

TOEFL TPO READINGS
CLASSIFIED
FROM TPO 1 TO 74

SCIENCE.....	1
16 – 2 DEVELOPMENT OF THE PERIODIC TABLE.....	1
37 – 1 THALES AND THE MILESIAINS.....	2
37 – 1 MICROSCOPES	3
ENERGY / INDUSTRY / TECHNOLOGY.....	4
06 – 1 POWERING THE INDUSTRIAL REVOLUTION.....	4
71 – 1 ELECTRICAL ENERGY FROM THE OCEAN.....	5
04 – 3 PETROLEUM RESOURCES	6
18 – 1 INDUSTRIALIZATION IN THE NETHERLANDS AND SCANDINAVIA	7
21 – 1 GEOTHERMAL ENERGY	8
26 – 1 ENERGY AND THE INDUSTRIAL REVOLUTION.....	9
29 – 3 THE HISTORY OF WATERPOWER.....	10
30 – 3 THE INVENTION OF THE MECHANICAL CLOCK	11
34 – 2 THE DEVELOPMENT OF STEAM POWER	12
36 – 3 INDUSTRIAL MELANISM: THE CASE OF THE PEPPERED MOTH	13
39 – 3 FOREST FIRE SUPPRESSION.....	14
43 – 2 THE ORIGIN OF PETROLEUM.....	15
49 – 3 BACKGROUND FOR THE INDUSTRIAL REVOLUTION	16
59 – 1 BUILDING MATERIALS.....	17
AGRICULTURE.....	18
21 – 2 THE ORIGINS OF AGRICULTURE	18
23 – 2 SEVENTEENTH-CENTURY DUTCH AGRICULTURE	19
33 – 2 RAILROADS AND COMMERCIAL AGRICULTURE IN NINETEENTH-CENTURY UNITED STATES.....	20
74 – 2 THE COMMERCIALIZATION OF PEARL RIVER AGRICULTURE	21
65 – 3 PASTORALISM AND AGRICULTURE IN IRAN.....	22
67 – 1 CROP ENGINEERING	23
71 – 3 THE PRODUCTIVITY OF WET RICE FARMING.....	24
EARTH / WATER / ENVIRONMENT / MOUNTAINS / CLIMATE	25
01 – 1 GROUNDWATER	25
03 – 2 DEPLETION OF THE OGALLALA AQUIFER	26
02 – 1 DESERT FORMATION.....	27
01 – 3 TIMBERLINE VEGETATION ON MOUNTAINS.....	28
03 – 3 THE LONG-TERM STABILITY OF ECOSYSTEMS.....	29
07 – 1 THE GEOLOGIC HISTORY OF THE MEDITERRANEAN	30
10 – 2 VARIATIONS IN THE CLIMATE.....	31
12 – 3 WATER IN THE DESERT	32
15 – 3 GLACIER FORMATION.....	33
18 – 3 LIGHTNING	34
19 – 3 DISCOVERING THE ICE AGES	35
22 – 3 THE ALLENDE METEORITE.....	36
23 – 1 URBAN CLIMATES	37

24 – 1 LAKE WATER	38
27 – 2 THE FORMATION OF VOLCANIC ISLANDS	39
28 – 1 GROUNDWATER	40
31 – 1 SPECIATION IN GEOGRAPHICALLY ISOLATED POPULATIONS	41
35 – 1 EARTH'S AGE	42
36 – 1 SOIL FORMATION	43
36 – 2 EARLY IDEAS ABOUT DEEP-SEA BIOLOGY	44
40 – 4 POPULATION AND CLIMATE	45
40 – 2 LATITUDE AND BIODIVERSITY	46
43 – 3 EL NIÑO	47
46 – 3 ECOSYSTEM DIVERSITY AND STABILITY	48
45 – 2 WIND POLLINATION	49
48 – 3 CLIMATE AND URBAN DEVELOPMENT	50
49 – 1 ANCIENT COASTLINES	51
52 – 1 STREAM DEPOSIT	52
54 – 3 ELEMENTS OF LIFE	53
55 – 1 MESOLITHIC COMPLEXITY IN SCANDINAVIA	54
56 – 3 URBAN HEAT ISLANDS	55
56 – 1 ALFRED WEGENER'S THEORY OF CONTINENTAL DRIFT	56
59 – 2 DEALING WITH EXTREME COLD	57
60 – 3 EARTH'S CORE	58
61 – 2 CONDITIONS ON EARTH WHEN LIFE BEGAN	59
60 – 1 UNDERGROUND LIFE	60
61 – 1 PHYSICAL PROPERTIES OF MINERALS	61
63 – 2 STRUCTURE AND COMPOSITION OF COMETS	62
66 – 2 VISIONS OF THE LAND	63
66 – 4 GEOLOGY AND LANDSCAPE	64
65 – 2 EARLY RESEARCH ON AIR	65
69 – 2 PACIFIC ECOSYSTEMS	66
68 – 2 PREDICTING VOLCANIC ERUPTIONS	67
70 – 2 NINETEENTH-CENTURY THEORIES OF MOUNTAIN FORMATION	68
SPACE	69
08 – 3 RUNNING WATER ON MARS	69
25 – 1 THE SURFACE OF MARS	70
72 – 2 THE IMPACT ORIGIN OF LUNAR CRATERS	71
73 – 3 STELLAR REMNANTS	72
74 – 3 LUNAR HISTORY	73
42 – 3 CALLISTO AND GANYMEDE	74
41 – 2 CLIMATE OF VENUS	75
16 – 3 PLANETS IN OUR SOLAR SYSTEM	76
50 – 3 STAR DEATH	77
51 – 2 SURFACE FLUIDS ON VENUS AND EARTH	78
PLANTS / BOTANY / MONTOLGY (OROLOGY) / HYDROLOGY / HYDROGRAPHY / HYDROGEOLOGY	79

05 – 1 MINERALS AND PLANTS	79
09 – 3 THE ARRIVAL OF PLANT LIFE IN HAWAII	80
19 – 2 SUCCESSION, CLIMAX, AND ECOSYSTEMS	81
22 – 1 SPARTINA	82
25 – 3 THE EVOLUTIONARY ORIGIN OF PLANTS	83
31 – 3 SAVANNA FORMATION.....	84
32 – 1 PLANT COLONIZATION.....	85
34 – 3 PROTECTION OF PLANTS BY INSECTS	86
38 – 3 TRANSGENIC PLANTS.....	87
72 – 1 INVADING ALGAE	88
73 – 1 HOW EFFECTIVE ARE STRUCTURAL PLANT DEFENSES?	89
44 – 3 SEAGRASSES	90
44 – 4 WHAT IS A COMMUNITY?	91
47 – 3 CORAL REEFS.....	92
50 – 4 FOREST SUCCESSION	93
53 – 2 RAIN FOREST SOILS	94
55 – 2 EVOLUTION OF THE FLOWERING PLANTS	95
61 – 3 GRASSES OF THE PRAIRIE	96
62 – 4 WHAT CONTROLS FLOWERING	97
62 – 1 PLANT ADAPTATIONS TO COOL ENVIRONMENTS.....	98
67 – 3 TAXONOMY OF ORGANISMS.....	99

ANIMALS / SPECIES / ZOOLOGY..... 100

02 – 3 THE ORIGINS OF CETACEANS	100
04 – 1 DEER POPULATIONS OF THE PUGET SOUND	101
05 – 3 THE CAMBRIAN EXPLOSION	102
08 – 2 EXTINCTION OF THE DINOSAURS	103
15 – 2 MASS EXTINCTIONS	104
11 – 2 ORIENTATION AND NAVIGATION	105
11 – 3 BEGGING BY NESTLINGS	106
13 – 2 BIOLOGICAL CLOCKS	107
15 – 1 A WARM-BLOODED TURTLE	108
17 – 2 ANIMAL SIGNALS IN THE RAIN FOREST	109
17 – 3 SYMBIOTIC RELATIONSHIPS	110
20 – 3 FOSSIL PRESERVATION	111
26 – 2 SURVIVAL OF PLANTS AND ANIMALS IN DESERT CONDITIONS	112
27 – 3 PREDATOR-PREY CYCLES.....	113
28 – 3 BUCK RUBS AND BUCK SCRAPES.....	114
29 – 2 COMPETITION.....	115
30 – 2 THE PACE OF EVOLUTIONARY CHANGE	116
32 – 3 DISTRIBUTIONS OF TROPICAL BEE COLONIES	117
33 – 3 EXTINCTION EPISODES OF THE PAST.....	118
35 – 3 SEASONAL SUCCESSION IN PHYTOPLANKTON	119
37 – 2 DIRECT SPECIES TRANSLOCATION	120
38 – 2 THE RACCOONS' SUCCESS	121
39 – 2 THE EXTINCTION OF MOA.....	122

74 – 1	EARLY HORSES.....	123
40 – 3	AMPHIBIAN THERMOREGULATION.....	124
42 – 1	GEOGRAPHIC ISOLATION OF SPECIES.....	125
42 – 2	EXPLAINING DINOSAUR EXTINCTION.....	126
45 – 3	FEEDING STRATEGIES IN THE OCEAN.....	127
44 – 1	FROM FISH TO TERRESTRIAL VERTEBRATES.....	128
46 – 4	CROWN OF THORNS STARFISH AND CORAL REEFS.....	129
45 – 1	MICROSCOPES THE BERINGIA LANDSCAPE.....	130
47 – 2	TERMITE INGENUITY.....	131
48 – 2	DETERMINING DINOSAUR DIET.....	132
54 – 2	OVERKILL OF THE NORTH AMERICAN MEGAFUNA.....	133
55 – 3	HOW HERDING CAN PROVIDE SAFETY.....	134
56 – 4	HABITATS AND CHIPMUNK SPECIES.....	135
57 – 1	PESTS AND PESTICIDES.....	136
57 – 2	THE DEBATE OVER SPONTANEOUS GENERATION.....	137
58 – 2	PINYON PINES AND PINYON JAYS.....	138
58 – 4	CETACEAN INTELLIGENCE.....	139
64 – 1	CHARACTERISTICS OF PTEROSAURS.....	140
66 – 3	WHAT IS CORAL?.....	141
65 – 1	THE PIT ORGANS OF SNAKES.....	142
69 – 1	WHY SNAKES HAVE FORKED TONGUES.....	143

PHOTOGRAPHY / THEATER / CINAMA 144

01 – 2	THE ORIGINS OF THEATER.....	144
02 – 3	EARLY CINEMA.....	145
12 – 2	TRANSITION TO SOUND IN FILM.....	146
22 – 2	THE BIRTH OF PHOTOGRAPHY.....	147
44 – 2	THE USE OF THE CAMERA OBSCURA.....	148
57 – 3	THE PROBLEM OF NARRATIVE CLARITY IN SILENT FILMS.....	149
66 – 1	THE ACTOR AND THE AUDIENCE.....	150

ART / ARCHITECTURE / CULTURE / ART HISTORY / ARCHEOLOGY / SCULPTURE / VISUAL ARTS 151

03 – 1	ARCHITECTURE.....	151
04 – 2	CAVE ART IN EUROPE.....	152
10 – 1	CHINESE POTTERY.....	153
11 – 1	ANCIENT EGYPTIAN SCULPTURE.....	154
23 – 3	ROCK ART OF THE AUSTRALIA ABORIGINES.....	155
24 – 3	MOVING INTO PUEBLOS.....	156
27 – 1	CRAFTS IN THE ANCIENT NEAR EAST.....	157
29 – 1	CHARACTERISTICS OF ROMAN POTTERY.....	158
34 – 1	ISLAMIC ART AND THE BOOK.....	159
37 – 3	MODERN ARCHITECTURE IN THE UNITED STATES.....	160
41 – 1	NAVAJO ART.....	161
47 – 1	ROMAN CULTURAL INFLUENCE ON BRITAIN.....	162

52 – 2 NATUFIAN CULTURE	163
53 – 3 PALEOLITHIC CAVE PAINTINGS	164
59 – 3 ANCIENT GREEK POTTERY	165
58 – 1 THE DEVELOPMENT OF INSTRUMENTAL MUSIC	166
62 – 2 THE USE OF JADE AND BRONZE IN CHINA	167
69 – 3 ANCIENT SOUTHWESTERN CULTURES	168
WRITING / PRINTING	169
39 – 1 EARLY WRITING SYSTEMS.....	169
73 – 2 WRITING HISTORY	170
46 – 1 THE ORIGINS OF WRITING	171
49 – 2 MOVABLE TYPE	172
53 – 1 EVIDENCE OF THE EARLIEST WRITING	173
60 – 2 THE REVOLUTION OF CHEAP PRINT	174
HISTORY / SOCIETIES / ECONOMY / BIOGRAPHY / ANCIENT SOCIETIES.....	175
07 – 2 ANCIENT ROME AND GREECE	175
07 – 3 AGRICULTURE, IRON, AND THE BANTU PEOPLES	176
08 – 1 THE RISE OF TEOTIHUACÁN.....	177
09 – 1 COLONIZING THE AMERICAS VIA THE NORTHWEST COAST	178
10 – 3 SEVENTEENTH-CENTURY EUROPEAN ECONOMIC GROWTH.....	179
14 – 2 MAYA WATER PROBLEMS	180
14 – 3 PASTORALISM IN ANCIENT INNER EURASIA	181
05 – 2 THE ORIGIN OF THE PACIFIC ISLAND PEOPLE	182
06 – 2 WILLIAM SMITH	183
16 – 1 TRADE AND THE ANCIENT MIDDLE EAST	184
17 – 1 EUROPE'S EARLY SEA TRADE WITH ASIA.....	185
19 – 1 THE ROMAN ARMY'S IMPACT ON BRITAIN.....	186
20 – 1 WESTWARD MIGRATION	187
20 – 2 EARLY SETTLEMENTS IN THE SOUTHWEST ASIA	188
25 – 2 THE DECLINE OF VENETIAN SHIPPING	189
26 – 3 SUMER AND THE FIRST CITIES OF THE ANCIENT NEAR EAST	190
28 – 2 EARLY SAHARAN PASTORALISTS	191
32 – 2 SIAM, 1851-1910.....	192
33 – 1 THE FIRST CIVILIZATIONS	193
72 – 3 CITIES AND KINGDOMS IN ALEXANDER'S EMPIRE	194
71 – 2 MINOAN PALACES.....	195
41 – 3 TRADE AND EARLY STATE FORMATION	196
40 – 1 ANCIENT ATHENS	197
42 – 4 EUROPE IN THE TWELFTH CENTURY	198
43 – 1 THE EMPIRE OF ALEXANDER THE GREAT	199
46 – 2 THE COMMERCIAL REVOLUTION IN MEDIEVAL EUROPE	200
48 – 4 THE RISE OF MOSCOW	201
48 – 1 CHINESE POPULATION GROWTH	202

50 – 2 THE ACHIEVEMENT OF BRAZILIAN INDEPENDENCE	203
51 – 3 POPULATION GROWTH IN NINETEENTH-CENTURY EUROPE	204
50 – 1 AMERICAN RAILROADS.....	205
52 – 3 EARLY FOOD PRODUCTION IN SUB-SAHARAN AFRICA.....	206
52 – 4 ENGLAND’S ECONOMY IN SIXTEENTH CENTURY	207
51 – 1 MEMPHIS: UNITED EGYPT’S FIRST CAPITAL	208
54 – 4 DOCUMENTING THE INCAS.....	209
54 – 1 THE COMMERCIALIZATION OF LUMBER	210
56 – 2 REPRESENTATIVE GOVERNMENT IN COLONIAL NORTH AMERICA	211
58 – 3 THE RISE OF CLASSIC MAYA CIVILIZATION.....	212
60 – 4 A MODEL OF URBAN EXPANSION	213
63 – 1 THE SUMERIANS AND REGIONAL INTERDEPENDENCE.....	214
63 – 3 THE ROMAN EMPIRE.....	215
64 – 3 THE DEVELOPMENT OF COMPLEX SOCIETIES IN ANCIENT MEXICO	216
64 – 4 NINETEEN-CENTURY POLITICS IN THE UNITED STATES.....	217
67 – 2 HUMAN ACTIVITY AND THE ARCHAEOLOGICAL RECORD.....	218
68 – 1 SALT AND THE RISE OF VENICE.....	219
70 – 1 THE DEVELOPMENT OF CHINESE DYNASTIES	220
HUMAN BIOLOGY	221
18 – 2 THE MYSTERY OF YAWNING	221
24 – 2 BREATHING DURING SLEEP	222
HUMAN BEING / ANTHROPOLOGY / PSYCHOLOGY / MEMORY /	
ADULTHOOD AND IMMATURITY	223
06 – 3 INFANTILE AMNESIA	223
12 – 1 WHICH HAND DID THEY USE?	224
13 – 1 TYPES OF SOCIAL GROUPS	225
13 – 3 METHODS OF STUDYING INFANT PERCEPTION.....	226
14 – 1 CHILDREN AND ADVERTISING.....	227
21 – 3 AUTOBIOGRAPHICAL MEMORY	228
30 – 1 ROLE OF PLAY IN DEVELOPMENT	229
31 – 2 EARLY CHILDHOOD EDUCATION.....	230
35 – 2 THE DEVELOPMENT OF SOCIAL COMPLEXITY	231
62 – 3 IS PERCEPTUAL DEVELOPMENT AN INNATE OR SOCIALLY ACQUIRED PROCESS?.....	232
64 – 2 AUDITORY PERCEPTION IN INFANCY	233
68 – 3 RESEARCH INTO AGING AND EXTENDING LIFE SPAN.....	234
70 – 3 FUNCTIONALISM	235
JOB / CAREER	236
09 – 2 REFLECTION IN TEACHING	236

SCIENCE

16 – 2 Development of the Periodic Table

The periodic table is a chart that reflects the periodic recurrence of chemical and physical properties of the elements when the elements are arranged in order of increasing atomic number (the number of protons in the nucleus). It is a monumental scientific achievement, and its development illustrates the essential interplay between observation, prediction, and testing required for scientific progress. In the 1800s scientists were searching for new elements. By the late 1860s more than 60 chemical elements had been identified, and much was known about their descriptive chemistry. Various proposals were put forth to arrange the elements into groups based on similarities in chemical and physical properties. The next step was to recognize a connection between group properties (physical or chemical similarities) and atomic mass (the measured mass of an individual atom of an element). When the elements known at the time were ordered by increasing atomic mass, it was found that successive elements belonged to different chemical groups and that the order of the groups in this sequence was fixed and repeated itself at regular intervals. Thus when the series of elements was written so as to begin a new horizontal row with each alkali metal, elements of the same groups were automatically assembled in vertical columns in a periodic table of the elements. This table was the forerunner of the modern table.

When the German chemist Lothar Meyer and (independently) the Russian Dmitry Mendeleev first introduced the periodic table in 1869–70, one-third of the naturally occurring chemical elements had not yet been discovered. Yet both chemists were sufficiently farsighted to leave gaps where their analyses of periodic physical and chemical properties indicated that new elements should be located. Mendeleev was bolder than Meyer and even assumed that if a measured atomic mass put an element in the wrong place in the table, the atomic mass was wrong. In some cases this was true. Indium, for example, had previously been assigned an atomic mass between those of arsenic and selenium. Because there is no space in the periodic table between these two elements, Mendeleev suggested that the atomic mass of indium be changed to a completely different value, where it would fill an empty space between cadmium and tin. In fact, subsequent work has

shown that in a periodic table, elements should not be ordered strictly by atomic mass. For example, tellurium comes before iodine in the periodic table, even though its atomic mass is slightly greater. Such anomalies are due to the relative abundance of the "isotopes" or varieties of each element. All the isotopes of a given element have the same number of protons, but differ in their number of neutrons, and hence in their atomic mass. The isotopes of a given element have the same chemical properties but slightly different physical properties. We now know that atomic number (the number of protons in the nucleus), not atomic mass number (the number of protons and neutrons), determines chemical behavior.

Mendeleev went further than Meyer in another respect: he predicted the properties of six elements yet to be discovered. For example, a gap just below aluminum suggested a new element would be found with properties analogous to those of aluminum. Mendeleev designated this element "eka-aluminum" (eka is the Sanskrit word for "next") and predicted its properties. Just five years later an element with the proper atomic mass was isolated and named gallium by its discoverer. The close correspondence between the observed properties of gallium and Mendeleev's predictions for eka-aluminum lent strong support to the periodic law. Additional support came in 1885 when eka-silicon, which had also been described in advance by Mendeleev, was discovered and named germanium.

The structure of the periodic table appeared to limit the number of possible elements. It was therefore quite surprising when John William Strutt (Lord Rayleigh), discovered a gaseous element in 1894 that did not fit into the previous classification scheme. A century earlier, Henry Cavendish had noted the existence of a residual gas when oxygen and nitrogen are removed from air, but its importance had not been realized. Together with William Ramsay, Rayleigh isolated the gas (separating it from other substances into its pure state) and named it argon. Ramsay then studied a gas that was present in natural gas deposits and discovered that it was helium, an element whose presence in the Sun had been noted earlier in the spectrum of sunlight but that had not previously been known on Earth. Rayleigh and Ramsay postulated the existence of a new group of elements, and in 1898 other members of the series (neon, krypton, and xenon) were isolated.

37 – 1 Thales and the Milesians

While many other observers and thinkers had laid the groundwork for science, Thales (circa 624 B.C.E.-ca 547 B.C.E.), the best known of the earliest Greek philosophers, made the first steps toward a new, more objective approach to finding out about the world. He posed a very basic question: "What is the world made of?" Many others had asked the same question before him, but Thales based his answer strictly on what he had observed and what he could reason out-not on imaginative stories about the gods or the supernatural. He proposed water as the single substance from which everything in the world was made and developed a model of the universe with Earth as a flat disk floating in water.

Like most of the great Greek philosophers, Thales had an influence on others around him. His two best-known followers, though there were undoubtedly others who attained less renown, were Anaximander and Anaximenes. Both were also from Miletus (located on the southern coast of present-day Turkey) and so, like Thales, were members of the Milesian School. Much more is known about Anaximander than about Anaximenes, probably because Anaximander, who was born sometime around 610 B.C.E., ambitiously attempted to write a comprehensive history of the universe. As would later happen between another teacher-student pair of philosophers, Plato and Aristotle, Anaximander disagreed with his teacher despite his respect for him. He doubted that the world and all its contents could be made of water and proposed instead a formless and unobservable substance he called "apeiron" that was the source of all matter.

Anaximander's most important contributions, though, were in other areas. Although he did not accept that water was the prime element, he did believe that all life originated in the sea, and he was thus one of the first to conceive of this important idea. Anaximander is credited with drawing up the first world map of the Greeks and also with recognizing that Earth's surface was curved. He believed, though, that the shape of Earth was that of a cylinder rather than the sphere that later Greek philosophers would conjecture. Anaximander, observing the motions of the heavens around the polestar, was probably the first of the Greek philosophers to picture the sky as sphere completely surrounding Earth-an idea that, elaborated upon later, would prevail until the

advent of the Scientific Revolution in the seventeenth century.

Unfortunately, most of Anaximander's written history of the universe was lost, and only a few fragments survive today. Little is known about his other ideas. Unfortunately, too, most of the written work for Anaximenes, who may have been Anaximander's pupil, has also been lost. All we can say for certain about Anaximenes, who was probably born around 560 B.C.E., is that following in the tradition of Anaximander, he also disagreed with his mentor. The world, according to Anaximenes, was not composed of either water or apeiron, but air itself was the fundamental element of the universe. Compressed, it became water and earth, and when rarefied or thinned out, it heated up to become fire. Anaximenes may have also been the first to study rainbows and speculate upon their natural rather than supernatural cause.

With the door opened by Thales and the other early philosophers of Milestus, Greek thinkers began to speculate about the nature of the universe. This exciting burst of intellectual activity was for the most part purely creative. The Greeks, from Thales to Plato and Aristotle, were philosophers and not scientists in today's sense. It is possible for anyone to create "ideas" about the nature and structure of the universe, for instance, and many times these ideas can be so consistent and elaborately structured, or just so apparently obvious, that they can be persuasive to many people. A scientific theory about the universe, however, demands much more than the various observations and analogies that were woven together to form systems of reasoning, carefully constructed as they were, that would eventually culminate in Aristotle's model of the world and the universe. Without experimentation and objective, critical testing of their theories, the best these thinkers could hope to achieve was some internally consistent speculation that covered all the bases and satisfied the demands of reason.

37 – 1 Microscopes

Before microscopes were first used in the seventeenth century, no one knew that living organisms were composed of cells. The first microscopes were light microscopes, which work by passing visible light through a specimen. Glass lenses in the microscope bend the light to magnify the image of the specimen and project the image into the viewer's eye or onto photographic film. Light microscopes can magnify objects up to 1,000 times without causing blurriness.

Magnification, the increase in the apparent size of an object, is one important factor in microscopy. Also important is resolving power, a measure of the clarity of an image. Resolving power is the ability of an optical instrument to show two objects as separate. For example, what looks to the unaided eye like a single star in the sky may be resolved as two stars with the help of a telescope. Any optical device is limited by its resolving power. The light microscope cannot resolve detail finer than 0.2 micrometers, about the size of the smallest bacterium; consequently, no matter how many times its image of such a bacterium is magnified, the light microscope cannot show the details of the cell's internal structure.

From the year 1665, when English microscopist Robert Hooke discovered cells, until the middle of the twentieth century, biologists had only light microscopes for viewing cells. But they discovered a great deal, including the cells composing animal and plant tissues, microscopic organisms, and some of the structures within cells. By the mid-1800s, these discoveries led to the cell theory, which states that all living things are composed of cells and that all cells come from other cells.

Our knowledge of cell structure took a giant leap forward as biologists began using the electron microscope in the 1950s. Instead of light, the electron microscope uses a beam of electrons and has a much higher resolving power than the light microscope. In fact, the most powerful modern electron microscopes can distinguish objects as small as 0.2 nanometers, a thousandfold improvement over the light microscope. The period at the end of this sentence is about a million times bigger than an object 0.2 nanometers in diameter, which is the size of a large atom. Only under special conditions can electron microscopes detect individual atoms. However, cells, cellular

organelles, and even molecules like DNA and protein are much larger than single atoms.

Biologists use the scanning electron microscope to study the detailed architecture of cell surfaces. It uses an electron beam to scan the surface of a cell or group of cells that have been coated with metal. The metal stops the beam from going through the cells. When the metal is hit by the beam, it emits electrons. The electrons are focused to form an image of the outside of the cells. The scanning electron microscope produces images that look three-dimensional.

The transmission electron microscope, on the other hands, is used to study the details of internal cell structure. Specimens are cut into extremely thin sections, and the transmission electron microscope aims an electron beam through a section, just as a light microscope aims a beam of light through a specimen. However, instead of lenses made of glass, the transmission electron microscope uses electromagnets as lenses, as do all electron microscopes. The electromagnets bend the electron beam to magnify and focus an image onto a viewing screen or photographic film.

Electron microscopes have truly revolutionized the study of cells and cell organelles. Nonetheless, they have not replaced the light microscope. One problem with electron microscopes is that they cannot be used to study living specimens because the specimen must be held in a vacuum chamber; that is, all the air and liquid must be removed. For a biologist studying a living process, such as the whirling movement of a bacterium, a light microscope equipped with a video camera might be better than either a scanning electron microscope or a transmission electron microscope. Thus, the light microscope remains a useful tool, especially for studying living cells. The size of a cell often determines the type of microscope a biologist uses to study it.

ENERGY / INDUSTRY / TECHNOLOGY

06 – 1 Powering the Industrial Revolution

In Britain one of the most dramatic changes of the Industrial Revolution was the harnessing of power. Until the reign of George III (1760 – 1820), available sources of power for work and travel had not increased since the Middle Ages. There were three sources of power: animal or human muscles; the wind, operating on sail or windmill; and running water. Only the last of these was suited at all to the continuous operating of machines, and although waterpower abounded in Lancashire and Scotland and ran grain mills as well as textile mills, it had one great disadvantage: streams flowed where nature intended them to, and water-driven factories had to be located on their banks whether or not the location was desirable for other reasons. Furthermore, even the most reliable waterpower varied with the seasons and disappeared in a drought. The new age of machinery, in short, could not have been born without a new source of both movable and constant power.

The source had long been known but not exploited. Early in the eighteenth century, a pump had come into use in which expanding steam raised a piston in a cylinder, and atmospheric pressure brought it down again when the steam condensed inside the cylinder to form a vacuum. This “atmospheric engine,” invented by Thomas Savery and vastly improved by his partner, Thomas Newcomen, embodied revolutionary principles, but it was so slow and wasteful of fuel that it could not be employed outside the coal mines for which it had been designed. In the 1760s, James Watt perfected a separate condenser for the steam, so that the cylinder did not have to be cooled at every stroke; then he devised a way to make the piston turn a wheel and thus convert reciprocating (back and forth) motion into rotary motion. He thereby transformed an inefficient pump of limited use into a steam engine of a thousand uses. The final step came when steam was introduced into the cylinder to drive the piston backward as well as forward, thereby increasing the speed of the engine and cutting its fuel consumption.

Watt's steam engine soon showed what it could do. It liberated industry from dependence on running water. The engine eliminated water in the mines by driving efficient pumps, which made possible deeper and deeper mining. The ready availability of

coal inspired William Murdoch during the 1790s to develop the first new form of nighttime illumination to be discovered in a millennium and a half. Coal gas rivaled smoky oil lamps and flickering candles, and early in the new century, well-to-do Londoners grew accustomed to gaslit houses and even streets. Iron manufacturers, which had starved for fuel while depending on charcoal, also benefited from ever-increasing supplies of coal: blast furnaces with steam-powered bellows turned out more iron and steel for the new machinery. Steam became the motive force of the Industrial Revolution as coal and iron ore were the raw materials.

By 1800 more than a thousand steam engines were in use in the British Isles, and Britain retained a virtual monopoly on steam engine production until the 1830s. Steam power did not merely spin cotton and roll iron; early in the new century, it also multiplied ten times over the amount of paper that a single worker could produce in a day. At the same time, operators of the first printing presses run by steam rather than by hand found it possible to produce a thousand pages in an hour rather than thirty. Steam also promised to eliminate a transportation problem not fully solved by either canal boats or turnpikes. Boats could carry heavy weights, but canals could not cross hilly terrain; turnpikes could cross the hills, but the roadbeds could not stand up under great weights. These problems needed still another solution, and the ingredients for it lay close at hand. In some industrial regions, heavily laden wagons, with flanged wheels, were being hauled by horses along metal rails; and the stationary steam engine was puffing in the factory and mine. Another generation passed before inventors succeeded in combining these ingredients, by putting the engine on wheels and the wheels on the rails, so as to provide a machine to take the place of the horse. Thus the railroad age sprang from what had already happened in the eighteenth century.

71 – 1 Electrical Energy from the Ocean

Solar energy reaching Earth is responsible for differential heating of the atmosphere and thus air circulation as wind. Some of the energy of wind is transferred to the oceans, where it causes waves and is partly responsible for oceanic currents, although Earth's rotation also plays a role in currents. Gravitational attraction between Earth and the Sun and Moon generates tides and, along with Earth's rotation, causes most coastal areas to experience a twice-daily rise and fall of sea level. In short, the oceans possess a tremendous reservoir of largely untapped energy.

If we could effectively harness the energy possessed by the oceans, an almost limitless, largely nonpolluting energy supply would be ensured. Unfortunately, ocean energy is diffuse, meaning that the amount of energy for a given volume of water is small and thus difficult to concentrate and use. Several ways of using ocean energy are being considered or are under development, and one is currently in use, although it accounts for only a tiny proportion of all energy production. Of the several sources of ocean energy—temperature differences with depth, currents, waves, and tides—only the last shows much promise for the near future.

Ocean water at depth might be as much as 25° C colder than surface water, a difference that allows for ocean thermal energy conversion (OTEC). OTEC exploits this temperature difference to run turbines and generate electricity. The amount of energy available is enormous, but a number of practical problems must be solved before it can be used. For one thing, any potential site must be close to land and also have a sufficiently rapid change with depth to result in the required temperature difference. Furthermore, enormous quantities of warm and cold seawater would have to circulate through an electrical-generating plant, thus requiring that large surface areas be devoted to this purpose.

The concept of OTEC is more than a century old, but despite several decades of research, no commercial OTEC plants are operating or even under construction, although small experimental ones have been tested in Hawaii and Japan.

Wind-generated ocean currents, such as the Gulf Stream, which flows along the east coast of North America, also possess energy that might be tapped to generate electricity. Unlike streams that can be

dammed to impound a reservoir, any electrical generating facility exploiting oceanic currents would have to concentrate currents' diffuse energy and contend with any unpredictable changes in direction. In addition, whereas hydroelectric generating plants on land depend on the rapid movement of water from a higher elevation to the turbines, the energy of ocean currents comes from their flow velocity, which is at most a few kilometers per hour.

The most obvious form of energy in the oceans lies in waves. Harnessing wave energy and converting it to electricity is not a new idea, and it has been used on an extremely limited scale. Unfortunately, the energy possessed by a wave is distributed along its crest and is difficult to concentrate. Furthermore, any facility would have to be designed to withstand the effects of storms and saltwater corrosion. The Japanese have developed wave-energy devices to power lighthouses and buoys, and a facility capable of providing power to about 300 homes began operating in Scotland during September 2000.

Perhaps tidal power is the most promising form of ocean energy. In fact, it has been used for centuries in some coastal areas to run mills, but its use at present for electrical generation is limited. Most coastal areas experience a twice-daily rise and fall of tides, but only a few areas are suitable for exploiting this energy source. One limitation is that the tidal range must be at least five meters, and there must also be a coastal region where water can be stored following high tide.

Suitable sites for using tidal power are limited not only by tidal range but also by location. Many areas along the U. S. Gulf Coast would certainly benefit from tidal power plants, but a tide range of generally less than one meter precludes the possibility of development. Even areas with an appropriate tidal range such as the Arctic islands of Canada offer little potential because of their great distances from population centers.

04 – 3 Petroleum Resources

Petroleum, consisting of crude oil and natural gas, seems to originate from organic matter in marine sediment. Microscopic organisms settle to the seafloor and accumulate in marine mud. The organic matter may partially decompose, using up the dissolved oxygen in the sediment. As soon as the oxygen is gone, decay stops and the remaining organic matter is preserved.

Continued sedimentation—the process of deposits' settling on the sea bottom—buries the organic matter and subjects it to higher temperatures and pressures, which convert the organic matter to oil and gas. As muddy sediments are pressed together, the gas and small droplets of oil may be squeezed out of the mud and may move into sandy layers nearby. Over long periods of time (millions of years), accumulations of gas and oil can collect in the sandy layers. Both oil and gas are less dense than water, so they generally tend to rise upward through water-saturated rock and sediment.

Oil pools are valuable underground accumulations of oil, and oil fields are regions underlain by one or more oil pools. When an oil pool or field has been discovered, wells are drilled into the ground. Permanent towers, called derricks, used to be built to handle the long sections of drilling pipe. Now portable drilling machines are set up and are then dismantled and removed. When the well reaches a pool, oil usually rises up the well because of its density difference with water beneath it or because of the pressure of expanding gas trapped above it. Although this rise of oil is almost always carefully controlled today, spouts of oil, or gushers, were common in the past. Gas pressure gradually dies out, and oil is pumped from the well. Water or steam may be pumped down adjacent wells to help push the oil out. At a refinery, the crude oil from underground is separated into natural gas, gasoline, kerosene, and various oils. Petrochemicals such as dyes, fertilizer, and plastic are also manufactured from the petroleum.

As oil becomes increasingly difficult to find, the search for it is extended into more-hostile environments. The development of the oil field on the North Slope of Alaska and the construction of the Alaska pipeline are examples of the great expense and difficulty involved in new oil discoveries. Offshore drilling platforms extend the search for oil to the ocean's continental shelves—those gently sloping submarine regions at the

edges of the continents. More than one-quarter of the world's oil and almost one-fifth of the world's natural gas come from offshore, even though offshore drilling is six to seven times more expensive than drilling on land. A significant part of this oil and gas comes from under the North Sea between Great Britain and Norway.

Of course, there is far more oil underground than can be recovered. It may be in a pool too small or too far from a potential market to justify the expense of drilling. Some oil lies under regions where drilling is forbidden, such as national parks or other public lands. Even given the best extraction techniques, only about 30 to 40 percent of the oil in a given pool can be brought to the surface. The rest is far too difficult to extract and has to remain underground.

Moreover, getting petroleum out of the ground and from under the sea and to the consumer can create environmental problems anywhere along the line. Pipelines carrying oil can be broken by faults or landslides, causing serious oil spills. Spillage from huge oil-carrying cargo ships, called tankers, involved in collisions or accidental groundings (such as the one off Alaska in 1989) can create oil slicks at sea. Offshore platforms may also lose oil, creating oil slicks that drift ashore and foul the beaches, harming the environment. Sometimes, the ground at an oil field may subside as oil is removed. The Wilmington field near Long Beach, California, has subsided nine meters in 50 years; protective barriers have had to be built to prevent seawater from flooding the area. Finally, the refining and burning of petroleum and its products can cause air pollution. Advancing technology and strict laws, however, are helping control some of these adverse environmental effects.

18 – 1 Industrialization in the Netherlands and Scandinavia

While some European countries, such as England and Germany, began to industrialize in the eighteenth century, the Netherlands and the Scandinavian countries of Denmark, Norway, and Sweden developed later. All four of these countries lagged considerably behind in the early nineteenth century. However, they industrialized rapidly in the second half of the century, especially in the last two or three decades. In view of their later start and their lack of coal—undoubtedly the main reason they were not among the early industrializers—it is important to understand the sources of their success.

All had small populations. At the beginning of the nineteenth century, Denmark and Norway had fewer than 1 million people, while Sweden and the Netherlands had fewer than 2.5 million inhabitants. All exhibited moderate growth rates in the course of the century (Denmark the highest and Sweden the lowest), but all more than doubled in population by 1900. Density varied greatly. The Netherlands had one of the highest population densities in Europe, whereas Norway and Sweden had the lowest. Denmark was in between but closer to the Netherlands.

Considering human capital as a characteristic of the population, however, all four countries were advantaged by the large percentages of their populations who could read and write. In both 1850 and 1914, the Scandinavian countries had the highest literacy rates in Europe, or in the world, and the Netherlands was well above the European average. This fact was of enormous value in helping the national economies find their niches in the evolving currents of the international economy.

Location was an important factor for all four countries. All had immediate access to the sea, and this had important implications for a significant international resource, fish, as well as for cheap transport, merchant marines, and the shipbuilding industry. Each took advantage of these opportunities in its own way. The people of the Netherlands, with a long tradition of fisheries and mercantile shipping, had difficulty in developing good harbors suitable for steamships: eventually they did so at Rotterdam and Amsterdam, with exceptional results for transit trade with Germany and central Europe and for the processing of overseas foodstuffs and raw materials (sugar,

tobacco, chocolate, grain, and eventually oil). Denmark also had an admirable commercial history, particularly with respect to traffic through the Sound (the strait separating Denmark and Sweden). In 1857, in return for a payment of 63 million kronor from other commercial nations, Denmark abolished the Sound toll dues, the fees it had collected since 1497 for the use of the Sound. This, along with other policy shifts toward free trade, resulted in a significant increase in traffic through the Sound and in the port of Copenhagen.

The political institutions of the four countries posed no significant barriers to industrialization or economic growth. The nineteenth century passed relatively peacefully for these countries, with progressive democratization taking place in all of them. They were reasonably well governed, without notable corruption or grandiose state projects, although in all of them the government gave some aid to railways, and in Sweden the state built the main lines. As small countries dependent on foreign markets, they followed a liberal trade policy in the main, though a protectionist movement developed in Sweden. In Denmark and Sweden agricultural reforms took place gradually from the late eighteenth century through the first half of the nineteenth, resulting in a new class of peasant landowners with a definite market orientation.

The key factor in the success of these countries (along with high literacy, which contributed to it) was their ability to adapt to the international division of labor determined by the early industrializers and to stake out areas of specialization in international markets for which they were especially well suited. This meant a great dependence on international commerce, which had notorious fluctuations; but it also meant high returns to those factors of production that were fortunate enough to be well placed in times of prosperity. In Sweden exports accounted for 18 percent of the national income in 1870, and in 1913, 22 percent of a much larger national income. In the early twentieth century, Denmark exported 63 percent of its agricultural production: butter, pork products, and eggs. It exported 80 percent of its butter, almost all to Great Britain, where it accounted for 40 percent of British butter imports.

21 – 1 Geothermal Energy

Earth's internal heat, fueled by radioactivity, provides the energy for plate tectonics and continental drift, mountain building, and earthquakes. It can also be harnessed to drive electric generators and heat homes. Geothermal energy becomes available in a practical form when underground heat is transferred by water that is heated as it passes through a subsurface region of hot rocks (a heat reservoir) that may be hundreds or thousands of feet deep. The water is usually naturally occurring groundwater that seeps down along fractures in the rock; less typically, the water is artificially introduced by being pumped down from the surface. The water is brought to the surface, as a liquid or steam, through holes drilled for the purpose.

By far the most abundant form of geothermal energy occurs at the relatively low temperatures of 80° to 180° centigrade. Water circulated through heat reservoirs in this temperature range is able to extract enough heat to warm residential, commercial, and industrial spaces. More than 20,000 apartments in France are now heated by warm underground water drawn from a heat reservoir in a geologic structure near Paris called the Paris Basin. Iceland sits on a volcanic structure known as the Mid-Atlantic Ridge. Reykjavik, the capital of Iceland, is entirely heated by geothermal energy derived from volcanic heat.

Geothermal reservoirs with temperatures above 180° centigrade are useful for generating electricity. They occur primarily in regions of recent volcanic activity as hot, dry rock; natural hot water; or natural steam. The latter two sources are limited to those few areas where surface water seeps down through underground faults or fractures to reach deep rocks heated by the recent activity of molten rock material. The world's largest supply of natural steam occurs at The Geysers, 120 kilometers north of San Francisco, California. In the 1990s enough electricity to meet about half the needs of San Francisco was being generated there. This facility was then in its third decade of production and was beginning to show signs of decline, perhaps because of over development. By the late 1990s some 70 geothermal electric-generating plants were in operation in California, Utah, Nevada, and Hawaii, generating enough power to supply about a million people. Eighteen countries now generate electricity using geothermal heat.

