also suggested that the highest mountains should correlate with the greatest depth of adjacent ocean floor, and this, too, seemed to be confirmed by his observations of the Andes on the Wilkes expedition. Because continents and oceans were primordial and not interchangeable features, the theory implied that the continents would be composed of different material from the ocean floor; this seemed to be confirmed by what little was known about the ocean floor from dredging and from the study of volcanic islands. On the other hand, unlike Suess's later account, Dana's theory had no explanation for why fossil assemblages should be so alike in distant places.

Because continents and oceans were not interchangeable in Dana's version of contraction, the concept came to be known as "permanence theory." Although clearly a type of contraction theory, resting as it did on the premise of global cooling, Dana's work was perceived by his American colleagues as standing in opposition to the "contraction" hypothesis articulated by Suess. Although starting from the same premise, Suess and Dana had come to very different conclusions about the present state of the globe. Permanence theory became widely accepted in the United States; in the early twentieth century, the permanence of ocean basins was considered by most geologists to be a fact. In 1910, Bailey Willis, a well-known professor of geology at Stanford University, wrote emphatically in the journal *Science* that *"the great ocean basins are permanent features of the earth's surface and they have existed, where they are now, with moderate changes of outline, since the waters first gathered."*[30]

In some ways "permanence" theory was a misnomer, because Dana believed that the earth was still contracting and therefore still changing. But how was it changing? Dana had still to address the central question of tectonics—the origins of mountains. To do this, he had to incorporate, and to some extent subsume, the work of his foremost adversary, James Hall Jr.

Permanent Oceans and the Geosynclinal Theory

The principal mountain belt of North American experience was the Appalachians, and any American theory of mountain-building would have to account for its geology. Many of the details of Appalachian geology had been worked out by two of the United States' most famous early geologists, brothers Henry Darwin Rogers and William B. Rogers, who mapped large portions of the Appalachians in the 1830s and 1840s.[31] One of the principal conclusions of their work was that the asymmetric pattern of deformation and overthrusting evident throughout the Appalachians indicated a compression from east to west—suggesting large-scale tangential movement of the type advocated by De la Beche and Elie de Beaumont as the consequence of global contraction. Rather than interpreting these asymmetric folds as the result of thermal contraction, however, the Rogers brothers attributed them to the disturbing effects of thermal expansion of molten material and gases beneath the crust.

> We suppose the strata of [a folded] region to have been subjected to excessive upward tension, arising from the expansion of the molten matter and gaseous vapors, the tension relieved by linear fissures, through which much elastic vapor escaped, the sudden release of pressure adjacent to the lines of fracture, producing violent

> pulsations on the surface of the liquid below. This oscillating movement of the [underlying] fluid mass would communicate a series of temporary flexures to the overlying crust and those flexures would be rendered permanent . . . by the intrusion of molten matter. If, during this oscillation, we conceive the whole heaving tract to have been shoved (or floated) bodily forward in the direction of the advancing waves, the union of the tangential with the vertical wave-like movement will explain the peculiar steepening of the front side of each flexure, while a repetition of similar operations would occasion the folding under, or inversion, visible in the more compressed districts.[32]

The Rogers' complex interpretive framework was not widely accepted by their contemporaries, but it did alert people to the "stupendous mechanical problems" involved in interpreting the complex fold-thrust belts of Appalachian geology.[33] It also motivated James Hall and James Dana to consider these problems in greater detail.

James Hall (1811–1898), State Paleontologist of New York, was the first president of the Geological Society of America (1889), a charter member of the U.S. National Academy of Sciences (1863), and an early president of the American Association for the Advancement of Science.[34] Hall emphatically rejected the Rogers' dynamic interpretation as unsupported by empirical evidence. However, he was greatly impressed by their documentation of the thick sequences of sedimentary rock that had been deformed to create the Appalachian chain and set out to explain this important observation. Borrowing an idea first proposed by John Herschel, Hall proposed the concept that Dana later named a "geosyncline": an area of subsidence and large-scale sedimentary accumulation adjacent to the continents.

Hall suggested that large accumulations of sediments along continental margins contributed to the depression of that continental margin in a kind of positive-feedback loop that resulted in the gradual buildup of a huge sedimentary pile. Sediments, then, accumulated along continental margins; the weight of these sediments depressed the crust along those margins; and this permitted more sediments to accumulate. Eventually, the pile would become so thick that the sedimentary material would be heated and compressed, and, by some mechanism not exactly explicated, it would be uplifted and deformed into a marginal mountain range. Thus the origin of mountains was causally linked to the accumulation of sediments along continental margins, and any apparent periodicity of mountain-building was a consequence of the time required for sedimentary accumulation.[35]

Dana famously criticized Hall's hypothesis as a theory of the elevation of mountains in which "the elevation of mountains is left out"—because the actual mechanism of uplift remained unexplained—and sought to replace Hall's interpretation of the Appalachians with his own.[36] He argued that thick sedimentary piles were not the *cause* of depressions along continental margins but the *result* of them. The Rogers' pattern of asymmetric folding from east to west was clear evidence of lateral compression generated along the American continental margin by the differential thermal contraction of the Atlantic basin and the adjacent North American continent. The zone of sedimentary accumulation formed in response to stresses caused by this differential contraction (not the weight of the sediments themselves); its subsequent deformation was the consequence of the same effect. Although Hall vigorously opposed Dana's theory as excessively speculative, Dana's subsumation of Hall's work was highly effective, and later American colleagues came to see their views as complementary, referring to the "geosynclinal theory of Hall and Dana."[37]