Extracting heat from very hot, dry rocks presents a more difficult problem: the rocks must be fractured to permit the circulation of water, and the water must be provided artificially. The rocks are fractured by water pumped down at very high pressures. Experiments are under way to develop technologies for exploiting this resource.

Like most other energy sources, geothermal energy presents some environmental problems. The surface of the ground can sink if hot groundwater is withdrawn without being replaced. In addition, water heated geothermally can contain salts and toxic materials dissolved from the hot rock. These waters present a disposal problem if they are not returned to the ground from which they were removed.

The contribution of geothermal energy to the world's energy future is difficult to estimate. Geothermal energy is in a sense not renewable, because in most cases the heat would be drawn out of a reservoir much more rapidly than it would be replaced by the very slow geological processes by which heat flows through solid rock into a heat reservoir. However, in many places (for example, California, Hawaii, the Philippines, Japan, Mexico, the rift valleys of Africa) the resource is potentially so large that its future will depend on the economics of production. At present, we can make efficient use of only naturally occurring hot water or steam deposits. Although the potential is enormous, it is likely that in the near future geothermal energy can make important local contributions only where the resource is close to the user and the economics are favorable, as they are in California, New Zealand, and Iceland. Geothermal energy probably will not make large-scale contributions to the world energy budget until well into the twenty-first century, if ever.

26 – 1 Energy and the Industrial Revolution

For years historians have sought to identify crucial elements in the eighteenth-century rise in industry, technology, and economic power known as the Industrial Revolution, and many give prominence to the problem of energy. Until the eighteenth century, people relied on energy derived from plants as well as animal and human muscle to provide power. Increased efficiency in the use of water and wind helped with such tasks as pumping, milling, or sailing. However, by the eighteenth century, Great Britain in particular was experiencing an energy shortage. Wood, the primary source of heat for homes and industries and also used in the iron industry as processed charcoal, was diminishing in supply. Great Britain had large amounts of coal; however, there were not yet efficient means by which to produce mechanical energy or to power machinery. This was to occur with progress in the development of the steam engine.

In the late 1700s James Watt designed an efficient and commercially viable steam engine that was soon applied to a variety of industrial uses as it became cheaper to use. The engine helped solve the problem of draining coal mines of groundwater and increased the production of coal needed to power steam engines elsewhere. A rotary engine attached to the steam engine enabled shafts to be turned and machines to be driven, resulting in mills using steam power to spin and weave cotton. Since the steam engine was fired by coal, the large mills did not need to be located by rivers, as had mills that used water-driven machines. The shift to increased mechanization in cotton production is apparent in the import of raw cotton and the sale of cotton goods. Between 1760 and 1850, the amount of raw cotton imported increased 230 times. Production of British cotton goods increased sixtyfold, and cotton cloth became Great Britain's most important product, accounting for one-half of all exports. The success of the steam engine resulted in increased demands for coal, and the consequent increase in coal production was made possible as the steam-powered pumps drained water from the ever-deeper coal seams found below the water table.

The availability of steam power and the demands for new machines facilitated the transformation of the iron industry. Charcoal, made from wood and thus in limited supply, was replaced with coal-derived coke (substance left after coal is heated) as steam-driven bellows came into use for

producing of raw iron. Impurities were burnt away with the use of coke, producing a high-quality refined iron. Reduced cost was also instrumental in developing steam-powered rolling mills capable of producing finished iron of various shapes and sizes. The resulting boom in the iron industry expanded the annual iron output by more than 170 times between 1740 and 1840, and by the 1850s Great Britain was producing more tons of iron than the rest of the world combined. The developments in the iron industry were in part a response to the demand for more machines and the ever-widening use of higher-quality iron in other industries.

Steam power and iron combined to revolutionize transport, which in turn had further implications. Improvements in road construction and sailing had occurred, but shipping heavy freight over land remained expensive, even with the use of rivers and canals wherever possible. Parallel rails had long been used in mining operations to move bigger loads, but horses were still the primary source of power. However, the arrival of the steam engine initiated a complete transformation in rail transportation, entrenching and expanding the Industrial Revolution. As transportation improved, distant and larger markets within the nation could be reached, thereby encouraging the development of larger factories to keep pace with increasing sales. Greater productivity and rising demands provided entrepreneurs with profits that could be reinvested to take advantage of new technologies to further expand capacity, or to seek alternative investment opportunities. Also, the availability of jobs in railway construction attracted many rural laborers accustomed to seasonal and temporary employment. When the work was completed, many moved to other construction jobs or to factory work in cities and towns, where they became part of an expanding working class.

29 – 3 The History of Waterpower

Moving water was one of the earliest energy sources to be harnessed to reduce the workload of people and animals. No one knows exactly when the waterwheel was invented, but irrigation systems existed at least 5,000 years ago, and it seems probable that the earliest waterpower device was the noria, a waterwheel that raised water for irrigation in attached jars. The device appears to have evolved no later than the fifth century B.C., perhaps independently in different regions of the Middle and Far East.

The earliest waterpower mills were probably vertical-axis mills for grinding corn, known as Norse or Greek mills, which seem to have appeared during the first or second century B.C. in the Middle East and a few centuries later in Scandinavia. In the following centuries, increasingly sophisticated waterpower mills were built throughout the Roman Empire and beyond its boundaries in the Middle East and northern Europe. In England, the Saxons are thought to have used both horizontal-axis and vertical-axis wheels. The first documented English mill was in the eighth century, but three centuries later about 5,000 were recorded, suggesting that every settlement of any size had its mill.

Raising water and grinding corn were by no means the only uses of the waterpower mill, and during the following centuries, the applications of waterpower kept pace with the developing technologies of mining, iron working, paper making, and the wool and cotton industries. Water was the main source of mechanical power, and by the end of the seventeenth century, England alone is thought to have had some 20,000 working mill.

There was much debate on the relative efficiencies of different types of waterwheels. The period from about 1650 until 1800 saw some excellent scientific and technical investigations of different designs. They revealed output powers ranging from about 1 horsepower to perhaps 60 for the largest wheels and confirmed that for maximum efficiency, the water should pass across the blades as smoothly as possible and fall away with minimum speed, having given up almost all of its kinetic energy. (They also proved that, in principle, the overshot wheel, a type of wheel in which an overhead stream of water powers the wheel, should win the efficiency competition.)

But then steam power entered the scene, putting the whole future of waterpower in doubt. An energy

analyst writing in the year 1800 would have painted a very pessimistic picture of the future for waterpower. The coal-fired steam engine was taking over, and the waterwheel was fast becoming obsolete. However, like many later experts, this one would have suffered from an inability to see into the future. A century later the picture was completely different: by then, the world had an electric industry, and a quarter of its generating capacity was water powered.

30 – 3 The Invention of the Mechanical Clock

In Europe, before the introduction of the mechanical clock, people told time by sun (using, for example, shadow sticks or sun dials) and water clocks. Sun clocks worked, of course, only on clear days; water clocks misbehaved when the temperature fell toward freezing, to say nothing of long-run drift as the result of sedimentation and clogging. Both these devices worked well in sunny climates; but in northern Europe the sun may be hidden by clouds for weeks at a time, while temperatures vary not only seasonally but from day to night.

Medieval Europe gave new importance to reliable time. The Catholic Church had its seven daily prayers, one of which was at night, requiring an alarm arrangement to waken monks before dawn. And then the new cities and towns, squeezed by their walls, had to know and order time in order to organize collective activity and ration space. They set a time to go to work, to open the market, to close the market, to leave work, and finally a time to put out fires and to go to sleep. All this was compatible with older devices so long as there was only one authoritative timekeeper; but with urban growth and the multiplication of time signals, discrepancy brought discord and strife. Society needed a more dependable instrument of time measurement and found it in the mechanical clock.

We do not know who invented this machine, or where. It seems to have appeared in Italy and England (perhaps simultaneous invention) between 1275 and 1300. Once known, it spread rapidly, driving out water clocks but not solar dials, which were needed to check the new machines against the timekeeper of last resort. These early versions were rudimentary, inaccurate, and prone to breakdown.

Ironically, the new machine tended to undermine Catholic Church authority. Although church ritual had sustained an interest in timekeeping throughout the centuries of urban collapse that followed the fall of Rome, church time was nature's time. Day and night were divided into the same number of parts, so that except at the equinoxes, days and night hours were unequal; and then of course the length of these hours varied with the seasons. But the mechanical clock kept equal hours, and this implied a new time reckoning. The Catholic Church resisted, not coming over to the new hours for about a century. From the start,

however, the towns and cities took equal hours as their standard, and the public clocks installed in town halls and market squares became the very symbol of a new, secular municipal authority. Every town wanted one; conquerors seized them as especially precious spoils of war; tourists came to see and hear these machines the way they made pilgrimages to sacred relics.

The clock was the greatest achievement of medieval mechanical ingenuity. Its general accuracy could be checked against easily observed phenomena, like the rising and setting of the sun. The result was relentless pressure to improve technique and design. At every stage, clockmakers led the way to accuracy and precision; they became masters of miniaturization, detectors and correctors of error, searchers for new and better. They were thus the pioneers of mechanical engineering and served as examples and teachers to other branches of engineering.

The clock brought order and control, both collective and personal. Its public display and private possession laid the basis for temporal autonomy: people could now coordinate comings and goings without dictation from above. The clock provided the punctuation marks for group activity, while enabling individuals to order their own work (and that of others) so as to enhance productivity. Indeed, the very notion of productivity is a by-product of the clock: once one can relate performance to uniform time units, work is never the same. One moves from the task-oriented time consciousness of the peasant (working on job after another, as time and light permit) and the time-filling busyness of the domestic servant (who always had something to do) to an effort to maximize product per unit of time.

34 – 2 The Development of Steam Power

By the eighteenth century, Britain was experiencing a severe shortage of energy. Because of the growth of population, most of the great forests of medieval Britain had long ago been replaced by fields of grain and hay. Wood was in ever-shorter supply, yet it remained tremendously important. It served as the primary source of heat for all homes and industries and as a basic raw material. Processed wood (charcoal) was the fuel that was mixed with iron ore in the blast furnace to produce pig iron (raw iron). The iron industry's appetite for wood was enormous, and by 1740 the British iron industry was stagnating. Vast forests enabled Russia to become the world's leading producer of iron, much of which was exported to Britain. But Russia's potential for growth was limited too, and in a few decades Russia would reach the barrier of inadequate energy that was already holding England back.

As this early energy crisis grew worse, Britain looked toward its abundant and widely scattered reserves of coal as an alternative to its vanishing wood. Coal was first used in Britain in the late Middle Ages as a source of heat. By 1640 most homes in London were heated with it, and it also provided heat for making beer, glass, soap, and other products. Coal was not used, however, to produce mechanical energy or to power machinery. It was there that coal's potential was enormous.

As more coal was produced, mines were dug deeper and deeper and were constantly filling with water. Mechanical pumps, usually powered by hundreds of horses walking in circles at the surface, had to be installed. Such power was expensive and bothersome. In an attempt to overcome these disadvantages, Thomas Savery in 1698 and Thomas Newcomen in 1705 invented the first primitive steam engines. Both engines were extremely inefficient. Both burned coal to produce steam, which was then used to operate a pump. However, by the early 1770s, many of the Savery engines and hundreds of the Newcomen engines were operating successfully, though inefficiently, in English and Scottish mines.

In the early 1760s, a gifted young Scot named James Watt was drawn to a critical study of the steam engine. Watt was employed at the time by the University of Glasgow as a skilled crafts worker making scientific instruments. In 1763, Watt was called on to repair a Newcomen engine being used

in a physics course. After a series of observations, Watt saw that the Newcomen's waste of energy could be reduced by adding a separate condenser. This splendid invention, patented in 1769, greatly increased the efficiency of the steam engine. The steam engine of Watt and his followers was the technological advance that gave people, at least for a while, unlimited power and allowed the invention and use of all kinds of power equipment.

The steam engine was quickly put to use in several industries in Britain. It drained mines and made possible the production of ever more coal to feed steam engines elsewhere. The steam power plant began to replace waterpower in the cotton-spinning mills as well as other industries during the 1780s, contributing to a phenomenal rise in industrialization. The British iron industry was radically transformed. The use of powerful, steam-driven bellows in blast furnaces helped iron makers switch over rapidly from limited charcoal to unlimited coke (which is made from coal) in the smelting of pig iron (the process of refining impure iron) after 1770. In the 1780s, Henry Cort developed the puddling furnace, which allowed pig iron to be refined in turn with coke. Cort also developed heavy-duty, steam-powered rolling mills, which were capable of producing finished iron in every shape and form.

The economic consequence of these technical innovations in steam power was a great boom in the British iron industry. In 1740 annual British iron production was only 17,000 tons, but by 1844, with the spread of coke smelting and the impact of Cort's inventions, it had increased to 3,000,000 tons. This was a truly amazing expansion. Once scarce and expensive, iron became cheap, basic, and indispensable to the economy.

36 – 3 Industrial Melanism: The Case of the Peppered Moth

The idea of natural selection is that organisms in a species that have characteristics favoring survival are most likely to survive and produce offspring with the same characteristics. Because the survival of organisms with particular characteristics is favored over the survival of other organisms in the same species that lack these characteristics, future generations of the species are likely to include more organisms with the favorable characteristics.

One of the most thoroughly analyzed examples of natural selection in operation is the change in color that has occurred in certain populations of the peppered moth, *Biston betularia*, in industrial regions of Europe during the past 100 years. Originally moths were uniformly pale gray or whitish in color; dark-colored (melanic) individuals were rare and made up less than 2 percent of the population. Over a period of decades, dark-colored forms became an increasingly large fraction of some populations and eventually came to dominate peppered moth populations in certain areas—especially those of extreme industrialization such as the Ruhr Valley of Germany and the Midlands of England. Coal from industry released large amounts of black soot into the environment, but the increase of the dark-colored forms was not due to genetic mutations caused by industrial pollution. For example, caterpillars that feed on soot-covered leaves did not give rise to dark-colored adults. Rather, pollution promoted the survival of dark forms on soot-covered trees. Melanics were normally quickly eliminated in nonindustrial areas by adverse selection; birds spotted them easily. This phenomenon, an increase in the frequency of dark-colored mutants in polluted areas, is known as industrial melanism. The North American equivalent of this story is another moth, the swettaria form of *Biston cognataria*, first noticed in industrialized areas such as Chicago and New York City in the early 1900s. By 1961 it constituted over 90 percent of the population in parts of Michigan.

The idea that natural selection was responsible for the changing ratio of dark- to light-colored peppered moths was developed in the 1950s by H. B. D. Kettlewell of Oxford University. If natural selection was the explanation, then there should be different survival rates for dark- and light-colored moths. To determine whether this was true, Kettlewell released thousands of light and dark moths (each marked with a paint spot) into rural

and industrialized areas. In the nonindustrial area of Dorset, he recaptured 14.6 percent of the pale forms but only 4.7 percent of the dark forms. In the industrial area of Birmingham, the situation was reversed: 13 percent of pale forms but 27.5 percent of dark forms were recaptured.

Clearly some environmental factor was responsible for the greater survival rates of dark moths. Birds were predators of peppered moths. Kettlewell hypothesized that the normal pale forms are difficult to see when resting on lichen-covered trees, whereas dark forms are conspicuous. In industrialized areas, lichens are destroyed by pollution, tree barks become darker, and dark moths are the ones birds have difficulty detecting. As a test, Kettlewell set up hidden observation positions and watched birds voraciously eat moths placed on tree trunks of a contrasting color. The action of natural selection in producing a small but highly significant step of evolution was seemingly demonstrated, with birds as the selecting force.

Not every researcher has been convinced that natural selection by birds is the only explanation of the observed frequencies of dark and light peppered moths. More recent data, however, provide additional support for Kettlewell's ideas about natural selection. The light-colored form of the peppered moth is making a strong comeback. In Britain, a Clean Air Act was passed in 1965. Sir Cyril Clarke has been trapping moths at his home in Liverpool, Merseyside, since 1959. Before about 1975, 90 percent of the moths were dark, but since then there has been a steep decline in melanic forms, and in 1989 only 29.6 percent of the moths caught were melanic. The mean concentration of sulphur dioxide pollution fell from about 300 micrograms per cubic meter in 1970 to less than 50 micrograms per cubic meter in 1975 and has remained fairly constant since then. If the spread of the light-colored form of the moth continues at the same speed as the melanic form spread in the last century, soon the melanic form will again be only an occasional resident of the Liverpool area.

39 – 3 Forest Fire Suppression

Forest fires have recently increased in intensity and extent in some forest types throughout the western United States. This recent increase in fires has resulted partly from climate change the recent trend toward hot, dry summers and partly from human activities, for complicated reasons that foresters came increasingly to understand about 30 years ago but whose relative importance is still debated. One factor is the direct effect of logging, which often turns a forest into something approximating a huge pile of kindling (wood for burning): the ground in a logged forest may remain covered with branches and treetops, left behind when the valuable trunks are carted away; a dense growth of new vegetation springs up, further increasing the forest's fuel loads; and the trees logged and removed are of course the biggest and most fire-resistant individuals, leaving behind smaller and more flammable trees.

Another factor is that the United States Forest Service in the first decade of the 1900s adopted the policy of fire suppression (attempting to put out forest fires) for the obvious reason that it did not want valuable timber to go up in smoke, or people's homes and lives to be threatened. The Forest Service's announced goal became "Put out every forest fire by 10:00 A. M. on the morning after the day when it is first reported." Firefighters became much more successful at achieving that goal after 1945, thanks to improved firefighting technology. For a few decades the amount of land burnt annually decreased by 80 percent. That happy situation began to change in the 1980s, due to the increasing frequency of large forest fires that were essentially impossible to extinguish unless rain and low winds combined to help. People began to realize that the United States federal government's fire-suppression policy was contributing to those big fires and that natural fires caused by lightning had previously played an important role in maintaining forest structure.

The natural role of fire varies with altitude, tree species, and forest type. To make Montana's low-altitude ponderosa pine forest as an example, historical records, plus counts of annual tree rings and datable fire scars on tree stumps, demonstrated that a ponderosa pine forest experiences a lightning-lit fire about once a decade under natural conditions (i.e., before fire suppression began around 1910 and became effective after 1945). The mature ponderosa trees

have bark two inches thick and are relatively resistant to fire, which instead burns out the understory - the lower layer - of fire-sensitive Douglas fir seedlings that have grown up since the previous fire. But after only a decade's growth until the next fire, those young seedling plants are still too low for fire to spread from them into the crowns of the ponderosa pine trees. Hence the fire remains confined to ground and understory. As a result, many natural ponderosa pine forests have a parklike appearance, with low fuel loads, big trees spaced apart, and a relatively clear understory.

However, loggers concentrated on removing those big, old, valuable, fire-resistant ponderosa pines, while fire suppression for decades let the understory fill up with Douglas fir saplings that would in turn become valuable when full-grown. Tree densities increased from 30 to 200 trees per acre, the forest's fuel load increased by a factor of 6, and the government repeatedly failed to appropriate money to thin out the saplings. When a fire finally does start in a sapling-choked forest, whether due to lightning or human carelessness or (regrettably often) intentional arson, the dense, tall saplings (young trees) may become a ladder that allows the fire to jump into the crowns of the trees. The outcome is sometimes an unstoppable inferno.

Foresters now identify the biggest problem in managing Western forests as what to do with those increased fuel loads that built up during the previous half century of effective fire suppression. In the wetter eastern United States, dead trees rot away more quickly than in the drier West, where more dead trees persist like giant matchsticks. In an ideal world, the Forest Service would manage and restore the forests, thin them out, and remove the dense understory by cutting or by controlled small fires. But no politician or voter wants to spend what it would cost to do that.

43 – 2 The Origin of Petroleum

Petroleum is defined as a gaseous, liquid, and semisolid naturally occurring substance that consists chiefly of hydrocarbons (chemical compounds of carbon and hydrogen). Petroleum is therefore a term that includes both oil and natural gas. Petroleum is nearly always found in marine sedimentary rocks. In the ocean, microscopic phytoplankton (tiny floating plants) and bacteria (simple, single-celled organisms) are the principal sources of organic matter that is trapped and buried in sediment. Most of the organic matter is buried in clay that is slowly converted to a fine-grained sedimentary rock known as shale. During this conversion, organic compounds are transformed to oil and natural gas.

Sampling on the continental shelves and along the base of the continental slopes has shown that fine muds beneath the seafloor contain up to 8 percent organic matter. Two additional kinds of evidence support the hypothesis that petroleum is a product of the decomposition of organic matter: oil possesses optical properties known only in hydrocarbons derived from organic matter, and oil contains nitrogen and certain compounds believed to originate only in living matter. A complex sequence of chemical reactions is involved in converting the original solid organic matter to oil and gas, and additional chemical changes may occur in the oil and gas even after they have formed.

It is now well established that petroleum migrates through aquifers and can become trapped in reservoirs. Petroleum migration is analogous to groundwater migration. When oil and gas are squeezed out of the shale in which they originated and enter a body of sandstone or limestone somewhere above, they migrate readily because sandstones (consisting of quartz grains) and limestones (consisting of carbonate minerals) are much more permeable than any shale. The force of molecular attraction between oil and quartz or carbonate minerals is weaker than that between water and quartz or carbonate minerals. Hence, because oil and water do not mix, water remains fastened to the quartz or carbonate grains, while oil occupies the central parts of the larger openings in the porous sandstone or limestone. Because oil is lighter than water, it tends to glide upward past the carbonate- and quartz-held water. In this way, oil becomes segregated from the water; when it encounters a trap, it can form a pool.

Most of the petroleum that forms in sediments does not find a suitable trap and eventually makes its way, along with groundwater, to the surface of the sea. It is estimated that no more than 0.1 percent of all the organic matter originally buried in a sediment is eventually trapped in an oil pool. It is not surprising, therefore, that the highest ratio of oil and gas pools to volume of sediment is found in rock no older than 2.5 million years—young enough so that little of the petroleum has leaked away—and that nearly 60 percent of all oil and gas discovered so far has been found in strata that formed in the last 65 million years. This does not mean that older rocks produced less petroleum; it simply means that oil in older rocks has had a longer time in which to leak away.

How much oil is there in the world? This is an extremely controversial question. Many billions of barrels of oil have already been pumped out of the ground. A lot of additional oil has been located by drilling but is still waiting to be pumped out. Possibly a great deal more oil remains to be found by drilling. Unlike coal, the volume of which can be accurately estimated, the volume of undiscovered oil can only be guessed at. Guesses involve the use of accumulated experience from a century of drilling. Knowing how much oil has been found in an intensively drilled area, such as eastern Texas, experts make estimates of probable volumes in other regions where rock types and structures are similar to those in eastern Texas. Using this approach and considering all the sedimentary basins of the world, experts estimate that somewhere between 1,500 and 3,000 billion barrels of oil will eventually be discovered.

49 – 3 Background for the Industrial Revolution

The Industrial Revolution had several roots, one of which was a commercial revolution that, beginning as far back as the sixteenth century, accompanied Europe's expansion overseas. Both exports and imports showed spectacular growth, particularly in England and France. An increasingly larger portion of the stepped-up commercial activity was the result of trade with overseas colonies. Imports included a variety of new beverages, spices, and ship's goods around the world and brought money flowing back. Europe's economic institutions, particularly those in England, were strong, had wealth available for new investment, and seemed almost to be waiting for some technological breakthrough that would expand their profit-making potential even more.

The breakthrough came in Great Britain, where several economic advantages created a climate especially favorable to the encouragement of new technology. One was its geographic location at the crossroads of international trade. Internally, Britain was endowed with easily navigable natural waterway, which helped its trade and communication with the world. Beginning in the 1770's, it enjoyed a boom in canal building, which helped make its domestic market more accessible. Because water transportation was the cheapest means of carrying goods to market, canals reduced prices and thus increased consumer demand. Great Britain also had rich deposits of coal that fed the factories springing up in industrial and consumer goods.

Another advantage was Britain's large population of rural, agricultural wage earners, as well as cottage workers, who had the potential of being more mobile than peasants of some other countries. Eventually they found their way to the cities or mining communities and provided the human power upon which the Industrial Revolution was built. The British people were also consumers; the absence of internal tariffs, such as those that existed in France or Italy or between the German states, made Britain the largest free-trade area in Europe. Britain's relatively stable government also helped create an atmosphere conducive to industrial progress.

Great Britain's better-developed banking and credit system also helped speed the industrial progress, as did the fact that it was the home of an impressive array of entrepreneurs and inventors.

Among them were a large number of nonconformists whose religious principles encouraged thrift and industry rather than luxurious living and who tended to pour their profits back into their business, thus providing the basis for continued expansion.

A precursor to the Industrial Revolution was a revolution in agricultural techniques. Ideas about agricultural reform developed first in Holland, where as early as the mid-seventeenth century, such modern methods as crop rotation, heavy fertilization, and diversification were all in use. Dutch peasant farmers were known throughout Europe for their agricultural innovations, but as British markets and opportunities grew, the English quickly learned from them. As early as the seventeenth century the Dutch were helping them drain marshes and fens where, with the help of advanced techniques, they grew new crops. By the mid-eighteenth century new agricultural methods as well as selective breeding of livestock had caught on throughout the country.

Much of the increased production was consumed by Great Britain's burgeoning population. At the same time, people were moving to the city, partly because of the enclosure movement; that is, the fencing of common fields and pastures in order to provide more compact, efficient privately held agricultural parcels that would produce more goods and greater profits. In the sixteenth century enclosures were usually used for creating sheep pastures, but by the eighteenth century new farming techniques made it advantageous for large landowners to seek enclosures in order to improve agricultural production. Between 1714 and 1820 over 6 million acres of English land were enclosed. As a result, many small, independent farmers were forced to sell out simply because they could not compete. Non-landholding peasants and cottage workers, who worked for wages and grazed cows or pigs on the village common, were also hurt when the common was no longer available. It was such people who began to flock to the cities seeking employment and who found work in the factories that would transform the nation and, the world.

59 – 1 Building Materials

It is difficult to exaggerate the radical change that the new industrial world of the nineteenth century brought to architectural materials. Since the beginning of architectural history, the same basic substances had been employed. They were provided directly by nature and used in their natural or near-natural state, only cut, shaped, and dried into the functional forms of timbers, stone blocks, and clay bricks. The exceptions were lime mortar and Roman concrete. Metals, which had the tensile strength that masonry materials lacked, were employed in minor and supplementary ways. Bronze was expensive as well as brittle. Iron, the structurally more important metal, was available in limited quantities and uneven quality, and was too easily converted to rust by the elements. It was, therefore, restricted in use to things such as tie-rods and chains, and, along with bronze, to masonry clamps and decoration. Prior to the nineteenth century, the structural presence of iron in architecture was scarcely noticeable.

The Industrial Revolution changed all that. Iron materials became available in such large quantities that they could play far more than a minor architectural role. In 1800, the world production of iron stood at 825,000 tons, by 1830 it was 1,825,000 tons, and nearly 40,000,000 in 1900—an increase of almost 50-fold over the century. The growth was not only in quantity, but quality as well. Iron, found bountifully in Earth's crust as an oxide, is a material of almost protean variability. It is not simply pure or impure, but can be made hard or soft, brittle or ductile, strong or weak. These qualities depend on carbon content, freedom from impurities (slag), and heating and cooling treatments of the refined metal. Traditionally, three versions existed: cast iron, wrought iron, and steel. Cast iron is the crudest form, containing the most impurities and thus being extremely brittle. Wrought iron, because it includes almost no carbon, is highly malleable (hence its name), but also comparatively soft. The optimum material is steel, which incorporates a restricted amount of carbon for hardness but is otherwise free of impurities, giving it great strength, and which, as a result of tempering treatments, is also malleable.

Cast and wrought iron came into prodigious manufacture in the early and mid-nineteenth century as a result of rapid growth in demand, new means to transport materials, and more efficient iron-founding techniques. But the mass production

of steel required further technological innovations to rid the metal of weakening impurities and to control more perfectly the amount of carbon added. Such advances were made with the Bessemer process (put into use in 1860) and the open-hearth process of 1864, scientific iron metallurgy in the last third of the century perfected these techniques.

A significant but little-known fact is that the main form of increased steel production was technically not steel but a kind of wrought iron. It lacked a crucial property of true steel—its hardening power—yet it differed from the older forms of wrought iron because it was free from the weakening presence of slag, at the same time being malleable (unlike cast iron). It was called steel only because the name carried the status of a high-quality and high-priced product.

Another alternative new building substance was concrete, composed of an aggregate of broken stone, gravel, or other small chunks of hard matter embedded in a matrix of lime, sand, and water. First used in Roman times, its modern revival depended on the invention of portland cement in 1824, a substance of many times greater strength, durability, and fire resistance than ancient lime cement. Mass-produced concrete began to come into widespread use in the 1850s and 1860s, in the construction of the sewers of Paris, for example. However, even with portland cement, the use of concrete was still severely restricted by its low tensile strength, but the remedy was at hand in the newly available iron and steel; their properties complemented those of concrete. The latter material was cheap, easily molded into large structural forms with great compressive but little tensile strength. Iron and steel, on the other hand, were expensive, difficult to shape, yet endowed with extreme tensile strength and easily procurable in the simple form of long, thin bars.

AGRICULTURE

21 – 2 The Origins of Agriculture

How did it come about that farming developed independently in a number of world centers (the Southeast Asian mainland, Southwest Asia, Central America, lowland and highland South America, and equatorial Africa) at more or less the same time? Agriculture developed slowly among populations that had an extensive knowledge of plants and animals. Changing from hunting and gathering to agriculture had no immediate advantages. To start with, it forced the population to abandon the nomad's life and become sedentary, to develop methods of storage and, often, systems of irrigation. While hunter-gatherers always had the option of moving elsewhere when the resources were exhausted, this became more difficult with farming. Furthermore, as the archaeological record shows, the state of health of agriculturalists was worse than that of their contemporary hunter-gatherers.

Traditionally, it was believed that the transition to agriculture was the result of a worldwide population crisis. It was argued that once hunter-gatherers had occupied the whole world, the population started to grow everywhere and food became scarce; agriculture would have been a solution to this problem. We know, however, that contemporary hunter-gatherer societies control their population in a variety of ways. The idea of a world population crisis is therefore unlikely, although population pressure might have arisen in some areas.

Climatic changes at the end of the glacial period 13,000 years ago have been proposed to account for the emergence of farming. The temperature increased dramatically in a short period of time (years rather than centuries), allowing for a growth of the hunting-gathering population due to the abundance of resources. There were, however, fluctuations in the climatic conditions, with the consequences that wet conditions were followed by dry ones, so that the availability of plants and animals oscillated brusquely.

It would appear that the instability of the climatic conditions led populations that had originally been nomadic to settle down and develop a sedentary style of life, which led in turn to population growth and to the need to increase the amount of food available. Farming originated in these conditions. Later on, it became very difficult to change

because of the significant expansion of these populations. It could be argued, however, that these conditions are not sufficient to explain the origins of agriculture. Earth had experienced previous periods of climatic change, and yet agriculture had not been developed.

It is archaeologist Steven Mithen's thesis, brilliantly developed in his book *The Prehistory of the Mind* (1996), that approximately 40,000 years ago the human mind developed cognitive fluidity, that is, the integration of the specializations of the mind: technical, natural history (geared to understanding the behavior and distribution of natural resources), social intelligence, and the linguistic capacity. Cognitive fluidity explains the appearance of art, religion, and sophisticated speech. Once humans possessed such a mind, they were able to find an imaginative solution to a situation of severe economic crisis such as the farming dilemma described earlier. Mithen proposes the existence of four mental elements to account for the emergence of farming: (1) the ability to develop tools that could be used intensively to harvest and process plant resources; (2) the tendency to use plants and animals as the medium to acquire social prestige and power; (3) the tendency to develop "social relationships" with animals structurally similar to those developed with people — specifically, the ability to think of animals as people (anthropomorphism) and of people as animals (totemism); and (4) the tendency to manipulate plants and animals.

The fact that some societies domesticated animals and plants, discovered the use of metal tools, became literate, and developed a state should not make us forget that others developed pastoralism or horticulture (vegetable gardening) but remained illiterate and at low levels of productivity; a few entered the modern period as hunting and gathering societies. It is anthropologically important to inquire into the conditions that made some societies adopt agriculture while others remained hunter-gatherers or horticulturalists. However, it should be kept in mind that many societies that knew of agriculture more or less consciously avoided it. Whether Mithen's explanation is satisfactory is open to contention, and some authors have recently emphasized the importance of other factors.

23 – 2 Seventeenth-Century Dutch Agriculture

Agriculture and fishing formed the primary sector of the economy in the Netherlands in the seventeenth century. Dutch agriculture was modernized and commercialized. New crops and agricultural techniques raised levels of production so that they were in line with market demands, and cheap grain was imported annually from the Baltic region in large quantities. According to estimates, about 120,000 tons of imported grain fed about 600,000 people: that is about a third of the Dutch population. Importing the grain, which would have been expensive and time consuming for the Dutch to have produced themselves, kept the price of grain low and thus stimulated individual demand for other foodstuffs and consumer goods.

Apart from this, being able to give up labor-intensive grain production freed both the land and the workforce for more productive agricultural divisions. The peasants specialized in livestock husbandry and dairy farming as well as in cultivating industrial crops and fodder crops: flax, madder, and rape were grown, as were tobacco, hops, and turnips. These products were bought mostly by urban businesses. There was also a demand among urban consumers for dairy products such as butter and cheese, which, in the sixteenth century, had become more expensive than grain. The high prices encouraged the peasants to improve their animal husbandry techniques; for example, they began feeding their animals indoors in order to raise the milk yield of their cows.

In addition to dairy farming and cultivating industrial crops, a third sector of the Dutch economy reflected the way in which agriculture was being modernized-horticulture. In the sixteenth century, fruit and vegetables were to be found only in gardens belonging to wealthy people. This changed in the early part of the seventeenth century when horticulture became accepted as an agricultural sector. Whole villages began to cultivate fruit and vegetables. The produce was then transported by water to markets in the cities, where the consumption of fruit and vegetables was no longer restricted to the wealthy.

As the demand for agricultural produce from both consumers and industry increased, agricultural land became more valuable and people tried to work the available land more intensively and to reclaim more land from wetlands and lakes. In order to increase

production on existing land, the peasants made more use of crop rotation and, in particular, began to apply animal waste to the soil regularly, rather than leaving the fertilization process up to the grazing livestock. For the first time industrial waste, such as ash from the soap-boilers, was collected in the cities and sold in the country as artificial fertilizer. The increased yield and price of land justified reclaiming and draining even more land.

The Dutch battle against the sea is legendary. Noorderkwartier in Holland, with its numerous lakes and stretches of water, was particularly suitable for land reclamation and one of the biggest projects undertaken there was the draining of the Beemster lake which began in 1608. The richest merchants in Amsterdam contributed money to reclaim a good 7,100 hectares of land. Forty-three windmills powered the drainage pumps so that they were able to lease the reclamation to farmers as early as 1612, with the investors receiving annual leasing payments at an interest rate of 17 percent. Land reclamation continued, and between 1590 and 1665, almost 100,000 hectares were reclaimed from the wetland areas of Holland, Zeeland, and Friesland. However, land reclamation decreased significantly after the middle of the seventeenth century because the price of agricultural products began to fall, making land reclamation far less profitable in the second part of the century.

Dutch agriculture was finally affected by the general agricultural crisis in Europe during the last two decades of the seventeenth century. However, what is astonishing about this is not that Dutch agriculture was affected by critical phenomena such as a decrease in sales and production, but the fact that the crisis appeared only relatively late in Dutch agriculture. In Europe as a whole, the exceptional reduction in the population and the related fall in demand for grain since the beginning of the seventeenth century had caused the price of agricultural products to fall. Dutch peasants were able to remain unaffected by this crisis for a long time because they had specialized in dairy farming industrial crops, and horticulture. However, toward the end of the seventeenth century, they too were overtaken by the general agricultural crisis.

33 – 2 Railroads and Commercial Agriculture in Nineteenth-Century United States

By 1850 the United States possessed roughly 9,000 miles of railroad track; ten years later it had over 30,000 miles, more than the rest of the world combined. Much of the new construction during the 1850s occurred in the west of the Appalachian Mountains – over 2,000 miles in the states of Ohio and Illinois alone.

The effect of the new railroad lines rippled outward through the economy. Farmers along the tracks began to specialize in crops that they could market in distant locations. With their profits they purchased manufactured goods that earlier they might have made at home. Before the railroad reached Tennessee, the state produced about 25,000 bushels (or 640 tons) of wheat, which sold for less than 50 cents a bushel. Once the railroad came, farmers in the same counties grew 400,000 bushels (over 10,000 tons) and sold their crop at a dollar a bushel.

The new railroad networks shifted the direction of western trade. In 1840 most northwestern grain was shipped south down the Mississippi River to the bustling port of New Orleans. But low water made steamboat travel hazardous in summer, and ice shut down traffic in winter. Products such as lard, tallow, and cheese quickly spoiled if stored in New Orleans' hot and humid warehouses. Increasingly, traffic from the Midwest flowed west to east, over the new rail lines. Chicago became the region's hub, linking the farms of the upper Midwest to New York and other eastern cities by more than 2,000 miles of track in 1855. Thus while the value of goods shipped by river to New Orleans continued to increase, the South's overall share of western trade dropped dramatically.

A sharp rise in demand for grain abroad also encouraged farmers in the Northeast and Midwest to become more commercially oriented. Wheat, which in 1845 commanded \$1.08 a bushel in New York City, fetched \$2.46 in 1855; in similar fashion the price of corn nearly doubled. Farmers responded by specializing in cash crops, borrowing to purchase more land, and investing in equipment to increase productivity.

As railroad lines fanned out from Chicago, farmers began to acquire open prairie land in Illinois and then Iowa, putting the fertile, deep black soil into production. Commercial agriculture transformed this remarkable treeless environment. To settlers

accustomed to eastern woodlands, the thousands of square miles of tall grass were an awesome sight. Indian grass, Canada wild rye, and native big bluestem all grew higher than a person. Because eastern plows could not penetrate the densely tangled roots of prairie grass, the earliest settlers erected farms along the boundary separating the forest from the prairie. In 1837, however, John Deere patented a sharp-cutting steel plow that sliced through the sod without soil sticking to the blade. Cyrus McCormick refined a mechanical reaper that harvested fourteen times more wheat with the same amount of labor. By the 1850s McCormick was selling 1,000 reapers a year and could not keep up with demand, while Deere turned out 10,000 plows annually.

The new commercial farming fundamentally altered the Midwestern landscape and the environment. Native Americans had grown corn in the region for years, but never in such large fields as did later settlers who became farmers, whose surpluses were shipped east. Prairie farmers also introduced new crops that were not part of the earlier ecological system, notably wheat, along with fruits and vegetables. Native grasses were replaced by a small number of plants cultivated as commodities. Corn had the best yields, but it was primarily used to feed livestock. Because bread played a key role in the American and European diet, wheat became the major cash crop. Tame grasses replaced native grasses in pastures for making hay.

Western farmers altered the landscape by reducing the annual fires that had kept the prairie free from trees. In the absence of these fires, trees reappeared on land not in cultivation and, if undisturbed, eventually formed woodlots. The earlier unbroken landscape gave way to independent farms, each fenced off in a precise checkerboard pattern. It was an artificial ecosystem of animals, woodlots, and crops, whose large, uniform layout made western farms more efficient than the more-irregular farms in the East.

74 – 2 The Commercialization of Pearl River Agriculture

The Pearl River Delta of southern China was an agriculturally rich region. Formed of alluvial soils that had been captured from the silt flowing down from the major rivers, these sand flats, as the Chinese called them, were worked and reworked until they became very productive rice paddies that by the sixteenth century annually produced two crops of rice and one of vegetables and wheat. Peasant farmers produced on their own plots much of the food the family consumed. However, agriculture in the region could not be called subsistence agriculture, for many nonfood commercial crops were grown and exchanged in markets that dotted the countryside. Besides rice, peasant farmers in the Pearl River Delta grew sugarcane, hemp, cotton, and mulberries to feed silkworms, among other crops.

But while commerce and market exchanges were an important part of the rural economy until about 1550, the rural economy was not commercialized. As the population grew from 1400 to 1550, the gross volume of crops marketed and the number of rural markets both increased, but they did so at the rate of population growth. The proportion of agricultural land devoted to commercial crops in 1550 was about what it had been in 1400. But after 1550 the agricultural economy of the region became highly commercialized. Markets and marketing activity expanded at a rate faster than that of the population.

The most immediate stimulus for the commercialization of the economy was the new demand for Chinese goods—in particular silk and porcelains—from European traders who arrived in the South China Sea in the sixteenth century. For the Pearl River region, the critical trade flow was triangular, linked by the Portuguese. Either with their own funds or with silver that Japanese merchants supplied, Portuguese traders headed up the Pearl River to biannual fairs in the city of Guangzhou where they exchanged silver for export goods. These goods were loaded onto ships bound for the Japanese city of Nagasaki. Japanese merchants in Nagasaki paid in silver for the Chinese goods, launching another round of trade. Trade through Manila, in the Philippines, also brought silver into the South China economy. Chinese merchants from Guangdong and Fujian provinces sailed to Manila with their goods, which

the Spanish bought with silver. From there the silver flowed back to China as the Chinese commodities found their way to Europe. By 1600 this trade resulted in an annual inflow of perhaps 200,000 kilograms of silver into the coastal economies of south and southeast China.

In the Pearl River Delta, the silk industry developed on a base that was first created by the sand-flat fields and then a particular combination of fish ponds with fruit trees. In the fifteenth century peasant farmers in the Pearl River Delta began replacing some of their sand-flat rice fields with fishponds, probably in response to increased demand from Guangzhou. On the embankments, peasant farmers in the early Ming Dynasty (1368-1644) mostly planted fruit trees, giving rise to the fruit tree and fishpond combination. The carp fed on organic matter that either dropped or was thrown into the pond, while the mud scooped up from the pond fertilized the fruit trees and the rice fields and added height to the embankments and more protection for the fishponds.