The details of geosyncline theory would be debated well into the twentieth century, but agreement had been forged on a number of counts. Thick piles of sediments accumulated on continental margins where they were altered by heat and pressure—that much was clear—and somehow they became deformed and uplifted into mountain ranges along those margins. Dana's theory gave an account—albeit incomplete—of why this was so: differential contraction produced lateral pressure along the continental margins, which alternately resulted in downfolds and sedimentary accumulation, as well as uplift and deformation of the same. Thus the tallest mountains occurred next to the deepest oceans, and nearly all mountain chains occurred along continental margins.

Dana's account also explained the occurrence of multiple episodes of deformation within a given mountain chain. Continual contraction resulted in repeated episodes of downwarping, sedimentary accumulation, compression, and uplift, and the result was the gradual growth of the continents through accretion of successive mountain belts. Dana's was a unifying theory, bringing together the best of American field work with the long-standing European tradition of interpretation based on the premise of secular cooling.

Two Continents, Two (or More) Theories

At the end of the nineteenth century, geology was an international science, and geologists frequently traveled abroad to visit colleagues and to see sites of interest. Nevertheless, European and American geologists found themselves subscribing to incompatible views of earth evolution. From the same starting point—the secular cooling of the earth—two different pictures emerged. In the European view, the earth was in a state of continual flux with complete interchangeability of its parts. Ocean basins could be elevated into continents, continents could collapse to form ocean basins, and change occurred across the globe. In the American view, the basic outlines of the earth had been set at the beginning of geological time and had not changed fundamentally since then. Continents were always continents, oceans were always oceans, and change was confined to discrete zones at the interface between them. The two theories also differentially weighed the available facts. The American perspective emphasized the physical properties of minerals, the contrasting compositions of continental rocks and the ocean floor, and the asymmetry of folding in the Appalachians. The European view emphasized the biogeographical patterns, the stratigraphic evidence of interchangeability of land and sea, and the diverse patterns of folding in European and African mountain belts.

In Great Britain—which had little in the way of mountains to explain—neither theory was entirely accepted. A tension persisted between the uniformitarian and progressionist perspectives, in which field-oriented geologists leaned mostly toward the former and mathematically oriented physical geologists held to the latter. On the face of it, geologists who accepted the uniformitarian view might have been inclined toward Suess, whose earth history was less obviously directional than Dana's. But the *tone* of Suess's theory was cataclysmic—with its collapsing continents and "breaking up of the terrestrial globe." In contrast, Dana's theory was explicitly progressionist—Dana believed that the earth's heat was running down and the globe would ultimately decay to darkness—but the tone of his theory was gradualistic, if not nearly steady-state.[38] The major patterns of the earth had been established at the beginning of geological time, and the subsequent changes were local and gradual. For progressionists, Dana's theory was sufficiently progressive, but for uniformitarians, it was easily interpreted as portraying an essentially steady-state earth.

Thus, by the start of the twentieth century, two different versions of earth history were being handed to the next generation: Europeans inherited a vision of a constantly shifting earth—with an unsteady history and an uncertain future—whereas Americans inherited a vision of a much more stable place.

2

The Collapse of Thermal Contraction

In 1901, Karl Zittel, president of the Bavarian Royal Academy of Sciences, declared that "Suess has secured almost general recognition for the contraction theory" of mountain-building.[1] This was wishful thinking. Suess's *Das Antlitz der Erde* was indeed an influential work, but by the time Suess finished the final volume (1904), the thermal contraction theory was under serious attack. Problems were evident from three different but equally important quarters.

Horizontal—Not Vertical—Displacement in the Swiss Alps

The most obvious problem for contraction theory arose from field studies of mountains themselves. As early as the 1840s, it had been recognized that the Swiss Alps contained large slabs of rock that appeared to have been transported laterally over enormous distances.[2] These slabs consisted of nearly flat-lying rocks that might be construed as undisplaced, except that they lay on top of younger rocks. In the late nineteenth century, several prominent geologists, most notably Albert Heim (1849–1937), undertook extensive field work in the Alps to attempt to resolve their structure. Heim's detailed field work, beautiful maps, and elegant prose convinced geological colleagues that the Alpine strata had been displaced horizontally over enormous distances. In some cases, the rocks had been accordioned so tightly that layers that previously extended horizontally for hundreds of kilometers were now reduced to distances of a few kilometers. But in even more startling cases, the rocks were scarcely folded at all, as if huge slabs of rocks had been simply lifted up from one area of the crust and laid down in another.

Heim interpreted the slabs of displaced rock in his own Glarus district as a huge double fold with missing lower limbs, but in 1884 the French geologist Marcel Bertrand (1847–1907) argued that these displacements were not folds but faults. Large segments of the Alps were the result of huge faults that had thrust strata from south to north, over and on top of younger rocks.[3] August Rothpletz (1853–1918), an Austrian geologist, realized that the Alpine thrust faults were similar to those that had been earlier described by the Rogers brothers in the Appalachians.[4] By the late 1880s, thrust faults had been mapped in detail in North America, Scotland, and Scandinavia.

In many cases the thrust faults appeared to have formed along the bottom of huge recumbent folds—folds whose axial planes were horizontal, rather than vertical, implying that they had been formed by compression on one side only (figure 2.1). Apparently, these folds had been subjected to so much lateral compression that they had rolled over on their sides and dislocated from the underlying strata. By the end of the century, the existence of such nappes, as they came to be called, was considered established, and in 1903 Pierre Termier (1859–1930), formerly a student of Bertrand, used the theory of nappes to produce the first unified geological interpretation of the European Alps.[5]

Although considerable argument ensued over the mechanism of nappe formation, most Alpinists accepted the reality of their existence. The most widely cited account of how they might form came from Heim. On the basis of detailed microscopic examination, Heim demonstrated that fossil molluscs in rocks from the basal portions of the Glarus nappes had been flattened and stretched up to ten times their original lengths. Heim suggested that during recrystallization the rocks had behaved in a plastic manner, thus accounting for the stretched rather than broken fossils, allowing the

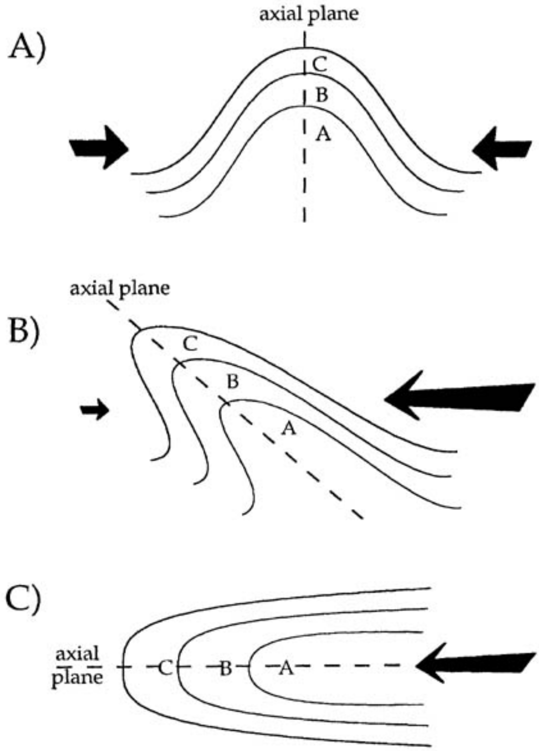


Figure 2.1. The formation of nappes. (**A**) Most geological folds have near-vertical axial planes, with strata sloping outward on both sides, implying bilateral compression. (**B**)However, if compression is asymmetrical, the fold will incline away from the direction of maximum stress and the dip of the strata will be steeper on one side of the fold than on the other. (**C**) If the asymmetry of compression is great enough, the fold will "roll over" onto its side; in this case, the axial plane and the strata will lie in a nearly horizontal position—this is a nappe. Note that the sequence of strata on the upper limb is the same as if the rocks had not been folded at all, but on the lower limb the order is reversed. (Illustration by Melody Brown.)

rocks to flow rather than crumple under pressure. In his classic work on the geology of the Alps, *Untersuchungen über den Mechanismus der Gebirgsbildung,* published in 1878, Heim postulated an upper "zone of fracture" in the crust underlain by a lower "zone of flow." But regardless of the exact mechanism, the geological evidence of lateral displacement was clear. So were the theoretical implications: to account for these huge displacements by terrestrial shrinkage would have required an impossibly large primordial earth.[6]

On Heim's original double-fold theory, the strata of the Glarus district were interpreted to have been compressed to 50 percent of their original length; now it appeared that they had been reduced to 20 percent or less. Such extreme shortening could not be a consequence of terrestrial contraction. Contrary to conventional geological wisdom, there was simply no way to shrink the earth enough to generate this much compression. Some other mechanism was required. Dana's version of contraction theory could perhaps withstand this attack: his theory limited mountains to the edges of ocean basins, and extreme shortening in focused regions was perhaps plausible; it need not be extrapolated to the entire earth. But Suess's version of global generation of mountains by contraction appeared to be completely at odds with these data, gathered in his own geological backyard. It was becoming evident that mountains were not caused by vertical movements of the crust, as contraction theory would have it, but by focused horizontal shortening. And this was as true in the Appalachians as it was in the Alps. Speaking on the occasion of his presidential address to the Geological Society of America in 1893, J. William Dawson declared, "[Both] the Alps and the Appalchians are mountains of crumpling, showing evidence of enormous lateral pressure proceeding from the adjoining sea basins, and to this, it is now almost universally admitted, their elevation must in great part be due."[7] Vertical uplift was a side effect of horizontal compression, not the reverse.