The fruit tree and fishpond culture provided a ready-made base for expansion of the silk industry. As the demand for silk increased, peasant farmers replaced the fruit trees with mulberry trees (silkworms feed on mulberry leaves) and then began digging up even more rice paddies to expand this system. By 1581 in the Longshan area of Shunde county, for example, eighteen percent of the productive land was fishponds and, combined with the mulberry trees on the embankments, accounted for about thirty percent of the cultivated land area.

65 – 3 Pastoralism and Agriculture in Iran

Geographical constraints have had important consequences for the economy and society of Iran. Where rainfall is adequate, there are fertile valleys and grasslands suitable for grazing animals. However, since the natural vegetation tends to be sparse, it is difficult for such animals to remain in one place for any length of time. Thus nomadic pastoralism-keeping livestock (such as sheep and goats) by wandering from place to place-was one of the first and most persistent human economic activities to flourish in this area. This nomadic movement was often of the vertical variety, with people and animals moving from lowlands in wintertime to highlands in summer. The animals raised by the pastoralists provided not only food but also material for crafts such as the making of carpets, thick felt cloth, and tents. The pastoralists were typically organized into large tribal confederations capable of controlling the vast territories needed for maintaining their herds.

The tribes were a powerful social and political factor throughout Iranian history. The skills necessary for herding animals, hunting and chasing off predators, directing migrations, disciplining tribesmen, and protecting lands and animals from rivals could be easily adapted and directed toward military purposes as well. It was typically the tribes that produced the soldiers and rulers of the country and provided the power base for most of its dynasties. Once established, governments needed to cultivate the support of friendly tribal groups and tried to control hostile tribes by combat, deportation, or forcible settlement. At the beginning of the twentieth century, approximately one-fourth of the population were tribal peoples, and they were a potent force in Iranian affairs. With the advent of mechanized armies in the 1930s, however, there were systematic efforts to break the power of the tribes and to coerce the tribal population into a sedentary way of life. These efforts have been largely successful, and the tribes are no longer so significant a force in either the Iranian economy or society. Less than 5 percent of the population now consists of nomadic pastoralists.

The aridity of the Iranian plateau retarded its agricultural development in comparison to adjacent regions such as Mesopotamia, which had great rivers to draw upon for a supply of water. Eventually, at some uncertain date probably about 26 centuries ago, there was a technological

breakthrough that made it possible to farm crops outside the few oases, streams, and other places with sufficient rainfall for agriculture. This was the development of underground canals known as qanats. The qanat system took advantage of the natural slope (inclination) of the plateau basins. A well would be dug in the foothills to reach a water source, usually water from melting snow that had seeped underground. Then a sequence of wells and shafts connected by underground canals would be constructed to transport the water to an area suitable for cultivation where it could support the needs of one or more villages. The slope of the underground canals had to be controlled carefully to prevent erosion, and the interior surface of wells and shafts needed to be kept under constant maintenance to prevent them from collapsing.

Since the canals were underground, loss from evaporation was minimized. Gravity provided the means of moving the water, so no mechanical energy was required to operate the system. The numerous wells and shafts kept the length of the tricky underground canal short and facilitated repairs of each segment. Built up over the centuries, the system eventually became immense. It has been estimated that the total length of the qanat system today, counting wells, shafts, and canals, is in excess of 300,000 kilometers (almost the distance from Earth to the Moon!), which gives some idea of the tremendous investment in money and labor power it represents. Yet the type of agriculture that developed around the qanat system gave modest yields and required hard work from the peasant farmers, who received only a small share of the agricultural produce.

67 – 1 Crop Engineering

Our current ability to precisely engineer crop genomes was preceded by a long history of genetic manipulation in agriculture. Human impact and its accompanying effects began early in our history at many tropical and subtropical sites around the globe. Our ancestors were omnivores, consuming whatever plant or animal material they fortuitously encountered. Even then, humans had considerable effects on the environment, reducing and even driving to extinction populations of the animal species they hunted and expanding the distribution of plants by accidentally distributing seeds as they migrated.

Humans probably first realized that seeds could yield a stable food supply through agriculture when they observed plants arising from refuse or wasteland, perhaps fruit trees growing along forest and jungle paths from discarded or defecated seeds or else vegetables sprouting in garbage dumps at temporary settlements. A more organized approach to agriculture began about eight to ten thousand years ago coincidentally at a number of locations around the globe. The most diverse farming developed in the Near East, with legumes, cereals flax, sesame, and fruit trees. At about the same time, New World residents were growing beans, maize, squashes, and potatoes, and Asian farmers were beginning to cultivate rice.

These early domesticated crops foreshadowed the overwhelming changes contemporary agriculture has wrought in plants. Humans soon learned to separate varieties that could be grown as crops from wild types in order to prevent characteristics undesirable for cultivation from mingling with those selected for farming. Continued selection of crops with desirable characteristics increased the separation between feral (wild) and managed plants and accelerated the diminishing diversity and more limited variation found in today's crops.

The simplest way to select crops is to save seeds preferentially from plants with beneficial traits, and the first farmers selected for large seeds and fruit, increased seed production, lack of dormancy, faster germination, higher annual yield, and reduced seed scattering. The success of this early selection resulted in an accelerating impact of agriculture on crop diversity and feral plants. Crops quickly became commodities, moved and traded over a rapidly widening area, so that many plants

were distributed well beyond their previous ranges, and some throughout the globe.

Three phenomena have characterized the more recent impact of agriculture on Earth. The first was the increase in human population, which has doubled at shorter and shorter intervals over the last thousand years. The result was increased acreage under cultivation and a fundamental remodeling of the globe toward managed rather than wild ecosystems. By 1998 there were 3,410,523,800 acres of land under cultivation worldwide, an area larger than the United States. Entire ecosystems have disappeared, others remain but are threatened, and the sheer volume of people and area of farmland have been major forces of biological change.

The second event through which agriculture modified our planet was European colonization. Previously, migration and trade had moved crops between countries and continents, but the Europeans inaugurated an unprecedented dispersal of biological material worldwide. Maize, tomatoes, and potatoes were transported from the New World to the Old; wheat, rye, and barley were carried from the Old World to the New; and rice, soybeans, and alfalfa were moved from their Asian sources to every arable continent. Each of these and innumerable other introductions conveyed not only unique material but also assemblages of introduced plant pests and diseases that today cause the majority of pest-management problems around the world.

The third factor shaping the nature of agriculture and the environment alike is the increasing precision with which we have selected and bred crops. This acuity stemmed from many advances, but at its heart lies the work of two men—one, the English naturalist Charles Darwin, and the other, an Austrian monk, Gregor Mendel. The concepts of evolution and genetics were not their work alone, but both of them were decades ahead of their colleagues in synthesizing the companion concepts of natural selection and inheritance that are at the core of all contemporary biological science and that form the substrate upon which biotechnology grew.

71 – 3 The Productivity of Wet Rice Farming

The wet rice fields of Asia are the most productive of all preindustrial intensive agricultural systems. As the term "wet rice" suggests, the method involves flooding the fields for most of the growing season. In some areas the water comes entirely from the natural rainfall of the monsoons, but the water level often is controlled by artificial irrigation. Most wet rice is, therefore, grown on the floodplains of rivers. In parts of Southeast Asia, southern China, and the Himalayan fringes, wet rice is grown on terraced hillsides originally constructed centuries ago.

Perhaps the most remarkable feature of wet rice cultivation is the capability of a plot to yield a harvest year in and year out, with little or no need to be left uncultivated for extended periods in order to recover its fertility. Indeed, where the seasonal availability of water is adequate, the same plot may yield two and even three harvests in a single year, being under almost continuous cultivation. This capability has been attributed to four factors: the protection from erosion the water gives the soil, the high water table of the rivers' floodplains that reduces the leaching of nutrients from the soil, the replacement of soil nutrients by the silt carried in the rivers' flooding, and nitrogen fixation (the conversion of atmospheric nitrogen into forms usable by plants) from the blue-green algae that live in the floodwater. The chief benefit of wet rice—its ability to grow in the same field year after year—is then related to the practice of flooding the fields and allowing the water to remain throughout most of the maturation period.

Despite the advantages of their methods, wet rice farmers must solve problems relating to the management of water, because the proper supply of water at the proper time is so important for a good yield. It is more efficient if storage ponds or tanks, canals, ditches, and other improvements are held and managed communally, because less labor is required to construct and maintain them than if each family owned its own. Water may be stored in small lakes or ponds or captured and held in tanks until it is used to flood the fields. The water level must be controlled fairly precisely while the rice is growing. If it rises too high, the plants will be unable to photosynthesize and if it falls too low, the benefits of having standing water in the fields will be reduced. Unless the land is naturally flat, the plots, therefore, must be leveled to keep the height of the water constant in different areas of the field.

The water level must be controlled by low dikes, usually constructed of earth, which can be easily breached to drain the plot for the harvest.

Managing the supply of the water that is so critical in growing wet rice requires sophisticated organization, especially with respect to cooperative labor patterns and rights of access to both land and water. Each family may work its lands on its own, but who will construct and maintain public irrigation works? Who will decide when dikes, dams, ditches, and tanks need to be repaired, and how will the labor to repair them be coordinated? The water used on the fields of one family is not available to other families. How shall the consumption of water be coordinated? Who will get how much water and when? How, in short, will use rights to water be determined?

In one example of organizing the allocation of water to the fields, the Sinhalese of the dry zone of Sri Lanka draw water from a collectively owned rain-fed reservoir constructed of earth. All of the cultivated land of a village is in a single area, below the reservoir. The entire field is laid out into diked plots of equal size. Each plot receives a fixed amount of water from the communal reservoir, though the number of families working a single plot varies. Poor families must share water rights of access to one plot with other families, whereas richer families have access to several plots. Assigning each plot (rather than each family) the same proportion of irrigation water reduces the chance that a given plot will receive insufficient moisture.

01 – 1 Groundwater

Groundwater is the word used to describe water that saturates the ground, filling all the available spaces. By far the most abundant type of groundwater is meteoric water; this is the groundwater that circulates as part of the water cycle. Ordinary meteoric water is water that has soaked into the ground from the surface, from precipitation (rain and snow) and from lakes and streams. There it remains, sometimes for long periods, before emerging at the surface again. At first thought it seems incredible that there can be enough space in the “solid” ground underfoot to hold all this water.

The necessary space is there, however, in many forms. The commonest spaces are those among the particles—sand grains and tiny pebbles—of loose, unconsolidated sand and gravel. Beds of this material, out of sight beneath the soil, are common. They are found wherever fast rivers carrying loads of coarse sediment once flowed. For example, as the great ice sheets that covered North America during the last ice age steadily melted away, huge volumes of water flowed from them. The water was always laden with pebbles, gravel, and sand, known as glacial outwash, that was deposited as the flow slowed down.

The same thing happens to this day, though on a smaller scale, wherever a sediment-laden river or stream emerges from a mountain valley onto relatively flat land, dropping its load as the current slows: the water usually spreads out fanwise, depositing the sediment in the form of a smooth, fan-shaped slope. Sediments are also dropped where a river slows on entering a lake or the sea, the deposited sediments are on a lake floor or the seafloor at first, but will be located inland at some future date, when the sea level falls or the land rises, and such beds are sometimes thousands of meters thick.

In lowland country almost any spot on the ground may overlie what was once the bed of a river that has since become buried by soil; if they are now below the water’s upper surface (the water table),

the gravels and sands of the former riverbed, and its sandbars, will be saturated with groundwater.

So much for unconsolidated sediments. Consolidated (or cemented) sediments, too, contain millions of minute water-holding pores. This is because the gaps among the original grains are often not totally plugged with cementing chemicals; also, parts of the original grains may become dissolved by percolating groundwater, either while consolidation is taking place or at any time afterwards. The result is that sandstone, for example, can be as porous as the loose sand from which it was formed.

Thus a proportion of the total volume of any sediment, loose or cemented, consists of empty space. Most crystalline rocks are much more solid; a common exception is basalt, a form of solidified volcanic lava, which is sometimes full of tiny bubbles that make it very porous.

The proportion of empty space in a rock is known as its porosity. But note that porosity is not the same as permeability, which measures the ease with which water can flow through a material; this depends on the sizes of the individual cavities and the crevices linking them.

Much of the water in a sample of water-saturated sediment or rock will drain from it if the sample is put in a suitable dry place. But some will remain, clinging to all solid surfaces. It is held there by the force of surface tension without which water would drain instantly from any wet surface, leaving it totally dry. The total volume of water in the saturated sample must therefore be thought of as consisting of water that can, and water that cannot, drain away.

The relative amount of these two kinds of water varies greatly from one kind of rock or sediment to another, even though their porosities may be the same. What happens depends on pore size. If the pores are large, the water in them will exist as drops too heavy for surface tension to hold, and it will drain away; but if the pores are small enough, the water in them will exist as thin films, too light to overcome the force of surface tension holding them in place; then the water will be firmly held.

03 – 2 Depletion of the Ogallala Aquifer

The vast grasslands of the High Plains in the central United States were settled by farmers and ranchers in the 1880s. This region has a semiarid climate, and for 50 years after its settlement, it supported a low-intensity agricultural economy of cattle ranching and wheat farming. In the early twentieth century, however, it was discovered that much of the High Plains was underlain by a huge aquifer (a rock layer containing large quantities of groundwater). This aquifer was named the Ogallala Aquifer after the Ogallala Sioux Indians, who once inhabited the region.

The Ogallala Aquifer is a sandstone formation that underlies some 583,000 square kilometers of land extending from northwestern Texas to southern South Dakota. Water from rains and melting snows has been accumulating in the Ogallala for the past 30,000 years. Estimates indicate that the aquifer contains enough water to fill Lake Huron, but unfortunately, under the semiarid climatic conditions that presently exist in the region, rates of addition to the aquifer are minimal, amounting to about half a centimeter a year.

The first wells were drilled into the Ogallala during the drought years of the early 1930s. The ensuing rapid expansion of irrigation agriculture, especially from the 1950s onward, transformed the economy of the region. More than 100,000 wells now tap the Ogallala. Modern irrigation devices, each capable of spraying 4.5 million liters of water a day, have produced a landscape dominated by geometric patterns of circular green islands of crops. Ogallala water has enabled the High Plains region to supply significant amounts of the cotton, sorghum, wheat, and corn grown in the United States. In addition, 40 percent of American grain-fed beef cattle are fattened here.

This unprecedented development of a finite groundwater resource with an almost negligible natural recharge rate—that is, virtually no natural water source to replenish the water supply—has caused water tables in the region to fall drastically. In the 1930s, wells encountered plentiful water at a depth of about 15 meters; currently, they must be dug to depths of 45 to 60 meters or more. In places, the water table is declining at a rate of a meter a year, necessitating the periodic deepening of wells and the use of ever-more-powerful pumps. It is estimated that at current withdrawal rates, much of the aquifer will run dry within 40 years. The

situation is most critical in Texas, where the climate is driest, the greatest amount of water is being pumped, and the aquifer contains the least water. It is projected that the remaining Ogallala water will, by the year 2030, support only 35 to 40 percent of the irrigated acreage in Texas that is supported in 1980.

The reaction of farmers to the inevitable depletion of the Ogallala varies. Many have been attempting to conserve water by irrigating less frequently or by switching to crops that require less water. Others, however, have adopted the philosophy that it is best to use the water while it is still economically profitable to do so and to concentrate on high-value crops such as cotton. The incentive of the farmers who wish to conserve water is reduced by their knowledge that many of their neighbors are profiting by using great amounts of water, and in the process are drawing down the entire region's water supplies.

In the face of the upcoming water supply crisis, a number of grandiose schemes have been developed to transport vast quantities of water by canal or pipeline from the Mississippi, the Missouri, or the Arkansas rivers. Unfortunately, the cost of water obtained through any of these schemes would increase pumping costs at least tenfold, making the cost of irrigated agricultural products from the region uncompetitive on the national and international markets. Somewhat more promising have been recent experiments for releasing capillary water (water in the soil) above the water table by injecting compressed air into the ground. Even if this process proves successful, however, it would almost triple water costs. Genetic engineering also may provide a partial solution, as new strains of drought-resistant crops continue to be developed. Whatever the final answer to the water crisis may be, it is evident that within the High Plains, irrigation water will never again be the abundant, inexpensive resource it was during the agricultural boom years of the mid-twentieth century.

02 – 1 Desert Formation

The deserts, which already occupy approximately a fourth of the Earth's land surface, have in recent decades been increasing at an alarming pace. The expansion of desert-like conditions into areas where they did not previously exist is called desertification. It has been estimated that an additional one-fourth of the Earth's land surface is threatened by this process.

Desertification is accomplished primarily through the loss of stabilizing natural vegetation and the subsequent accelerated erosion of the soil by wind and water. In some cases the loose soil is blown completely away, leaving a stony surface. In other cases, the finer particles may be removed, while the sand-sized particles are accumulated to form mobile hills or ridges of sand.

Even in the areas that retain a soil cover, the reduction of vegetation typically results in the loss of the soil's ability to absorb substantial quantities of water. The impact of raindrops on the loose soil tends to transfer fine clay particles into the tiniest soil spaces, sealing them and producing a surface that allows very little water penetration. Water absorption is greatly reduced; consequently runoff is increased, resulting in accelerated erosion rates. The gradual drying of the soil caused by its diminished ability to absorb water results in the further loss of vegetation, so that a cycle of progressive surface deterioration is established.

In some regions, the increase in desert areas is occurring largely as the result of a trend toward drier climatic conditions. Continued gradual global warming has produced an increase in aridity for some areas over the past few thousand years. The process may be accelerated in subsequent decades if global warming resulting from air pollution seriously increases.

There is little doubt, however, that desertification in most areas results primarily from human activities rather than natural processes. The semiarid lands bordering the deserts exist in a delicate ecological balance and are limited in their potential to adjust to increased environmental pressures. Expanding populations are subjecting the land to increasing pressures to provide them with food and fuel. In wet periods, the land may be able to respond to these stresses. During the dry periods that are common phenomena along the desert margins, though, the pressure on the land is often far in excess of its diminished capacity, and desertification results.

Four specific activities have been identified as major contributors to the desertification processes:

overcultivation, overgrazing, firewood gathering, and overirrigation. The cultivation of crops has expanded into progressively drier regions as population densities have grown. These regions are especially likely to have periods of severe dryness, so that crop failures are common. Since the raising of most crops necessitates the prior removal of the natural vegetation, crop failures leave extensive tracts of land devoid of a plant cover and susceptible to wind and water erosion.

The raising of livestock is a major economic activity in semiarid lands, where grasses are generally the dominant type of natural vegetation. The consequences of an excessive number of livestock grazing in an area are the reduction of the vegetation cover and the trampling and pulverization of the soil. This is usually followed by the drying of the soil and accelerated erosion.

Firewood is the chief fuel used for cooking and heating in many countries. The increased pressures of expanding populations have led to the removal of woody plants so that many cities and towns are surrounded by large areas completely lacking in trees and shrubs. The increasing use of dried animal waste as a substitute fuel has also hurt the soil because this valuable soil conditioner and source of plant nutrients is no longer being returned to the land.

The final major human cause of desertification is soil salinization resulting from overirrigation. Excess water from irrigation sinks down into the water table. If no drainage system exists, the water table rises, bringing dissolved salts to the surface. The water evaporates and the salts are left behind, creating a white crustal layer that prevents air and water from reaching the underlying soil.

The extreme seriousness of desertification results from the vast areas of land and the tremendous numbers of people affected, as well as from the great difficulty of reversing or even slowing the process. Once the soil has been removed by erosion, only the passage of centuries or millennia will enable new soil to form. In areas where considerable soil still remains, though, a rigorously enforced program of land protection and cover-crop planting may make it possible to reverse the present deterioration of the surface.

01 – 3 Timberline Vegetation on Mountains

The transition from forest to treeless tundra on a mountain slope is often a dramatic one. Within a vertical distance of just a few tens of meters, trees disappear as a life-form and are replaced by low shrubs, herbs, and grasses. This rapid zone of transition is called the upper timberline or tree line. In many semiarid areas there is also a lower timberline where the forest passes into steppe or desert at its lower edge, usually because of a lack of moisture.

The upper timberline, like the snow line, is highest in the tropics and lowest in the Polar Regions. It ranges from sea level in the Polar Regions to 4,500 meters in the dry subtropics and 3,500-4,500 meters in the moist tropics. Timberline trees are normally evergreens, suggesting that these have some advantage over deciduous trees (those that lose their leaves) in the extreme environments of the upper timberline. There are some areas, however, where broadleaf deciduous trees form the timberline. Species of birch, for example, may occur at the timberline in parts of the Himalayas.

At the upper timberline the trees begin to become twisted and deformed. This is particularly true for trees in the middle and upper latitudes, which tend to attain greater heights on ridges, whereas in the tropics the trees reach their greater heights in the valleys. This is because middle- and upper- latitude timberlines are strongly influenced by the duration and depth of the snow cover. As the snow is deeper and lasts longer in the valleys, trees tend to attain greater heights on the ridges, even though they are more exposed to high-velocity winds and poor, thin soils there. In the tropics, the valleys appear to be more favorable because they are less prone to dry out, they have less frost, and they have deeper soils.

There is still no universally agreed-on explanation for why there should be such a dramatic cessation of tree growth at the upper timberline. Various environmental factors may play a role. Too much snow, for example, can smother trees, and avalanches and snow creep can damage or destroy them. Late-lying snow reduces the effective growing season to the point where seedlings cannot establish themselves. Wind velocity also increases with altitude and may cause serious stress for trees, as is made evident by the deformed shapes at high altitudes. Some scientists have proposed that the presence of increasing

levels of ultraviolet light with elevation may play a role, while browsing and grazing animals like the ibex may be another contributing factor. Probably the most important environmental factor is temperature, for if the growing season is too short and temperatures are too low, tree shoots and buds cannot mature sufficiently to survive the winter months.

Above the tree line there is a zone that is generally called alpine tundra. Immediately adjacent to the timberline, the tundra consists of a fairly complete cover of low-lying shrubs, herbs, and grasses, while higher up the number and diversity of species decrease until there is much bare ground with occasional mosses and lichens and some prostrate cushion plants. Some plants can even survive in favorable microhabitats above the snow line. The highest plants in the world occur at around 6,100 meters on Makalu in the Himalayas. At this great height, rocks, warmed by the sun, melt small snowdrifts.

The most striking characteristic of the plants of the alpine zone is their low growth form. This enables them to avoid the worst rigors of high winds and permits them to make use of the higher temperatures immediately adjacent to the ground surface. In an area where low temperatures are limiting to life, the importance of the additional heat near the surface is crucial. The low growth form can also permit the plants to take advantage of the insulation provided by a winter snow cover. In the equatorial mountains the low growth form is less prevalent.

03 – 3 The Long-Term Stability of Ecosystems

Plant communities assemble themselves flexibly, and their particular structure depends on the specific history of the area. Ecologists use the term “succession” to refer to the changes that happen in plant communities and ecosystems over time. The first community in a succession is called a pioneer community, while the long-lived community at the end of succession is called a climax community. Pioneer and successional plant communities are said to change over periods from 1 to 500 years. These changes—in plant numbers and the mix of species—are cumulative. Climax communities themselves change but over periods of time greater than about 500 years.

An ecologist who studies a pond today may well find it relatively unchanged in a year’s time. Individual fish may be replaced, but the number of fish will tend to be the same from one year to the next. We can say that the properties of an ecosystem are more stable than the individual organisms that compose the ecosystem.

At one time, ecologists believed that species diversity made ecosystems stable. They believed that the greater the diversity the more stable the ecosystem. Support for this idea came from the observation that long-lasting climax communities usually have more complex food webs and more species diversity than pioneer communities. Ecologists concluded that the apparent stability of climax ecosystems depended on their complexity. To take an extreme example, farmlands dominated by a single crop are so unstable that one year of bad weather or the invasion of a single pest can destroy the entire crop. In contrast, a complex climax community, such as a temperate forest, will tolerate considerable damage from weather to pests.

The question of ecosystem stability is complicated, however. The first problem is that ecologists do not all agree what “stability” means. Stability can be defined as simply lack of change. In that case, the climax community would be considered the most stable, since, by definition, it changes the least over time. Alternatively, stability can be defined as the speed with which an ecosystem returns to a particular form following a major disturbance, such as a fire. This kind of stability is also called resilience. In that case, climax communities would be the most fragile and the least stable, since they

can require hundreds of years to return to the climax state.

Even the kind of stability defined as simple lack of change is not always associated with maximum diversity. At least in temperate zones, maximum diversity is often found in mid-successional stages, not in the climax community. Once a redwood forest matures, for example, the kinds of species and the number of individuals growing on the forest floor are reduced. In general, diversity, by itself, does not ensure stability. Mathematical models of ecosystems likewise suggest that diversity does not guarantee ecosystem stability—just the opposite, in fact. A more complicated system is, in general, more likely than a simple system to break down. A fifteen-speed racing bicycle is more likely to break down than a child’s tricycle.

Ecologists are especially interested to know what factors contribute to the resilience of communities because climax communities all over the world are being severely damaged or destroyed by human activities. The destruction caused by the volcanic explosion of Mount St. Helens, in the northwestern United States, for example, pales in comparison to the destruction caused by humans. We need to know what aspects of a community are most important to the community’s resistance to destruction, as well as its recovery.

Many ecologists now think that the relative long-term stability of climax communities comes not from diversity but from the “patchiness” of the environment, an environment that varies from place to place supports more kinds of organisms than an environment that is uniform. A local population that goes extinct is quickly replaced by immigrants from an adjacent community. Even if the new population is of a different species, it can approximately fill the niche vacated by the extinct population and keep the food web intact.

07 – 1 The Geologic History of the Mediterranean

In 1970 geologists Kenneth J. Hsu and William B.F. Ryan were collecting research data while aboard the oceanographic research vessel *Glomar Challenger*. An objective of this particular cruise was to investigate the floor of the Mediterranean and to resolve questions about its geologic history. One question was related to evidence that the invertebrate fauna (animals without spines) of the Mediterranean had changed abruptly about 6 million years ago. Most of the older organisms were nearly wiped out, although a few hardy species survived. A few managed to migrate into the Atlantic. Somewhat later, the migrants returned, bringing new species with them. Why did the near extinction and migrations occur?

Another task for the *Glomar Challenger*'s scientists was to try to determine the origin of the domelike masses buried deep beneath the Mediterranean seafloor. These structures had been detected years earlier by echo-sounding instruments, but they had never been penetrated in the course of drilling. Were they salt domes such as are common along the United States Gulf Coast, and if so, why should there have been so much solid crystalline salt beneath the floor of the Mediterranean?

With question such as these clearly before them, the scientists aboard the *Glomar Challenger* processed to the Mediterranean to search for the answers. On August 23, 1970, they recovered a sample. The sample consisted of pebbles of hardened sediment that had once been soft, deep-sea mud, as well as granules of gypsum and fragments of volcanic rock. Not a single pebble was found that might have indicated that the pebbles came from the nearby continent. In the days following, samples of solid gypsum were repeatedly brought on deck as drilling operations penetrated the seafloor. Furthermore, the gypsum was found to possess peculiarities of composition and structure that suggested it had formed on desert flats. Sediment above and below the gypsum layer contained tiny marine fossils, indicating open-ocean conditions. As they drilled into the central and deepest part of the Mediterranean basin, the scientists took solid, shiny, crystalline salt from the core barrel. Interbedded with the salt were thin layers of what appeared to be windblown silt.

The time had come to formulate a hypothesis. The investigators theorized that about 20 million years

ago, the Mediterranean was a broad seaway linked to the Atlantic by two narrow straits. Crustal movements closed the straits, and the landlocked Mediterranean began to evaporate. Increasing salinity caused by the evaporation resulted in the extermination of scores of invertebrate species. Only a few organisms especially tolerant of very salty conditions remained. As evaporation continued, the remaining brine (salt water) became so dense that the calcium sulfate of the hard layer was precipitated. In the central deeper part of the basin, the last of the brine evaporated to precipitate more soluble sodium chloride (salt). Later, under the weight of overlying sediments, this salt flowed plastically upward to form salt domes. Before this happened, however, the Mediterranean was a vast desert 3,000 meters deep. Then, about 5.5 million years ago came the deluge. As a result of crustal adjustments and faulting, the Strait of Gibraltar, where the Mediterranean now connects to the Atlantic, opened, and water cascaded spectacularly back into the Mediterranean. Turbulent waters tore into the hardened salt flats, broke them up, and ground them into the pebbles observed in the first sample taken by the *Challenger*. As the basin was refilled, normal marine organisms returned. Soon layer of oceanic ooze began to accumulate above the old hard layer. The salt and gypsum, the faunal changes, and the unusual gravel provided abundant evidence that the Mediterranean was once a desert.

10 – 2 Variations in the Climate

One of the most difficult aspects of deciding whether current climatic events reveal evidence of the impact of human activities is that it is hard to get a measure of what constitutes the natural variability of the climate. We know that over the past millennia the climate has undergone major changes without any significant human intervention. We also know that the global climate system is immensely complicated and that everything is in some way connected, and so the system is capable of fluctuating in unexpected ways. We need therefore to know how much the climate can vary of its own accord in order to interpret with confidence the extent to which recent changes are natural as opposed to being the result of human activities.

Instrumental records do not go back far enough to provide us with reliable measurements of global climatic variability on timescales longer than a century. What we do know is that as we include longer time intervals, the record shows increasing evidence of slow swings in climate between different regimes. To build up a better picture of fluctuations appreciably further back in time requires us to use proxy records.

Over long period of time, substances whose physical and chemical properties change with the ambient climate at the time can be deposited in a systematic way to provide a continuous record of changes in those properties overtime, sometimes for hundreds or thousands of years. Generally, the layering occurs on an annual basis, hence the observed changes in the records can be dated. Information on temperature, rainfall, and other aspects of the climate that can be inferred from the systematic changes in properties is usually referred to as proxy data. Proxy temperature records have been reconstructed from ice core drilled out of the central Greenland ice cap, calcite shells embedded in layered lake sediments in Western Europe, ocean floor sediment cores from the tropical Atlantic Ocean, ice cores from Peruvian glaciers, and ice cores from eastern Antarctica. While these records provide broadly consistent indications that temperature variations can occur on a global scale, there are nonetheless some intriguing differences, which suggest that the pattern of temperature variations in regional climates can also differ significantly from each other.

What the proxy records make abundantly clear is that there have been significant natural changes in the climate over timescales longer than a few thousand years. Equally striking, however, is the relative stability of the climate in the past 10,000 years (the Holocene period).

To the extent that the coverage of the global climate from these records can provide a measure of its true variability, it should at least indicate how all the natural causes of climate change have combined. These include the chaotic fluctuations of the atmosphere, the slower but equally erratic behavior of the oceans, changes in the land surfaces, and the extent of ice and snow. Also included will be any variations that have arisen from volcanic activity, solar activity, and, possibly, human activities.

One way to estimate how all the various processes leading to climate variability will combine is by using computer models of the global climate. They can do only so much to represent the full complexity of the global climate and hence may give only limited information about natural variability. Studies suggest that to date the variability in computer simulations is considerably smaller than in data obtained from the proxy records.

In addition to the internal variability of the global climate system itself, there is the added factor of external influences, such as volcanoes and solar activity. There is a growing body of opinion that both these physical variations have a measurable impact on the climate. Thus we need to be able to include these in our deliberations. Some current analyses conclude that volcanoes and solar activity explain quite a considerable amount of the observed variability in the period from the seventeenth to the early twentieth centuries, but that they cannot be invoked to explain the rapid warming in recent decades.

12 – 3 Water in the Desert

Rainfall is not completely absent in desert areas, but it is highly variable. An annual rainfall of four inches is often used to define the limits of a desert. The impact of rainfall upon the surface water and groundwater resources of the desert is greatly influenced by landforms. Flats and depressions where water can collect are common features, but they make up only a small part of the landscape.

Arid lands, surprisingly, contain some of the world's largest river systems, such as the Murray-Darling in Australia, the Rio Grande in North America, the Indus in Asia, and the Nile in Africa. These rivers and river systems are known as "exogenous" because their sources lie outside the arid zone. They are vital for sustaining life in some of the driest parts of the world. For centuries, the annual floods of the Nile, Tigris, and Euphrates, for example, have brought fertile silts and water to the inhabitants of their lower valleys. Today, river discharges are increasingly controlled by human intervention, creating a need for international river-basin agreements. The filling of the Ataturk and other dams in Turkey has drastically reduced flows in the Euphrates, with potentially serious consequences for Syria and Iraq.

The flow of exogenous rivers varies with the season. The desert sections of long rivers respond several months after rain has fallen outside the desert, so that peak flows may be in the dry season. This is useful for irrigation, but the high temperatures, low humidity, and different day lengths of the dry season, compared to the normal growing season, can present difficulties with some crops.

Regularly flowing rivers and streams that originate within arid lands are known as "endogenous". These are generally fed by groundwater springs, and many issue from limestone massifs, such as the Atlas Mountains in Morocco. Basaltic rocks also support springs, notably at the Jabal Al-Arab on the Jordan-Syria border. Endogenous rivers often do not reach the sea but drain into inland basins, where the water evaporates or is lost in the ground. Most desert streambeds are normally dry, but they occasionally receive large flows of water and sediment.

Deserts contain large amounts of groundwater when compared to the amounts they hold in surface stores such as lakes and rivers. But only a small fraction of groundwater enters the

hydrological cycle—feeding the flows of streams, maintaining lake levels, and being recharged (or refilled) through surface flows and rainwater. In recent years, groundwater has become an increasingly important source of freshwater for desert dwellers. The United Nations Environment Programme and the World Bank have funded attempts to survey the groundwater resources of arid lands and to develop appropriate extraction techniques. Such programs are much needed because in many arid lands there is only a vague idea of the extent of groundwater resources. It is known, however, that the distribution of groundwater is uneven, and that much of it lies at great depths.

Groundwater is stored in the pore spaces and joints of rocks and unconsolidated (unsolidified) sediments or in the openings widened through fractures and weathering. The water-saturated rock or sediment is known as an "aquifer". Because they are porous, sedimentary rocks, such as sandstones and conglomerates, are important potential sources of groundwater. Large quantities of water may also be stored in limestones when joints and cracks have been enlarged to form cavities. Most limestone and sandstone aquifers are deep and extensive but may contain groundwater that are not being recharged. Most shallow aquifers in sand and gravel deposits produce lower yields, but they can be rapidly recharged. Some deep aquifers are known as "fossil waters". The term "fossil" describes water that has been present for several thousand years. These aquifers became saturated more than 10,000 years ago and are no longer being recharged.

Water does not remain immobile in an aquifer but can seep out at springs or leak into other aquifers. The rate of movement may be very slow: in the Indus plain, the movement of saline (salty) groundwater has still not reached equilibrium after 70 years of being tapped. The mineral content of groundwater normally increases with the depth, but even quite shallow aquifers can be highly saline.

15 – 3 Glacier Formation

Glaciers are slowly moving masses of ice that have accumulated on land in areas where more snowfalls during a year than melts. Snow falls as hexagonal crystals, but once on the ground, snow is soon transformed into a compacted mass of smaller, rounded grains. As the air space around them is lessened by compaction and melting, the grains become denser. With further melting, refreezing, and increased weight from newer snowfall above, the snow reaches a granular recrystallized stage intermediate between flakes and ice known as firn. With additional time, pressure, and refrozen meltwater from above, the small firn granules become larger, interlocked crystals of blue glacial ice. When the ice is thick enough, usually over 30 meters, the weight of the snow and firn will cause the ice crystals toward the bottom to become plastic and to flow outward or downward from the area of snow accumulation.

Glaciers are open systems, with snow as the system's input and meltwater as the system's main output. The glacial system is governed by two basic climatic variables: precipitation and temperature. For a glacier to grow or maintain its mass, there must be sufficient snowfall to match or exceed the annual loss through melting, evaporation, and calving, which occurs when the glacier loses solid chunks as icebergs to the sea or to large lakes. If summer temperatures are high for too long, then all the snowfall from the previous winter will melt. Surplus snowfall is essential for a glacier to develop. A surplus allows snow to accumulate and for the pressure of snow accumulated over the years to transform buried snow into glacial ice with a depth great enough for the ice to flow. Glaciers are sometimes classified by temperature as faster-flowing temperate glaciers or as slower-flowing polar glaciers.

Glaciers are part of Earth's hydrologic cycle and are second only to the oceans in the total amount of water contained. About 2 percent of Earth's water is currently frozen as ice. Two percent may be a deceiving figure, however, since over 80 percent of the world's freshwater is locked up as ice in glaciers, with the majority of it in Antarctica. The total amount of ice is even more awesome if we estimate the water released upon the hypothetical melting of the world's glaciers. Sea level would rise about 60 meters. This would change the geography of the planet considerably. In contrast, should another ice age occur, sea level

would drop drastically. During the last ice age, sea level dropped about 120 meters.

When snow falls on high mountains or in Polar regions, it may become part of the glacial system. Unlike rain, which returns rapidly to the sea or atmosphere, the snow that becomes part of a glacier is involved in a much more slowly cycling system. Here water may be stored in ice form for hundreds or even hundreds of thousands of years before being released again into the liquid water system as meltwater. In the meantime, however, this ice is not static. Glaciers move slowly across the land with tremendous energy, carving into even the hardest rock formations and thereby reshaping the landscape as they engulf, push, drag, and finally deposit rock debris in places far from its original location. As a result, glaciers create a great variety of landforms that remain long after the surface is released from its icy covering.

Throughout most of Earth's history, glaciers did not exist, but at the present time about 10 percent of Earth's land surface is covered by glaciers. Present-day glaciers are found in Antarctica, in Greenland, and at high elevations on all the continents except Australia. In the recent past, from about 2.4 million to about 10,000 years ago, nearly a third of Earth's land area was periodically covered by ice thousands of meters thick. In the much more distant past, other ice ages have occurred.

18 – 3 Lightning

Lightning is a brilliant flash of light produced by an electrical discharge from a storm cloud. The electrical discharge takes place when the attractive tension between a region of negatively charged particles and a region of positively charged particles becomes so great that the charged particles suddenly rush together. The coming together of the oppositely charged particles neutralizes the electrical tension and releases a tremendous amount of energy, which we see as lightning. The separation of positively and negatively charged particles takes place during the development of the storm cloud.

The separation of charged particles that forms in a storm cloud has a sandwich-like structure. Concentrations of positively charged particles develop at the top and bottom of the cloud, but the middle region becomes negatively charged. Recent measurements made in the field together with laboratory simulations offer a promising explanation of how this structure of charged particles forms. What happens is that small (millimeter-to-centimeter-size) pellets of ice form in the cold upper regions of the cloud. When these ice pellets fall, some of them strike much smaller ice crystals in the center of the cloud. The temperature at the center of the cloud is about -15°C or lower. At such temperatures, the collision between the ice pellets and the ice crystals causes electrical charges to shift so that the ice pellets acquire a negative charge and the ice crystals become positively charged. Then updraft wind currents carry the light, positively charged ice crystals up to the top of the cloud. The heavier negatively charged ice pellets are left to concentrate in the center. This process explains why the top of the cloud becomes positively charged, while the center becomes negatively charged. The negatively charged region is large: several hundred meters thick and several kilometers in diameter. Below this large, cold, negatively charged region, the cloud is warmer than -15°C , and at these temperatures, collisions between ice crystals and falling ice pellets produce positively charged ice pellets that then populate a small region at the base of the cloud.

Most lightning takes place within a cloud when the charge separation within the cloud collapses. However, as the storm cloud develops, the ground beneath the cloud becomes positively charged and lightning can take place in the form of an electrical discharge between the negative charge of the

cloud and the positively charged ground. Lightning that strikes the ground is the most likely to be destructive, so even though it represents only 20 percent of all lightning, it has received a lot of scientific attention.

Using high-speed photography, scientists have determined that there are two steps to the occurrence of lightning from a cloud to the ground. First, a channel, or path, is formed that connects the cloud and the ground. Then a strong current of electrons follows that path from the cloud to the ground, and it is that current that illuminates the channel as the lightning we see.

The formation of the channel is initiated when electrons surge from the cloud base toward the ground. When a stream of these negatively charged electrons comes within 100 meters of the ground it is met by a stream of positively charged particles that comes up from the ground. When the negatively and positively charged streams meet, a complete channel connecting the cloud and the ground is formed. The channel is only a few centimeters in diameter, but that is wide enough for electrons to follow the channel to the ground in the visible form of a flash of lightning. The stream of positive particles that meets the surge of electrons from the cloud often arises from a tall pointed structure such as a metal flagpole or a tower. That is why the subsequent lightning that follows the completed channel often strikes a tall structure.

Once a channel has been formed, it is usually used by several lightning discharges, each of them consisting of a stream of electrons from the cloud meeting a stream of positive particles along the established path. Sometimes, however, a stream of electrons following an established channel is met by a positive stream making a new path up from the ground. The result is a forked lightning that strikes the ground in two places.

19 – 3 Discovering the Ice Ages

In the middle of the nineteenth century, Louis Agassiz, one of the first scientists to study glaciers, immigrated to the United States from Switzerland and became a professor at Harvard University, where he continued his studies in geology and other sciences. For his research, Agassiz visited many places in the northern parts of Europe and North America, from the mountains of Scandinavia and New England to the rolling hills of the American Midwest. In all these diverse regions, Agassiz saw signs of glacial erosion and sedimentation. In flat plains country, he saw moraines (accumulations of earth and loose rock that form at the edges of glaciers) that reminded him of the terminal moraines found at the end of valley glaciers in the Alps. The heterogeneous material of the drift (sand, clay, and rocks deposited there) convinced him of its glacial origin.