The Idea of Isostasy

A second major problem arose from the rapidly expanding fields of geodesy and geophysics. While Heim and his colleagues were mapping the details of the Swiss Alps, workers with the Great Trigonometrical Survey of India were undertaking extensive geodetic measurements to produce accurate maps of British colonial holdings.[8] In the early 1850s, Colonel (later Sir) George Everest, the Surveyor-General of India, discovered a discrepancy in the measured distance between two geodetic stations there, Kaliana and Kalianpur, 370 miles apart. When measured on the basis of triangulation, the latitude difference was five seconds greater than when computed on the basis of astronomical observation. Everest thought that the difference might be due to the gravitational attraction of the Himalayas on surveyors' plumb bobs and enlisted John Pratt, a Cambridge-trained mathematician and the Archdeacon of Calcutta, to look more closely at the problem. Pratt computed the expected deflection based on the observable mass of the Himalayas and discovered that the discrepancy was actually *less* than it should have been: it was as if part of the mass of the Himalayas were somehow missing. When Pratt pursued the question, he found that the converse was true too: in coastal regions, plumb bobs were deflected toward the oceans, implying that the force of gravity over the oceans was not as low as expected considering the low density of water (figure 2.2).[9]

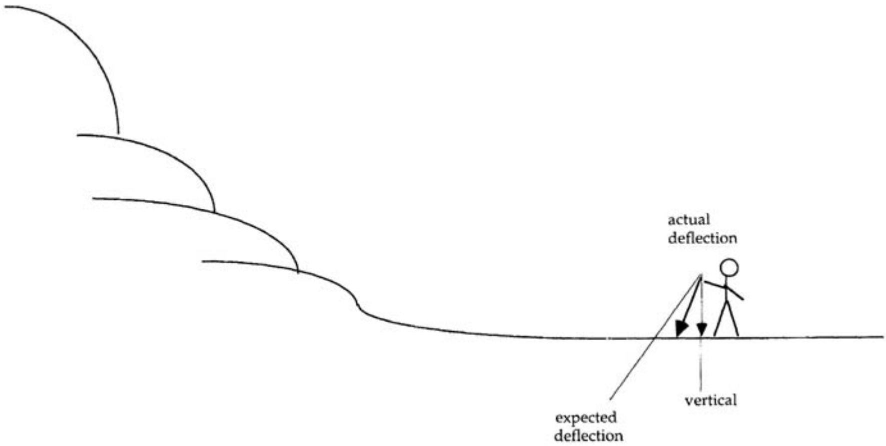


Figure 2.2. The detection of isostatic compensation. Adjacent to mountain ranges, a surveyor expects the plumb bob to be deflected by the gravitational attraction of the adjacent rock masses. However, surveyors in the Himalayas found the plumb bob deflected far less than expected. This implied that the surface mass excess of the Himalayas was somehow compensated by a subterranean mass deficit. The gravitational attraction of the Himalayas was first computed by John Henry Pratt in 1855.

Pratt published these results in the *Philosophical Transactions of the Royal Society* in 1855, where the paper was reviewed by George Biddell Airy, the Astronomer Royal of the United Kingdom. In a follow-up paper, Airy suggested that the discrepancy could be explained if the observed surficial mass of the Himalayas was gravitationally compensated by a subterranean mass deficit. Compensation could be achieved if a low-density crust, overlying a denser substrate, were thickened under high mountains, allowing the continents to float in the heavier substrate like icebergs at sea (figure 2.3).[10] But Pratt later put forward an alternative explanation: that compensation was achieved by underground differences in density that make up for above-ground topography. The crust, in this view, has constant thickness but varies markedly in density; compensation occurs because rocks comprising mountain belts are less dense than those comprising low lands and ocean deeps. "The density of the crust beneath the mountains," he wrote, "must be less than that below the plains, and still less than that below the ocean beds."[11] In a view rather similar to Dana's, Pratt supposed that these density differentials were produced by differential radial contraction during the early history of the globe: the areas that had contracted the most were the densest; those that had contracted the least were the lightest.[12]

One observational problem had engendered two different solutions. The observed phenomena could be explained by variations in either the thickness or the density of the crust. For Airy, mountains were like icebergs, supported by invisible roots beneath, and the size of the roots was proportional to the height of the mountains. For Pratt, mountains were like dough that was well risen—flatlands were a leaden loaf—and there was no need for roots below them. In either model, there would be some level within the earth where the total overlying mass was the same; this came to be known as the depth of compensation, referring to the point at which the extra mass above was

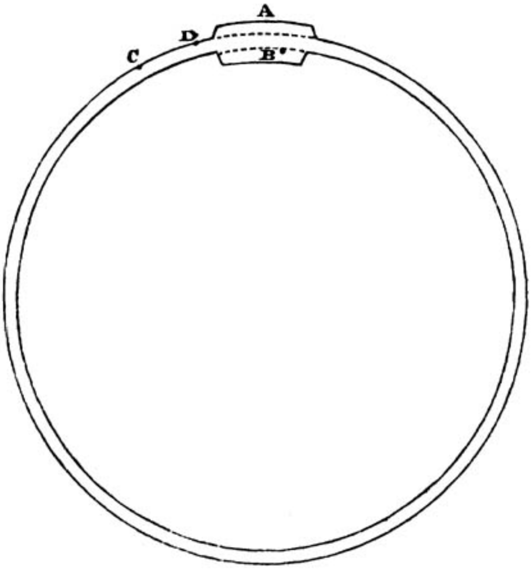


Figure 2.3. Airy's roots of mountains hypothesis. George Biddell Airy reviewed Pratt's calculations and suggested that the missing gravitational effect could be explained if the low-density earth crust floated in a higher density liquid substrate and if mountainous areas were supported by hidden "roots" that compensated for the extra mass above. Point A represents a continental block, B is the root below it. C and D are points on the oceanic crust, where no compensation is required. (From G. B. Airy 1855.)

compensated, or balanced out, by a lack of mass below. In Airy's model, the depth of compensation varied according to the depth of the roots; in Pratt's model the depth of compensation was everywhere the same. Thus the two models became known as the "roots of mountains" and the "uniform depth of compensation" hypotheses, respectively, or the Airy and Pratt models for short.[13] In either model, at some unknown depth, the weight of the overlying rocks would be the same everywhere, a condition that the American geologist Clarence Dutton named *isostasy*—meaning equal standing. Isostasy implied that the continents and oceans were balanced in a delicate equilibrium. Collectively, these ideas came to be known as the *theory* of isostasy—essentially a restatement of Archimedes' principle applied to the structure of the earth. To maintain the earth's surface features in a condition of hydrostatic equilibrium, elevated areas such as mountain ranges had to be compensated by a mass deficit below them, and lowlands had to be compensated by a mass surfeit.

Isostasy as a Theory of the Earth: Osmond Fisher

The phenomenon of isostatic compensation was discovered in the utilitarian context of geodetic mapping and surveying, but geologists soon considered its theoretical implications. Among these were the Reverend Osmond Fisher (1817–1914), the rector at the parish of Harlton, near Cambridge, a Fellow of the Geological Society of London and a close friend of Cambridge geologist Adam Sedgwick. Fisher was one of a number of British scientists in the late nineteenth century who were attempting to quantify questions pertaining to the structure and history of the earth.[14] In a series of papers published in the 1870s, Fisher attempted to mathematicize the concept of contraction and thereby demonstrate its sufficiency as an explanation for the earth's surface features (figure 2.4). Instead, he proved the reverse: mathematical analysis showed that thermal contraction was incapable of causing observed differences in ele-

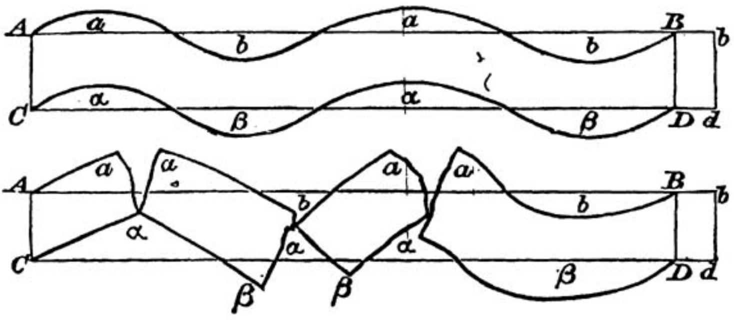


Figure 2.4. Fisher's geometric representation of the problem of terrestrial contraction. In the mid-nineteenth century, Osmond Fisher attempted to interpret the theory of earth contraction mathematically to demonstrate its sufficiency as a mechanism of surface deformation. His goal was to find "some simple laws which must govern the disturbed strata in spite of the confusion which appears to reign among them." (From Fisher 1881, p. 46.)

vation around the globe. "To my excessive surprise," he declared, "the result showed the utter inadequacy of the contraction hypothesis."[15]

The assumption of a solid earth was premised on the arguments of William Thomson (later Lord Kelvin) that the earth's internal pressure would render it solid.[16] Therefore contraction had to be evaluated in the context of conductive cooling of a solid sphere. Upon failure of the model of a solid earth, Fisher worked with ideas of a formerly liquid earth, a partially liquid earth losing water vapor, and a cooling crust resting on a thin fluid layer, but none of these modifications appeared adequate to save the hypothesis.[17] In 1881, Fisher laid out his refutation of the contraction hypothesis in *Physics of the Earth's Crust*—the first book in English dedicated to what would come to be called theoretical geophysics.[18] Whether solid, partially liquid, previously liquid, or partially gaseous, the earth simply could not contract sufficiently to do the work required of it. At best, contraction would produce elevation differentials of eight to nine hundred feet.[19]

Fisher was sensitive to the assumptions involved in his calculations and the uncertainties they introduced, and he tried to differentiate between constraints that were measurable—or could be calculated from other measurements, such as the mean density of the earth—and those that were purely guesswork. Ideally, the model should be built as far as possible on the former. The problem was that even "known" constraints could often "be satisfied in more ways than one."[20] An example was the constraint of rigidity. Kelvin had forcefully argued that the earth must be not only solid throughout but also fairly rigid.[21] His argument was based on the tides: if the earth were mostly fluid with only a thin solid crust, then it would respond to the tidal force in a similar manner as the surface waters, and there would be no tides, or only very small ones. That the oceans ebbed and flowed, while the earth beneath stayed still, was proof for him of a solid earth.[22] Fisher responded that the tides proved only that the earth was *mostly* solid. If the crust were sufficiently thick and rigid to resist the tidal force, the possibility remained of a fluid or plastic layer beneath it. Fisher pointed out that if the crust were solid due to its low temperature and the core were solid due to its high pressure, then at some intermediate level there could be a crossover zone where temperatures were high enough to cause melting but pressures were low enough to sustain a liquid.[23] Kelvin's objections notwithstanding, the earth might yet contain an internal fluid layer somewhere beneath the crust, and Fisher calculated that it could occur at a depth of twenty to twenty-five miles.