The areas covered by this material were so vast that the ice that deposited it must have been a continental glacier larger than Greenland or Antarctica. Eventually, Agassiz and others convinced geologists and the general public that a great continental glaciation had extended the polar ice caps far into regions that now enjoy temperate climates. For the first time, people began to talk about ice ages. It was also apparent that the glaciation occurred in the relatively recent past because the drift was soft, like freshly deposited sediment. We now know the age of the glaciation accurately from radiometric dating of the carbon-14 in logs buried in the drift. The drift of the last glaciation was deposited during one of the most recent epochs of geologic time, the Pleistocene, which lasted from 1.8 million to 10,000 years ago. Along the east coast of the United States, the southernmost advance of this ice is recorded by the enormous sand and drift deposits of the terminal moraines that form Long Island and Cape Cod.

It soon became clear that there were multiple glacial ages during the Pleistocene, with warmer interglacial intervals between them. As geologists mapped glacial deposits in the late nineteenth century, they became aware that there were several layers of drift, the lower ones corresponding to earlier ice ages. Between the older layers of glacial material were well-developed soils containing fossils of warm-climate plants. These soils were evidence that the glaciers retreated as the climate warmed. By the early part

of the twentieth century, scientists believed that four distinct glaciations had affected North America and Europe during the Pleistocene epoch.

This idea was modified in the late twentieth century, when geologists and oceanographers examining oceanic sediment found fossil evidence of warming and cooling of the oceans. Ocean sediments presented a much more complete geologic record of the Pleistocene than continental glacial deposits did. The fossils buried in Pleistocene and earlier ocean sediments were of foraminifera—small, single-celled marine organisms that secrete shells of calcium carbonate, or calcite. These shells differ in their proportion of ordinary oxygen (oxygen-16) and the heavy oxygen isotope (oxygen-18). The ratio of oxygen-16 to oxygen-18 found in the calcite of a foraminifer's shell depends on the temperature of the water in which the organism lived. Different ratios in the shells preserved in various layers of sediment reveal the temperature changes in the oceans during the Pleistocene epoch.

Isotopic analysis of shells allowed geologists to measure another glacial effect. They could trace the growth and shrinkage of continental glaciers, even in parts of the ocean where there may have been no great change in temperature—around the equator, for example. The oxygen isotope ratio of the ocean changes as a great deal of water is withdrawn from it by evaporation and is precipitated as snow to form glacial ice. During glaciations, the lighter oxygen-16 has a greater tendency to evaporate from the ocean surface than the heavier oxygen-18 does. Thus, more of the heavy isotope is left behind in the ocean and absorbed by marine organisms. From this analysis of marine sediments, geologists have learned that there were many shorter, more regular cycles of glaciation and deglaciation than geologists had recognized from the glacial drift of the continents alone.

22 – 3 The Allende Meteorite

Sometime after midnight on February 8, 1969, a large, bright meteor entered Earth's atmosphere and broke into thousands of pieces, plummeted to the ground, and scattered over an area 50 miles long and 10 miles wide in the state of Chihuahua in Mexico. The first meteorite from this fall was found in the village of Pueblito de Allende. Altogether, roughly two tons of meteorite fragments were recovered, all of which bear the name Allende for the location of the first discovery.

Individual specimens of Allende are covered with a black, glassy crust that formed when their exteriors melted as they were slowed by Earth's atmosphere. When broken open, Allende stones are revealed to contain an assortment of small, distinctive objects, spherical or irregular in shape and embedded in a dark gray matrix (binding material), which were once constituents of the solar nebula — the interstellar cloud of gas and dust out of which our solar system was formed.

The Allende meteorite is classified as a chondrite. Chondrites take their name from the Greek word *chondros* — meaning “seed” — an allusion to their appearance as rocks containing tiny seeds. These seeds are actually chondrules: millimeter-sized melted droplets of silicate material that were cooled into spheres of glass and crystal. A few chondrules contain grains that survived the melting event, so these enigmatic chondrules must have formed when compact masses of nebular dust were fused at high temperatures — approaching 1,700 degrees Celsius — and then cooled before these surviving grains could melt. Study of the textures of chondrules confirms that they cooled rather quickly, in times measured in minutes or hours, so the heating events that formed them must have been localized. It seems very unlikely that large portions of the nebula were heated to such extreme temperatures, and huge nebula areas could not possibly have lost heat so fast. Chondrules must have been melted in small pockets of the nebula that were able to lose heat rapidly. The origin of these peculiar glassy spheres remains an enigma.

Equally perplexing constituents of Allende are the refractory inclusions: irregular white masses that tend to be larger than chondrules. They are composed of minerals uncommon on Earth, all rich in calcium, aluminum, and titanium, the most refractory (resistant to melting) of the major elements in the nebula. The same minerals that

occur in refractory inclusions are believed to be the earliest-formed substances to have condensed out of the solar nebula. However, studies of the textures of inclusions reveal that the order in which the minerals appeared in the inclusions varies from inclusion to inclusion, and often does not match the theoretical condensation sequence for those metals.

Chondrules and inclusions in Allende are held together by the chondrite matrix, a mixture of fine-grained, mostly silicate minerals that also includes grains of iron metal and iron sulfide. At one time it was thought that these matrix grains might be pristine nebular dust, the sort of stuff from which chondrules and inclusions were made. However, detailed studies of the chondrite matrix suggest that much of it, too, has been formed by condensation or melting in the nebula, although minute amounts of surviving interstellar dust are mixed with the processed materials.

All these diverse constituents are aggregated together to form chondritic meteorites, like Allende, that have chemical compositions much like that of the Sun. To compare the compositions of a meteorite and the Sun, it is necessary that we use ratios of elements rather than simply the abundances of atoms. After all, the Sun has many more atoms of any element, say iron, than does a meteorite specimen, but the ratios of iron to silicon in the two kinds of matter might be comparable. The compositional similarity is striking. The major difference is that Allende is depleted in the most volatile elements, like hydrogen, carbon, oxygen, nitrogen, and the noble gases, relative to the Sun. These are the elements that tend to form gases even at very low temperatures. We might think of chondrites as samples of distilled Sun, a sort of solar sludge from which only gases have been removed. Since practically all the solar system's mass resides in the Sun, this similarity in chemistry means that chondrites have average solar system composition, except for the most volatile elements; they are truly lumps of nebular matter, probably similar in composition to the matter from which planets were assembled.

23 – 1 Urban Climates

The city is an extraordinary processor of mass and energy and has its own metabolism. A daily input of water, food, and energy of various kinds is matched by an output of sewage, solid waste, air pollutants, energy, and materials that have been transformed in some way. The quantities involved are enormous. Many aspects of this energy use affect the atmosphere of a city, particularly in the production of heat.

In winter the heat produced by a city can equal or surpass the amount of heat available from the Sun. All the heat that warms a building eventually transfers to the surrounding air, a process that is quickest where houses are poorly insulated. But an automobile produces enough heat to warm an average house in winter, and if a house were perfectly insulated, one adult could also produce more than enough heat to warm it. Therefore, even without any industrial production of heat, an urban area tends to be warmer than the countryside that surrounds it.

The burning of fuel, such as by cars, is not the only source of this increased heat. Two other factors contribute to the higher overall temperature in cities. The first is the heat capacity of the materials that constitute the city, which is typically dominated by concrete and asphalt. During the day, heat from the Sun can be conducted into these materials and stored—to be released at night. But in the countryside materials have a significantly lower heat capacity because a vegetative blanket prevents heat from easily flowing into and out of the ground. The second factor is that radiant heat coming into the city from the Sun is trapped in two ways: (1) by a continuing series of reflection among the numerous vertical surfaces that buildings present and (2) by the dust dome, the cloudlike layer of polluted air that most cities produce. Shortwave radiation from the Sun passes through the pollution dome more easily than outgoing longwave radiation does; the latter is absorbed by the gaseous pollutants of the dome and reradiated back to the urban surface.

Cities, then, are warmer than the surrounding rural areas, and together they produce a phenomenon known as the urban heat island. Heat islands develop best under particular conditions associated with light winds, but they can form almost any time. The precise configuration of a heat island depends on several factors. For example, the wind can

make a heat island stretch in the direction it blows. When a heat island is well developed, variations can be extreme; in winter, busy streets in cities can be 1.7°C warmer than the side streets. Areas near traffic lights can be similarly warmer than the areas between them because of the effect of cars standing in traffic instead of moving. The maximum differences in temperature between neighboring urban and rural environments is called the heat-island intensity for that region. In general, the larger the city, the greater its heat-island intensity. The actual level of intensity depends on such factors as the physical layout, population density, and productive activities of a metropolis.

The surface-atmosphere relationships inside metropolitan areas produce a number of climatic peculiarities. For one thing, the presence or absence of moisture is affected by the special qualities of the urban surface. With much of the built-up landscape impenetrable by water, even gentle rain runs off almost immediately from rooftops, streets, and parking lots. Thus, city surfaces, as well as the air above them, tend to be drier between episodes of rain; with little water available for the cooling process of evaporation, relative humidity is usually lower. Wind movements are also modified in cities because buildings increase the friction on air flowing around them. This friction tends to slow the speed of winds, making them far less efficient at dispersing pollutants. On the other hand, air turbulence increases because of the effect of skyscrapers on airflow. Rainfall is also increased in cities. The cause appears to be in part greater turbulence in the urban atmosphere as hot air rises from the built-up surface.

24 – 1 Lake Water

Where does the water in a lake come from, and how does water leave it? Water enters a lake from inflowing rivers, from underwater seeps and springs, from overland flow off the surrounding land, and from rain falling directly on the lake surface. Water leaves a lake via outflowing rivers, by soaking into the bed of the lake, and by evaporation. So much is obvious.

The questions become more complicated when actual volumes of water are considered: how much water enters and leaves by each route? Discovering the inputs and outputs of rivers is a matter of measuring the discharges of every inflowing and outflowing stream and river. Then exchanges with the atmosphere are calculated by finding the difference between the gains from rain, as measured (rather roughly) by rain gauges, and the losses by evaporation, measured with models that correct for the other sources of water loss. For the majority of lakes, certainly those surrounded by forests, input from overland flow is too small to have a noticeable effect. Changes in lake level not explained by river flows plus exchanges with the atmosphere must be due to the net difference between what seeps into the lake from the groundwater and what leaks into the groundwater. Note the word "net": measuring the actual amounts of groundwater seepage into the lake and out of the lake is a much more complicated matter than merely inferring their difference.

Once all this information has been gathered, it becomes possible to judge whether a lake's flow is mainly due to its surface inputs and outputs or to its underground inputs and outputs. If the former are greater, the lake is a surface-water-dominated lake; if the latter, it is a seepage-dominated lake. Occasionally, common sense tells you which of these two possibilities applies. For example, a pond in hilly country that maintains a steady water level all through a dry summer in spite of having no streams flowing into it must obviously be seepage dominated. Conversely, a pond with a stream flowing in one end and out the other, which dries up when the stream dries up, is clearly surface water dominated.

By whatever means, a lake is constantly gaining water and losing water: its water does not just sit there, or, anyway, not for long. This raises the matter of a lake's residence time. The residence time is the average length of time that any

particular molecule of water remains in the lake, and it is calculated by dividing the volume of water in the lake by the rate at which water leaves the lake. The residence time is an average; the time spent in the lake by a given molecule (if we could follow its fate) would depend on the route it took: it might flow through as part of the fastest, most direct current, or it might circle in a backwater for an indefinitely long time.

Residence times vary enormously. They range from a few days for small lakes up to several hundred years for large ones; Lake Tahoe, in California, has a residence time of 700 years. The residence times for the Great Lakes of North America, namely, Lakes Superior, Michigan, Huron, Erie, and Ontario, are, respectively, 190, 100, 22, 2.5, and 6 years. Lake Erie's is the lowest: although its area is larger than Lake Ontario's, its volume is less than one-third as great because it is so shallow — less than 20 meters on average.

A given lake's residence time is by no means a fixed quantity. It depends on the rate at which water enters the lake, and that depends on the rainfall and the evaporation rate. Climatic change (the result of global warming?) is dramatically affecting the residence times of some lakes in northwestern Ontario, Canada. In the period 1970 to 1986, rainfall in the area decreased from 1,000 millimeters to 650 millimeters per annum, while above-average temperatures speeded up the evapotranspiration rate (the rate at which water is lost to the atmosphere through evaporation and the processes of plant life). The result has been that the residence time of one of the lakes increased from 5 to 18 years during the study period. The slowing down of water renewal leads to a chain of further consequences; it causes dissolved chemicals to become increasingly concentrated, and this, in turn, has a marked effect on all living things in the lake.

27 – 2 The Formation of Volcanic Islands

Earth's surface is not made up of a single sheet of rock that forms a crust but rather a number of "tectonic plates" that fit closely, like the pieces of a giant jigsaw puzzle. Some plates carry islands or continents; others form the seafloor. All are slowly moving because the plates float on a denser semi-liquid mantle, the layer between the crust and Earth's core. The plates have edges that are spreading ridges (where two plates are moving apart and new seafloor is being created), subduction zones (where two plates collide and one plunges beneath the other), or transform faults (where two plates neither converge nor diverge but merely move past one another). It is at the boundaries between plates that most of Earth's volcanism and earthquake activity occur.

Generally speaking, the interiors of plates are geologically uneventful. However, there are exceptions. A glance at a map of the Pacific Ocean reveals that there are many islands far out at sea that are actually volcanoes - many no longer active, some overgrown with coral - that originated from activity at points in the interior of the Pacific Plate that forms the Pacific seafloor.

How can volcanic activity occur so far from a plate boundary? The Hawaiian Islands provide a very instructive answer. Like many other island groups, they form a chain. The Hawaiian Islands Chain extends northwest from the island of Hawaii. In the 1840s American geologist James Daly observed that the different Hawaii islands seem to share a similar geologic evolution but are progressively more eroded, and therefore probably older, toward the northwest. Then in 1963, in the early days of the development of the theory of plate tectonics, Canadian geophysicist Tuzo Wilson realized that this age progression could result if the islands were formed on a surface plate moving over a fixed volcanic source in the interior. Wilson suggested that the long chain of volcanoes stretching northwest from Hawaii is simply the surface expression of a long-lived volcanic source located beneath the tectonic plate in the mantle. Today's most northwest island would have been the first to form. Then as the plate moved slowly northwest, new volcanic islands would have formed as the plate moved over the volcanic source. The most recent island, Hawaii, would be at the end of the chain and is now over the volcanic source.

Although this idea was not immediately accepted, the dating of lavas in the Hawaii (and other) chains showed that their ages increase away from the presently active volcano, just as Daly had suggested. Wilson's analysis of these data is now a central part of plate tectonics. Most volcanoes that occur in the interiors of plates are believed to be produced by mantle plumes, columns of molten rock that rise from deep within the mantle. A volcano remains an active "hot spot" as long as it is over the plume. The plumes apparently originate at great depths, perhaps as deep as the boundary between the core and the mantle, and many have been active for a very long time. The oldest volcanoes in the Hawaii hot-spot trail have ages close to 80 million years. Other islands, including Tahiti and Easter Islands in the Pacific, Reunion and Mauritius in the India Ocean, and indeed most of the large islands in the world's oceans, owe their existence to mantle plumes.

The oceanic volcanic islands and their hot-spot trails are thus especially useful for geologists because they record the past locations of the plate over a fixed source. They therefore permit the reconstruction of the process of seafloor spreading, and consequently of the geography of continents and of ocean basins in the past. For example, given the current position of the Pacific Plate, Hawaii is above the Pacific Ocean hot spot. So the position of The Pacific Plate 50 million years ago can be determined by moving it such that a 50-million-year-old volcano in the hot-spot trail sits at the location of Hawaii today. However because the ocean basins really are short-lived features on geologic times scale, reconstruction of the world's geography by backtracking along the hot-spot trail works only for the last 5 percent or so of geologic time.

28 – 1 Groundwater

Most of the world's potable water — freshwater suitable for drinking — is accounted for by groundwater, which is stored in the pores and fractures in rocks. There is more than 50 times as much freshwater stored underground than in all the freshwater rivers and lakes at the surface. Nearly 50 percent of all groundwater is stored in the upper 1,000 meters of Earth. At greater depths within Earth, the pressure of the overlying rock causes pores and cracks to close, reducing the space that pore water can occupy, and almost complete closure occurs at a depth of about 10 kilometers. The greatest water storage, therefore, lies near the surface.

Aquifers, Porosity and Permeability

Groundwater is stored in a variety of rock types. A groundwater reservoir from which water can be extracted is called an aquifer. We can effectively think of an aquifer as a deposit of water. Extraction of water depends on two properties of the aquifer: porosity and permeability. Between sediment grains are spaces that can be filled with water. This pore space is known as porosity and is expressed as a percentage of the total rock volume. Porosity is important for water-storage capacity, but for water to flow through rocks, the pore spaces must be connected. The ability of water, or other fluids, to flow through the interconnected pore spaces in rocks is termed permeability. In the intergranular spaces of rocks, however, fluid must flow around and between grains in a tortuous path; this winding path causes a resistance to flow. The rate at which the flowing water overcomes this resistance is related to the permeability of rock.

Sediment sorting and compaction influence permeability and porosity. The more poorly sorted or the more tightly compacted a sediment is, the lower its porosity and permeability. Sedimentary rocks — the most common rock type near the surface — are also the most common reservoirs for water because they contain the most space that can be filled with water. Sandstones generally make good aquifers, while finer-grained mudstones are typically impermeable. Impermeable rocks are referred to as aquicludes. Igneous and metamorphic rocks are more compact, commonly crystalline, and rarely contain spaces between grains. However, even igneous and metamorphic rocks may act as groundwater reservoirs if

extensive fracturing occurs in such rocks and if the fracture system is interconnected.

The Water Table

The water table is the underground boundary below which all the cracks and pores are filled with water. In some cases, the water table reaches Earth's surface, where it is expressed as rivers, lakes and marshes. Typically, though, the water table may be tens or hundreds of meters below the surface. The water table is not flat but usually follows the contours of the topography. Above the water table is the vadose zone, through which rainwater percolates. Water in the vadose zone drains down to the water table, leaving behind a thin coating of water on mineral grains. The vadose zone supplies plant roots near the surface with water.

Because the surface of the water table is not flat but instead rises and falls with topography, groundwater is affected by gravity in the same fashion as surface water. Groundwater flows downhill to topographic lows. If the water table intersects the land surface, groundwater will flow out onto the surface at springs, either to be collected there or to subsequently flow farther along a drainage. Groundwater commonly collects in stream drainages but may remain entirely beneath the surface of dry stream-beds in arid regions. In particularly wet years, short stretches of an otherwise dry stream-bed may have flowing water because the water table rises to intersect the land surface.

31 – 1 Speciation in Geographically Isolated Populations

Evolutionary biologists believe that speciation, the formation of a new species, often begins when some kind of physical barrier arises and divides a population of a single species into separate subpopulations. Physical separation between subpopulations promotes the formation of new species because once the members of one subpopulation can no longer mate with members of another subpopulation, they cannot exchange variant genes that arise in one of the subpopulations. In the absence of gene flow between the subpopulations, genetic differences between the groups begin to accumulate. Eventually the subpopulations become so genetically distinct that they cannot interbreed even if the physical barriers between them were removed. At this point the subpopulations have evolved into distinct species. This route to speciation is known as allopatry ("allo" means "different", and "patria" means "homeland").

Allopatric speciation may be the main speciation route. This should not be surprising, since allopatry is pretty common. In general, the subpopulations of most species are separated from each other by some measurable distance. So even under normal situations the gene flow among the subpopulations is more of an intermittent trickle than a steady stream. In addition, barriers can rapidly arise and shut off the trickle. For example, in the 1800s a monstrous earthquake changed the course of the Mississippi River, a large river flowing in the central part of the United States of America. The change separated populations of insects now living along opposite shore, completely cutting off gene flow between them.

Geographic isolation also can proceed slowly, over great spans of time. We find evidence of such extended events in the fossil record, which affords glimpses into the breakup of formerly continuous environments. For example, during past ice ages, glaciers advanced down through North America and Europe and gradually cut off parts of populations from one another. When the glacier retreated, the separated populations of plants and animals came into contact again. Some groups that had descended from the same parent population were no longer reproductively compatible - they had evolved into separate species. In other groups, however, genetic divergences had not proceeded so far, and the descendants could still interbreed -

for them, reproductive isolation was not completed, and so speciation had not occurred.

Allopatric speciation can also be brought by the imperceptibly slow but colossal movements of the tectonic plates that make up Earth's surface. About 5 million years ago such geologic movements created the land bridge between North America and South America that we call the Isthmus of Panama. While previously the gap between the continents had allowed a free flow of water, now the isthmus presented a barrier that divided the Atlantic Ocean from the Pacific Ocean. This division set the stage for allopatric speciation among populations of fishes and other marine species.

In the 1980s, John Graves studied two populations of closely related fishes, one population from the Atlantic side of isthmus, the other from the Pacific side. He compared four enzymes found in the muscles of each population. Graves found that all four Pacific enzymes function better at lower temperatures than the four Atlantic versions of the same enzymes. This is significant because Pacific seawater is typically 2 to 3 degrees cooler than seawater on the Atlantic side of isthmus. Analysis by gel electrophoresis revealed slight differences in amino acid sequence of the enzymes of two of the four pairs. This is significant because the amino acid sequence of an enzyme is determined by genes.

Graves drew two conclusions from these observations. First, at least some of the observed differences between the enzymes of the Atlantic and Pacific fish populations were not random but were the result of evolutionary adaptation. Second, it appears that closely related populations of fishes on both sides of the isthmus are starting to genetically diverge from each other. Because Graves's study of geographically isolated populations of isthmus fishes offers a glimpse of the beginning of a process of gradual accumulation of mutations that are neutral or adaptive, divergences here might be evidence of allopatric speciation in process.

35 – 1 Earth's Age

One of the first recorded observers to surmise a long age for Earth was the Greek historian Herodotus, who lived from approximately 480 B.C. to 425 B.C. He observed that the Nile River Delta was in fact a series of sediment deposits built up in successive floods. By noting that individual floods deposit only thin layers of sediment, he was able to conclude that the Nile Delta had taken many thousands of years to build up. More important than the amount of time Herodotus computed, which turns out to be trivial compared with the age of Earth, was the notion that one could estimate ages of geologic features by determining rates of the processes responsible for such features, and then assuming the rates to be roughly constant over time. Similar applications of this concept were to be used again and again in later centuries to estimate the ages of rock formations and, in particular, of layers of sediment that had compacted and cemented to form sedimentary rocks.

It was not until the seventeenth century that attempts were made again to understand clues to Earth's history through the rock record. Nicolaus Steno (1638-1686) was the first to work out principles of the progressive depositing of sediment in Tuscany. However, James Hutton (1726-1797), known as the founder of modern geology, was the first to have the important insight that geologic processes are cyclic in nature. Forces associated with subterranean heat cause land to be uplifted into plateaus and mountain ranges. The effects of wind and water then break down the masses of uplifted rock, producing sediment that is transported by water downward to ultimately form layers in lakes, seashores, or even oceans. Over time, the layers become sedimentary rock. These rocks are then uplifted sometime in the future to form new mountain ranges, which exhibit the sedimentary layers (and the remains of life within those layers) of the earlier episodes of erosion and deposition.

Hutton's concept represented a remarkable insight because it unified many individual phenomena and observations into a conceptual picture of Earth's history. With the further assumption that these geologic processes were generally no more or less vigorous than they are today, Hutton's examination of sedimentary layers led him to realize that Earth's history must be enormous, that geologic time is an abyss and human history a speck by comparison.

After Hutton, geologists tried to determine rates of sedimentation so as to estimate the age of Earth from the total length of the sedimentary or stratigraphic record. Typical numbers produced at the turn of the twentieth century were 100 million to 400 million years. These underestimated the actual age by factors of 10 to 50 because much of the sedimentary record is missing in various locations and because there is a long rock sequence that is older than half a billion years that is far less well defined in terms of fossils and less well preserved.

Various other techniques to estimate Earth's age fell short, and particularly noteworthy in this regard were flawed determinations of the Sun's age. It had been recognized by the German philosopher Immanuel Kant (1724-1804) that chemical reactions could not supply the tremendous amount of energy flowing from the Sun for more than about a millennium. Two physicists during the nineteenth century both came up with ages for the Sun based on the Sun's energy coming from gravitational contraction. Under the force of gravity, the compression resulting from a collapse of the object must release energy. Ages for Earth were derived that were in the tens of millions of years, much less than the geologic estimates of the time.

It was the discovery of radioactivity at the end of the nineteenth century that opened the door to determining both the Sun's energy source and the age of Earth. From the initial work came a suite of discoveries leading to radio isotopic dating, which quickly led to the realization that Earth must be billions of years old, and to the discovery of nuclear fusion as an energy source capable of sustaining the Sun's luminosity for that amount of time. By the 1960s, both analysis of meteorites and refinements of solar evolution models converged on an age for the solar system, and hence for Earth, of 4.5 billion years.

36 – 1 Soil Formation

Living organisms play an essential role in soil formation. The numerous plants and animals living in the soil release minerals from the parent material from which soil is formed, supply organic matter, aid in the translocation (movement) and aeration of the soil, and help protect the soil from erosion. The types of organisms growing or living in the soil greatly influence the soil's physical and chemical characteristics. In fact, for mature soils in many parts of the world, the predominant type of natural vegetation is considered the most important direct influence on soil characteristics. For this reason, a soil scientist can tell a great deal about the attributes of the soil in any given area simply from knowing what kind of flora the soil supports. Thus prairies and tundra regions, which have characteristic vegetations, also have characteristic soils.

The quantity and total weight of soil flora generally exceed that of soil fauna. By far the most numerous and smallest of the plants living in soil are bacteria. Under favorable conditions, a million or more of these tiny, single-celled plants can inhabit each cubic centimeter of soil. It is the bacteria, more than any other organisms, that enable rock or other parent material to undergo the gradual transformation to soil. Some bacteria produce organic acids that directly attack parent material, breaking it down and releasing plant nutrients. Others decompose organic litter (debris) to form humus (nutrient-rich organic matter). A third group of bacteria inhabits the root systems of plants called legumes. These include many important agricultural crops, such as alfalfa, clover, soybeans, peas, and peanuts. The bacteria that legumes host within their root nodules (small swellings on the root) change nitrogen gas from the atmosphere into nitrogen compounds that plants are able to metabolize, a process, known as nitrogen fixation, that makes the soil more fertile. Other microscopic plants also are important in soil development. For example, in highly acidic soils where few bacteria can survive, fungi frequently become the chief decomposers of organic matter.

More complex forms of vegetation play several vital roles with respect to the soil. Trees, grass, and other large plants supply the bulk of the soil's humus. The minerals released as these plants decompose on the surface constitute an important nutrient source for succeeding generations of plants as well as for other soil organisms. In

addition, trees can extend their roots deep within the soil and bring up nutrients from far below the surface. These nutrients eventually enrich the surface soil when the tree drops its leaves or when it dies and decomposes. Finally, trees perform the vital function of slowing water runoff and holding the soil in place with their root systems, thus combating erosion. The increased erosion that often accompanies agricultural use of sloping land is principally caused by the removal of its protective cover of natural vegetation.

Animals also influence soil composition. The faunal counterparts of bacteria are protozoa. These single-celled organisms are the most numerous representatives of the animal kingdom, and, like bacteria, a million or more can sometimes inhabit each cubic centimeter of soil. Protozoa feed on organic matter and hasten its decomposition. Among other soil-dwelling animals, the earthworm is probably the most important. Under exceptionally favorable conditions, up to a million earthworms (with a total body weight exceeding 450 kilograms) may inhabit an acre of soil. Earthworms ingest large quantities of soil, chemically alter it, and excrete it as organic matter called casts. The casts form a high-quality natural fertilizer. In addition, earthworms mix of soil both vertically and horizontally, improving aeration and drainage.

Insects such as ants and termites also can be exceedingly numerous under favorable climatic and soil conditions. In addition, mammals such as moles, field mice, gophers, and prairie dogs sometimes are present in sufficient numbers to have significant impact on the soil. These animals primarily work the soil mechanically. As a result, the soil is aerated broken up, fertilized, and brought to the surface, hastening soil development.

36 – 2 Early Ideas About Deep-sea Biology

In 1841 Edward Forbes was offered the chance to serve as naturalist aboard HMS Beacon, an English Royal Navy ship assigned to survey the Aegean Sea. For a year and a half, the Beacon crisscrossed the Aegean waters. During that time Forbes was able to drag this small, triangular dredge—a tool with a leather net for capturing creatures along the sea bottom—at a hundred locations, at depths ranging from 6 to 1380 feet. He collected hundreds of different species of animals, and he saw that they were distributed in eight different depth zones, each containing its own distinct assemblage of animal life, the way zones of elevation on the side of a mountain are populated by distinct sets of plants.

Forbes also thought he saw, as he later told the British Association, that "the number of species and individuals diminishes as we descend, pointing to a zero in the distribution of animal life as yet unvisited." This zero, Forbes casually speculated—he simply extended a line on his graph of animal number versus depth—probably began at a depth of 1,800 feet. Below that was the final zone in Forbes's scheme, zone nine, a zone that covered most of the ocean floor and thus most of the solid surface of Earth: Forbes called this the azoic zone, where no animal, to say nothing of plants, could survive.

Forbes's azoic zone was entirely plausible at the time, and it was certainly far from the strangest idea that was then entertained about the deep sea. In the first decade of the nineteenth century, a French naturalist named Francois Peron had sailed around the world measuring the temperature of the ocean. He found that the deeper the water, the colder it got, and he concluded that the seafloor was covered with a thick layer of ice. Peron ignored the fact that water expands when it freezes and that ice therefore floats. A more popular belief at the time was that water at great depth would be compressed to such a density that nothing could sink through it. This ignored the fact that water is all but incompressible. But even the more sensible naturalists of the day were guilty of a similar misconception. They imagined the deep sea as being filled with an unmoving and undisturbable pool of cold, dense water. In reality the deep is always being refreshed by cold water sinking from above.

The central implication of all these misconceptions was that nothing could live in the abyss (deep), just as Forbes's observations seemed to indicate. But Forbes erred in two ways. One was the particular study site he happened to use as a springboard for his sweeping postulate of a lifeless abyss. Although the Aegean had been the birthplace of marine biology, its depths are now known to be exceptionally lacking in animal diversity. Moreover, through no fault of his own, Forbes was not particularly successful at sampling such life as did exist at the bottom of the Aegean. It was his dredge that was inadequate. Its opening was so small and the holes in the net so large that the dredge inevitably missed animals. Many of those it did catch must have poured out of its open mouth when Forbes reeled it in. His azoic zone, then, was a plausible but wild extrapolation from pioneering but feeble data.

As it turned out, the existence of the azoic zone had been disproved even before Forbes suggested it, and the theory continued to be contradicted regularly throughout its long and influential life. Searching for the Northwest Passage from the Atlantic to the Pacific in 1818, Sir John Ross had lowered his "deep-sea clam"—a sort of bivalved sediment scoop—into the water of Baffin Bay (an inlet between the Atlantic and Arctic oceans), which he determined to be more than a thousand fathoms deep in some places. Modern soundings indicate he overestimated his depths by several hundred fathoms, but in any case Ross's clam dove several times deeper than Forbes's dredge. It brought back mud laced with worms, and starfish that had entangled themselves in the line at depths well below the supposed boundary of the azoic zone.

40 – 4 Population and Climate

The human population on Earth has grown to the point that it is having an effect on Earth's atmosphere and ecosystems. Burning of fossil fuels, deforestation, urbanization, cultivation of rice and cattle, and the manufacture of chlorofluorocarbons (CFCs) for propellants and refrigerants are increasing the concentration of carbon dioxide, methane, nitrogen oxides, sulphur oxides, dust, and CFCs in the atmosphere. About 70 percent of the Sun's energy passes through the atmosphere and strikes Earth's surface. This radiation heats the surface of the land and ocean, and these surfaces then reradiate infrared radiation back into space. This allows Earth to avoid heating up too much. However, not all of the infrared radiation makes it into space; some is absorbed by gases in the atmosphere and is reradiated back to Earth's surface. A greenhouse gas is one that absorbs infrared radiation and then reradiates some of this radiation back to Earth. Carbon dioxide, CFCs, methane, and nitrogen oxides are greenhouse gases. The natural greenhouse effect of our atmosphere is well established. In fact, without greenhouse gases in the atmosphere, scientists calculate that Earth would be about 33°C cooler than it currently is.

The current concentration of carbon dioxide in the atmosphere is about 360 parts per million. Human activities are having a major influence on atmospheric carbon dioxide concentrations, which are rising so fast that current predictions are that atmospheric concentrations of carbon dioxide will double in the next 50 to 100 years. The Intergovernmental Panel on Climate Change (IPCC) report in 1992, which represents a consensus of most atmospheric scientists, predicts that a doubling of carbon dioxide concentration would raise global temperatures anywhere between 1.4°C and 4.5°C. The IPCC report issued in 2001 raised the temperature prediction almost twofold. The suggested rise in temperature is greater than the changes that occurred in the past between ice ages. The increase in temperatures would not be uniform, with the smallest changes at the equator and changes two or three times as great at the poles. The local effects of these global changes are difficult to predict, but it is generally agreed that they may include alterations in ocean currents, increased winter flooding in some areas of the Northern Hemisphere, a higher incidence of summer drought in some areas, and rising sea levels, which may flood low-lying countries.

Scientists are actively investigating the feedback mechanism within the physical, chemical, and biological components of Earth's climate system in order to make accurate predictions of the effects the

rise in greenhouse gases will have on future global climates. Global circulation models are important tools in this process. These models incorporate current knowledge on atmospheric circulation patterns, ocean currents, the effect of landmasses, and the like to predict climate under changed conditions. There are several models, and all show agreement on a global scale. For example, all models show substantial changes in climate when carbon dioxide concentration is doubled. However, there are significant differences in the regional climates predicted by different models. Most models project greater temperature increases in mid-latitude regions and in mid-continental regions relative to the global average. Additionally, changes in precipitation patterns are predicted, with decreases in mid-latitude regions and increased rainfall in some tropical areas. Finally, most models predict that there will be increased occurrences of extreme events, such as extended periods without rain (drought), extreme heat waves, greater seasonal variation in temperatures, and increases in the frequency and magnitude of severe storms. Plants and animals have strong responses to virtually every aspect of these projected global changes.

The challenge of predicting organismal responses to global climate change is difficult. Partly, this is due to the fact that there are more studies of short-term, individual organism responses than there are of long-term, systemwide studies. It is extremely difficult, both monetarily and physically, for scientists to conduct field studies at spatial and temporal scales that are large enough to include all the components of real-world systems, especially ecosystems with large, freely ranging organisms. One way paleobiologists try to get around this limitation is to attempt to reconstruct past climates by examining fossil life.

The relative roles that abiotic and biotic factors play in the distribution of organisms is especially important now, when the world is confronted with the consequences of a growing human population. Changes in climate, land use, and habitat destruction are currently causing dramatic decreases in biodiversity throughout the world. An understanding of climate-organism relationships is essential to efforts to preserve and manage Earth's biodiversity.

40 – 2 Latitude and Biodiversity

When we look at the way in which biodiversity (biological diversity) is distributed over the land surface of the planet, we find that it is far from even. The tropics contain many more species overall than an equivalent area at the higher latitudes. This seems to be true for many different groups of animals and plants.

Why is it that higher latitudes have lower diversities than the tropics? Perhaps it is simply a matter of land area. The tropics contain a larger surface area of land than higher latitudes—a fact that is not always evident when we examine commonly used projections of Earth's curved surface, since this tends to exaggerate the areas of land in the higher latitudes—and some biogeographers regard the differences in diversity as a reflection of this effect. But an analysis of the data by biologist Klaus Rohde does not support this explanation. Although area may contribute to biodiversity, it is certainly not the whole story; otherwise, large landmasses would always be richer in species.

Productivity seems to be involved instead, though perhaps its influence is indirect. Where conditions are most suitable for plant growth—that is, where temperatures are relatively high and uniform and where there is an ample supply of water—one usually finds large masses of vegetation. This leads to a complex structure in the layers of plant material. In a tropical rain forest, for example, a very large quantity of plant material builds up above the surface of the ground. There is also a large mass of material, developed below ground as root tissues, but this is less apparent. Careful analysis of the above ground material reveals that it is arranged in a series of layers, the precise number of layers varying with age and the nature of the forest. The arrangement of the biological mass ("biomass") of the vegetation into layered forms is termed its "structure" (as opposed to its "composition," which refers to the species of organisms forming the community). Structure is essentially the architecture of vegetation, and as in the case of tropical forests, it can be extremely complicated. In a mature floodplain tropical forest in the Amazon River basin, the canopy (the uppermost layers of a forest, formed by the crowns of trees) takes on a stratified structure. There are three clear peaks in leaf cover at heights of approximately 3, 6, and 30 meters above the ground; and the very highest layer, at 50 meters, corresponds to the very tall trees that stand free of the main canopy and form an open layer of

their own. So, such a forest contains essentially four layers of canopy. Forests in temperate lands often have just two canopy layers, so they have much less complex architecture.

Structure has a strong influence on the animal life inhabiting a site. It forms the spatial environment within which an animal feeds, moves around shelters, lives, and breeds. It even affects the climate on a very local level (the "microclimate") by influencing light intensity, humidity, and both the range and extremes of temperature. An area of grassland vegetation with very simple structure, for example, has a very different microclimate at the ground level from that experienced in the upper canopy. Wind speeds are lower, temperatures are lower during the day (but warmer at night), and the relative humidity is much greater near the ground. The complexity of the microclimate is closely related to the complexity of structure in vegetation, and generally speaking, the more complex the structure of vegetation, the more species of animal are able to make a living there. The high plant biomass of the tropics leads to a greater spatial complexity in the environment, and this leads to a higher potential for diversity in the living things that can occupy a region. The climates of the higher latitudes are generally less favorable for the accumulation of large quantities of biomass; hence, the structure of vegetation is simpler and the animal diversity is consequently lower.

43 – 3 El Niño

The cold Humboldt Current of the Pacific Ocean flows toward the equator along the coasts of Ecuador and Peru in South America. When the current approaches the equator, the westward-flowing trade winds cause nutrient-rich cold water along the coast to rise from deeper depths to more shallow ones. This upwelling of water has economic repercussions. Fishing, especially for anchovies, is a major local industry.

Every year during the months of December and January, a weak, warm countercurrent replaces the normally cold coastal waters. Without the upwelling of nutrients from below to feed the fish, fishing comes to a standstill. Fishers in this region have known the phenomenon for hundreds of years. In fact, this is the time of year they traditionally set aside to tend to their equipment and await the return of cold water. The residents of the region have given this phenomenon the name of El Niño, which is Spanish for "the child," because it occurs at about the time of the celebration of birth of the Christ child.

While the warm-water countercurrent usually lasts for two months or less, there are occasions when the disruption to the normal flow lasts for many months. In these situations, water temperatures are raised not just along the coast, but for thousands of kilometers offshore. Over the last few decades, the term El Niño has come to be used to describe these exceptionally strong episodes and not the annual event. During the past 60 years, at least ten El Niño have been observed. Not only do El Niño affect the temperature of the equatorial Pacific, but the strongest of them impact global weather.

The processes that interact to produce an El Niño involve conditions all across the Pacific, not just in the waters off South America. Over 60 years ago, Sir Gilbert Walker, a British scientist, discovered a connection between surface pressure readings at weather stations on the eastern and western sides of the Pacific. He noted that a rise in atmospheric pressure in the eastern Pacific is usually accompanied by a fall in pressure in the western Pacific and vice versa. He called this seesaw pattern the Southern Oscillation. It was later realized that there is a close link between El Niño and the Southern Oscillation. In fact, the link between the two is so great that they are often referred to jointly as ENSO (El Niño-Southern Oscillation).

During a typical year, the eastern Pacific has a higher pressure than the western Pacific does. This east-to-west pressure gradient enhances the trade winds over the equatorial waters. This results in a warm surface current that moves east to west at the equator. The western Pacific develops a thick, warm layer of water while the eastern Pacific has the cold Humboldt Current enhanced by upwelling. However, in other years the Southern Oscillation, for unknown reasons, swings in the opposite direction, dramatically changing the usual conditions described above, with pressure increasing in the western Pacific and decreasing in the eastern Pacific. This change in the pressure gradient causes the trade winds to weaken or, in some cases, to reverse. This then causes the warm water in the western Pacific to flow eastward, increasing sea-surface temperatures in the central and eastern Pacific. The eastward shift signals the beginning of an El Niño.

Scientists try to document as many past El Niño events as possible by piecing together bits of historical evidence, such as sea-surface temperature records, daily observations of atmospheric pressure and rainfall, fisheries' records from South America, and the writings of Spanish colonists dating back to the fifteenth century. From such historical evidence we know that El Niño have occurred as far back as records go. It would seem that they are becoming more frequent. Records indicate that during the sixteenth century, an El Niño occurred on average every six years. Evidence gathered over the past few decades indicates that El Niños are now occurring on average a little over every two years. Even more alarming is the fact that they appear to be getting stronger. The 1997-1998 El Niño brought copious and damaging rainfall to the southern United States, from California to Florida. Snowstorms in the northeast portion of the United States were more frequent and intense than in most years.

46 – 3 Ecosystem Diversity and Stability

Conservation biologists have long been concerned that species extinction could have significant consequences for the stability of entire ecosystems—groups of interacting organisms and the physical environment that they inhabit. An ecosystem could survive the loss of some species, but if enough species were lost, the ecosystem would be severely degraded. In fact, it is possible that the loss of a single important species could start a cascade of extinctions that might dramatically change an entire ecosystem. A good illustration of this occurred after sea otters were eliminated from some Pacific kelp (seaweed) bed ecosystems: the kelp beds were practically obliterated too because in the absence of sea otter predation, sea urchin populations exploded and consumed most of the kelp and other macroalgae.

It is usually claimed that species-rich ecosystems tend to be more stable than species-poor ecosystems. Three mechanisms by which higher diversity increases ecosystem stability have been proposed. First, if there are more species in an ecosystem, then its food web will be more complex, with greater redundancy among species in terms of their nutritional roles. In other words, in a rich system if a species is lost, there is a good chance that other species will take over its function as prey, predator, producer, decomposer, or whatever role it played. Second, diverse ecosystems may be less likely to be invaded by new species, notably exotics (foreign species living outside their native range), that would disrupt the ecosystem's structure and function. Third, in a species-rich ecosystem, diseases may spread more slowly because most species will be relatively less abundant, thus increasing the average distance between individuals of the same species and hampering disease transmission among individuals.

Scientific evidence to illuminate these ideas has been slow in coming, and many shadows remain. One of the first studies to provide data supporting a relationship between diversity and stability examined how grassland plants responded to a drought. Researchers D. Tilman and J. A. Downing used the ratio of above-ground biomass in 1988 (after two years of drought) to that in 1986 (predrought) in 207 plots in a grassland field in the Cedar Creek Natural History Area in Minnesota as an index of ecosystem response to disruption by drought. In an experiment that began in 1982, they compared these values with the number of plant

species in each plot and discovered that the plots with a greater number of plant species experienced a less dramatic reduction in biomass. Plots with more than ten species had about half as much biomass in 1988 as in 1986, whereas those with fewer than five species only produced roughly one-eighth as much biomass after the two-year drought. Apparently, species-rich plots were likely to contain some drought-resistant plant species that grew better in drought years, compensating for the poor growth of less-tolerant species.

To put this result in more general terms, a species-rich ecosystem may be more stable because it is more likely to have species with a wide array of responses to variable conditions such as droughts. Furthermore, a species-rich ecosystem is more likely to have species with similar ecological functions, so that if a species is lost from an ecosystem, another species, probably a competitor, is likely to flourish and occupy its functional role. Both of these, variability in responses and functional redundancy, could be thought of as insurance against disturbances.

The Minnesota grassland research has been widely accepted as strong evidence for the diversity-stability theory; however, its findings have been questioned, and similar studies on other ecosystems have not always found a positive relationship between diversity and stability. Clearly, this is a complex issue that requires further field research with a broad spectrum of ecosystems and species: grassland plants and computer models will only take us so far. In the end, despite insightful attempts to detect some general patterns, we may find it very difficult to reduce this topic to a simple, universal truth.

45 – 2 Wind pollination

Pollen, a powdery substance, which is produced by flowering plants and contains male reproductive cells, is usually carried from plant to plant by insects or birds, but some plants rely on the wind to carry their pollen. Wind pollination is often seen as being primitive and wasteful in costly pollen and yet it is surprisingly common, especially in higher latitudes. Wind is very good at moving pollen a long way; pollen can be blown for hundreds of kilometers, and only birds can get pollen anywhere near as far. The drawback is that wind is obviously unspecific as to where it takes the pollen. It is like trying to get a letter to a friend at the other end of the village by climbing onto the roof and throwing an armful of letters into the air and hoping that one will end up in the friend's garden. For the relatively few dominant tree species that make up temperate forests, where there are many individuals of the same species within pollen range, this is quite a safe gamble. If a number of people in the village were throwing letters off roofs, your friend would be bound to get one. By contrast, in the tropics, where each tree species has few, widely scattered individuals, the chance of wind blowing pollen to another individual is sufficiently slim that animals are a safer bet as transporters of pollen. Even tall trees in the tropics are usually not wind pollinated despite being in windy conditions. In a similar way, trees in temperate forests that are insect pollinated tend to grow as solitary, widely spread individuals.

Since wind-pollinated flowers have no need to attract insects or other animals, they have dispensed with bright petals, nectar, and scent. These are at best a waste and at worst an impediment to the transfer of pollen in the air. The result is insignificant-looking flowers and catkins (dense cylindrical clusters of small, petalless flowers).

Wind pollination does, of course, require a lot of pollen. Birch and hazel trees can produce 5.5 and 4 million grains per catkin, respectively. There are various adaptations to help as much of the pollen go as far as possible. Most deciduous wind-pollinated trees (which shed their leaves every fall) produce their pollen in the spring while the branches are bare of leaves to reduce the surrounding surfaces that "compete" with the stigmas (the part of the flower that receives the pollen) for pollen. Evergreen conifers, which do not shed their leaves, have less to gain from spring

flowering, and, indeed, some flower in the autumn or winter.

Pollen produced higher in the top branches is likely to go farther: it is windier (and gustier) and the pollen can be blown farther before hitting the ground. Moreover, dangling catkins like hazel hold the pollen in until the wind is strong enough to bend them, ensuring that pollen is only shed into the air when the wind is blowing hard. Weather is also important. Pollen is shed primarily when the air is dry to prevent too much sticking to wet surfaces or being knocked out of the air by rain. Despite these adaptations, much of the pollen fails to leave the top branches, and only between 0.5 percent and 40 percent gets more than 100 meters away from the parent. But once this far, significant quantities can go a kilometer or more. Indeed, pollen can travel many thousands of kilometers at high altitudes. Since all this pollen is floating around in the air, it is no wonder that wind-pollinated trees are a major source of allergies.

Once the pollen has been snatched by the wind, the fate of the pollen is obviously up to the vagaries of the wind, but not everything is left to chance. Windborne pollen is dry, rounded, smooth, and generally smaller than that of insect-pollinated plants. But size is a two-edged sword. Small grains may be blown farther but they are also more prone to be whisked past the waiting stigma because smaller particles tend to stay trapped in the fast-moving air that flows around the stigma. But stigmas create turbulence, which slows the air speed around them and may help pollen stick to them.

48 – 3 Climate and Urban Development

For more than a hundred years, it has been known that cities are generally warmer than surrounding rural areas. This region of city warmth, known as the urban heat island, can influence the concentration of air pollution. However, before we look at its influence, let's see how the heat island actually forms.

The urban heat island is due to industrial and urban development. In rural areas, a large part of the incoming solar energy is used in evaporating water from vegetation and soil. In cities, where less vegetation and exposed soil exist, the majority of the Sun's energy is absorbed by urban structures and asphalt. Hence, during warm daylight hours, less evaporative cooling in cities allows surface temperatures to rise higher than in rural areas. The cause of the urban heat island is quite involved. Depending on the location, time of year, and time of day, any or all of the following differences between cities and their surroundings can be important: albedo (reflectivity of the surface), surface roughness, emissions of heat, emissions of moisture, and emissions of particles that affect net radiation and the growth of cloud droplets.

At night, the solar energy (stored as vast quantities of heat in city buildings and roads) is slowly released into the city air. Additional city heat is given off at night (and during the day) by vehicles and factories, as well as by industrial and domestic heating and cooling units. The release of heat energy is retarded by the tall vertical city walls that do not allow infrared radiation to escape as readily as does the relatively level surface of the surrounding countryside. The slow release of heat tends to keep nighttime city temperatures higher than those of the faster-cooling rural areas. Overall, the heat island is strongest (1) at night when compensating sunlight is absent; (2) during the winter, when nights are longer and there is more heat generated in the city; and (3) when the region is dominated by a high-pressure area with light winds, clear skies, and less humid air. Over time, increasing urban heat islands affect climatological temperature records, producing artificial warming in climatic records taken in cities. This warming, therefore, must be accounted for in interpreting climate change over the past century.

The constant outpouring of pollutants into the environment may influence the climate of the city. Certain particles reflect solar radiation, thereby

reducing the sunlight that reaches the surface. Some particles serve as nuclei upon which water and ice form. Water vapor condenses onto these particles when the relative humidity is as low as 70 percent, forming haze that greatly reduces visibility. Moreover, the added nuclei increase the frequency of city fog.

Studies suggest that precipitation may be greater in cities than in the surrounding countryside; this phenomenon may be due in part to the increased roughness of city terrain, brought on by large structures that cause surface air to slow and gradually converge. This piling up of air over the city then slowly rises, much like toothpaste does when its tube is squeezed. At the same time, city heat warms the surface air, making it more unstable, which enhances rising air motions, which, in turn, aids in forming clouds and thunderstorms. This process helps explain why both tend to be more frequent over cities.

On clear still nights when the heat island is pronounced, a small thermal low-pressure area forms over the city. Sometimes a light breeze—called a country breeze—blows from the countryside into the city. If there are major industrial areas along the outskirts, pollutants are carried into the heart of town, where they tend to concentrate. Such an event is especially probable if vertical mixing and dispersion of pollutants are inhibited. Pollutants from urban areas may even affect the weather downwind from them.

49 – 1 Ancient Coastlines

Information on past climates is of primary relevance to archaeology because of what it tells us about the effects on the land and on the resources that people needed to survive. The most crucial effect of climate was on the sheer quantity of land available in each period, measurable by studying ancient coastlines. These have changed constantly through time, even in relatively recent periods, as can be seen from the Neolithic stone circle of Er Lannic, in Brittany, France (once inland but now half submerged on an island) or medieval villages in east Yorkshire, England, that have tumbled into the sea in the last few centuries as the North Sea gnaws its way westward and erodes the cliffs. Conversely, silts deposited by rivers sometimes push the sea farther back, creating new land, as at Ephesus in western Turkey, a port on the coast in Roman times but today some five kilometers inland.

Nevertheless, for archeologists concerned with the long periods of time of the Paleolithic period there are variations in coastlines of much greater magnitude to consider. The expansion and contraction of the continental glaciers caused huge and uneven rises and falls in sea levels worldwide. When the ice sheets grew, the sea level would drop as water became locked up in the glaciers; when the ice melted, the sea level would rise again. Falls in sea level often exposed a number of important land bridges, such as those linking Alaska to northeast Asia and Britain to northwest Europe, a phenomenon with far-reaching effects not only on human colonization of the globe but also on the environment as a whole—the flora and fauna of isolated or insular areas were radically and often irreversibly affected. Between Alaska and Asia today lies the Bering Strait, which is so shallow that a fall in sea level of only four meters would turn it into a land bridge. When the ice sheets were at their greatest extent some 18,000 years ago (the glacier maximum), it is thought that the fall was about 120 meters, which therefore created not merely a bridge but a vast plain, 1,000 kilometers from the north to the south, which has been called Beringia. The existence of Beringia (and the extent to which it could have supported human life) is one of the crucial pieces of evidence in the continuing debate about the likely route and date of human colonization of the New World.

The assessment of past rises and falls in sea level requires study of submerged land surfaces off the

coast and of raised or elevated beaches on land. Raised beaches are remnants of former coastlines at higher levels relative to the present shoreline and visible, for instance, along the Californian coast north of San Francisco. The height of a raised beach above the present shoreline, however, does not generally give a straightforward indication of the height of a former sea level. In the majority of cases, the beaches lie at a higher level because the land has been raised up through isostatic uplift or tectonic movement. Isostatic uplift of the land occurs when the weight of ice is removed as temperatures rise, as at the end of an ice age; it has affected coastlines, for example, in Scandinavia, Scotland, Alaska, and Newfoundland during the postglacial period. Tectonic movements involve displacements in the plates that make up Earth's crust. Middle and Late Pleistocene raised beaches in the Mediterranean are one instance of such movements.

Raised beaches often consist of areas of sand, pebbles, or dunes, sometimes containing seashells or piles of debris comprising shells and bones of marine animals used by humans. In Tokyo Bay, for example, shell mounds of the Jomon period (about 10,000 to 300 B.C.E.) mark the position of the shoreline at a time of maximum inundation by the sea (6,500-5,500 years ago), when, through tectonic movement, the sea was three to five meters higher in relation to the contemporary landmass of Japan than at present. Analysis of the shells themselves has confirmed the changes in marine topography, for it is only during the maximum phase that subtropical species of mollusc are present, indicating a higher water temperature.

52 – 1 Stream Deposit

A large, swift stream or river can carry all sizes of particles, from clay to boulders. When the current slows down, its competence (how much it can carry) decreases and the stream deposits the largest particles in the streambed. If current velocity continues to decrease—as a flood wanes, for example—finer particles settle out on top of the large ones. Thus, a stream sorts its sediment according to size. A waning flood might deposit a layer of gravel, overlain by sand and finally topped by silt and clay. Streams also sort sediment in the downstream direction. Many mountain streams are choked with boulders and cobbles, but far downstream, their deltas are composed mainly of fine silt and clay. This downstream sorting is curious because stream velocity generally increases in the downstream direction. Competence increases with velocity, so a river should be able to transport larger particles than its tributaries carry. One explanation for downstream sorting is that abrasion wears away the boulders and cobbles to sand and silt as the sediment moves downstream over the years. Thus, only the fine sediment reaches the lower parts of most rivers.

A stream deposits its sediment in three environments: Alluvial fans and deltas form where stream gradient (angle of incline) suddenly decreases as a stream enters a flat plain, a lake, or the sea; floodplain deposits accumulate on a floodplain adjacent to the stream channel; and channel deposits form in the stream channel itself. Bars, which are elongated mounds of sediment, are transient features that form in the stream channel and on the banks. They commonly form in one year and erode the next. Rivers used for commercial navigation must be recharged frequently because bars shift from year to year. Imagine a winding stream. The water on the outside of the curve moves faster than the water on the inside. The stream erodes its outside bank because the current's inertia drives it into the outside bank. At the same time, the slower water on the inside point of the bend deposits sediment, forming a point bar. A mid-channel bar is a sandy and gravelly deposit that forms in the middle of a stream channel.

Most streams flow in a single channel. In contrast, a braided stream flows in many shallow, interconnecting channels. A braided stream forms where more sediment is supplied to a stream than

it can carry. The stream dumps the excess sediment, forming mid-channel bars. The bars gradually fill a channel, forcing the stream to overflow its banks and erode new channels. As a result, a braided stream flows simultaneously in several channels and shifts back and forth across its floodplain. Braided streams are common in both deserts and glacial environments because both produce abundant sediment. A desert yields large amounts of sediment because it has little or no vegetation to prevent erosion. Glaciers grind bedrock into fine sediment, which is carried by streams flowing from the melting ice. If a steep mountain stream flows onto a flat plain, its gradient and velocity decrease sharply. As a result, it deposits most of its sediment in a fan-shaped mound called an alluvial fan. Alluvial fans are common in many arid and semiarid mountainous regions.

A stream also slows abruptly where it enters the still water of a lake or ocean. The sediment settles out to form a nearly flat landform called a delta. Part of the delta lies above water level, and the remainder lies slightly below water level. Deltas are commonly fan-shaped, resembling the Greek letter "delta" (Δ). Both deltas and alluvial fans change rapidly. Sediment fills channels (waterways), which are then abandoned while new channels develop as in a braided stream. As a result, a stream feeding a delta or fan splits into many channels called distributaries. A large delta may spread out in this manner until it covers thousands of square kilometers. Most fans, however, are much smaller, covering a fraction of a square kilometer to a few square kilometers. The Mississippi River has flowed through seven different delta channels during the past 5,000 to 6,000 years. But in recent years, engineers have built great systems of levees (retaining walls) in attempts to stabilize the channels.

54 – 3 Elements of Life

The creation of life requires a set of chemical elements for making the components of cells. Life on Earth uses about 25 of the 92 naturally occurring chemical elements, although just 4 of these elements—oxygen, carbon, hydrogen, and nitrogen—make up about 96 percent of the mass of living organisms. Thus, a first requirement for life might be the presence of most or all of the elements used by life.

Interestingly, this requirement can probably be met by almost any world. Scientists have determined that all chemical elements in the universe besides hydrogen and helium (and a trace amount of lithium) were produced by stars. These are known as heavy elements because they are heavier than hydrogen and helium. Although all of these heavy elements are quite rare compared to hydrogen and helium, they are found just about everywhere.

Heavy elements are continually being manufactured by stars and released into space by stellar deaths, so their amount compared to hydrogen and helium gradually rises with time. Heavy elements make up about 2 percent of the chemical content (by mass) of our solar system, the other 98 percent is hydrogen and helium. In some very old star systems, which formed before many heavy elements were produced, the heavy-element share may be less than 0.1 percent. Nevertheless, every star system studied has at least some amount of all the elements used by life. Moreover, when planetesimals—small, solid objects formed in the early solar system that may accumulate to become planets—condense within a forming star system, they are inevitably made from heavy elements because the more common hydrogen and helium remain gaseous. Thus, planetesimals everywhere should contain the elements needed for life, which means that objects built from planetesimals—planets, moons, asteroids, and comets—also contain these elements. The nature of solar-system formation explains why Earth contains all the elements needed for life, and it is why we expect these elements to be present on other worlds throughout our solar system, galaxy, and universe.

Note that this argument does not change, even if we allow for life very different from life on Earth. Life on Earth is carbon based, and most biologists believe that life elsewhere is likely to be carbon based as well. However, we cannot absolutely rule

out the possibility of life with another chemical basis, such as silicon or nitrogen. The set of elements (or their relative proportions) used by life based on some other element might be somewhat different from that used by carbon-based life on Earth. But the elements are still products of stars and would still be present in planetesimals everywhere. No matter what kinds of life we are looking for, we are likely to find the necessary elements on almost every planet, moon, asteroid, and comet in the universe.

A somewhat stricter requirement is the presence of these elements in molecules that can be used as ready-made building blocks for life, just as early Earth probably had an organic soup of amino acids and other complex molecules. Earth's organic molecules likely came from some combination of three sources: chemical reactions in the atmosphere, chemical reactions near deep-sea vents in the oceans, and molecules carried to Earth by asteroids and comets. The first two sources can occur only on worlds with atmospheres or oceans, respectively. But the third source should have brought similar molecules to nearly all worlds in our solar system.

Studies of meteorites and comets suggest that organic molecules are widespread among both asteroids and comets. Because each body in the solar system was repeatedly struck by asteroids and comets during the period known as the heavy bombardment (about 4 billion years ago), each body should have received at least some organic molecules. However, these molecules tend to be destroyed by solar radiation on surfaces unprotected by atmospheres. Moreover, while these molecules might stay intact beneath the surface (as they evidently do on asteroids and comets), they probably cannot react with each other unless some kind of liquid or gas is available to move them about. Thus, if we limit our search to worlds on which organic molecules are likely to be involved in chemical reactions, we can probably rule out any world that lacks both an atmosphere and a surface or subsurface liquid medium, such as water.

55 – 1 Mesolithic Complexity in Scandinavia

The European Mesolithic (roughly the period from 8000 B.C. to 2700 B.C.) testifies to a continuity in human culture from the times of the Ice Age. This continuity, however, was based on continuous adjustment to environmental changes following the end of the last glacial period (about 12,500 years ago). Three broad subdivisions within the northern Mesolithic are known in Scandinavia. The Maglemose Period (7500 -5700 B.C.) was a time of seasonal exploitation of rivers and lakes, combined with terrestrial hunting and foraging. The sites from the Kongemose Period (5700-4600 B.C.) are mainly on the Baltic Sea coasts, along bays and near lagoons, where the people exploited both marine and terrestrial resources. Many Kongemose sites are somewhat larger than Maglemose ones. The Ertebolle Period (4600-3200 B.C.) was the culmination of Mesolithic culture in southern Scandinavia.

By the Ertebolle Period, the Scandinavians were occupying coastal settlements year-round and subsisting off a very wide range of food sources. These included forest game and waterfowl, shellfish, sea mammals, and both shallow-water and deepwater fish. There were smaller seasonal coastal sites, too: for specific activities such as deepwater fishing, sealing, or hunting of migratory birds. One such site, the Aggersund site in Denmark, was occupied for short periods of time in the autumn, when the inhabitants collected oysters and hunted some game, especially migratory swans. Ertebolle technology was far more elaborate than that of its Mesolithic predecessors; a wide variety of antler, bone, and wood tools for specialized purposes such as fowling and sea-mammal hunting were developed, including dugout canoes up to ten meters long.

With sedentary settlement comes evidence of greater social complexity in the use of cemeteries for burials and changes in burial practices. The trend toward more sedentary settlement, the cemeteries, and the occasional social differentiation revealed by elaborate burials are all reflections of an intensified use of resources among these relatively affluent hunter-gatherers of 3000 B.C. Mesolithic societies intensified the food quest by exploiting many more marine species, making productive use of migratory waterfowl and their breeding grounds, and collecting shellfish in enormous numbers. This intensification is also

reflected in a much more elaborate and diverse technology, more exchange of goods and materials between neighbors, greater variety in settlement types, and a slowly rising population throughout southern Scandinavia. These phenomena may, in part, be a reflection of rising sea levels throughout the Mesolithic that flooded many cherished territories. There are signs, too, of regional variations in artifact forms and styles, indicative of cultural differences between people living in well-delineated territories and competing for resources.

Mesolithic cultures are much less well-defined elsewhere in Europe, partly because the climatic changes were less extreme than in southern Scandinavia and because there were fewer opportunities for coastal adaptation. In much of central Europe, settlement was confined to lakeside and riverside locations, widely separated from one another by dense forests. Many Mesolithic lakeside sites were located in transitional zones between different environments so that the inhabitants could return to a central base location, where for much of the year they lived close to predictable resources such as lake fish. However, they would exploit both forest game and other seasonal resources from satellite camps. For example, the archaeologist Michael Jochim believes that some groups lived during most of the year in camps along the Danube River in central Europe, moving to summer encampments on the shores of neighboring lakes. In areas like Spain, there appears to have been intensified exploitation of marine and forest resources. There was a trend nearly everywhere toward greater variety in the diet, with more attention being paid to less obvious foods and to those that require more complex processing methods than do game and other such resources.

Thus, in parts of Europe, there was a long-term trend among hunter-gatherer societies toward a more extensive exploitation of food resources, often within the context of a strategy that sought ways to minimize the impact of environmental uncertainty. In more favored southern Scandinavia, such societies achieved a new level of social complexity that was to become commonplace among later farming peoples, and this preadaptation proved an important catalyst for rapid economic and social change when farming did come to Europe.

56 – 3 Urban Heat Islands

Climatic changes such as changes in temperature, precipitation, humidity, or wind speed that are produced by urbanization involve all major surface conditions. Some of these changes are quite obvious and relatively easy to measure. Others are more subtle and sometimes difficult to measure. The amount of change in any of these elements, at any time, depends on several variables, including the extent of the urban complex, the nature of industry, site factors such as topography and proximity to water bodies, time of day, and existing weather conditions.

The most studied and well-documented urban climatic effect is the urban heat island. The term simply refers to the fact that temperatures within cities are generally higher than in rural areas. The heat island is evident when temperature data are examined. For example, the distribution of average minimum temperatures in the Washington, D C., metropolitan area for the three-month winter period (December through February) over a five-year span, clearly represents a well-developed heat island. The warmest winter temperatures occurred in the heart of the city, while the suburbs and surrounding countryside experienced average minimum temperatures that were as much as 3 3°C lower. Remember that these temperatures are averages, because on many clear, calm nights the temperature difference between the city center and the countryside was considerably greater, often 11°C or more. Conversely, on many overcast or windy nights the temperature differential approached zero degrees.

The radical change in the surface that results when rural areas are transformed into cities is a significant cause of the urban heat island. First, the tall buildings and the concrete and asphalt of the city absorb and store greater quantities of solar radiation than do the vegetation and soil typical of rural areas. In addition, because the city surface is impermeable, the runoff of water following a rain is rapid, resulting in a severe reduction in the evaporation rate. Hence, heat that once would have been used to convert liquid water to a gas now goes to increase the surface temperature. At night, while both the city and the countryside cool by radiative losses, the stone-like surface of the city gradually releases the additional heat accumulated during the day, keeping the urban air warmer than that of the outlying areas.

A portion of the urban temperature rise must also be attributed to waste heat from sources such as home heating and air conditioning, power generation, industry, and transportation. Many studies have shown that the magnitude of human-generated energy in metropolitan areas is great when compared to the amount of energy received from the Sun at the surface. For example, investigations in Sheffield, England, and Berlin, Germany, showed that the annual heat production in those cities was equal to approximately one-third of that received from solar radiation. Another study of densely built-up Manhattan in New York City revealed that during the winter, the quantity of heat produced by combustion alone was 2 times greater than the amount of solar energy reaching the ground. In summer, the figure dropped to 1/6.

There are other, somewhat less influential, causes of the heat island. For example, the blanket of pollutants over a city contributes to the heat island by absorbing a portion of the upward-directed long-wave radiation emitted at the surface and re-emitting some of it back to the ground. A somewhat similar effect results from the complex three-dimensional structure of the city. The vertical walls of office buildings, stores, and apartments do not allow radiation to escape as readily as in outlying rural areas where surfaces are relatively flat. As the sides of these structures emit their stored heat, a portion is re-radiated between buildings rather than upward, and is therefore slowly dissipated. In addition to re-radiating the heat loss from the city, tall buildings also alter the flow of air. Because of the greater surface roughness, wind speeds within an urban area are reduced. Estimates from available records suggest a decrease on the order of about 25 percent from rural values. The lower wind speeds decrease the city's ventilation by inhibiting the movement of cooler outside air which, if allowed to penetrate, would reduce the higher temperatures of the city center.

56 – 1 Alfred Wegener's Theory of Continental Drift

In 1912, the German geologist Alfred Wegener proposed that Earth's continents are mobile rafts of lighter crust that have shifted over time by plowing their way through the denser crust of the oceans. The theory, called continental drift, was partly motivated by the apparent fit, like puzzle pieces, of the coastlines of South America and Africa. Wegener first presented his theory of continental drift at a meeting and in a paper, and then as a book. *The Origin of Continents and Oceans*, published in 1915. He continued to write updated versions of this work until his death in an ill-fated expedition to Greenland in 1930. Wegener maintained that Earth is composed of concentric shells of increasing density from crust to core. The outermost shell is not continuous but made of continental blocks of lighter rock called sial (an acronym for silica- and aluminum-rich rock) floating in the denser sima (silica- and magnesia-rich rock) underlying the oceans. All the continents had been joined in the supercontinent of Pangaea. As the continent broke up, the pieces moving apart left bits behind, explaining the presence of nonvolcanic islands and island chains, according to Wegener. Where the moving pieces collided, mountains formed. They were thrust up either by the plowing of the continents through the sima, as in the case of the Andes, or by the colliding of two blocks of sial, as in the case of the Himalayas. As for the force driving continental drift, Wegener initially invoked *Polfluchtkraft*-a force causing flight from the poles as a result of Earth's rotation-and later the tidal force resulting from the gravitational attraction between Earth and the Moon and Sun.

One of the most influential geologists to join the mobilist camp, as the drifters school became known, was Arthur Holmes. Holmes recognized the importance of radioactive heat-which had recently been discovered-and realized that there must be a mechanism to remove it from Earth's interior. That mechanism, he argued, is convection-the rising of less dense material and the sinking of more dense material. He went on to propose that the mantle (the part of Earth's interior below the outer crust and above the core) convects in large, circulating patterns, and that this motion carries the continents across Earth's surface. He also related crustal movement and mantle convection to the evolution of mountain belts. Wegener adopted Holmes's mechanism in the last rendition of his theory

Holmes, for his part, presented his grand concept of a dynamic Earth in his influential and popular text *Principles of Physical Geology*, published in 1944.

Although it was eventually supplanted by the theory of plate tectonics, Wegener's theory of continental drift influenced science because it explained disparate observations, because it was placed in the context of existing theories, and because it offered a coherent view of Earth's evolution. For example, Wegener showed not only that the coastlines on opposite sides of the Atlantic fit together, but that geologic features on the different continents fit as well. He asserted that the Appalachians, which can be traced northward through the Canadian Maritime Provinces, match the Caledonian Mountains in Scotland and Norway. He marshaled evidence from the distributions of fossil and living species to argue that land bridges joining continents were less likely than a single continent. The example commonly cited is that of *Mesosaurus*, a shoreline scavenger reptile that lived in the Permian period and is found as fossils in rocks on both sides of the South Atlantic Ocean. *Mesosaurus* was thought not to be a great swimmer, certainly not able to cross an ocean.

Wegener also found supporting evidence in ancient climates. Mounting observations indicated that the past climate of many regions was much different from the present climate. In the tropics, geologists had found sand and gravel left by ancient glaciers, and in rainy regions they had located prehistoric deserts. Then there was the discovery, by a British expedition, of plant fossils only 600 kilometers (370 miles) from the South Pole. Particularly puzzling was evidence suggesting that widely different climates in different regions had occurred at the same time, so one could not account for different climates by claiming simply that the whole of Earth was once hotter or once cooler than now. Wegener solved this dilemma by showing that observations of paleoclimate could be explained if the positions of the continents had shifted relative to those of the poles.

59 – 2 Dealing with Extreme Cold

There are a number of environments in which organisms are exposed to temperatures below 0°C and thus the risk of freezing. In polar regions, terrestrial organisms are exposed to freezing temperatures for most of the year. In more temperate regions, they may have to tolerate several months of winter, when subzero temperatures persist for long periods of time. High mountains are another place where there is permanent snow and ice, even at the equator. Exposure to subzero temperatures may occur on a daily and/or seasonal basis.

Endothermic animals (warm-blooded animals) can stop their bodies from freezing by generating their own heat. They retain heat because of the insulation provided by feathers or fur, and the layer of fat beneath the skin. Other heat conservation measures include huddling together, recovering heat from exhaled breath, and recovering heat from the extremities of the body. Endotherms can remain active in the cold if they can find enough food, or they can reduce their metabolism and lie dormant until warmer conditions return. Although air temperatures may be low, the temperature beneath an insulating layer of snow, under the ground, or at the bottom of a lake may remain above 0°C. Most organisms, however, can neither generate their own heat nor avoid the freezing temperatures, and for them, the choice is to survive ice formation within their bodies or to prevent their bodies from freezing.

Organisms run the risk of freezing at temperatures that are below the melting point of their body and cell fluids. There are two main responses: either they can survive ice forming within them (they are freezing tolerant) or they have mechanisms that ensure that their fluids remain liquid at temperatures that are below the freezing point of water and the melting point of their body fluids (they are freeze avoiding). The strategy that an organism uses depends on the structure and physiology it has developed during its evolutionary history and on the particular demands of its environment. If the organism is living in a wet or damp environment, ice is likely to make contact with its surface when its surroundings freeze. This may cause its body fluids to freeze by the ice traveling across the cell or body wall, or through body openings—a process known as inoculative freezing. Most organisms surviving low temperatures in such environments are thus likely

to be freezing tolerant, since inoculative freezing will cause their bodies to freeze. Some, however, may have a structure such as a cuticle, eggshell, cocoon, or sheath that allows them to prevent inoculative freezing by acting as a barrier to the spread of ice into their bodies. This allows them to maintain their body or cell fluids as liquids, despite the fact that their surfaces are in contact with external ice, and enables them to avoid inoculative freezing. In a situation where the organism is likely to be exposed to subzero temperatures with little or no water in contact with its surface (many terrestrial insects, for example), it does not have the problem of inoculative freezing and it is perhaps easier for it to maintain its body fluids in a liquid state at low temperatures and thus survive by avoiding freezing.

The two strategies of cold survival are, however, not always mutually exclusive. There have been a few reports of insects that were apparently freezing tolerant switching to being freeze avoiding. The overwintering larvae of a beetle from northern Indiana, when studied in the winters of 1977-1979, froze at -8°C to -12°C but survived down to -28°C. When examined again in 1982, however, they froze and were killed at -26°C, apparently switching from a freezing-tolerant to a freeze-avoiding strategy during the intervening years. There are adaptations in common between freeze-avoiding and freezing-tolerant insects that may make it easy to switch between the two strategies. It must be said, however, there has been only one other report of an insect, another beetle, displaying a shift in strategy of this sort. One Antarctic nematode is freezing tolerant when immersed in water but, when free of surface water, there is, of course, no inoculative freezing and it can survive by avoiding freezing. The cold-tolerance strategy displayed thus depends on the particular characteristics of the animal's microenvironment.

60 – 3 Earth's Core

Knowledge of Earth's deep interior is derived from the study of the waves produced by earthquakes, called seismic waves. Among the various kinds of seismic waves are primary waves (P-waves) and secondary waves (S-waves). Primary and secondary waves pass deep within Earth and therefore are the most instructive. Study of abrupt changes in the characteristics of seismic waves at different depths provides the basis for a threefold division of Earth into a central core; a thick, overlying mantle; and a thin, enveloping crust. Sudden changes in seismic wave velocities and angles of transmission are termed discontinuities.

One of the discontinuities is the Gutenberg discontinuity, which is located nearly halfway to the center of Earth at a depth of 2,900 kilometers and marks the outer boundary of Earth's core. At that depth, the S-waves cannot propagate, while at the same time P-wave velocity is drastically reduced. S-waves are unable to travel through fluids. Thus, if S-waves were to encounter a fluidlike region of Earth's interior, they would be absorbed there and would not be able to continue. Geophysicists believe this is what happens to S-waves as they enter the outer core. As a result, the S-waves generated on one side of Earth fail to appear at seismograph stations on the opposite side of Earth, and this observation is the principal evidence of an outer core that behaves as a fluid. Unlike S-waves, P-waves are able to pass through liquids. They are, however, abruptly slowed and sharply refracted (bent) as they enter a fluid medium. Therefore, as P-waves encounter the molten outer core of Earth, their velocity is reduced and they are refracted downward.

The radius of the core is about 3,500 kilometers. The inner core is solid and has a radius of about 1,220 kilometers, which makes this inner core slightly larger than the Moon. Most geologists believe that the inner core has the same composition as the outer core and that it can only exist as a solid because of the enormous pressure at the center of Earth. Evidence of the existence of a solid inner core is derived from the study of hundreds of records of seismic waves produced over several years. These studies showed that the inner core behaves seismically as if it were a solid.

Earth has an overall density of 5.5 grams per cubic centimeter, yet the average density of rocks at the surface is less than 3.0 grams per cubic centimeter.

This difference indicates that materials of high density must exist in the deep interior of the planet to achieve the 5.5 grams per cubic centimeter overall density. Under the extreme pressure conditions that exist in the region of the core, iron mixed with nickel would very likely have the required high density. Laboratory experiments, however, suggest that a highly pressurized iron-nickel alloy might be too dense and that minor amounts of such elements such as silicon, sulfur, carbon, or oxygen may also be present to lighten the core material.

Support for the theory that the core is composed of iron (85 percent) with lesser amounts of nickel has come from the study of meteorites. Many of these samples of solar system materials are iron meteorites that consist of metallic iron alloyed with a small percentage of nickel. Some geologists suspect that iron meteorites may be fragments from the core of a shattered planet. The presence of iron meteorites in our solar system suggests that the existence of an iron-nickel core for Earth is plausible.

There is further evidence that Earth may have a metallic core. Anyone who understands the functioning of a compass is aware that Earth has a magnetic field. The planet itself behaves as if there was a great bar magnet embedded within it. A magnetic field is developed by the flow of electric charges and requires good electrical conductors. Silicate rocks, such as those in the mantle and crust, do not conduct electricity very well, whereas metals such as iron and nickel are good conductors. Heat-driven movements in the outer core, coupled with movements induced by Earth's spin, are thought to provide the necessary flow of electrons (very small particles that carry a negative charge) around the inner core that produces the magnetic field. Without a metallic core, Earth's magnetic field would not be possible.

61 – 2 Conditions on Earth When Life Began

In the 1920s, Aleksandr Oparin, a Russian biochemist, proposed and developed the idea that life originated in the warm, watery environment of early Earth's surface, under an atmosphere mostly composed of methane. The early seas were believed by Oparin to be rich in simple organic molecules, which reacted to form more complex molecules, eventually leading to proteins and life. Then, almost 30 years after Oparin published his ideas, Stanley Miller demonstrated that amino acids, the building blocks of the proteins necessary for life, could form under conditions thought to prevail on early Earth. Miller's experiment was elegant. He passed electric discharges through a mixture of methane, hydrogen, ammonia, and steam, and when he analyzed the results, found that he had made amino acids. The discharges were a proxy for lightning, the gas mixture an educated guess about what the early atmosphere may have been like. Amino acids cannot replicate themselves, and are not themselves alive. Nevertheless, this experiment has long been recognized as a landmark for understanding a process that must have been one of the important steps in the evolution of life on Earth, the natural synthesis of amino acids. However, it now seems likely that Miller's experiments may not be directly applicable to the events of the early Archean (that is, early in the geologic eon that lasted from Earth's formation until about 2.5 billion years ago).

One of the problems hindering understanding of the origin of life is that environmental conditions on early Earth are not known with any certainty. It is possible to make only reasoned estimates. For example, for some fairly long period of time after formation, perhaps as much as several hundred million years, the surface must have been much hotter than it is today. Continued impacts of meteorites, large and small, would have added further heat energy, and in the earliest part of Earth history the larger impacting bodies may have broken through the cooling crust to expose underlying molten material. Large quantities of volcanic gases would have been released into the atmosphere as lavas erupted onto the surface, producing a greenhouse effect far more severe than anything likely to result from human activity. It is quite possible that the early atmosphere was many times as dense as today's, and that the seas and oceans were hot. Some have even suggested that because of the high atmospheric pressure, the

oceans could have been hotter than the boiling point of water today. However, life as we know it is quite sensitive to temperature, and no modern organisms are known to survive much above 100°C. It is unlikely that life became established until surface temperatures had decreased to this level, or lower.

Although we do not know the precise composition of the early atmosphere, there has been enough progress made on this subject in recent years that it is possible to say with some certainty that the methane-rich composition envisioned by Oparin, and the methane-ammonia-hydrogen mixture used by Miller in his experiments, are probably not very realistic. Based on studies of our closest neighbor planets, Mars and Venus, and also considering evidence from Earth's sedimentary rocks, it seems probable that Earth's early atmosphere was rich in carbon dioxide rather than methane. On both Mars and Venus, carbon dioxide is by far the most abundant gas in the atmosphere. On Earth it is a minor constituent. But there is an enormous amount of this compound buried in the sedimentary rocks of Earth's crust, enough so that, if it were all released, our atmosphere would be much more like those of our neighboring planets. How did carbon dioxide gas end up in the crust? The answer lies in what geologists refer to as the carbon cycle. Through a series of chemical reactions, carbon dioxide from the atmosphere finds itself, in dissolved form, in the oceans. In seawater it combines with calcium to precipitate as calcium carbonate, the main constituent of limestone. Over geologic time so much carbon dioxide from the atmosphere has been converted to limestone in this fashion that there is more than 100,000 times as much stored as limestone as there is in the atmosphere.

60 – 1 Underground Life

Until about the late 1980s, most scientists believed that life was restricted to the top few meters of the soil or ocean sediments. The few reports of organisms being recovered from great depths within Earth were dismissed as contamination with material from the surface layers. Two technical developments changed this view. The first was the development of drilling techniques that gave confidence that samples could be retrieved from depth without contamination. Samples were recovered using a diamond-studded drill bit that headed a great length of rotating steel pipe from a drilling derrick. A concentrated tracer material was added to the lubricating fluid so that when a deep sample of rock was removed, any contaminated material could be identified and cut away to leave a pristine sample of rock from deep within Earth. The second development was the advent of techniques for identifying microorganisms without having to grow them in culture. All organisms contain DNA, and their presence can be revealed by dyes that either stain DNA directly or can be attached to nucleic acid probes. By varying the nucleic acid probe, scientists can demonstrate the presence of different types of microorganisms.

The first scientists to use these techniques were involved in the Subsurface Science Program of the United States Department of Energy (DOE). They were interested in the possibility that if organisms existed in the depths of Earth, they might degrade organic pollutants and help maintain the purity of groundwater or, rather less usefully, degrade the containers in which the DOE was proposing to deposit the radioactive waste from nuclear facilities. They demonstrated the presence of many different types of microorganisms in rocks at depths down to 500 meters beneath the surface. Since then, microbes have been discovered in many different types of rocks and deep within ocean sediments. The record depth at which life has been found is at the bottom of a South African gold mine, 3.5 kilometers below ground. Pressure and temperature increase as you go deeper into Earth. Some scientists think that subsurface bacteria could withstand temperatures as high as 150°C. This would allow organisms to exist to depths of about 7 kilometers beneath the seafloor and to 4 kilometers below the surface of the land. Although the organisms are often sparsely distributed, this is such an enormous volume that it has been estimated that the total biomass of deep

subsurface organisms exceeds that of those living on, or just below, the surface.

Bacteria are the most numerous of these subsurface organisms, but there are also fungi and protozoa. Some 10,000 strains of microorganism have been isolated from subsurface cores. Each gram of rock contains anything from 100 bacteria to 10 million bacteria (compared with more than 1 billion per gram in agricultural soils); ocean sediments contain even higher numbers. The protozoa feed on the bacteria, forming part of a simple subterranean food chain, but what do the bacteria feed on? Sedimentary rocks are formed from sands and from ocean, river, or lake sediments that have organic material trapped within them. Microbes living in pores within the sediments can utilize these ancient nutrients and grow. As sedimentary rocks are buried more deeply, they become increasingly compacted and their pores filled with minerals. The distribution of microorganisms is thus likely to become more patchy, condensed into the remaining pores and concentrations of nutrients. The bulk of Earth's crust, however, consists of igneous rocks, such as granite and basalt, which are solidified from molten magma. These rocks were too hot to support life when they were first formed; the organisms that inhabit cracks and fissures within the rocks are carried there by the groundwater flowing through them. Subsurface bacteria do not just rely on nutrients trapped within the rock or carried there by groundwater. Some are chemotrophs, deriving their energy from the oxidation of iron or sulfur compounds and building organic material directly from the carbon dioxide and hydrogen gas dissolved in the rock. These bacteria excrete organic compounds that are then utilized by other types of bacteria. These ecosystems based on chemotrophic bacteria are completely independent of material and solar energy from the surface.

61 – 1 Physical Properties of Minerals

A mineral is a naturally occurring solid formed by inorganic processes. Since the internal structure and chemical composition of a mineral are difficult to determine without the aid of sophisticated tests and apparatus, the more easily recognized physical properties are used in identification.

Most people think of a crystal as a rare commodity, when in fact most inorganic solid objects are composed of crystals. The reason for this misconception is that most crystals do not exhibit their crystal form: the external form of a mineral that reflects the orderly internal arrangement of its atoms. Whenever a mineral forms without space restrictions, individual crystals with well-formed crystal faces will develop. Some crystals, such as those of the mineral quartz, have a very distinctive crystal form that can be helpful in identification. However, most of the time, crystal growth is interrupted because of competition for space, resulting in an intergrown mass of crystals, none of which exhibits crystal form.