At this point in his analysis, Fisher turned from mathematical and physical considerations to geological evidence. Did geology support the notion of a fluid substrate? Yes—and more—it *required* it. The most conspicuous feature of mountain belts, Fisher argued, was the widepread evidence of lateral compression: this recognition had motivated the contraction theory in the first place. Yet, how could rocks be transported laterally without being crushed altogether? The answer lay with Fisher's fluid substratum. "The shifting of the crust towards a mountain range," he wrote, "which is testified by the corrugation of the rocks of which it is formed, requires a more or less fluid substratum to admit of it."[24] Rocks could move laterally if the substrate beneath them were yielding. *"What is required to explain the phenomenon is a liquid, or at least a plastic substratum for the crust to rest on, which will allow it to undergo a certain amount of lateral shift,"* Fisher wrote emphatically. A fluid substrate was also required to explain the accumulation of thick piles of sediment in coastal regions: the solid crust could sink into the fluid substrate, thereby making room for further accumulation of surface sediment. "The sinking of areas such as deltas, and other regions of deposition, demands [that] the crust . . . be in a condition of approximate hydrostatical equilibrium, such that any considerable addition of load will cause any region to sink, or any considerable amount denuded off an area will cause it to rise." Once the notion of a fluid substratum was accepted, Fisher concluded, "many of the facts [of geology] are more easily explained." A fluid substrate was not only a deductive consequence of physical principles but also inductively "necessary for the explanation of the phenomena of geology."[25]

The idea of a fluid substrate was also necessary for Airy's version of isostatic compensation, and Fisher presented Airy's model and the Himalayan data on which it was based as confirmatory evidence for his views. A fluid substrate would permit the continents to float in just the manner required by the roots of mountains hypothesis. Although Fisher quoted extensively from Pratt's analysis of the Himalayan data, he disagreed with Pratt's argument that isostatic compensation "may be produced in an infinite variety of ways and therefore, without data, it is useless to speculate regarding the arrangement of matter which actually exists in the solid parts below."[26] Although the effect could indeed be produced in a variety of ways, Fisher argued that there *were* relevant data—the data of observational geology—and that only the Airy explanation fitted them.

His reasoning here was quite specific:

> [Because] the crust . . . must accommodate itself to compression by being crushed together and thickened in place, . . . the very important consequence follows that elevations above the datum level will be accomodated by depressions beneath. The anticlinals will not be filled with fluid from below, but will be the upper portions of double bulges, which will dip into the fluid below, as well as rise into the air above. . . . It is analogous to the case of a broken-up area of ice, re-frozen and floating upon the water. The thickened parts which stand higher above the general surface also project deeper into the liquid below.[27]

Roots of mountains were the inevitable result of lateral compression of a rigid crust floating above a partially fluid substrate. In a world without compression, the Pratt model might hold, but that was not the world that geology revealed.

This conclusion led Fisher back to the question that had motivated Dana and Hall and virtually all geologists interested in the large scale features of the crust: What causes lateral compression? Pursuing the iceberg analogy, Fisher argued that the roots of mountains would be larger than their exposed portions. Denudation of the exposed tops would lead to gradual uplift of the hidden roots. Concurrently, sediments, carried by streams draining the mountains, would be deposited in adjacent coastal areas, and the result would be a net transfer of material from mountains to coastal plains. But, because mountains are not entirely free-floating, but are attached to their adjacent oceanic crust, the crust would be stressed along the boundary between the sinking ocean and the floating continent—the same boundary where Hall had placed geosynclines. Consequently, the crust must fissure, and magmas (either already liquid or liquified by pressure release) would escape from the fluid substrate and intrude into the crust along the lines of fissures, thus explaining the occurrence of volcanoes and ocean islands near continental margins. The ocean islands, having no roots themselves, would tend to sink (which, it was well known from the work of Charles Darwin, they did), while the igneous intrusions on the continents would compress the zone around them to produce mountain belts. Meanwhile, the gradual melting of the mountain roots would provide a mechanism for further generation of igneous rocks.

Although Fisher credited his ideas to Airy's insight, his vision of the earth was a blend of the Pratt and Airy models. The Airy model assumed uniform density for the entire crust, but Fisher held out for an oceanic crust that was denser than the continental crust. Otherwise, he argued, there was no way to explain why water pools in ocean basins, rather than being spread uniformly across the earth. This blended view of isostasy implied that the earth was fundamentally heterogeneous in a manner inconsistent with the theory of the interchangeability of continents and oceans, and it left Fisher with the unresolved question of what gave rise to the continents and oceans in the first instance. Fisher addressed this issue in a paper in *Nature* published in 1882. He was particularly troubled by the enormous size of the Pacific basin, which covers nearly half the globe. Concluding that such a "remarkable circumstance" warrants a remarkable cause, he lent his support to to the fissiparturition theory of his Cambridge colleague George H. Darwin (son of Charles), namely, that the moon had formed by separation from the earth early in planetary history.

In 1879, Darwin *fils* had proposed that the primordial earth, partially consolidated and covered by a thin crust, and subject to large tidal forces as it rotated rapidly, had split, and the moon had gradually receded from the slightly larger earth. Darwin called this process fissiparturition—birth through fission; Fisher extended Darwin's idea to argue that the scar left where the moon had broken off became the Pacific basin. Magmas welled up into the fractured region, producing a secondary crust of somewhat greater density than the original. The fragmented primary crust became the continents, preserving a "rude parallelism" of the coastlines; the secondary crust became the ocean basins. The major features of the earth were relics of this primordial event. The overall result was a geological model consistent with Airy isostasy: relics of low density primordial continental crust floating in a younger, denser oceanic substrate.[28]

Fisher's theory of volcanic compression did not have a profound impact on his colleagues, but his refutation of contraction and insights about isostasy did. Forty years later, John Joly, Professor of Geology at Trinity College, Dublin, and a Fellow of the Royal Society of London, concluded that Fisher wrote with a "perfect vision of what is involved in this theory [of isostasy] and of the essential simplicity and adequacy of the condition as a physical basis of the structure of the earth."[29] Writing more immediately in the United States, Clarence Dutton proclaimed that Fisher "has rendered most effectual service in utterly destroying the [contraction] hypothesis." Anyone who read Fisher's book, Dutton argued, would see that the contraction hypothesis was "nothing but a delusion and a snare, and that the quicker it is thrown aside and abandoned the better it will be for geological science."[30] But having dispensed with the contraction hypothesis, what was one to replace it with? Dutton was working on an answer of his own.

Isostasy as a Theory of the Earth: Clarence Dutton

One of the first Americans to challenge contraction theory was Clarence Edward Dutton (1841–1912), who is credited with coining the term *isostasy* and bringing the subject to the attention of American geologists.[31] Known in retrospect as one of the first American exponents of a quantitative approach to geological problems, Dutton was known in his own time as both a theoretician and an outstanding field geologist. In a review in *Nature* in 1880 by A. Geikie, Dutton's most famous work, his *Report on the Geology of the High Plateaus of Utah,* was praised as "one of the very best of the many admirable contributions to geology which have recently been made by the official surveys of the United States."[32]

After graduating from Yale University in 1860, C. E. Dutton joined the United States Army to fight in the American Civil War and then served as an officer in Army Ordnance. In 1867, Dutton joined G. K. Gilbert and John Wesley Powell in the *United States Geographical and Geological Survey of the Rocky Mountain Region,* one of the precursors of the U.S. Geological Survey (subsequently founded in 1879). Powell had already become famous for his dramatic explorations of the Colorado Plateau and its Grand Canyon, the "last completely blank area on the country's map."[33]

Dutton was deeply impressed by the spectacular geology of the Western regions, much of which did not fit standard American views based on Appalachian geology. His reports described abundant evidence of active and recent vulcanism, hot springs which indicated water at depth, and, above all, the Colorado Plateau and its Grand Canyon, which revealed the surprising power of erosional processes and the possibility of large-scale uplift without compressive deformation. Dutton was particularly impressed by the much greater prominence of igneous rocks in the Sierra Nevada than in the Appalachians, and the former gradually replaced the latter for him as the archetypical mountain belt.

Beginning in the early 1870s, Dutton delivered a series of papers developing isostasy in lieu of contraction as a framework for approaching large-scale geological problems. His most pointed attack, "A criticism upon the contractional hypothesis," was delivered in J. D. Dana's own journal, the *American Journal of Science.*[34] While accepting the reality of secular cooling, Dutton provided a devastating critique of its ability to affect the earth's surface processes. As it was for Fisher, Dutton's starting point was

Kelvin's analysis not only that the earth must be solid but also that the bulk of its cooling must necessarily take place in a relatively thin outer layer. Below some finite depth, the increase of temperature with depth must be virtually nil, and contraction equally nil. Given actual measurements of geothermal gradients taken from mines, and measurements of the thermal conductivity of rocks recently completed by Kelvin and J. D. Forbes, one could calculate how old the earth was: approximately 100 million years.[35]

Dutton favorably summarized Kelvin's argument but added the crucial caveat that the accuracy of the calculations depended critically on one's estimates of thermal conductivity and near-surface geothermal gradient. Or, as he sagely understated it, "The only ground of controversy must be the values to be assigned to the constants [in Fourier's theorem]." Dutton repeated the calculations using a range of plausible values for these constants and came to a rather different result from Kelvin's: the earth was anywhere between 98 million and 2.5 billion years old. But for Dutton the age of the earth was not the central issue. The central issue was that, given any reasonable geological estimate of thermal conductivity and geothermal gradient, secular cooling was an ineffectual mechanism of surface change. Significant contraction was limited to a relatively thin outer crust. "The unavoidable deduction from [Fourier's] theorem," he wrote, " is that the greatest possible contraction due to secular cooling is insufficient in amount to account for the phenomena attributed to it by the contractional hypothesis."[36]