Although color is an obvious feature of a mineral, it is often an unreliable diagnostic property. Slight impurities in the common mineral quartz, for example, give it a variety of colors, including pink, purple (amethyst), white, and even black. When a mineral, such as quartz, exhibits a variety of colors, it is said to possess exotic coloration. Exotic coloration is usually caused by the inclusion of impurities, such as foreign ions, in the crystalline structure. Other minerals—for example, sulfur, which is yellow, and malachite, which is bright green—are said to have inherent coloration because their color is a consequence of their chemical makeup and does not vary significantly.

Streak is the color of a mineral in its powdered form and is obtained by rubbing a mineral across a plate of unglazed porcelain. Whereas the color of a mineral often varies from sample to sample, the streak usually does not and is therefore the more reliable property.

Luster is the appearance or quality of light reflected from the surface of a mineral. Minerals that have the appearance of metals, regardless of color, are said to have a metallic luster. Minerals with a nonmetallic luster are described by various adjectives, including vitreous (glassy), pearly, silky, resinous, and earthy (dull).

One of the most useful diagnostic properties of a mineral is hardness, the resistance of a mineral to abrasion or scratching. This property is determined by rubbing a mineral of unknown hardness against one of known hardness, or vice versa. A numerical value can be obtained by using Mohs' scale of hardness, which consists of ten minerals arranged in order from talc, the softest, at number one, to diamond, the hardest, at number ten. Any mineral of unknown hardness can be compared with these or with other objects of known hardness. For example, a fingernail has a hardness of 2.5, a copper penny 5, and a piece of glass 5.5. The mineral gypsum, which has a hardness of two, can be easily scratched with your fingernail. On the other hand, the mineral calcite, which has a hardness of three, will scratch your fingernail but will not scratch glass. Quartz, the hardest of the common minerals, will scratch a glass plate.

The tendency of a mineral to break along planes of weak bonding is called cleavage. Minerals that possess cleavage are identified by the smooth, flat surfaces produced when the mineral is broken. The simplest type of cleavage is exhibited by the micas. Because the micas have excellent cleavage in one direction, they break to form thin, flat sheets. Some minerals have several cleavage planes, which produce smooth surfaces when broken, while others exhibit poor cleavage, and still others exhibit no cleavage at all. When minerals break evenly in more than one direction, cleavage is described by the number of planes exhibited and the angles at which they meet. Cleavage should not be confused with crystal form. When a mineral exhibits cleavage, it will break into pieces that have the same configuration as the original sample does. By contrast, quartz crystals do not have cleavage, and if broken, would shatter into shapes that do not resemble each other or the original crystals. Minerals that do not exhibit cleavage are said to fracture when broken. Some break into pieces with smooth curved surfaces resembling broken glass. Others break into splinters or fibers, but most fracture irregularly.

63 – 2 Structure and Composition of Comets

Astronomers now have a fairly good idea of what a comet really is. When it is far from the Sun, it is a very small object only a few kilometers across. It consists mainly of ices (water, methane, ammonia) with bits of dust embedded in it—a kind of dirty ice ball. As it approaches the Sun, radiation from the Sun vaporizes the icy matter and releases some of the dust. This forms a gigantic halo around the ice ball. This halo—called the coma—extends out tens of thousands of kilometers from the icy core, which is the nucleus of the comet. Sunlight reflected off the dust particles makes the coma visible to observers on Earth. Ultraviolet radiation from the Sun breaks down the vapor molecules into their constituents. These components can be excited by absorbing radiation from the Sun. In returning to lower-energy states, the excited atoms and ions emit light, contributing to the luminosity of the coma.

When the comet gets even closer to the Sun, one of its most spectacular parts begins to form—the tail. Actually, there are two kinds of tails: the dust tail and the ion tail. The dust tail is produced by the light from the Sun reflecting off the dust particles in the coma. A photon carries momentum. In bouncing off a dust particle, it imparts a tiny, but perceptible, momentum change to the dust particle, driving it away from the coma. As the comet sweeps along its orbit, it leaves a curving trail of dust behind in its path. This visible dust tail can extend for tens or hundreds of millions of kilometers out from the nucleus. The dust tail is characterized by its gently curving shape and its yellowish color.

A different mechanism is responsible for the ion tail. Near the Sun, ultraviolet radiation from the Sun (solar wind) ionizes and excites the atoms in the coma. As the solar wind sweeps through the coma, the high-velocity charged particles of the solar wind interact with the electrically charged excited ions in the coma, driving them away from the head of the comet. In returning to lower-energy states, these excited ions emit photons and form a luminous, bluish-colored tail extending out from the comet directly away from the Sun. Since both kinds of tails are produced by radiation streaming out from the Sun, they extend out from the coma in the general direction away from the Sun. A comet may exhibit several tails of each kind.

Although the nucleus is of the order of a few kilometers in size, the diameter of the coma may be

tens or hundreds of thousands of kilometers, the tails typically extend out tens or hundreds of millions of kilometers away from the coma.

A comet leaves a trail of matter behind it as it moves through the inner solar system. Some of this debris may get strewn across Earth's orbit around the Sun. When Earth passes through this part of its annual path, it sweeps through the dust trail. The particles enter Earth's atmosphere at high velocity. The air friction can cause one of these bits of matter to produce a brief streak of light as it burns up in the atmosphere.

Since a comet loses matter on each pass by the Sun, eventually it will be depleted to the point where it is no longer visible. Comets that approach the Sun have finite lifetimes. Given the typical sizes of comets and the typical rates at which they lose matter, astronomers have concluded that the lifetimes of comets with orbits that bring them near enough to the Sun to be seen from Earth are very much shorter than the age of the solar system. Where do the new comets come from to replace the old ones that dissipate and vanish from view?

Dutch astronomer Jan Oort proposed that a giant cloud of matter left over from the formation of the solar system surrounds the Sun and extends out to about 50,000 astronomical units. This cloud contains large chunks of matter like the nuclei of comets. The gravitational influence of a passing star can be sufficient to perturb the orbit of one of these chunks to send it toward the inner solar system and bring it near the Sun.

66 – 2 Visions of the Land

Successive generations of North Americans have viewed their continent's natural environment in different ways. From the vantage point of the present it is clear that perceptions of the land have changed dramatically from the first years of settlement to the Civil War. Not only have such visions often shifted, but also different peoples have used their particular perspective to reshape the land itself and make it fit their own sense of what nature should be. If the consequences of some changes, such as cutting forests and filling in lowlands, have been deliberate and purposeful-to open the landscape and create sweeping vistas, for example-other human undertakings, such as mining and dam building, have brought results neither anticipated nor intended. Native peoples, no less than the first colonists and subsequent immigrants to North America, have reshaped the natural environment to meet their physical wants and spiritual needs. Indeed, much of the landscape we know today reflects patterns of use and abuse that began several centuries ago.

Long before the first European settlers reached the continent's eastern shores, native peoples had developed agricultural practices that had changed the face of the land. By cutting away the bark to kill trees selectively, Indians in the Virginia tidewater (low, coastal land) and much of the Northeast had cleared space to plant small gardens of corn, squash, beans, and melons. Although the first English immigrants described the countryside as almost entirely wooded, the forests provided canopies of large, well-spaced trees under which a horse and rider could pass unhindered. By frequently moving their garden plots to find more-fertile soil and by periodically burning the undergrowth, Indians had further opened the land, in this way facilitating their hunting of deer and other game. Native American visions of the landscape not surprisingly featured people living in harmony with nature, whose riches they celebrated in seasonal rituals and through time-honored practices.

In contrast, the European colonists who intruded on this harmonious world often viewed it as alien and menacing; some called it, in the language of the Bible, a "howling wilderness." The newcomers to America brought with them agricultural practices and preconceptions about nature based on their experiences in England. They saw uncultivated lands as "wastes" that needed to be "broken,"

"dressed." and "improved" . In New England, transplanted English settlers attempted to subdue what they considered a fearsome wilderness by mapping the countryside, draining marshlands, clearing pastures, fencing particular parcels, and planting wheat and other familiar crops. Within twenty years of the initial Puritan settlement, Edward Johnson boasted of the newcomers' achievements: "This remote, rocky, barren, bushy, wild-woody wilderness, a receptacle for lions, wolves, bears, foxes, racoons, beavers, otters, and all kind of wild creatures, a place that never afforded the Natives better than the flesh of a few wild creatures and parched Indian corn inched out with chestnuts and bitter acorns, now through the mercy of Christ [has] become a second England for fertility in so short a space, that it is indeed the wonder of the world".

So, rather than adapting to their new land, the English either changed it by cutting trees, building farms, and plowing or searched for soil and landscape features that reminded them of the English countryside. Seeking to tame the land and to conquer their fear of it, generations of seventeenth- and eighteenth-century settlers nevertheless failed to gain the mastery they desired. In part, this failure resulted from their custom of settling along waterways. To expedite travel and facilitate the shipment of agricultural produce, newcomers invariably built their homes along rivers. Despite the colonists' attempts to control waterways through dams, rivers never failed to remind them of nature's unpredictability and power. Rivers could and often did, change course abruptly or flood during sudden rainstorms. Not until the middle of the eighteenth century did the colonists begin to discard their negative view of the landscape as a wilderness to be feared and controlled, and to substitute the idea that nature could be as much useful as fearsome.

66 – 4 Geology and Landscape

Most people consider the landscape to be unchanging, but Earth is a dynamic body, and its surface is continually altering—slowly on the human time scale, but relatively rapidly when compared to the great age of Earth (about 4,500 billion years). There are two principal influences that shape the terrain: constructive processes such as uplift, which create new landscape features, and destructive forces such as erosion, which gradually wear away exposed landforms.

Hills and mountains are often regarded as the epitome of permanence, successfully resisting the destructive forces of nature, but in fact they tend to be relatively short-lived in geological terms. As a general rule, the higher a mountain is, the more recently it was formed; for example, the high mountains of the Himalayas are only about 50 million years old. Lower mountains tend to be older, and are often the eroded relics of much higher mountain chains. About 400 million years ago, when the present-day continents of North America and Europe were joined, the Caledonian mountain chain was the same size as the modern Himalayas. Today, however, the relics of the Caledonian orogeny (mountain-building period) exist as the comparatively low mountains of Greenland, the northern Appalachians in the United States, the Scottish Highlands, and the Norwegian coastal plateau.

The Earth's crust is thought to be divided into huge, movable segments, called plates, which float on a soft plastic layer of rock. Some mountains were formed as a result of these plates crashing into each other and forcing up the rock at the plate margins. In this process, sedimentary rocks that originally formed on the seabed may be folded upwards to altitudes of more than 26,000 feet. Other mountains may be raised by earthquakes, which fracture the Earth's crust and can displace enough rock to produce block mountains. A third type of mountain may be formed as a result of volcanic activity which occurs in regions of active fold mountain belts, such as in the Cascade Range of western North America. The Cascades are made up of lavas and volcanic materials. Many of the peaks are extinct volcanoes.

Whatever the reason for mountain formation, as soon as land rises above sea level it is subjected to destructive forces. The exposed rocks are attacked by the various weather processes and gradually

broken down into fragments, which are then carried away and later deposited as sediments. Thus, any landscape represents only a temporary stage in the continuous battle between the forces of uplift and those of erosion.

The weather, in its many forms, is the main agent of erosion. Rain washes away loose soil and penetrates cracks in the rocks. Carbon dioxide in the air reacts with the rainwater, forming a weak acid (carbonic acid) that may chemically attack the rocks. The rain seeps underground and the water may reappear later as springs. These springs are the sources of streams and rivers, which cut through the rocks and carry away debris from the mountains to the lowlands.

Under very cold conditions, rocks can be shattered by ice and frost. Glaciers may form in permanently cold areas, and these slowly moving masses of ice cut out valleys, carrying with them huge quantities of eroded rock debris. In dry areas the wind is the principal agent of erosion. It carries fine particles of sand, which bombard exposed rock surfaces, thereby wearing them into yet more sand. Even living things contribute to the formation of landscapes. Tree roots force their way into cracks in rocks and, in so doing, speed their splitting. In contrast, the roots of grasses and other small plants may help to hold loose soil fragments together, thereby helping to prevent erosion by the wind.

65 – 2 Early Research on Air

In the field of chemistry, the understanding of the word “air” has undergone radical change. Air for John Mayow, a seventeenth-century chemist, was essentially a receptacle for airborne particles, and through them manifested a variety of chemical properties. But although Mayow and a few other chemists did detect specific chemical properties in what we call gases (including our carbon dioxide), most chemists left them unaccounted for until the beginning of the eighteenth century. As chemists became aware that the atmosphere itself (and not just particles within it) had a role to play in combustion, respiration, and other reactions, they did not attribute this to the chemical properties of air but rather to substances that air could absorb and release according to circumstances. Thus, air provided a physical environment in which some reactions took place.

In the early 1700s, the air was widely seen as just such an environment, and “air” and “the air” were one and the same thing. Chemists were not in the habit of regarding airs or gases as having different chemical properties. There was simply air. One obvious reason for this was practical. Chemists could examine solids and liquids, exposing them to a variety of tests and seeing how they contributed to assorted reactions. Chemists had, however, no comparable way of examining air, and they came to view chemistry as the sum total of the reactions of solids and liquids, excluding gases. Chemists stressed chemical qualities over physical properties like weight and let physicists deal with air. Chemists generally did not examine air, and they did not try to weigh it. That does not mean that chemists did not weigh substances. They did a lot of weighing, and pharmacists and metallurgists did more. But weighing gases was outside their brief. In the *Encyclopedia of Diderot and d'Alembert*, published between 1751 and 1775, readers were told that “the incoercibility of gases will remove them from our researches for a long time to come.”

By the time of the *Encyclopedia*, however, this had begun to change. One of the first and key sources of change was the invention by the Reverend Stephen Hales of a new instrument, the pneumatic trough. This instrument is important for what it made possible in the handling of air. The history of its invention and early use illustrates the difference there may be between the motives for inventing a device and the ways in which that device is used.

Hales was a botanist and chemist as well as a physiologist. He wrote a book in 1727 investigating mechanical subjects like the pressure of sap in plants. But Hales went further, addressing chemical as well as physiological questions. He urged chemists to consider air chemically. He described an instrument for washing the air produced in the course of a chemical reaction. He wanted to get rid of impurities in the air by letting it pass through water. Air passed from a reaction vessel through water in a trough (or tube) and then into a second vessel that was partly filled with water and that could capture air.

In devising this apparatus, Hales had coincidentally furnished an instrument for catching and holding air, which could then be subjected to various tests. Used in this way the apparatus became known as the pneumatic trough half a century after its invention, it became a staple of the chemical laboratory. It also became one of the key instruments in the reform of chemistry that we know as the chemical revolution because it was essential to incorporating a whole new state of matter, the gaseous state, into chemistry, alongside the already studied solid and liquid states. Once that step had been taken, it was possible to speculate and then to demonstrate that the gaseous state, like the solid and liquid states, could contain a variety of chemical substances. This was an enormous step, and it did not happen overnight. Hales had shown that air could be contained, washed, and purified, and tested chemically as well as physically. This, however, did not lead him to think that there was more than one kind of air. Air for him remained air, not one of a number of airs. Other chemists would take that essential step.

69 – 2 Pacific Ecosystems

The Pacific Ocean accounts for one-third of Earth's surface and half of the world's ocean area. It has about 25,000 islands, of which about 7,500 are oceanic, being relatively far from a continental shore. The great majority of all Pacific islands were born barren of life; hard, dense, volcanic rock pimples on the sea's surface. New Zealand is the chief exception. It is among the continental islands of the western Pacific, together with Fiji, the Solomons, and others to the west.

Life arrived on most other islands by accident or by drift. Some plants arrived by air transport; seeds carried in the digestive tracts of birds account for nearly 40 percent of Hawaii's early plants. The first invaders were either creatures that could float well enough, in air or water to cross stretches of ocean, or those whose seeds could survive a voyage in some bird's gut. At times of lower sea level, land bridges linked, or nearly linked, many islands in the far western Pacific, so some species colonized these islands without being notably good floaters or stowaways. In the eastern Pacific (Easter Island, for example) only the best floaters and travelers arrived and survived. Consequently, the western islands have far more species and far greater biodiversity than do the eastern islands of Polynesia. Mammals found it hard to get anywhere in the island Pacific, only bats and rats successfully colonized east of New Guinea. Almost all species derive from Asia, the early Pacific was an Asian lake, with only a tiny proportion of species from the Americas. As a rule of thumb, the further from Indonesia, the more impoverished the plant and animal life and, in consequence, the less stable and resilient in the face of disturbance. This attenuation is strong for land species, less strong for marine species, and nonexistent for oceanic birds although fairly strong for land birds.

Pacific ecosystems evolved in relative (but differential) isolation from the continental crucibles of biological evolution. This meant opportunities for speciation – the development of new species occupying ecological niches elsewhere – were already filled. The finches described by the naturalist Charles Darwin on the Galapagos Islands – birds that divided into different species, each specialized for a narrow niche – are the classic example. On islands that had no mammals, reptiles and birds took their place. Thus the Galapagos have giant tortoises, and New Zealand once had giant birds that functioned more or less like

browsing or grazing mammals. Throughout most of the Pacific, the paucity of grazing animals meant that plants developed no defenses, such as spines, poisonous chemicals, or bitterness. The remote islands had a very high proportion of endemism – that is, of species that existed only there. In the case of Hawaii, as many as 99 percent of the species were endemic. All this led to a certain biological vulnerability among the terrestrial island species, should they ever be obliged to compete for niche space with the winners of the more intense continental competitions for survival. This vulnerability increased toward the east and toward the remote corners of the Pacific along a gradient defined chiefly by the degree of isolation.

A second source of vulnerability, perhaps more decisive arose from the arrival of humankind in the Pacific. Island animals evolved with no experience of the ways of humankind or indeed of any large terrestrial predators. As a result, they had to immunities to the predators or the effects of human action. Pacific animals were often unwary and easy prey. At the extreme again the Galapagos, Darwin found many birds almost tame, so naively trusting that they would allow him to get within arm's reach. Pacific plants had little experience with fire, because natural fires were very rare: except in a few places. Thus few plants were well adapted to fire and most proved vulnerable to it. In contrast, continental species that had evolved in the presence of humankind or in places where natural fire is much more common, could recover easily after burns, and some could flourish as a result of fire. In short, Pacific ecosystems were different from continental ones on account of their isolation. They were well adapted to their circumstances but very vulnerable to alien invasion and human impact.

68 – 2 Predicting Volcanic Eruptions

Volcanoes are the landforms created when molten rock, or magma, escapes from vents in the Earth's surface and then solidifies around these vents. In any given year, roughly 50 of Earth's active volcanoes erupt - usually with some warning. Before they blow, they typically shake, swell, warm up and belch a variety of gases. Scientists from the Volcano Disaster Assistance Program (VDAP)-part of the US. Geological Survey and based at the US. Geological Survey's Cascades Volcano Observatory in Vancouver, Washington -are always on call, ready when summoned to rush at a moment's notice to an awakening volcano anywhere in the world; armed with the latest in lasers, seismometers, and other monitoring devices, they can assess the volcano's potential for violence and predict when it might ignite.

Geologists have enjoyed fair success in predicting individual eruptive episodes when they concentrate on a specific volcano after an eruptive phase has begun. These monitoring efforts involve carefully measuring changes in a volcano's surface temperature, watching for the slightest expansion in its slope, and keeping track of regional earthquake activity. A laboratory at the University of Washington in Seattle is staffed 24 hours a day to monitor the rumblings of Mount St. Helens. Even with the advances brought by today's technology, however, the art of volcano prediction has not been fully mastered. The US. Geological Survey missed the call on Mount St. Helens' 1980 blast despite the fact that the mountain was being watched closely by a large team of scientists armed with the latest in prediction technology. It did successfully predict the eruption of Mount Pinatubo in the Philippines evacuating virtually everyone within 25 kilometers (15 miles) before the volcano's powerful blast on May 17, 1991.

Before a volcano erupts, hot magma rises toward the surface, so any local manifestation of increasing heat may signal an impending event. Ongoing surveys can identify new surface hot springs and take the temperature of the water and steam in existing ones. If the escaping steam isn't much hotter than the boiling point of water, then surface water is probably seeping into the mountain and being heated by contact with hot subsurface rocks, and all is well for the time being. If the steam is superheated, with temperatures as high as 500°C(900°F), then it probably derives from

shallow water-rich magma, a sign that an eruption may be brewing. As magma rises, the volcanic cone itself begins to heat up. The overall temperature of a volcanic cone can be monitored from an orbiting satellite equipped with infrared heat sensors to detect the slightest change in surface temperature. This high-altitude technology serves as a simultaneous early-warning system for most of Earth's 600 or so active volcanoes. Impending eruptions may also be predicted by increased gas emissions from rising magmas. For this reason, volcanologists continuously monitor sulfur dioxide and carbon dioxide emissions from potentially active volcanoes.

Active volcanoes expand in volume as they acquire new supplies of magma from below. As a result, an increase in the steepness or bulging of a volcano's slope may signal an impending eruption. To detect the inflation of a volcanic cone, a tilt meter, a device like a carpenter's level, is used. As magma rises, it pushes aside fractured rock, enlarging the fractures as it moves. Because this type of fracturing causes earthquakes, eruptions are often preceded by a distinctive pattern of earthquake activity called harmonic tremors, a continuous rhythmic rumbling. Sensitive equipment that monitors the location where these tremors occur can measure the increased height of rising magma. The rate at which the magma rises provides an estimate of when an eruption may occur. Indeed, it served as the principal means by which scientists accurately predicted recent eruptions of Mount St. Helens.

Efforts to predict eruptions are thwarted, however, when we are unaware of a site's volcanic potential. Occasionally, a new volcano appears suddenly and rather unexpectedly, as was the case in 1943, when the volcano Parícutin developed literally overnight in the Mexican state of Michoacán, 320 kilometers (200 miles) west of Mexico City. The surrounding area was known to be volcanic because of its geologic zone, but it was not possible to predict that the volcano would appear at this particular site. Our ability to predict volcanic eruptions continues to improve but is not yet as accurate as we need it to be.

70 – 2 Nineteenth-Century Theories of Mountain Formation

One of the central scientific questions of nineteenth-century geology was the origin of mountains. How were they formed? What process squeezed and folded rocks like bread dough? What made Earth's surface move? Most theories invoked terrestrial contraction as a causal force. It was widely believed that Earth had formed as a hot, incandescent body and had been steadily cooling since the beginning of geological time. Because most materials contract as they cool, it seemed logical to assume that Earth had been contracting as it cooled, too. As it did, its surface would have deformed, producing mountains.

In Europe, Austrian geologist Eduard Suess (1831 – 1914) popularized the image of Earth as a drying apple: as the planet contracted, its surface wrinkled to accommodate the diminished surface area. Suess assumed that Earth's initial crust was continuous but broke apart as the interior shrank. The collapsed portions formed the ocean basins, the remaining elevated portions formed the continents. With continued cooling, the original continents became unstable and collapsed to form the next generation of ocean floor, and what had formerly been ocean now became dry land. Over the course of geological history, there would be a continual interchange of land and sea, a periodic rearrangement of the landmasses.

The interchangeability of continents and oceans explained a number of other perplexing geological observations, such as the presence of marine fossils on land (which had long before puzzled Leonardo da Vinci) and the extensive interleaving of marine and terrestrial sediments in the stratigraphic record. Suess's theory also explained the striking similarities of fossils in parts of Africa and South America. Indeed, in some cases the fossils seemed to be identical, even though they were found thousands of miles apart. These similarities had been recognized since the mid-nineteenth century, but they had been made newly problematic by Darwin's theory of evolution. If plants and animals had evolved independently in different places within diverse environments, then why did they look so similar? Suess explained this conundrum by attributing these similar species to an early geological age when the continents were contiguous in an ancient supercontinent called Gondwanaland.

Suess's theory was widely discussed and to varying degrees accepted in Europe, but in North America geologist James Dwight Dana (1813 – 1895) had developed a different version of contraction theory. Dana suggested that the continents had formed early in Earth history, when low-temperature minerals such as quartz and feldspar had solidified. Then the globe continued to cool and contract, until the high-temperature minerals such as olivine and pyroxene finally solidified – on the Moon, to form the lunar craters; on Earth, to form the ocean basins. As contraction continued after Earth was solid, its surface began to deform. The boundaries between continents and oceans were most affected by the pressure, and so mountains began to form along continental margins. With continued contraction came continued deformation, but with the continents and oceans always in the same relative positions. Although Dana's theory was a version of contraction, it came to be known as permanence theory, because it viewed continents and oceans as globally permanent features.

In North America permanence theory was linked to the theory of subsidence (or sinking) of sedimentary basins along continental margins. This idea was developed primarily by paleontologist James Hall (1811 – 1898), who noted that beneath the forest cover, the Appalachian Mountains of North America were built up of folded layers of shallowwater sedimentary rocks, thousands of feet thick. How did these sequences of shallow-water deposits form? How were they folded and uplifted into mountains? Hall suggested that materials eroded off the continents accumulated in the adjacent marginal basins, causing the basins to subside. 'Subsidence allowed more sediment to accumulate, causing more subsidence, until finally the weight of the pile caused the sediments to be heated, converted to rock, and then uplifted into mountains. 'Dana modified Hall's view by arguing that thick sedimentary piles were not the cause of subsidence but the result of it. Either way, the theory provided a concise explanation of how thick sequences of shallow-water rocks could accumulate, but was vague on the question of how they were transformed into mountain belts.

08 – 3 Running Water on Mars

Photographic evidence suggests that liquid water once existed in great quantity on the surface of Mars. Two types of flow features are seen: runoff channels and outflow channels. Runoff channels are found in the southern highlands. These flow features are extensive systems—sometimes hundreds of kilometers in total length—of interconnecting, twisting channels that seem to merge into larger, wider channels. They bear a strong resemblance to river systems on Earth, and geologists think that they are dried-up beds of long-gone rivers that once carried rainfall on Mars from the mountains down into the valleys. Runoff channels on Mars speak of a time 4 billion years ago (the age of the Martian highlands), when the atmosphere was thicker, the surface warmer, and liquid water widespread.

Outflow channels are probably relics of catastrophic flooding on Mars long ago. They appear only in equatorial regions and generally do not form extensive interconnected networks. Instead, they are probably the paths taken by huge volumes of water draining from the southern highlands into the northern plains. The onrushing water arising from these flash floods likely also formed the odd teardrop-shaped “islands” (resembling the miniature versions seen in the wet sand of our beaches at low tide) that have been found on the plains close to the ends of the outflow channels. Judging from the width and depth of the channels, the flow rates must have been truly enormous—perhaps as much as a hundred times greater than the 105 tons per second carried by the great Amazon River. Flooding shaped the outflow channels approximately 3 billion years ago, about the same times as the northern volcanic plains formed.

Some scientists speculate that Mars may have enjoyed an extended early Period during which rivers, lakes, and perhaps even oceans adorned its surface. A 2003 Mars Global Surveyor image shows what mission specialists think may be a delta—a fan-shaped network of channels and sediments where a river once flowed into a larger body of water, in this case a lake filling a crater in the southern highlands. Other researchers go even further, suggesting that the data provide evidence for large open expanses of water on the early

Martian surface. A computer-generated view of the Martian north polar region shows the extent of what may have been an ancient ocean covering much of the northern lowlands. The Hellas Basin, which measures some 3,000 kilometers across and has a floor that lies nearly 9 kilometers below the basin’s rim, is another candidate for an ancient Martian sea.

These ideas remain controversial. Proponents point to features such as the terraced “beaches” shown in one image, which could conceivably have been left behind as a lake or ocean evaporated and the shoreline receded. But detractors maintain that the terraces could also have been created by geological activity, perhaps related to the geologic forces that depressed the Northern Hemisphere far below the level of the south, in which case they have nothing whatever to do with Martian water. Furthermore, Mars Global Surveyor data released in 2003 seem to indicate that the Martian surface contains too few carbonate rock layers—layers containing compounds of carbon and oxygen—that should have been formed in abundance in an ancient ocean. Their absence supports the picture of a cold, dry Mars that never experienced the extended mild period required to form lakes and oceans. However, more recent data imply that at least some parts of the planet did in fact experience long periods in the past during which liquid water existed on the surface.

Aside from some small-scale gullies (channels) found since 2000, which are inconclusive, astronomers have no direct evidence for liquid water anywhere on the surface of Mars today, and the amount of water vapor in the Martian atmosphere is tiny. Yet even setting aside the unproven hints of ancient oceans, the extent of the outflow channels suggests that a huge total volume of water existed on Mars in the past. Where did all the water go? The answer may be that virtually all the water on Mars is now locked in the permafrost layer under the surface, with more contained in the planet’s polar caps.

25 – 1 The surface of Mars

The surface of Mars shows a wide range of geologic features, including huge volcanoes - the largest known in the solar system - and extensive impact cratering. Three very large volcanoes are found on the Tharsis bulge, an enormous geologic area near Mars's equator. Northwest of Tharsis is the largest volcano of all: Olympus Mons, with a height of 25 kilometers and measuring some 700 kilometers in diameter at its base. The three large volcanoes on the Tharsis bulge are a little smaller - "mere" 18 kilometers high.

None of these volcanoes was formed as a result of collisions between plates of the Martian crust - there is no plate motion on Mars. Instead, they are shield volcanoes - volcanoes with broad, sloping sides formed by molten rock. All four show distinctive lava channels and other flow features similar to those found on shield volcanoes on Earth. Images of the Martian surface reveal many hundreds of volcanoes. Most of the largest volcanoes are associated with the Tharsis bulge, but many smaller ones are found in the northern plains.

The great height of Martian volcanoes is a direct consequence of the planet's low surface gravity. As lava flows and spreads to form a shield volcano, the volcano's eventual height depends on the new mountain's ability to support its own weight. The lower the gravity, the lesser the weight and the greater the height of the mountain. It is no accident that Maxwell Mons on Venus and the Hawaiian shield volcanoes on Earth rise to about the same height (about 10 kilometers) above their respective bases - Earth and Venus have similar surface gravity. Mars's surface gravity is only 40 percent that of Earth, so volcanoes rise roughly 2.5 times as high. Are the Martian shield volcanoes still active? Scientists have no direct evidence for recent or ongoing eruptions, but if these volcanoes were active as recently as 100 million years ago (an estimate of the time of last eruption based on the extent of impact cratering on their slopes), some of them may still be at least intermittently active. Millions of years, though, may pass between eruptions.

Another prominent feature of Mars's surface is cratering. The Mariner spacecraft found that the surface of Mars, as well as that of its two moons, is pitted with impact craters formed by meteoroids falling in from space. As on our Moon, the smaller

craters are often filled with surface matter - mostly dust - confirming that Mars is a dry desert world. However, Martian craters get filled in considerably faster than their lunar counterparts. On the Moon, ancient craters less than 100 meters across (corresponding to depths of about 20 meters) have been obliterated, primarily by meteoritic erosion. On Mars, there are relatively few craters less than 5 kilometers in diameter. The Martian atmosphere is an efficient erosive agent, with Martian winds transporting dust from place to place and erasing surface features much faster than meteoritic impacts alone can obliterate them.

As on the Moon, the extent of large impact cratering (i.e. craters too big to have been filled in by erosion since they were formed) serves as an age indicator for the Martian surface. Age estimates ranging from four billion years for Mars's southern highlands to a few hundred million years in the youngest volcanic areas were obtained in this way.

The detailed appearance of Martian impact craters provides an important piece of information about conditions just below the planet's surface. Martian craters are surrounded by ejecta (debris formed as a result of an impact) that looks quite different from its lunar counterparts. A comparison of the Copernicus crater on the Moon with the (fairly typical) crater Yuty on Mars demonstrates the differences. The ejecta surrounding the lunar crater is just what one would expect from an explosion ejecting a large volume of dust, soil, and boulders. However, the ejecta on Mars gives the distinct impression of a liquid that has splashed or flowed out of crater. Geologists think that this fluidized ejecta crater indicates that a layer of permafrost, or water ice, lies just a few meters under the surface. Explosive impacts heated and liquefied the ice, resulting in the fluid appearance of the ejecta.

72 – 2 The Impact Origin of Lunar Craters

From their discovery in 1609 until recently, lunar craters were almost unanimously thought to have volcanic origins, even though the topography of the Moon and its craters had been studied minutely by generations of observers. Dissenting voices, such as that of Robert Hooke, who postulated in 1665 that they were formed by the impact of meteors, did little to sway the majority opinion. The most serious challenge to the volcanic-formation hypothesis was made in 1893 by the geologist G. K. Gilbert. After analyzing the depth-to-diameter ratios of craters, Gilbert suggested that the craters could have been formed only by the impact of meteors. The central peaks that were commonly observed resulted from rebound of rock, in a similar manner to liquids (think of drops of water hitting a puddle). Bright rays surrounded many craters that were formed from material flung out during the impact---impact ejecta.

Although impacts with a surface can occur at all angles, from less than 5° to 90°, early experiments suggested that only vertical (90°) impacts were capable of producing circular craters. So the main argument against an impact mechanism for lunar crater formation was that only circular craters could be seen on the Moon. However, Gilbert himself showed that the average impact angle was 45° — in fact, very few impacts were close to vertical.

In 1916 E. J. Öpik published work that recognized that the impact of meteors at very high velocities would be fundamentally different from the impact at low velocities created in early laboratory experiments. Such high-velocity impacts would have a similar effect to an explosion and craters would be circular even at low impact angles. Nevertheless, the volcanic hypothesis remained intact, and there was significant unwillingness amongst the scientific community about accepting impact cratering. This began to change in the 1960s with the work of the American geologist Gene Shoemaker, who made a lifetime study of cratering.

Of the 170 known impact craters on the Earth, Meteor Crater in Arizona was one of the first to be recognized as being formed from an impact. In 1906 D. M. Barringer had provided good evidence for impact formation. However, Meteor Crater was not universally accepted as an impact crater until the 1960s. Its impact origin was finally demonstrated in papers published in 1960 and 1963

by Shoemaker. Three lines of evidence show that an impact, and not a volcanic eruption, formed Meteor Crater.

First, as Barringer had recognized, many large fragments from an impact have been found on the desert plains surrounding Meteor Crater. Impact structures that are larger than Meteor Crater do not usually preserve fragments of the original projectiles because they vaporize on impact.

In addition, most volcanic craters are formed by sustained compression of volcanic gases which blast out large volumes of fragments. Volcanic ejecta form simple aprons around craters, the first-erupted materials at the bottom and the last at the top. Shoemaker's field studies showed that a completely different situation exists at Meteor Crater, where the ejecta form an inverted flap flung out and overturned in a single blast. Thus, a drill hole through the ejecta on the rim would pass through layers of the same material twice. In Shoemaker's words, the strata appeared to "have been peeled back from the area of the crater, somewhat like petals of a flower blossoming. Similar structures are observed in laboratory experiments involving hypervelocity impacts.

The final piece of evidence comes from an investigation of the effects of the intense pressures involved in an impact. At Meteor Crater, shattered target rocks are cemented together by glass at depths of 200-400 meters. The glass is formed by melting of the target during the impact. Melting occurs principally as a result of the very high pressures involved during an impact, and because the pressure is applied almost instantaneously, it is referred to as shock pressure. Unusual high pressure minerals are often found in the rocks affected by impact. These minerals are produced by shock pressure during the impact. In 1962 Shoemaker made a case for the impact formation of Copernicus Crater, one of the most prominent craters on the near side of the Moon. Before the decade was out, most scientists had come to accept that lunar craters were formed by impact rather than volcanism.

73 – 3 Stellar Remnants

Eventually, all stars consume their nuclear fuel and collapse into one of three final states—white dwarf, neutron star, or black hole.

White dwarfs are extremely small stars with densities greater than any known terrestrial material. It is believed that white dwarfs began life as low-mass or medium-mass stars. As the star ages, it begins to collapse—become a "dwarf"—because it no longer produces sufficient radiant heat to counteract the contracting force of its gravity. Although some white dwarfs are no larger than Earth, the mass of such a dwarf can equal 1.4 times that of the Sun. A spoonful of such matter would weigh several tons. As a star contracts into a white dwarf, its surface becomes very hot and very bright. But, without a continuing source of energy, it will slowly become cooler and dimmer. Although none have been observed, the terminal stage of a white dwarf must be a small, cold, nonluminous body called a black dwarf.

A study of white dwarfs produced what might at first appear to be a surprising conclusion: the smallest white dwarfs are the most massive, and the largest white dwarfs are the least massive. The explanation for this is that a more massive star has a greater gravitational force, and that greater gravitational force contracts the star into a smaller, more densely packed object than one gets with a less massive star. Thus, the smallest white dwarfs came from the collapse of stars that were larger and more massive than those the large white dwarfs came from.

Extending this idea of larger stars collapsing to smaller stars leads to the prediction that very large stars would collapse to stars even smaller and more massive than white dwarfs. Named neutron stars, these objects are thought to be the remnants of supernova events (the collapse of a large, very massive star). If Earth were to collapse to the density of a neutron star, it would have a diameter equivalent to the length of a football field, and a pea-sized sample of this matter would weigh 100 million tons. This is approximately the density of an atomic nucleus; thus, neutron stars can be thought of as large atomic nuclei.

Although neutron stars have high surface temperatures, their small size would greatly limit their luminosity. Consequently, locating one visually would be extremely difficult. However, theory predicts that a neutron star would have a very

strong magnetic field. Further, as a star collapses, it will rotate faster, for the same reason ice-skaters rotate faster as they pull in their arms. If the Sun were to collapse to the size of a neutron star, it would increase its rate of rotation from once every 25 days to nearly 1,000 times per second. The radio waves generated by such rotating stars would be concentrated into two narrow beams, like the rotating light on a police car or ambulance. If Earth happened to be in the path of this beacon, the star would appear to blink on and off, or pulsate, as the radio waves swept past.

In the early 1970s, a source that radiates short pulses of radio energy, called a pulsar (pulsating radio source), was discovered in the Crab Nebula. Visual inspection of this radio source revealed it to be a small star centered in the nebula. The first neutron star had been discovered. Are neutron stars made of the densest matter possible? No. During a supernova event, remnants of stars greater than three solar masses apparently collapse into objects even smaller and denser than neutron stars. Even though these objects would be very hot, their gravity would be so immense that even light could not escape the surface. Consequently, they would literally disappear from sight. These incredible bodies have appropriately been named "black holes." Anything that moved too near a black hole would be swept in by its irresistible gravity and annihilated.

How can astronomers find an object whose gravitational field prevents the escape of all matter and energy? Theory predicts that as matter is pulled into a black hole, it should become very hot and emit a flood of x-rays before being engulfed. The first black hole x-ray sources were discovered in 1971 by detectors on satellites.

74 – 3 Lunar History

Much is still unknown about the Moon's origin. Until recently, the most widely held hypothesis argued that the formation of the Moon paralleled that of Earth and the other planets. That is, the Moon formed from minute rock fragments and gases that composed a disk-shaped structure (the solar nebula) that orbited the early Sun. Debris from this disk collided and accumulated into larger masses that, in turn, accreted into planetary-sized bodies.

A new hypothesis, which has recently gained support from many scientists, suggests that a giant body collided with Earth to produce the Moon. The explosion caused by the impact of a Mars-sized body with a semi-molten Earth is thought to have ejected huge quantities of mantle rock from the primordial Earth. A portion of this ejected material remained in orbit around Earth, while the remainder either escaped or impacted upon Earth's surface. In a manner similar to that proposed in the earlier hypothesis, the material orbiting Earth then began to accumulate, eventually producing the Moon. Though the giant impact hypothesis provides a plausible mechanism for the Moon's formation, many questions must be answered before this proposal can be considered viable.

Despite the fact that the origin of the Moon is still debated, planetary geologists have been able to work out some of the basic details of the Moon's history, using among other things variations in crater density (quantity per unit area). Simply stated, the higher the crater density, the longer the topographic feature has existed. During its early history, the Moon was continually impacted as it swept up debris from the solar nebula. This continuous bombardment and perhaps radioactive decay generated enough heat to melt the Moon's outer shell and quite possibly the rest of the Moon as well.

When a large percentage of the debris had been gathered, the outer layer of the Moon began to cool and form a crystalline crust. From samples obtained by Apollo astronauts, the rocks of the primitive lunar crust are thought to be composed of a high percentage of a calcium-rich feldspar (anorthosite). This feldspar mineral crystallized early and, because it was less dense than the remaining melt, floated to the top and formed a surface scum. While this process was taking place, iron and other heavy metals probably sank to form a small central core. Even after the crust had solidified, its

surface was continually bombarded. Remnants of the original crust occupy the densely cratered highlands, which have been estimated to be as much as 4.5 billion years old.

The last period of heavy bombardment recorded in the lunar highlands occurred almost 500 million years after the crust had formed. It is not known with certainty whether this final episode of bombardment was simply a clean-up phase where the remaining large particles in the Earth-Moon orbit were swept up or whether it was an influx of bodies from farther out in the solar system.

The next major event in the Moon's evolution was the formation of maria basins, which are large craters that filled with lava flowing up through cracks in the Moon's surface. The meteoroids that produced these huge pits ejected mountainous quantities of lunar rock into piles rising 5 kilometers or more. The Apennine mountain range, which typifies such an accumulation, was produced in conjunction with the formation of the Imbrium Basin, the site explored by the Apollo 15 astronauts. The crater density of the ejected material is greater than that of the surface of the associated basin, confirming that an appreciable time elapsed between the formation and filling of these basins. Radiometric dating of the maria basalts (a type of rock) puts their age between 3.2 and 3.8 billion years, somewhat younger than the initial crust. In places, the lava flows overlap the highlands, another testimonial to the lesser age of the maria deposits.

The last prominent features to form on the lunar surface were the rayed craters as exemplified by the crater Copernicus. Rays of material ejected from these young "depressions" are clearly seen blanketing the surface of the maria and many older rayless craters. By contrast, the older craters have rounded rims, and their rays have been erased by the impact of small debris. However, even a relatively young crater like Copernicus must be millions of years old. Had it formed on Earth, erosional forces would have long since obliterated it.