Fisher had linked his theory to data collected in government geodetic surveys of the Himalayas, and Dutton linked his ideas to physical evidence from geological surveys in the American West. For example, Fourier's theorem required that secular cooling be necessarily greater in early periods of earth history, but this was inconsistent with the Rocky Mountains and the Sierra Nevada, both of which showed evidence of considerable disturbance since the Cretaceous period. Contraction theory also implied that orogeny should be a general feature of the globe, but, as Dutton noted, even the older Appalachians consist of a highly focused zone of folding adjacent to a huge undeformed plain. Why were surface disturbances restricted to narrow zones? Why did North America have great mountain chains on either coast, with a great, yawning, flat divide between them? Returning to geodetic evidence, Dutton pointed out that if, by some as yet unexplained mechanism, the earth contracted only along narrow zones of weakness, then how did it maintain an ellipsoid of revolution? "It is here that the analogy of the withered apple fails," Dutton wrote decisively. "If [the earth] is corrugated irregularly by shrinkage, it fails to preserve its original figure; and conversely, if it preserves its original figure, it must be corrugated uniformly."[37] One could not have it both ways. Contraction theory was inconsistent with both theoretical geophysics and observational geology. Dutton set to work on an alternative account of the processes that cause surface geological change.

American geologists on the East Coast had traditionally viewed mountains primarily as folded sediments, but Dutton's experience in the West led him to view mountains as igneous rocks intruding into sedimentary sequences. He set aside the Appalachians as an orogenic archetype and focused instead on the Rockies and the Sierra Nevada. Generalizing from them, he wondered if magmas were not the driving force of orogeny. If they were, then the key question was not so much the origin of mountains but that of magmas. Dutton's thought moved to the role of heating and melting in forming mountain belts and changing the physical properties of their substrates. Rocks that were heated near their own melting points would necessarily become plastic and, upon melting, would increase in volume and decrease in density. Low-density molten rock would work its way toward the surface, disturbing the adjacent and superincumbent strata on its way. Left behind would be a plastic zone in which isostatic adjustment could occur. But what force would heat and melt rocks in the first place? The answer was the pressure of isostatic adjustment in response to erosion and sedimentation.

One point of agreement among virtually all North American geologists was that the thick geological strata of the Appalachians were mainly deposited in shallow waters. Therefore, as James Hall had argued, they must have subsided as they were deposited. But if they subsided, then they must have displaced the matter beneath them; "And what becomes of the displaced matter?" Dutton asked rhetorically.[38] Borrowing the notion of a plastic substrate introduced by Babbage and Herschel earlier in the century and developed further by Fisher, Dutton answered that the displaced material moved laterally, away from the zone of sedimentation and toward the uplifted zone of erosion. Sedimentary loading presses on the subjacent plastic zone and causes lateral displacement toward adjacent, less heavily weighted areas, in the process generating physical changes in the substrate. The result is igneous intrusion in uplifted areas and folding of the sinking sediments in the area from which substrate has been removed (figure 2.5). In a widely cited address to the Philosophical Society of Washington in 1889, Dutton closely parallelled Fisher's argument:

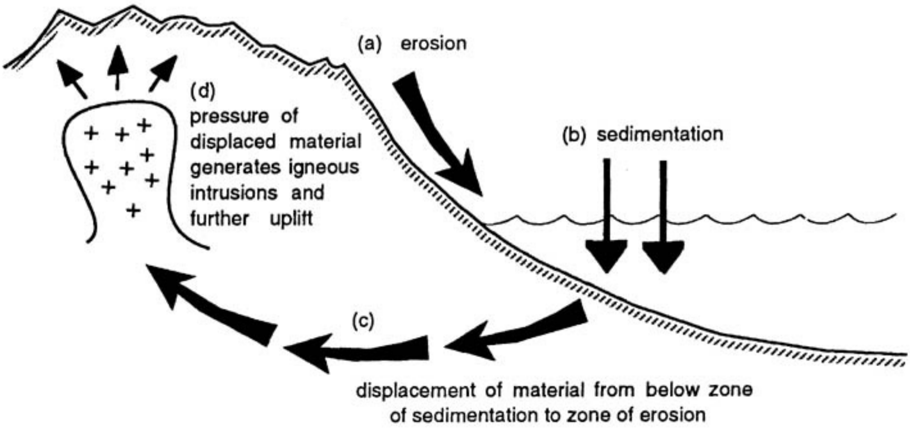


Figure 2.5. Dutton's isostatic model of crustal deformation. Clarence Dutton suggested that crustal deformation could be understood as a response to isostatic compensation. (**a**) The uplifted portions of the continents are eroded, and material is transported to coastal regions. (**b**) The weight of the deposited sediments causes subsidence along the continental margins, which causes displacement of material at depth. (**c**) The displaced material moves laterally and results in (**d**) igneous intrusions and further uplift of the continent, thus renewing the cycle. Dutton's idea draws freely on the work of Herschel, Dana, and Hall. (Illustration by Melody Brown.)