42 – 3 Callisto and Ganymede

From 1996 to 1999, the Galileo spacecraft passed through the Jovian system, providing much information about Jupiter's satellites. Callisto, the outermost of Jupiter's four largest satellites, orbits the planet in seventeen days at a distance from Jupiter of two million kilometers. Like our own Moon, Callisto rotates in the same period as it revolves, so it always keeps the same face toward Jupiter. Its noontime surface temperature is only about -140°C , so water ice is stable on its surface year-round. Callisto has a diameter of 4.820 kilometers, almost the same as that of Mercury. Its mass is only one-third as great, which means its density must be only one-third as great as well. This tells us that Callisto has far less of the rocky metallic materials found in the inner planets and must instead be an icy body through much of its interior.

Callisto has not fully differentiated, meaning separated into layers of different density materials. Astronomers can tell that it lacks a dense core from the details of its gravitational pull on the Galileo spacecraft during several very close flybys. This fact surprised scientists, who expected that all the big icy moons would be differentiated. It is much easier for an icy body to differentiate than for a rocky one, since the melting temperature of ice is so low. Only a little heating will soften the ice and get the process started, allowing the rock and metal to sink to the center and the slushy ice to float to the surface. Yet Callisto seems to have frozen solid before the process of differentiation was complete.

Like our Moon's highlands, the surface of Callisto is covered with impact craters. The survival of these craters tells us that an icy object can form and retain impact craters in its surface. In thinking of ice so far from the Sun, it is important not to judge its behavior from that of the much warmer ice we know on Earth; at the temperatures of the outer solar system, ice on the surface is nearly as hard as rock, and behaves similarly. Ice on Callisto does not deform or flow like ice in glaciers on Earth. Callisto is unique among the planet-sized objects of the solar system in its absence of interior forces to drive geological evolution. The satellite was born dead and has remained geologically dead for more than four billion years.

Ganymede, another of Jupiter's satellites and the largest in our solar system, is also cratered, but

less so than Callisto. About one-quarter of its surface seems to be as old and heavily cratered; the rest formed more recently, as we can tell by the sparse covering of impact craters as well as the relative freshness of the craters. Ganymede is a differentiated world, like the terrestrial planets. Measurements of its gravity field tell us that the rock and metal sank to form a core about the size of our Moon, with a mantle and crust of ice floating above it. In addition, the Galileo spacecraft discovered that Ganymede has a magnetic field, the signature of a partially molten interior. Ganymede is not a dead world, but rather a place of continuing geological activity powered by an internal heat source. Much of its surface may be as young as half a billion years.

The younger terrain is the result of tectonic and volcanic forces. Some features formed when the crust cracked, flooding many of the craters with water from the interior. Extensive mountain ranges were formed from compression of the crust, forming long ridges with parallel valleys spaced one to two kilometers apart. In some places older impact craters were split and pulled apart. There are even indications of large-scale crustal movements that are similar to the plate tectonics of Earth.

Why is Ganymede different from Callisto? Possibly the small difference in size and internal heating between the two led to this divergence in their evolution. But more likely the gravity of Jupiter is to blame for Ganymede's continuing geological activity. Ganymede is close enough to Jupiter that tidal forces from the giant planet may have episodically heated its interior and triggered major convulsions on its crust.

41 – 2 Climate of Venus

Earth has abundant water in its oceans but very little carbon dioxide in its relatively thin atmosphere. By contrast, Venus is very dry and its thick atmosphere is mostly carbon dioxide. The original atmospheres of both Venus and Earth were derived at least in part from gases spewed forth, or outgassed, by volcanoes. The gases that emanate from present-day volcanoes on Earth, such as Mount Saint Helens, are predominantly water vapor, carbon dioxide, and sulfur dioxide. These gases should therefore have been important parts of the original atmospheres of both Venus and Earth. Much of the water on both planets is also thought to have come from impacts from comets, icy bodies formed in the outer solar system.

In fact, water probably once dominated the Venusian atmosphere. Venus and Earth are similar in size and mass, so Venusian volcanoes may well have outgassed as much water vapor as on Earth, and both planets would have had about the same number of comets strike their surfaces. Studies of how stars evolve suggest that the early Sun was only about 70 percent as luminous as it is now, so the temperature in Venus' early atmosphere must have been quite a bit lower. Thus water vapor would have been able to liquefy and form oceans on Venus. But if water vapor and carbon dioxide were once so common in the atmospheres of both Earth and Venus, what became of Earth's carbon dioxide? And what happened to the water on Venus?

The answer to the first question is that carbon dioxide is still found in abundance on Earth, but now, instead of being in the form of atmospheric carbon dioxide, it is either dissolved in the oceans or chemically bound into carbonate rocks, such as the limestone and marble that formed in the oceans. If Earth became as hot as Venus, much of its carbon dioxide would be boiled out of the oceans and baked out of the crust. Our planet would soon develop a thick, oppressive carbon dioxide atmosphere much like that of Venus.

To answer the question about Venus' lack of water, we must return to the early history of the planet. Just as on present-day Earth, the oceans of Venus limited the amount of atmospheric carbon dioxide by dissolving it in the oceans and binding it up in carbonate rocks. But being closer to the Sun than Earth is, enough of the liquid water on Venus would have vaporized to create a thick cover of water

vapor clouds. Since water vapor is a greenhouse gas, this humid atmosphere—perhaps denser than Earth's present-day atmosphere, but far less dense than the atmosphere that envelops Venus today—would have efficiently trapped heat from the Sun. At first, this would have had little effect on the oceans of Venus. Although the temperature would have climbed above 100° C, the boiling point of water at sea level on Earth, the added atmospheric pressure from water vapor would have kept the water in Venus' oceans in the liquid state.

This hot and humid state of affairs may have persisted for several hundred million years. But as the Sun's energy output slowly increased over time, the temperature at the surface would eventually have risen above 374°C. Above this temperature, no matter what the atmospheric pressure, Venus' oceans would have begun to evaporate, and the added water vapor in the atmosphere would have increased the greenhouse effect. This would have made the temperature even higher and caused the oceans to evaporate faster, producing more water vapor. That, in turn, would have further intensified the greenhouse effect and made the temperature climb higher still.

Once Venus' oceans disappeared, so did the mechanism for removing carbon dioxide from the atmosphere. With no oceans to dissolve it, outgassed carbon dioxide began to accumulate in the atmosphere, intensifying the greenhouse effect even more. Temperatures eventually became high enough to "bake out" any carbon dioxide that was trapped in carbonate rocks. This liberated carbon dioxide formed the thick atmosphere of present-day Venus. Over time, the rising temperatures would have leveled off, solar ultraviolet radiation having broken down atmospheric water vapor molecules into hydrogen and oxygen. With all the water vapor gone, the greenhouse effect would no longer have accelerated.

16 – 3 Planets in Our Solar System

The Sun is the hub of a huge rotating system consisting of nine planets, their satellites, and numerous small bodies, including asteroids, comets, and meteoroids. An estimated 99.85 percent of the mass of our solar system is contained within the Sun, while the planets collectively make up most of the remaining 0.15 percent. The planets, in order of their distance from the Sun, are Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, Neptune, and Pluto. Under the control of the Sun's gravitational force, each planet maintains an elliptical orbit and all of them travel in the same direction.

The planets in our solar system fall into two groups: the terrestrial (Earth-like) planets (Mercury, Venus, Earth, and Mars) and the Jovian (Jupiter-like) planets (Jupiter, Saturn, Uranus, and Neptune). Pluto is not included in either category, because its great distance from Earth and its small size make this planet's true nature a mystery.

The most obvious difference between the terrestrial and the Jovian planets is their size. The largest terrestrial planet, Earth has a diameter only one quarter as great as the diameter of the smallest Jovian planet, Neptune, and its mass is only one seventeenth as great. Hence, the Jovian planets are often called giants. Also, because of their relative locations, the four Jovian planets are known as the outer planets, while the terrestrial planets are known as the inner planets. There appears to be a correlation between the positions of these planets and their sizes.

Other dimensions along which the two groups differ markedly are density and composition. The densities of the terrestrial planets average about 5 times the density of water, whereas the Jovian planets have densities that average only 1.5 times the density of water. One of the outer planets, Saturn, has a density of only 0.7 that of water, which means that Saturn would float in water. Variations in the composition of the planets are largely responsible for the density differences. The substances that make up both groups of planets are divided into three groups—gases, rocks, and ices—based on their melting points. The terrestrial planets are mostly rocks: dense rocky and metallic material, with minor amounts of gases. The Jovian planets, on the other hand, contain a large percentage of the gases hydrogen and helium, with

varying amounts of ices: mostly water, ammonia, and methane ices.

The Jovian planets have very thick atmospheres consisting of varying amounts of hydrogen, helium, methane, and ammonia. By comparison, the terrestrial planets have meager atmospheres at best. A planet's ability to retain an atmosphere depends on its temperature and mass. Simply stated, a gas molecule can "evaporate" from a planet if it reaches a speed known as the escape velocity. For Earth, this velocity is 11 kilometers per second. Any material, including a rocket, must reach this speed before it can leave Earth and go into space. The Jovian planets, because of their greater masses and thus higher surface gravities, have higher escape velocities (21-60 kilometers per second) than the terrestrial planets. Consequently, it is more difficult for gases to "evaporate" from them. Also, because the molecular motion of a gas depends on temperature, at the low temperatures of the Jovian planets even the lightest gases are unlikely to acquire the speed needed to escape. On the other hand, a comparatively warm body with a small surface gravity, like Earth's moon, is unable to hold even the heaviest gas and thus lacks an atmosphere. The slightly larger terrestrial planets Earth, Venus, and Mars retain some heavy gases like carbon dioxide, but even their atmospheres make up only an infinitesimally small portion of their total mass.

The orderly nature of our solar system leads most astronomers to conclude that the planets formed at essentially the same time and from the same material as the Sun. It is hypothesized that the primordial cloud of dust and gas from which all the planets are thought to have condensed had a composition somewhat similar to that of Jupiter. However, unlike Jupiter, the terrestrial planets today are nearly void of light gases and ices. The explanation may be that the terrestrial planets were once much larger and richer in these materials but eventually lost them because of these bodies' relative closeness to the Sun, which meant that their temperatures were relatively high.

50 – 3 Star Death

Until the early- to mid-twentieth century, scientists believed that stars generate energy by shrinking. As stars contracted, it was thought, they would get hotter and hotter, giving off light in the process. This could not be the primary way that stars shine, however. If it were, they would scarcely last a million years, rather than the billions of years in age that we know they are. We now know that stars are fueled by nuclear fusion. Each time fusion takes place, energy is released as a by-product. This energy, expelled into space, is what we see as starlight. The fusion process begins when two hydrogen nuclei smash together to form a particle called the deuteron (a combination of a positive proton and a neutral neutron). Deuterons readily combine with additional protons to form helium. Helium, in turn, can fuse together to form heavier elements, such as carbon. In a typical star, merger after merger takes place until significant quantities of heavy elements are built up.

We must distinguish, at this point, between two different stellar types: Population I and Population II, the latter being much older than the former. These groups can also be distinguished by their locations. Our galaxy, the Milky Way, is shaped like a flat disk surrounding a central bulge. Whereas Population I stars are found mainly in the galactic disk, Population II stars mostly reside in the central bulge of the galaxy and in the halo surrounding this bulge.

Population II stars date to the early stages of the universe. Formed when the cosmos was filled with hydrogen and helium gases, they initially contained virtually no heavy elements. They shine until their fusible material is exhausted. When Population II stars die, their material is spread out into space. Some of this dust is eventually incorporated into newly formed Population I stars. Though Population I stars consist mostly of hydrogen and helium gas, they also contain heavy elements (heavier than helium), which comprise about 1 or 2 percent of their mass. These heavier materials are fused from the lighter elements that the stars have collected. Thus, Population I stars contain material that once belonged to stars from previous generations. The Sun is a good example of a Population I star.

What will happen when the Sun dies? In several billion years, our mother star will burn much brighter. It will expend more and more of its nuclear

fuel, until little is left of its original hydrogen. Then, at some point in the far future, all nuclear reactions in the Sun's center will cease.

Once the Sun passes into its "postnuclear" phase, it will separate effectively into two different regions: an inner zone and an outer zone. While no more hydrogen fuel will remain in the inner zone, there will be a small amount left in the outer zone. Rapidly, changes will begin to take place that will serve to tear the Sun apart. The inner zone, its nuclear fires no longer burning, will begin to collapse under the influence of its own weight and will contract into a tiny hot core, dense and dim. An opposite fate will await the outer region, a loosely held-together ball of gas. A shock wave caused by the inner zone's contraction will send ripples through the dying star, pushing the stellar exterior's material farther and farther outward. The outer envelope will then grow rapidly, increasing, in a short interval, hundreds of times in size. As it expands, it will cool down by thousands of degrees. Eventually, the Sun will become a red giant star, cool and bright. It will be so large that it will occupy the whole space that used to be the Earth's orbit and so brilliant that it would be able to be seen with the naked eye thousands of light-years away. It will exist that way for millions of years, gradually releasing the material of its outer envelope into space. Finally, nothing will be left of the gaseous exterior of the Sun; all that will remain will be the hot, white core. The Sun will have become a white dwarf star. The core will shrink, giving off the last of its energy, and the Sun will finally die.

51 – 2 Surface Fluids on Venus and Earth

A fluid is a substance, such as a liquid or gas, in which the component particles (usually molecules) can move past one another. Fluids flow easily and conform to the shape of their containers. The geologic processes related to the movement of fluids on a planet's surface can completely resurface a planet many times. These processes derive their energy from the Sun and the gravitational forces of the planet itself. As these fluids interact with surface materials, they move particles about or react chemically with them to modify or produce materials. On a solid planet with a hydrosphere and an atmosphere, only a tiny fraction of the planetary mass flows as surface fluids. Yet the movements of these fluids can drastically alter a planet. Consider Venus and Earth, both terrestrial planets with atmospheres.

Venus and Earth are commonly regarded as twin planets but not identical twins. They are about the same size, are composed of roughly the same mix of materials, and may have been comparably endowed at their beginning with carbon dioxide and water. However, the twins evolved differently largely because of differences in their distance from the Sun. With a significant amount of internal heat, Venus may continue to be geologically active with volcanoes, rifting, and folding. However, it lacks any sign of a hydrologic system (water circulation and distribution): there are no streams, lakes oceans or glaciers. Space probes suggest that Venus may have started with as much water as Earth, but it was unable to keep its water in liquid form. Because Venus receives more heat from the Sun, water released from the interior evaporated and rose to the upper atmosphere where the Sun's ultraviolet rays broke the molecules apart. Much of the freed hydrogen escaped into space, and Venus lost its water. Without water, Venus became less and less like Earth and kept an atmosphere filled with carbon dioxide. The carbon dioxide acts as a blanket, creating an intense greenhouse effect and driving surface temperatures high enough to melt lead and to prohibit the formation of carbonate minerals. Volcanoes continually vented more carbon dioxide into the atmosphere. On Earth, liquid water removes carbon dioxide from the atmosphere and combines it with calcium, from rock weathering, to form carbonate sedimentary rocks. Without liquid water to remove carbon from the atmosphere, the level of carbon dioxide in the atmosphere of Venus remains high.

Like Venus, Earth is large enough to be geologically active and for its gravitational field to hold an atmosphere. Unlike Venus, it is just the right distance from the Sun so that temperature ranges allow water to exist as a liquid, a solid, and a gas. Water is thus extremely mobile and moves rapidly over the planet in a continuous hydrologic cycle. Heated by the Sun, the water moves in great cycles from the oceans to the atmosphere, over the landscape in river systems, and ultimately back to the oceans. As a result, Earth's surface has been continually changed and eroded into delicate systems of river valleys—a remarkable contrast to the surfaces of other planetary bodies where impact craters dominate. Few areas on Earth have been untouched by flowing water. As a result, river valleys are the dominant feature of its landscape. Similarly, wind action has scoured fine particles away from large areas, depositing them elsewhere as vast sand seas dominated by dunes or in sheets of loess (fine-grained soil deposits). These fluid movements are caused by gravity flow systems energized by heat from the Sun. Other geologic changes occur when the gases in the atmosphere or water react with rocks at the surface to form new chemical compounds with different properties. An important example of this process was the removal of most of Earth's carbon dioxide from its atmosphere to form carbonate rocks. However, if Earth were a little closer to the Sun, its oceans would evaporate; if it were farther from the Sun, the oceans would freeze solid. Because liquid water was present, self-replicating molecules of carbon, hydrogen, and oxygen developed life early in Earth's history and have radically modified its surface, blanketing huge parts of the continents with greenery. Life thrives on this planet, and it helped create the planet's oxygen- and nitrogen-rich atmosphere and moderate temperatures.

PLANTS / BOTANY / MONTOLOGY (OROLOGY) / HYDROLOGY / HYDROGRAPHY / HYDROGEOLOGY

05 – 1 Minerals and Plants

Research has shown that certain minerals are required by plants for normal growth and development. The soil is the source of these minerals, which are absorbed by the plant with the water from the soil. Even nitrogen, which is a gas in its elemental state, is normally absorbed from the soil as nitrate ions. Some soils are notoriously deficient in micro nutrients and are therefore unable to support most plant life. So-called serpentine soils, for example, are deficient in calcium, and only plants able to tolerate low levels of this mineral can survive. In modern agriculture, mineral depletion of soils is a major concern, since harvesting crops interrupts the recycling of nutrients back to the soil.

Mineral deficiencies can often be detected by specific symptoms such as chlorosis (loss of chlorophyll resulting in yellow or white leaf tissue), necrosis (isolated dead patches), anthocyanin formation (development of deep red pigmentation of leaves or stem), stunted growth, and development of woody tissue in an herbaceous plant. Soils are most commonly deficient in nitrogen and phosphorus. Nitrogen-deficient plants exhibit many of the symptoms just described. Leaves develop chlorosis; stems are short and slender, and anthocyanin discoloration occurs on stems, petioles, and lower leaf surfaces. Phosphorus-deficient plants are often stunted, with leaves turning a characteristic dark green, often with the accumulation of anthocyanin. Typically, older leaves are affected first as the phosphorus is mobilized to young growing tissue. Iron deficiency is characterized by chlorosis between veins in young leaves.

Much of the research on nutrient deficiencies is based on growing plants hydroponically, that is, in soilless liquid nutrient solutions. This technique allows researchers to create solutions that selectively omit certain nutrients and then observe the resulting effects on the plants. Hydroponics has applications beyond basic research, since it facilitates the growing of greenhouse vegetables during winter. Aeroponics, a technique in which plants are suspended and the roots misted with a nutrient solution, is another method for growing plants without soil.

While mineral deficiencies can limit the growth of plants, an over abundance of certain minerals can be toxic and can also limit growth. Saline soils, which have high concentrations of sodium chloride and other salts, limit plant growth, and research continues

to focus on developing salt-tolerant varieties of agricultural crops. Research has focused on the toxic effects of heavy metals such as lead, cadmium, mercury, and aluminum; however, even copper and zinc, which are essential elements, can become toxic in high concentrations. Although most plants cannot survive in these soils, certain plants have the ability to tolerate high levels of these minerals.

Scientists have known for some time that certain plants, called hyper accumulators, can concentrate minerals at levels a hundredfold or greater than normal. A survey of known hyper accumulators identified that 75 percent of them amassed nickel, cobalt, copper, zinc, manganese, lead, and cadmium are other minerals of choice. Hyper accumulators run the entire range of the plant world. They may be herbs, shrubs, or trees. Many members of the mustard family, spurge family, legume family, and grass family are top hyper accumulators. Many are found in tropical and subtropical areas of the world, where accumulation of high concentrations of metals may afford some protection against plant-eating insects and microbial pathogens.

Only recently have investigators considered using these plants to clean up soil and waste sites that have been contaminated by toxic levels of heavy metals—an environmentally friendly approach known as phytoremediation. This scenario begins with the planting of hyper accumulating species in the target area, such as an abandoned mine or an irrigation pond contaminated by runoff. Toxic minerals would first be absorbed by roots but later relocated to the stem and leaves. A harvest of the shoots would remove the toxic compounds off site to be burned or composted to recover the metal for industrial uses. After several years of cultivation and harvest, the site would be restored at a cost much lower than the price of excavation and reburial, the standard practice for remediation of contaminated soils. For examples, in field trials, the plant alpine penny cress removed zinc and cadmium from soils near a zinc smelter, and Indian mustard, native to Pakistan and India, has been effective in reducing levels of selenium salts by 50 percent in contaminated soils.

09 – 3 The Arrival of Plant Life in Hawaii

When the Hawaiian Islands emerged from the sea as volcanoes, starting about five million years ago, they were far removed from other landmasses. Then, as blazing sunshine alternated with drenching rains, the harsh, barren surfaces of the black rocks slowly began to soften. Winds brought a variety of life-forms.

Spores light enough to float on the breezes were carried thousands of miles from more ancient lands and deposited at random across the bare mountain flanks. A few of these spores found a toehold on the dark, forbidding rocks and grew and began to work their transformation upon the land. Lichens were probably the first successful flora. These are not single individual plants; each one is a symbiotic combination of an alga and a fungus. The algae capture the sun's energy by photosynthesis and store it in organic molecules. The fungi absorb moisture and mineral salts from the rocks, passing these on in waste products that nourish algae. It is significant that the earliest living things that built communities on these islands are examples of symbiosis, a phenomenon that depends upon the close cooperation of two or more forms of life and a principle that is very important in island communities.

Lichens helped to speed the decomposition of the hard rock surfaces, preparing a soft bed of soil that was abundantly supplied with minerals that had been carried in the molten rock from the bowels of Earth. Now, other forms of life could take hold: ferns and mosses (two of the most ancient types of land plants) that flourish even in rock crevices. These plants propagate by producing spores—tiny fertilized cells that contain all the instructions for making a new plant—but the spore are unprotected by any outer coating and carry no supply of nutrient. Vast numbers of them fall on the ground beneath the mother plants. Sometimes they are carried farther afield by water or by wind. But only those few spores that settle down in very favorable locations can start new life; the vast majority fall on barren ground. By force of sheer numbers, however, the mosses and ferns reached Hawaii, survived, and multiplied. Some species developed great size, becoming tree ferns that even now grow in the Hawaiian forests.

Many millions of years after ferns evolved (but long before the Hawaiian Islands were born from the sea), another kind of flora evolved on Earth: the

seed-bearing plants. This was a wonderful biological invention. The seed has an outer coating that surrounds the genetic material of the new plant, and inside this covering is a concentrated supply of nutrients. Thus the seed's chances of survival are greatly enhanced over those of the naked spore. One type of seed-bearing plant, the angiosperm, includes all forms of blooming vegetation. In the angiosperm the seeds are wrapped in an additional layer of covering. Some of these coats are hard—like the shell of a nut—for extra protection. Some are soft and tempting, like a peach or a cherry. In some angiosperms the seeds are equipped with gossamer wings, like the dandelion and milkweed seeds. These new characteristics offered better ways for the seed to move to new habitats. They could travel through the air, float in water, and lie dormant for many months.

Plants with large, buoyant seeds—like coconuts—drift on ocean currents and are washed up on the shores. Remarkably resistant to the vicissitudes of ocean travel, they can survive prolonged immersion in saltwater when they come to rest on warm beaches and the conditions are favorable, the seed coats soften. Nourished by their imported supply of nutrients, the young plants push out their roots and establish their place in the sun.

By means of these seeds, plants spread more widely to new locations, even to isolated islands like the Hawaiian archipelago, which lies more than 2,000 miles west of California and 3,500 miles east of Japan. The seeds of grasses, flowers, and blooming trees made the long trips to these islands. (Grasses are simple forms of angiosperms that bear their encapsulated seeds on long stalks.) In a surprisingly short time, angiosperms filled many of the land areas on Hawaii that had been bare.

19 – 2 Succession, Climax, and Ecosystems

In the late nineteenth century, ecology began to grow into an independent science from its roots in natural history and plant geography. The emphasis of this new “community ecology” was on the composition and structure of communities consisting of different species. In the early twentieth century, the American ecologist Frederic Clements pointed out that a succession of plant communities would develop after a disturbance such as a volcanic eruption, heavy flood, or forest fire. An abandoned field, for instance, will be invaded successively by herbaceous plants (plants with little or no woody tissue), shrubs, and trees, eventually becoming a forest. Light-loving species are always among the first invaders, while shade-tolerant species appear later in the succession.

Clements and other early ecologists saw almost lawlike regularity in the order of succession, but that has not been substantiated. A general trend can be recognized, but the details are usually unpredictable. Succession is influenced by many factors: the nature of the soil, exposure to sun and wind, regularity of precipitation, chance colonizations, and many other random processes.

The final stage of a succession, called the climax by Clements and early ecologists, is likewise not predictable or of uniform composition. There is usually a good deal of turnover in species composition, even in a mature community. The nature of the climax is influenced by the same factors that influenced succession. Nevertheless, mature natural environments are usually in equilibrium. They change relatively little through time unless the environment itself changes.

For Clements, the climax was a “superorganism”, an organic entity. Even some authors who accepted the climax concept rejected Clements’ characterization of it as a superorganism, and it is indeed a misleading metaphor. An ant colony may be legitimately called a superorganism because its communication system is so highly organized that the colony always works as a whole and appropriately according to the circumstances. But there is no evidence for such an interacting communicative network in a climax plant formation. Many authors prefer the term “association” to the term “community” in order to stress the looseness of the interaction.

Even less fortunate was the extension of this type of thinking to include animals as well as plants.

This resulted in the “biome”, a combination of coexisting flora and fauna. Though it is true that many animals are strictly associated with certain plants, it is misleading to speak of a “spruce-moose biome”, for example, because there is no internal cohesion to their association as in an organism. The spruce community is not substantially affected by either the presence or absence of moose. Indeed, there are vast areas of spruce forest without moose. The opposition to the Clementsian concept of plant ecology was initiated by Herbert Gleason, soon joined by various other ecologists. Their major point was that the distribution of a given species was controlled by the habitat requirements of that species and that therefore the vegetation types were a simple consequence of the ecologies of individual plant species.

With “climax”, “biome”, “superorganism”, and various other technical terms for the association of animals and plants at a given locality being criticized, the term “ecosystem” was more and more widely adopted for the whole system of associated organisms together with the physical factors of their environment. Eventually, the energy-transforming role of such a system was emphasized. Ecosystems thus involve the circulation, transformation, and accumulation of energy and matter through the medium of living things and their activities. The ecologist is concerned primarily with the quantities of matter and energy that pass through a given ecosystem, and with the rates at which they do so.

Although the ecosystem concept was very popular in the 1950s and 1960s, it is no longer the dominant paradigm. Gleason’s arguments against climax and biome are largely valid against ecosystems as well. Furthermore, the number of interactions is so great that they are difficult to analyze, even with the help of large computers. Finally, younger ecologists have found ecological problems involving behavior and life-history adaptations more attractive than measuring physical constants. Nevertheless, one still speaks of the ecosystem when referring to a local association of animals and plants, usually without paying much attention to the energy aspects.

22 – 1 *Spartina*

Spartina alterniflora, known as cordgrass, is a deciduous, perennial flowering plant native to the Atlantic coast and the Gulf Coast of the United States. It is the dominant native species of the lower salt marshes along these coasts, where it grows in the intertidal zone (the area covered by water some parts of the day and exposed others).

These natural salt marshes are among the most productive habitats in the marine environment. Nutrient-rich water is brought to the wetlands during each high tide, making a high rate of food production possible. As the seaweed and marsh grass leaves die, bacteria break down the plant material, and insects, small shrimplike organisms, fiddler crabs, and marsh snails eat the decaying plant tissue, digest it, and excrete wastes high in nutrients. Numerous insects occupy the marsh, feeding on living or dead cordgrass tissue, and redwing blackbirds, sparrows, rodents, rabbits, and deer feed directly on the cordgrass. Each tidal cycle carries plant material into the offshore water to be used by the subtidal organisms.

Spartina is an exceedingly competitive plant. It spreads primarily by underground stems; colonies form when pieces of the root system or whole plants float into an area and take root or when seeds float into a suitable area and germinate. *Spartina* establishes itself on substrates ranging from sand and silt to gravel and cobble and is tolerant of salinities ranging from that of near freshwater (0.05 percent) to that of salt water (3.5 percent). Because they lack oxygen, marsh sediments are high in sulfides that are toxic to most plants. *Spartina* has the ability to take up sulfides and convert them to sulfate, a form of sulfur that the plant can use; this ability makes it easier for the grass to colonize marsh environments. Another adaptive advantage is *Spartina*'s ability to use carbon dioxide more efficiently than most other plants.

These characteristics make *Spartina* a valuable component of the estuaries where it occurs naturally. The plant functions as a stabilizer and a sediment trap and as a nursery area for estuarine fish and shellfish. Once established, a stand of *Spartina* begins to trap sediment, changing the substrate elevation, and eventually the stand evolves into a high marsh system where *Spartina* is gradually displaced by higher-elevation, brackish-water species. As elevation increases, narrow,

deep channels of water form throughout the marsh. Along the east coast *Spartina* is considered valuable for its ability to prevent erosion and marshland deterioration; it is also used for coastal restoration projects and the creation of new wetland sites.

Spartina was transported to Washington State in packing materials for oysters transplanted from the east coast in 1894. Leaving its insect predators behind, the cordgrass has been spreading slowly and steadily along Washington's tidal estuaries on the west coast, crowding out the native plants and drastically altering the landscape by trapping sediment. *Spartina* modifies tidal mudflats, turning them into high marshes inhospitable to the many fish and waterfowl that depend on the mudflats. It is already hampering the oyster harvest and the Dungeness crab fishery, and it interferes with the recreational use of beaches and waterfronts. *Spartina* has been transplanted to England and to New Zealand for land reclamation and shoreline stabilization. In New Zealand the plant has spread rapidly, changing mudflats with marshy fringes to extensive salt meadows and reducing the number and kinds of birds and animals that use the marsh.

Efforts to control *Spartina* outside its natural environment have included burning, flooding, shading plants with black canvas or plastic, smothering the plants with dredged materials or clay, applying herbicide, and mowing repeatedly. Little success has been reported in New Zealand and England; Washington State's management program has tried many of these methods and is presently using the herbicide glyphosphate to control its spread. Work has begun to determine the feasibility of using insects as biological controls, but effective biological controls are considered years away. Even with a massive effort, it is doubtful that complete eradication of *Spartina* from nonnative habitats is possible, for it has become an integral part of these shorelines and estuaries during the last 100 to 200 years.

25 – 3 The Evolutionary Origin of Plants

The evolutionary history of plants has been marked by a series of adaptations. The ancestors of plants were photosynthetic single-celled organisms that gave rise to plants presumably lacked true roots, stems, leaves, and complex reproductive structures such as flowers. All of these features appeared later in the evolutionary history of plants. Of today's different groups of algae, green algae are probably the most similar to ancestral plants. This supposition stems from the close phylogenetic (natural evolutionary) relationship between the two groups. DNA comparisons have shown that green algae are plants' closest living relatives. In addition, other lines of evidence support the hypothesis that land plants evolved from ancestral green algae used the same type of chlorophyll and accessory pigments in photosynthesis as do land plants. This would not be true of red and brown algae. Green algae store food as starch, as do land plants and have cell walls made of cellulose, similar in composition to those of land plants. Again, the good storage and cell wall molecules of red and brown algae are different.

Today green algae live mainly in freshwater, suggesting that their early evolutionary history may have occurred in freshwater habitats. If so, the green algae would have been subjected to environmental pressures that resulted in adaptations that enhanced their potential to give rise to land-dwelling organisms.

The environmental conditions of freshwater habitats, unlike those of ocean habitats, are highly variable. Water temperature can fluctuate seasonally or even daily and changing level of rainfall can lead to fluctuations in the concentration of chemical in the water or even to period in which the aquatic habitat dries up. Ancient fresh water green algae must have evolved features that enable them to withstand extremes of temperature and periods of dryness. These adaptations served their descendant well as they invaded land.

The terrestrial world is green now, but it did not start out that way. When plants first made the transition ashore more than 400 million years ago, the land was barren and desolate, inhospitable to life. From a plant's evolutionary view point, however, it was also a land of opportunity, free of competitors and predators and full of carbon dioxide and sunlight (the raw materials for photosynthesis, which are present in far higher

concentrations in air than in water). So once natural selection had shaped the adaptations that helped plants overcome the obstacles to terrestrial living, plants prospered and diversified.

When plants pioneered the land, they faced a range of challenges posed by terrestrial environments. On land, the supportive buoyancy of water is missing, the plant is no longer bathed in a nutrient solution, and air tends to dry things out. These conditions favored the evolution of the structures that support the body, vessels that transport water and nutrients to all parts of plant, and structures that conserve water. The resulting adaptations to dry land include some structural features that arose early in plant evolution; now these features are common to virtually all land plant. They include roots or root like structures, a waxy cuticle that covers the surfaces of leaves and stems and limits the evaporation of water, and pores called stomata in leaves and stems that allow gas exchange but close when water is scarce, thus reducing water loss. Other adaptations occurred later in the transition to terrestrial life and now wide spread but not universal among plants. These include conducting vessels that transport water and minerals upward from the roots and that move the photosynthetic products from the leaves to the rest of the plant body and the stiffening substance lignin, which support the plant body, helping it expose maximum surface area to sunlight.

These adaptations allowed an increasing diversity of plant forms to exploit dry land. Life on land, however, also required new methods of transporting sperm to eggs. Unlike aquatic and marine forms, land plants cannot always rely on water currents to carry their sex cells and disperse their fertilized eggs. So the most successful groups of land plants are those that evolved methods of fertilized sex cell dispersal that are independent of water and structures that protect developing embryos from drying out. Protected embryos and waterless dispersal of sex cells were achieved with the origin of seed plants and the key evolutionary innovations that they introduced: pollen, seeds, and later, flowers and fruits.

31 – 3 Savanna Formation

Located in tropical areas at low altitudes, savannas are stable ecosystems, some wet and some dry consisting of vast grasslands with scattered trees and shrubs. They occur on a wide range of soil types and in extremes of climate. There is no simple or single factor that determines if a given site will be a savanna, but some factors seem to play important roles in their formation.

Savannas typically experience a rather prolonged dry season. One theory behind savanna formation is that wet forest species are unable to withstand the dry season, and thus savanna, rather than rain forest, is favored on the site. Savannas experience an annual rainfall of between 1,000 and 2,000 millimeters, most of it falling in a five- to eight-month wet season. Though plenty of rain may fall on a savanna during the year, for at least part of the year little does, creating the drought stress ultimately favoring grasses. Such conditions prevail throughout much of northern South America and Cuba, but many Central American savannas as well as coastal areas of Brazil and the island of Trinidad do not fit this pattern. In these areas, rainfall per month exceeds that in the above definition, so other factors must contribute to savanna formation.

In many characteristics, savanna soils are similar to those of some rain forests, though more extreme. For example, savanna soils, like many rain forest soils, are typically oxisols (dominated by certain oxide minerals) and ultisols (soils containing no calcium carbonate), with a high acidity and notably low concentrations of such minerals as phosphorus, calcium, magnesium, and potassium, while aluminum levels are high. Some savannas occur on wet, waterlogged soils; others on dry, sandy, well-drained soils. This may seem contradictory, but it only means that extreme soil conditions, either too wet or too dry for forests, are satisfactory for savannas. More moderate conditions support moist forests.

Waterlogged soils occur in areas that are flat or have poor drainage. These soils usually contain large amounts of clay and easily become water saturated. Air cannot penetrate between the soil particles, making the soil oxygen-poor. By contrast, dry soils are sandy and porous, their coarse textures permitting water to drain rapidly. Sandy soils are prone to the leaching of nutrients and minerals and so tend to be nutritionally poor.

Though most savannas are found on sites with poor soils (because of either moisture conditions or nutrient levels of both), poor soils can and do support lush rain forest.

Most savannas probably experience mild fires frequently and major burns every two years or so. Many savanna and dry-forest plant species are called pyrophytes, meaning they are adapted in various ways to withstand occasional burning. Frequent fire is a factor to which rain forest species seem unable to adapt, although ancient charcoal remains from Amazon forest soils dating prior to the arrival of humans suggest that moist forests also occasionally burn. Experiments suggest that if fire did not occur in savannas in the Americas, species composition would change significantly. When burning occurs, it prevents competition among plant species from progressing to the point where some species exclude others, reducing the overall diversity of the ecosystem. But in experimental areas protected from fire, a few perennial grass species eventually come to dominate, outcompeting all others. Evidence from other studies suggests that exclusion of fire results in markedly decreased plant-species richness, often with an increase in tree density. There is generally little doubt that fire is a significant factor in maintaining savanna, certainly in most regions.

On certain sites, particularly in South America, savanna formation seems related to frequent cutting and burning of moist forests for pastureland. Increase in pastureland and subsequent overgrazing have resulted in an expansion of savanna. The thin upper layer of humus (decayed organic matter) is destroyed by cutting and burning. Humus is necessary for rapid decomposition of leaves by bacteria and fungi and for recycling by surface roots. Once the humus layer disappears, nutrients cannot be recycled and leach from the soil, converting soil from fertile to infertile and making it suitable only for savanna vegetation. Forests on white, sandy soil are most susceptible to permanent alteration.

32 – 1 Plant Colonization

Colonization is one way in which plants can change the ecology of a site. Colonization is a process with two components: invasion and survival. The rate at which a site is colonized by plants depends on both the rate at which individual organisms (seeds, spores, immature or mature individuals) arrive at the site and their success at becoming established and surviving. Success in colonization depends to a great extent on there being a site available for colonization – a safe site where disturbance by fire or by cutting down of trees has either removed competing species or reduced levels of competition and other negative interactions to a level at which the invading species can become established. For a given rate of invasion, colonization of a moist, fertile site is likely to be much more rapid than that of a dry, infertile site because of poor survival on the latter. A fertile, plowed field is rapidly invaded by a large variety of weeds, whereas a neighboring construction site from which the soil has been compacted or removed to expose a coarse, infertile parent material may remain virtually free of vegetation for many months or even years despite receiving the same input of seeds as the plowed field.

Both the rate of invasion and the rate of extinction vary greatly among different plant species. Pioneer species – those that occur only in the earliest stages of colonization – tend to have high rates of invasion because they produce very large numbers of reproductive propagules (seeds, spores, and so on) and because they have an efficient means of dispersal (normally, wind).

If colonizers produce short-lived reproductive propagules, then they must produce very large numbers unless they have an efficient means of dispersal to suitable new habitats. Many plants depend on wind for dispersal and produce abundant quantities of small, relatively short-lived seeds to compensate for the fact that wind is not always a reliable means of reaching the appropriate type of habitat. Alternative strategies have evolved in some plants, such as those that produce fewer but larger seeds that are dispersed to suitable sites by birds or small mammals or those that produce long-lived seeds. Many forest plants seem to exhibit the latter adaptation, and viable seeds of pioneer species can be found in large numbers on some forest floors. For example, as many as 1,125 viable seeds per square meter were found in a 100-year-old Douglas fir/western

hemlock forest in coastal British Columbia. Nearly all the seeds that had germinated from this seed bank were from pioneer species. The rapid colonization of such sites after disturbance is undoubtedly in part a reflection of the large seed bank on the forest floor.

An adaptation that is well developed in colonizing species is a high degree of variation in germination (the beginning of a seed's growth). Seeds of a given species exhibit a wide range of germination dates, increasing the probability that at least some of the seeds will germinate during a period of favorable environmental conditions. This is particularly important for species that colonize an environment where there is no existing vegetation to ameliorate climatic extremes and in which there may be great climatic diversity.

Species succession in plant communities, i.e., the temporal sequence of appearance and disappearance of species is dependent on events occurring at different stages in the life history of a species. Variation in rates of invasion and growth plays an important role in determining patterns of succession, especially secondary succession. The species that are first to colonize a site are those that produce abundant seeds that are distributed successfully to new sites. Such species generally grow rapidly and quickly dominate new sites, excluding other species with lower invasion and growth rates. The first community that occupies a disturbed area therefore may be composed of species with the highest rate of invasion, whereas the community of the subsequent stage may consist of plants with similar survival rates but lower invasion rates.

34 – 3 Protection of Plants by Insects

Many plants – one or more species of at least 68 different families – can secrete nectar even when they have no blossoms, because they bear extrafloral nectaries (structures that produce nectar) on stems, leaves, leaf stems, or other structures. These plants usually occur where ants are abundant, most in the tropics but some in temperate areas. Among those of northeastern North America are various plums, cherries, roses, hawthorns, poplars, and oaks. Like floral nectar, extrafloral nectar consists mainly of water with a high content of dissolved sugars and, in some plants, small amounts of amino acids. The extrafloral nectaries of some plants are known to attract ants and other insects, but the evolutionary history of most plants with these nectaries is unknown. Nevertheless, most ecologists believe that all extrafloral nectaries attract insects that will defend the plant.

Ants are portably the most frequent and certainly the most persistent defenders of plants. Since the highly active worker ants require a great deal of energy, plants exploit this need by providing extrafloral nectar that supplies ants with abundant energy. To return this favor, ants guard the nectaries, driving away or killing intruding insects that might compete with ants for nectar. Many of these intruders are herbivorous and would eat the leaves of the plants.

Biologists once thought that secretion of extrafloral nectar has some purely internal physiological function, and that ants provide no benefit whatsoever to the plants that secrete it. This view and the opposing “protectionist” hypothesis that ants defend plants had been disputed for over a hundred years when, in 1910, a skeptical William Morton Wheeler commented on the controversy. He called for proof of the protectionist view: that visitations of the ants confer protection on the plants and that in the absence of the insects a much greater number would perish or fail to produce flowers or seeds than when the insects are present. That we now have an abundance of the proof that was called for was established when Barbara Bentley reviewed the relevant evidence in 1977, and since then many more observations and experiments have provided still further proof that ants benefit plants.

One example shows how ants attracted to extrafloral nectaries protect morning glories against

attacking insects. The principal insect enemies of the North American morning glory feed mainly on its flowers or fruits rather than its leaves. Grasshoppers feeding on flowers indirectly block pollination and the production of seeds by destroying the corolla or the stigma, which receives the pollen grains and on which the pollen germinates. Without their colorful corolla, flowers do not attract pollinators and are not fertilized. An adult grasshopper can consume a large corolla, about 2.5 inches long, in an hour. Caterpillars and seed beetles affect seed production directly. Caterpillars devour the ovaries, where the seeds are produced, and seed beetle larvae eat seeds as they burrow in developing fruits.

Extrafloral nectaries at the base of each sepal attract several kinds of insects, but 96 percent of them are ants, several different species of them. When buds are still small, less than a quarter of an inch long, the sepal nectaries are already present and producing nectar. They continue to do so as the flower develops and while the fruit matures. Observations leave little doubt that ants protect morning glory flowers and fruits from the combined enemy force of grasshoppers, caterpillars, and seed beetles. Bentley compares the seed production of six plants that grew where there were no ants with that of seventeen plants that were occupied by ants. Unprotected plants bore only 45 seeds per plant, but plants occupied by ants bore 211 seeds per plant. Although ants are not big enough to kill or seriously injure grasshoppers, they drive them away by nipping at their feet. Seed beetles are more vulnerable because they are much smaller than grasshoppers. The ants prey on the adult beetles, disturb females as they lay their eggs on developing fruits, and eat many of the eggs they do manage to lay.

38 – 3 Transgenic Plants

Genes from virtually any organism, from viruses to humans, can now be inserted into plants, creating what are known as transgenic plants. Now used in agriculture, there are approximately 109 million acres of transgenic crops grown worldwide, 68 percent of which are in the United States. The most common transgenic crops are soybeans, corn, cotton, and canola. Most often, these plants either contain a gene making them resistant to the herbicide glyphosate or they contain an insect-resistant gene that produces a protein called Bt toxin.

On the positive side, proponents of transgenic crops argue that these crops are environmentally friendly because they allow farmers to use fewer and less noxious chemicals for crop production. For example, a 21 percent reduction in the use of insecticide has been reported on Bt cotton (transgenic cotton that produces Bt toxin). In addition, when glyphosate is used to control weeds, other more persistent herbicides do not need to be applied.

On the negative side, opponents of transgenic crops suggest that there are many questions that need to be answered before transgenic crops are grown on a large scale. One question deals with the effects that Bt plants have on nontarget organisms such as beneficial insects, worms, and birds that consume the genetically engineered crop. For example, monarch caterpillars feeding on milkweed plants near Bt cornfields will eat some corn pollen that has fallen on the milkweed leaves. Laboratory studies indicate that caterpillars can die from eating Bt pollen. However, field tests indicate that Bt corn is not likely to harm monarchs. Furthermore, the application of pesticides (the alternative to growing Bt plants) has been demonstrated to cause widespread harm to nontarget insects.

Another unanswered question is whether herbicide-resistant genes will move into the populations of weeds. Crop plants are sometimes grown in areas where weedy relatives also live. If the crop plants hybridize and reproduce with weedy relatives, then this herbicide-resistant gene will be perpetuated in the offspring. In this way, the resistant gene can make its way into the weed population. If this happens, a farmer can no longer use glyphosate, for example, to kill those weeds. This scenario is not likely to occur in many instances because there

are no weedy relatives growing near the crop plant. However, in some cases, it may become a serious problem. For example, canola readily hybridizes with mustard weed species and could transfer its herbicide-resistant genes to those weeds.

We know that evolution will occur when transgenic plants grown on a large scale over a period of time. Of special concern is the development of insect populations resistant to the Bt toxin. This pesticide has been applied to plants for decades without the development of insect-resistant populations. However, transgenic Bt plants express the toxin in all tissues throughout growing season. Therefore, all insects carrying genes that make them susceptible to the toxin will die. That leaves only the genetically resistant insects alive to perpetuate the population. When these resistant insects mate, they will produce a high proportion of offspring capable of surviving in the presence of the Bt toxin. Farmers are attempting to slow the development of insect resistance in Bt crops by, for example, planting nontransgenic border rows to provide a refuge for susceptible insects. These insects may allow Bt susceptibility to remain in the population.

Perhaps the most serious concern about the transgenic crop plants currently in use is that they encourage farmers to move farther away from sustainable agricultural farming practices, meaning ones that allow natural resources to continually regenerate over the long run. Transgenics, at least superficially, simplify farming by reducing the choices made by the manager. Planting a glyphosate-resistant crop commits a farmer to using that herbicide for the reason, probably to the exclusion of all other herbicides and other weed-control practices. Farmers who use Bt transgenics may not feel that they need to follow through with integrated pest-management practices that use beneficial insects and timely applications of pesticides to control insect pests. A more sustainable approach would be to plant nontransgenic corn, monitor the fields throughout the growing season, and then apply a pesticide only if and when needed.

72 – 1 Invading Algae

Biological invasions by exotic species are an escalating threat to biodiversity around the globe. Invasive species can displace native ones and irrevocably alter ecosystems. Most imported species either die or have limited growth in their new home, but a small percentage of the newcomers are opportunists whose growth is dangerously out of control. The story of how *Caulerpa taxifolia* became famous as one of these species began in 1989, when Dr. Alexandre Meinesz, a marine biologist at the University of Nice in France, learned that this exotic alga, which is native to the Caribbean and other tropical seas, was thriving in the Mediterranean waters off the coast of Monaco. Upon investigation, Meinesz discovered that a mutant variety of *C. taxifolia* was dumped into the Mediterranean with other aquarium refuse by the prestigious Oceanographic Institute of Monaco. Eventually, the *C. taxifolia* was traced back to a stock developed at the Wilhelma Zoological and Botanical Gardens in Stuttgart, Germany. This stock of *C. taxifolia* had caught the attention of amateur and professional aquarists for its exceptional beauty and hardiness and had been widely distributed in public and private aquaria in the 1980s.

Instead of dying in the cold winter waters of the Mediterranean Sea as the tropical variety of *C. taxifolia* does, the mutant variety can withstand several months at temperatures that range between 10° and 13° C. In addition, the alga grows equally well on a variety of ocean substrates---sand, mud, or rocks--and from near the surface to depths of 50 meters. When temperatures exceed 18°C, the mutant *Caulerpa* grows at a remarkable rate, capable of elongating by 2 centimeters in a single day and forming a new frond every 2 days. A single square meter of seafloor can be matted with 5,500 leafy fronds. All of this growth is asexual. Unlike the tropical form, the mutated form of *C. taxifolia* produces only male gametes in the Mediterranean and thus replicates only by fragmentation. Just a small piece can regenerate and colonize an area. In fact, it is believed that fishing nets and boat anchors as well as the ripping action of storms have helped the alga spread to new areas.

With such a rapid rate of growth, the Mediterranean *Caulerpa* grows quickly over the sea bottom, blanketing all other sea life such as marine angiosperms, native algae, corals, and sea fans. It

cuts off vital essentials such as sunlight for photosynthesis and currents for filter feeding of nutrients. The area becomes a *Caulerpa* lawn while species diversity is lost. In addition, *C. taxifolia* is inedible to most marine herbivores because of the presence of toxic compounds, primarily caulerpenyne. With no predator to inhibit its phenomenal growth, the alga had encompassed the Mediterranean coasts of Monaco, France, Italy, Spain, and the Croatian coast on the Adriatic Sea covering 30,000 hectares by 2002. *Caulerpa* infestations have also been found in southern Australia, in estuaries near Sydney and Adelaide.

Various methods have been tried to eradicate this deadly invader. Divers have been employed to rip the alga out by hand, but care must be taken because even a small fragment is capable of colonizing new areas. Black plastic tarps have been placed over the alga in an attempt to deprive it of life-giving sunlight. Chemicals such as copper, a conventional algicide, and toxic levels of salt have also been applied to *Caulerpa* fields. There has been limited success but only if the area of invasion is small.

One hopeful avenue is biological control. In biological control, a disease or predator of the invasive organism is introduced and allowed to destroy it. In the native range of the tropical *C. taxifolia*, two species of sea slug are immune to its poisons and feed on the alga exclusively. They suck the cytoplasm out of the leaves and incorporate the algal chloroplasts, the part of the plant that contains green photosynthetic pigment, into their skin. The chloroplasts provide camouflage as the snails feed in the *Caulerpa* fields and, surprisingly, continue to photosynthesize, providing additional nutrition for the slugs. Dr. Meinesz is suggesting the importation of these slugs to the Mediterranean as the only way to eliminate the immense area that *Caulerpa* now covers. But many experts fear that another exotic importation may lead to more environmental damage.

73 – 1 How Effective Are Structural Plant Defenses?

Some of the least costly defenses, in terms of energy use, available to plants are structures that make penetration by predators difficult, if not impossible. These include tough leaves, spines, and epidermal hairs on leaves, which may trap, impale, or fence out insects and discourage browsing by vertebrate herbivores, as well as hard-coated seeds. These structures may have evolved early in the history of the plants, when they might have been subject to even greater predatory pressures. Because they represent little investment, plants still retain them.

Many seeds have thick, hard seed coats that provide protection from seed-eating animals. The problem with such seed defense is that the seeds need to be scarified—the hard seed needs scratching or scoring of the seed coat to weaken it—so the seedling itself can escape. If the seed is not scarified, the seedling embryo is sealed in, never to germinate. Many plants, however, have turned seed predation into a mechanism for seed dispersal, such as the transport and caching of seeds by squirrels, jays, mice, and ants.

The role of structural defenses in plants was largely presumed until experimental evidence was sought to demonstrate the effectiveness of such apparent defensive structures against grazing herbivores. Researchers investigated experimentally the effects of plant spinescence on the feeding habits of three large browsing mammalian herbivores of Africa. These were the kudu, a large African antelope attaining a female body weight of 180 kilograms; the impala, a medium-sized African antelope attaining a female body weight of 50 kg; and the Boer goat, a domestic ungulate weighing about 35 kg. The experimenters hand reared the antelope from calves to allow observation of feeding habits from very close range, 1 to 5 meters, under natural conditions, to determine biting rates the animals employed. They converted the bites to dry biomass by collecting samples of leaves and shoots of a size similar to those eaten and drying them to a constant weight. They calculated eating rate as the product of bite size (dry mass) and biting rate.

Another one of their experiments involved a detailed study to examine the influence of spinescence. They selected ten plants each of five species of trees at a height accessible to impalas

outside the enclosure. The woody plants exhibited three basic types of spinescence: paired prickles or thorns situated in or close to the leaf axils; short, sharp-tipped branchlets or spines, sometimes carrying small leaves; and prickles of various kinds on leaves. Thorns were either straight and long, up to 70 millimeters, or short and sharply curved (hooked). On each tree, two branches were matched for size, shape, density of leaf cover, and ease of access to impalas. These paired branches were labeled and the thorns were removed from one of the branches in each pair. Two months later, researchers visually estimated the relative loss of foliage from browsing.

Results clearly showed that thorns and spines affected the feeding behavior of the three ungulates. These structures restricted bite sizes to mostly single leaves or leaf clusters, and hooked thorns retarded biting rates. Acceptability of leaves of those plant species offering small leaf size along with prickles was lower, at least for kudu, than those of other palatable plant species. The inhibitory effect of prickles was greater for impalas and goats than for kudu, which bit off the shoot ends in spite of the prickles. For certain straight-thorned species, kudu compensated partially for their slow eating rates by spending more time gathering the leaves. Most spinescent species were similar to unarmed palatable species in their acceptability to the ungulates, even though the armed species had a higher crude protein to their foliage. Probably these spinescent species, especially species of *Acacia* plants, would be preferred over unarmed species but for the thorns.

The main effect of these armed structural defense features is to restrict bite size, thus increasing handling time. Thorns, spines, and prickles restrict foliage losses to large herbivores. In addition, the animals may incur scar tissue in the esophagus and scratches in the mouth and throat. These experiments show that at least in some cases structural plant defenses are effective at protecting plants.

44 – 3 Seagrasses

Many areas of the shallow sea bottom are covered with a lush growth of aquatic flowering plants adapted to live submerged in seawater. These plants are collectively called seagrasses. Seagrass beds are strongly influenced by several physical factors. The most significant is water motion: currents and waves. Since seagrass systems exist in both sheltered and relatively open areas, they are subject to differing amounts of water motion. For any given seagrass system, however, the water motion is relatively constant. Seagrass meadows in relatively turbulent waters tend to form a mosaic of individual mounds, whereas meadows in relatively calm waters tend to form flat, extensive carpets. The seagrass beds, in turn, dampen wave action, particularly if the blades reach the water surface. This damping effect can be significant to the point where just one meter into a seagrass bed the wave motion can be reduced to zero. Currents are also slowed as they move into the bed.

The slowing of wave action and currents means that seagrass beds tend to accumulate sediment. However, this is not universal and depends on the currents under which the bed exists. Seagrass beds under the influence of strong currents tend to have many of the lighter particles, including seagrass debris, moved out, whereas beds in weak current areas accumulate lighter detrital material. It is interesting that temperate seagrass beds accumulate sediments from sources outside the beds, whereas tropical seagrass beds derive most of their sediments from within.

Since most seagrass systems are depositional environments, they eventually accumulate organic material that leads to the creation of fine-grained sediments with a much higher organic content than that of the surrounding unvegetated areas. This accumulation, in turn, reduces the water movement and the oxygen supply. The high rate of metabolism (the processing of energy for survival) of the microorganisms in the sediments causes sediments to be anaerobic (without oxygen) below the first few millimeters. According to ecologist J. W. Kenworthy, anaerobic processes of the microorganisms in the sediment are an important mechanism for regenerating and recycling nutrients and carbon, ensuring the high rates of productivity—that is, the amount of organic material produced—that are measured in those beds. In contrast to other productivity in the ocean, which is confined to various species of algae and bacteria

dependent on nutrient concentrations in the water column, seagrasses are rooted plants that absorb nutrients from the sediment or substrate. They are, therefore, capable of recycling nutrients into the ecosystem that would otherwise be trapped in the bottom and rendered unavailable.

Other physical factors that have an effect on seagrass beds include light, temperature, and desiccation (drying out). For example, water depth and turbidity (density of particles in the water) together or separately control the amount of light available to the plants and the depth to which the seagrasses may extend. Although marine botanist W. A. Setchell suggested early on that temperature was critical to the growth and reproduction of eelgrass, it has since been shown that this particularly widespread seagrass grows and reproduces at temperatures between 2 and 4 degrees Celsius in the Arctic and at temperatures up to 28 degrees Celsius on the northeastern coast of the United States. Still, extreme temperatures, in combination with other factors, may have dramatic detrimental effects. For example, in areas of the cold North Atlantic, ice may form in winter. Researchers Robertson and Mann note that when the ice begins to break up, the wind and tides may move the ice around, scouring the bottom and uprooting the eelgrass. In contrast, at the southern end of the eelgrass range, on the southeastern coast of the United States, temperatures over 30 degrees Celsius in summer cause excessive mortality. Seagrass beds also decline if they are subjected to too much exposure to the air. The effect of desiccation is often difficult to separate from the effect of temperature. Most seagrass beds seem tolerant of considerable changes in salinity (salt levels) and can be found in brackish (somewhat salty) waters as well as in full-strength seawater.

44 – 4 What is a Community?

The Black Hills forest, the prairie riparian forest, and other forests of the western United States can be separated by the distinctly different combinations of species they comprise. It is easy to distinguish between prairie riparian forest and Black Hills forest—one is a broad-leaved forest of ash and cottonwood trees, the other is a coniferous forest of ponderosa pine and white spruce trees. One has kingbirds; the other juncos (birds with white outer tail feathers). The fact that ecological communities are indeed, recognizable clusters of species led some early ecologists, particularly those living in the beginning of the twentieth century, to claim that communities are highly integrated, precisely balanced assemblages. This claim harkens back to even earlier arguments about the existence of a balance of nature, where every species is there for a specific purpose, like a vital part in a complex machine. Such a belief would suggest that to remove any species, whether it be plant, bird, or insect, would somehow disrupt the balance, and the habitat would begin to deteriorate. Likewise, to add a species may be equally disruptive.

One of these pioneer ecologists was Frederick Clements, who studied ecology extensively throughout the Midwest and other areas in North America. He held that within any given region of climate, ecological communities tended to slowly converge toward a single endpoint, which he called the "climatic climax". This "climax" community was, in Clements's mind, the most well-balanced, integrated grouping of species that could occur within that particular region. Clements even thought that the process of ecological succession—the replacement of some species by others over time—was somewhat akin to the development of an organism, from embryo to adult. Clements thought that succession represented discrete stages in the development of the community (rather like infancy, childhood, and adolescence), terminating in the climatic "adult" stage, when the community became self-reproducing and succession ceased. Clements's view of the ecological community reflected the notion of a precise balance of nature.

Clements was challenged by another pioneer ecologist, Henry Gleason, who took the opposite view. Gleason viewed the community as largely a group of species with similar tolerances to the stresses imposed by climate and other factors typical of the region. Gleason saw the element of

chance as important in influencing where species occurred. His concept of the community suggests that nature is not highly integrated. Gleason thought succession could take numerous directions, depending upon local circumstances.

Who was right? Many ecologists have made precise measurements, designed to test the assumptions of both the Clements and Gleason models. For instance, along mountain slopes, does one life zone, or habitat type, grade sharply or gradually into another? If the divisions are sharp, perhaps the reason is that the community is so well integrated, so holistic, so like Clements viewed it, that whole clusters of species must remain together. If the divisions are gradual, perhaps, as Gleason suggested, each species is responding individually to its environment, and clusters of species are not so integrated that they must always occur together.

It now appears that Gleason was far closer to the truth than Clements. The ecological community is largely an accidental assemblage of species with similar responses to a particular climate. Green ash trees are found in association with plains cottonwood trees because both can survive well on floodplains and the competition between them is not so strong that only one can persevere. One ecological community often flows into another so gradually that it is next to impossible to say where one leaves off and the other begins. Communities are individualistic.

This is not to say that precise harmonies are not present within communities. Most flowering plants could not exist were it not for their pollinators—and vice versa. Predators, disease organisms, and competitors all influence the abundance and distribution of everything from oak trees to field mice. But if we see a precise balance of nature, it is largely an artifact of our perception, due to the illusion that nature, especially a complex system like a forest, seems so unchanging from one day to the next.

47 – 3 Coral Reefs

An important environment that is more or less totally restricted to the intertropical zone is the coral reef. Coral reefs are found where the ocean water temperature is not less than 21 °C, where there is a firm substratum, and where the seawater is not rendered too dark by excessive amounts of river-borne sediment. They will not grow in very deep water, so a platform within 30 to 40 meters of the surface is a necessary prerequisite for their development. Their physical structure is dominated by the skeletons of corals, which are carnivorous animals living off zooplankton. However, in addition to corals there are enormous quantities of algae, some calcareous, which help to build the reefs. The size of reefs is variable. Some atolls are very large—Kwajalein in the Marshall Islands of the South Pacific is 120 kilometers long and as much as 24 kilometers across—but most are very much smaller, and rise only a few meters above the water. The 2,000 kilometer complex of reefs known as the Great Barrier Reef, which forms a gigantic natural breakwater off the northeast coast of Australia, is by far the greatest coral structure on Earth.

Coral reefs have fascinated scientists for almost 200 years, and some of the most pertinent observations of them were made in the 1830s by Charles Darwin on the voyage of the *Beagle*. He recognized that there were three major kinds: fringing reefs, barrier reefs, and atolls; and he saw that they were related to each other in a logical and gradational sequence. A fringing reef is one that lies close to the shore of some continent or island. Its surface forms an uneven and rather rough platform around the coast, about the level of low water, and its outer edge slopes downwards into the sea. Between the fringing reef and the land there is sometimes a small channel or lagoon. When the lagoon is wide and deep and the reef lies at some distance from the shore and rises from deep water it is called a barrier reef. An atoll is a reef in the form of a ring or horseshoe with a lagoon in the center.

Darwin's theory was that the succession from one coral reef type to another could be achieved by the upward growth of coral from a sinking platform, and that there would be a progression from a fringing reef, through the barrier reef stage until, with the disappearance through subsidence (sinking) of the central island, only a reef-enclosed lagoon or atoll would survive. A long time after Darwin put forward

this theory, some deep boreholes were drilled in the Pacific atolls in the 1950s. The drill holes passed through more than a thousand meters of coral before reaching the rock substratum of the ocean floor, and indicated that the coral had been growing upward for tens of millions of years as Earth's crust subsided at a rate of between 15 and 51 meters per million years. Darwin's theory was therefore proved basically correct. There are some submarine islands called guyots and seamounts, in which subsidence associated with sea-floor spreading has been too speedy for coral growth to keep up.

Like mangrove swamps, coral reefs are extremely important habitats. Their diversity of coral genera is greatest in the warm waters of the Indian Ocean and the western Pacific. Indeed, they have been called the marine version of the tropical rain forest, rivaling their terrestrial counterparts in both richness of species and biological productivity. They also have significance because they provide coastal protection, opportunities for recreation, and are potential sources of substances like medicinal drugs. At present they are coming under a variety of threats, of which two of the most important are dredging and the effects of increased siltation brought about by accelerated erosion from neighboring land areas.

50 – 4 Forest Succession

Succession is a continuous change in the species composition, structure, and function of a forest through time following disturbance. Each stage of succession is referred to as a successional sere. The final stage of succession, which is generally self-replacing, is referred to as the climax sere. There are two major types of succession: primary and secondary. Primary succession is the establishment of vegetation on bare rocks or radically disturbed soil. Secondary succession is the reestablishment of vegetation following a disturbance that killed or removed the vegetation but did not greatly affect the soil. Volcanic eruptions, retreating glaciers, and bare sand dunes are examples of sites subject to primary succession, while clear-cutting of forests, wild fires, and hurricanes are examples of sites subject to secondary succession. Hundreds to thousands of years are required for primary succession to reach the climax sere, compared to decades to hundreds of years for it to occur in secondary succession. A longer time is needed to reach the climax sere for primary than secondary succession because soil development must first take place in primary succession. The rate of succession is dependent upon the extent of the disturbance and the availability of appropriate seeds for recolonization.

What morphological (structural) and ecophysiological characteristics determine the species composition and abundance in succession? In general, nitrogen fixing plants (plants that can make use of atmospheric nitrogen) are important early succession species in primary succession because nitrogen is not derived from the weathering of rock and little or no organic matter is present in the soil. Weedy plants are common early successional species because of their rapid growth and high reproductive rates, while stress-tolerant species are common late successional species.

The structure of a forest changes as well in secondary succession. Depending on the type and the severity of the disturbance, a moderate to large amount of dead organic matter from the previous forest remains on the site immediately from the disturbance. The leaf area of the forest is at a minimum and slowly increases as new vegetation occupies the site. Following a disturbance, such as a fire, the new canopy (the uppermost spreading and branching layer of a forest) is largely composed of similar-aged, or even-aged, trees.

Light, nutrient, and water availability are highest during the early successional sere because the vegetation has not completely occupied the site. Canopy closure, or maximum leaf area, can occur within several years after disturbance in some tropical forests, but may take three to fifty years in evergreen forests.

In the second stage of forest development there is tree mortality caused by competition for light, nutrients and water. The intense intraspecies (within a species) and interspecies (between species) competition for light, nutrients and water induces the mortality of plants that are shaded or have one or more life-history characteristics that are not well adapted to the changing environment. The third stage of forest development is characterized by openings in the overstory canopy, caused by tree mortality, and the renewed growth of understory in response to increased light reaching the forest floor. Consequently, the forest canopy becomes more complex, or multilayered. The final stage of forest development, the climax or old growth stage, is characterized by a species composition that in theory can continue to replace itself unless a catastrophic disturbance occurs. Unique characteristics of old growth forests include large accumulation of standing and fallen dead trees--referred to as coarse woody debris. Also, the annual input of forest litter is dominated by coarse woody debris compared to the earlier stages of forest development, when leaf and fine root debris were the dominant sources of nutrients and organic matter input into the soil.

Some ecosystems may never reach the latter stages of succession if natural disturbances (fire, flooding, hurricanes, etc.) are frequent. A pyric climax refers to an ecosystem that never reaches the potential climax vegetation defined by climate because of frequent fires. The ecotone, a boundary, between grassland and forest is a pyric climax, and only with fire suppression have woodlands and forests began to advance into these regions.

53 – 2 Rain Forest Soils

On viewing the lush plant growth of a tropical rain forest, most people would conclude that the soil beneath it is rich in nutrients. However, although rain forest soils are highly variable, they have in common the fact that abundant rainfall washes mineral nutrients out of them and into streams. This process is known as leaching. Because of rain leaching, most tropical rain forest soils have low to very low mineral nutrient content, in dramatic contrast to mineral-rich grassland soils. Tropical forest soils also often contain particular types of clays that, unlike the mineral-binding clays of temperate forest soils, do not bind mineral ions well. Aluminum is the dominant cation (positively charged ion) present in tropical soils; but plants do not require this element, and it is moderately toxic to a wide range of plants. Aluminum also reduces the availability of phosphorus, an element in high demand by plants.

High moisture and temperatures speed the growth of soil microbes that decompose organic compounds, so tropical soils typically contain far lower amounts of organic materials (humus) than do other forest or grassland soils. Because organic compounds help loosen compact clay soils, hold water, and bind mineral nutrients, the relative lack of organic materials in tropical soils is deleterious to plants. Plant roots cannot penetrate far into hard clay soils, and during dry periods, the soil cannot hold enough water to supply plant needs. Because the concentration of dark-colored organic materials is low in tropical soils, they are often colored red or yellow by the presence of iron, aluminum, and manganese oxides; when dry, these soils become rock hard. The famous Cambodian temples of Angkor Wat, which have survived for many centuries, were constructed from blocks of such hard rain forest soils.

Given such poor soils, how can lush tropical forests exist? The answer is that the forest's minerals are held in its living biomass—the trees and other plants and the animals. In contrast to grasslands, where a large proportion of plant biomass is produced underground, that of tropical forests is nearly all aboveground. Dead leaves, branches, and other plant parts, as well as the wastes and bodies of rain forest animals, barely reach the forest floor before they are rapidly decayed by abundant decomposers—bacterial and fungal. Minerals released by decay are quickly absorbed by multitudinous shallow, fine tree feeder roots and

stored in plant tissues. Many tropical rain forest plants (like those in other forests) have mycorrhizal (fungus-root) partners whose delicate hyphae spread through great volumes of soil, from which they release and absorb minerals and ferry them back to the host plant in exchange for needed organic compounds. The fungal hyphae are able to absorb phosphorus that plant roots could not themselves obtain from the very dilute soil solutions, and fungal hyphae can transfer mineral nutrients from one forest plant to another. Consequently, tropical rain forests typically have what are known as closed nutrient systems, in which minerals are handed off from one organism to another with little leaking through to the soil. When mineral nutrients do not spend much time in the soil, they cannot be leached into streams. Closed nutrient systems have evolved in response to the leaching effects of heavy tropical rainfall. Evidence for this conclusion is that nutrient systems are more open in the richest tropical soils and tightest in the poorest soils.

The growth of organisms is dependent on the availability of nutrients, none of which is more important than nitrogen. Although there is an abundant supply of nitrogen in Earth's atmosphere, it cannot be absorbed by plants unless it is "fixed," or combined chemically with other elements to form nitrogen compounds. Nitrogen-fixing bacteria help tropical rain forest plants cope with the poor soils there by supplying them with needed nitrogen. Many species of tropical rain forest trees belong to the legume family, which is known for associations of nitrogen-fixing bacteria within root nodules. Also, cycads (a type of tropical plant that resembles a palm tree) produce special aboveground roots that harbor nitrogen-fixing cyanobacteria. By growing above the ground, the roots are exposed to sunlight, which the cyanobacteria require for growth. Nitrogen fixation by free-living bacteria in tropical soils is also beneficial.

55 – 2 Evolution of the Flowering Plants

Many aspects of the history of flowering plants (angiosperms) remain mysterious. Evidence of the earliest angiosperms comes from fossilized leaves, stems, fruits, pollen, and, very rarely, flowers. In addition, there has been much study of modern plant morphology (structure) and genetics in order to determine which living species might be most closely related to the ancient ancestors of angiosperms. Despite intensive efforts for over 200 years, scientists have still not reached consensus on which type of plant was the ancestor to the angiosperms, and when and where the angiosperms first evolved. Indeed, Charles Darwin himself called the origin of the flowering plants an "abominable mystery."

What type of plant was the ancestor to the angiosperms? Most botanists now agree that the flowering plants are monophyletic in origin, meaning that they evolved from a common ancestor. Some paleontologists have suggested that the common ancestor may have been a type of cycad (palmlike tropical plants). Other paleontologists maintain that the angiosperms may have evolved from seed-bearing ferns. Finally, analysis of the morphological traits of some primitive living plants suggests that the ancestor may have been related to the modern pines. The question of angiosperm ancestry remains unresolved.

The time and place of the first appearance of flowering plants have long been a topic of great interest. There is good fossil evidence that early angiosperms, including a number resembling modern magnolias, were present in the Early Cretaceous geologic period (more than 100 million years ago). Angiosperms became increasingly abundant during this period. Between 100 million and 65 million years ago, a period known as the Late Cretaceous, angiosperms increased from less than 1 percent of flora (plant life) to well over 50 percent. Many of the modern plant families appeared during this time period. In the Early Tertiary period which followed: angiosperms increased to comprise 90 percent or more of Earth's total flora. Where did these successful plants first originate and spread from?

Analysis of the fossil leaf structure and geographic distribution of the earliest Cretaceous angiosperms has led many biogeographers to conclude that they evolved in the tropics and then migrated poleward.

It is known that angiosperms did not become dominant in the high latitudes until the Late Cretaceous. Paleontologists have recovered fossil angiosperm leaves, stems, and pollen from Early Cretaceous deposits in eastern South America and western Africa. These two continents were joined together as part of Gondwanaland, one of two supercontinents that existed at that time. The locations of these early angiosperm finds would have been close to the equator during the Early Cretaceous and are conformable with a model by which angiosperms spread from the tropics poleward.

Not all botanists agree with an African-South American center for the evolution and dispersal of the angiosperms, pointing out that many of the most primitive forms of flowering plants are found in the South Pacific, including portions of Fiji, New Caledonia, New Guinea, eastern Australia, and the Malay Archipelago. Recent genetic research has identified the rare tropical shrub *Amborella* as being the living plant most closely related to the ancient ancestor of all the angiosperms. This small shrub, which has tiny yellow-white flowers and red fruit; is found only on New Caledonia, a group of islands in the South Pacific. Many botanists conclude that the best explanation for the large numbers of primitive living angiosperms in the South Pacific region is that this is where the flowering plants first evolved and these modern species are relics of this early evolution. Comparisons of the DNA of *Amborella* and many hundreds of species of flowering plants suggest that the first angiosperm arose and the development of separate species occurred about 135 million years ago.

Recently discovered fossils complicate our understanding of the origin of the angiosperms even further. Paleontologists from China have found beautifully preserved fossils of an angiosperm plant, including flowers and seeds, in Jurassic period deposits from China. The site, which is about 130 million years old, is near modern Beijing. The new fossil plant found at the site is now the oldest known angiosperm. The age of the fossils and the very primitive features of the flowers have led the discoverers to suggest that the earliest flowering plants may have evolved in northern Asia.

61 – 3 Grasses of the Prairie

Around 140 species of grasses naturally occur in the arid Great Plains grasslands of North America. One key to the prairie grasses' success is their ability to conserve water in a dry environment. Like most plants, grasses take in water through their roots and lose it as water vapor through tiny mouth-shaped valves, or stomata, in their leaves. The larger the surface of the leaf and the more stomata it bears, the greater the risk that the plant will lose too much moisture through evaporation and collapse. Grasses are protected from this trauma by having a reduced number of stomata and by the design of their leaves, which take the form of narrow blades. What's more, the surfaces of these reduced leaves are often modified—corrugated with ridges or covered in hairs—so that the wind can't sweep across the surface and draw out moisture. The roughened surface holds a thin layer of humid air next to the leaf and thus helps to reduce the "evaporative demand," or drying power, of the atmosphere. Some grasses, including western wheatgrass, June grass, and blue grama, roll up the edges of their leaves during times of drought to help keep their tissues from drying out.

Why aren't the stomata kept tightly closed to seal moisture inside the leaf? The reason is that the stomata also supply plants with fresh air for photosynthesis. If plants sealed their stomata, this life-sustaining process would come to a halt for lack of carbon dioxide. But if the stomata are thrown wide open, the plants risk death due to the loss of moisture through their gaping valves. Prairie grasses resolve this dilemma by strategic scheduling. In the fierce blaze of the midday sun, the stomata close so that water vapor is held in and carbon dioxide is kept out. In this state, the leaf can capture solar energy and store it in energy-rich molecules (a process that requires sunlight but not carbon dioxide). Then, in the cool of the evening, when the evaporative demand drops off, the stomata snap open, letting water vapor trickle out but also permitting carbon dioxide to flood into the leaf. By mobilizing the energy that was stockpiled earlier in the day, the leaf uses this carbon dioxide to manufacture the sugars and other molecules that it needs for growth (a process that can be accomplished in total darkness). The result is that prairie grasses are partially nocturnal: they do most of their growing at night or in the early hours of the morning.

Prairie grasses also have another ingenious way of evading the demands of the sun. Like many grassland creatures (prairie dogs, ground squirrels, cottontails, badgers, and so on), they take refuge underground. What we think of as "grass"—the aboveground leaves and stems—actually constitutes less than half of the organism. Between 60 and 80 percent of the plant, by weight, typically grows below ground. A 3-meter stand of big bluestem is anchored underground by a mass of coarse, fibrous roots that reaches as much as 3.6 meters into the earth. Blue grama, for its part, seldom lifts its seed heads very far above the ground, but its network of fine, branching roots can sometimes probe the soil for water almost two meters down.

These extensive systems of roots push thirstily through the soil, intent on sucking up every available drop of water. But if the soil is very dry, as it is during periods of drought, the roots can't draw in enough moisture to keep pace with losses from the stomata. Grasses respond by transferring their most valuable resources (including sugars and proteins) from their leaves into their roots and, especially, into their rhizomes (underground stems). Dead to the world above ground—withered and crisp—the plants live frugally below the surface, drawing on their cached supplies and biding their time until the weather improves. When the rains eventually return, as they inevitably do, the grasses explode into action, sending out fresh rhizomes, which in turn put out fresh leaves and roots, to produce a burgeoning network of tender growth. The amazingly resilient blue grama can revive from dormancy, become green, and grow on as little as five millimeters of rainfall.

62 – 4 What Controls Flowering

The timing of flowering and seed production is precisely tuned to a plant's physiology and the rigors of its environment. In temperate climate plants lost flower early enough so that their seeds can mature before the deadly winds of autumn. Depending on how quickly the seed and food develop flowering may occur in spring as it does in oaks; in summer as in lettuces; or even in autumn as in asters.

What environmental cues do plants use to determine the seasons? Most cues such as temperature or water availability are quite variable: autumn can be warm; a late snow could fall in spring; also summer might be unusually cool and wet. So the only reliable cue is day length: longer days always mean that spring and summer are coming; shorter days foretell the onset of autumn and winter.

With respect to flowering botanists classify plants as day neutral, long day or short day. A day neutral plant flowers as soon as it has sufficiently grown and developed regardless of the length of day. The neutral plants include tomatoes, corn, snapdragons and roses. Although the naming is traditional, long day and short day plants are better described as short night and long night plants because their flowering actually depends on the duration of continuous darkness rather than on day length. Short night plants (which include lettuces, spinach, iris, clover and petunias) flower when the length of darkness is shorter than a species' specific critical period. Long night plants (including asters, potatoes, soy beans, goldenrod and cockleburs) flower when the length of uninterrupted darkness is longer than the species' specific critical period. Thus spinach is classified as a short night plant because it flowers only if the night is shorter than eleven hours (its critical period), and the cocklebur is a long night plant because it flowers only if an uninterrupted darkness lasts more than 8.5 hours. Both of these plants will flower with ten-hour nights.

Plant scientists can induce flowering. Plant scientists can induce flowering in the cocklebur by exposing leaves to long nights (longer than its 8.5 hour critical period) in a special chamber, while the rest of the plant continues to experience short nights. Clearly, a signal that induces flowering transmitted from the leaf to the flowering bud. Plant physiologists have been attempting for decades to isolate these elusive signaling molecule

often called florigen (literally, flowering maker). Some researchers believe they are close to demonstrating a flower's stimulating substance for specific type of plant. Using genetic manipulation, it is likely, however, that interactions among multiple and yet unidentified plant hormones stimulate or inhibit flowering, and that these chemicals may differ among plant species. Researchers have had more success in determining how plants measure the length of uninterrupted darkness, which is a crucial stimulus for producing whatever substance control flowering.

To measure continuous darkness, a plant needs two things: some sort of metabolic clock to measure time (the duration of darkness) and a light detecting system to set the clock. Virtually all organisms have an internal biological clock that measures the time even without environmental cues. In most organisms including plants, the biological clock is poorly understood, but we know that the environmental cues, particularly light, can reset the clock. How do plants detect light? The light detecting system of plants is a pigment in leaves called phytochrome (literally, plant color).

Plants seem to use the phytochrome system in combination with their internal biological clocks to detect the duration of continuous darkness. Cockleburs, for example, flower under the schedule of sixteen hours of darkness and eight hours of light. However, interrupting the middle of the dark period with just a minute or two of lights prevents flowering. Thus their flowering is controlled by the length of continuous darkness. It is evident that even brief exposure to sunlight or white light will reset their biological clocks. The color of the light used for the light exposure is also important. A nighttime flash of pure red light inhibits flowering, while flash of light at the far-red end of the spectrum has no effect on flowering, as if no light were detected.

62 – 1 Plant Adaptations to Cool Environments

There are many interesting adaptations that allow plants to survive in cool environments. One obvious strategy is dormancy (a suspension of activity) during the cold season. Most of the common trees in the forests of northeastern North America, western Europe, and eastern Asia, such as the maples, oaks, beech, birches, and ashes, are deciduous trees that lose their frost-sensitive leaves during the cold winter season. In most of these trees, the leaves suffer damage at temperatures of freezing or just below. The new leaves arise in the spring from winter buds that can remain viable at colder temperatures.

Most of the needle-leaved conifers of the northern and alpine forests, such as pines, spruces, and firs, do not lose their leaves during the winter. How do such evergreen plants escape intracellular freezing (freezing within cells) and tissue destruction when temperatures may drop to -40°C or colder? In these plants, the onset of cool temperatures causes physiological changes that allow plant tissue to either avoid freezing or restrict freezing to extracellular areas (those outside of cells). For plants to avoid freezing, they must chemically alter their liquids into a form that is analogous to antifreeze in automobiles. The liquids in these plants can be cooled far below 0°C and will not freeze. This process is called supercooling and is achieved by the metabolic synthesis of sugars and other molecules which, when in solution in the plant's tissue, lower the temperature for ice formation to far below 0°C . Supercooling seems to be the prevalent mechanism of frost resistance in herbs. For woody plants, supercooling is augmented by declines of cellular water content, greater cellular accommodation to deformation, and processes that allow water to accumulate and freeze in extracellular spaces. The loss of water from the cells to extracellular areas increases the solute content (the quantity of dissolved substances) of the remaining cell water, making it more resistant to freezing. The cell walls can accommodate the deformations caused by water freezing on the exterior of the cell. For northern and alpine evergreens such as pines and spruces, both supercooling and extracellular ice formation play a part in allowing the plants to withstand extremely cold temperatures. One interesting facet of these physiological adaptations to freezing is that most of these plants will still be damaged by cold temperatures if they do not have a period of cooling

in which to adjust to the onset of winter. This process of physiological preparation for the onset of winter cold is called frost hardening.

Some members of the cactus family appear to resist freezing during cool nights by radiating heat stored during the day in their thick, moist tissue. The greater the mass of the cactus, the more heat it can store and the less prone it will be to freezing damage during the night. How is it then that these cacti can survive cold temperatures when they are young and small? The giant saguaro cactus is perhaps the best-known symbol of the southwestern desert of North America. In the popular lore of North America, the distinctive shape of the multi-stemmed saguaro is a universally recognized icon that is used to represent deserts in movies, television, and comics. Yet, this distinctive and widely recognized plant is actually found only in the Sonoran Desert of California, Arizona, and adjacent Mexico. The saguaro cactus is damaged or killed if exposed to prolonged freezing temperatures. Desert climates in the northern Sonoran Desert are typified by warm days but sometimes experience nighttime temperatures that are below freezing in the winter. Young saguaros that survive are found sheltered beneath more frost-tolerant desert shrubs. The cover of these shrubs acts as a thermal blanket, capturing heat radiated from the ground and keeping the microclimate of the small saguaro warm at night. As the cactus grows, it eventually rises above the cover of the protective shrub. The radiation of heat from the stalk of the large mature cactus prevents freezing. This strategy works up to a point. The range of the saguaro is restricted to areas that do not experience more than about 12 to 24 continuous hours of air temperatures below 0°C . It appears that after 24 hours of freezing air temperatures, not enough heat reserve is left in the saguaro to keep the tissue from freezing.

67 – 3 Taxonomy of Organisms

Classification schemes are used by biologists to place the huge number of organisms on Earth into natural groupings. Ideally, these groupings are made by taxonomists on the basis of shared distinguishing features. Today taxonomists use such features as anatomy, developmental stages, and biochemical similarities to categorize organisms. Early classification schemes placed all organisms into either the plant or animal kingdom. Later, close examination of the unique structure of fungi and the diversity of single-celled organisms made it necessary to propose additional kingdoms that recognized the fundamental differences among plants, animals, fungi, and unicellular prokaryotes (organisms whose cells do not have a distinct membrane-bound nucleus) and eukaryotes (organisms whose cells have a distinct membrane-bound nucleus). The current scheme consisting of five kingdoms—Monera, Protista, Animalia, Fungi, and Plantae—was devised in response to this need. Scientists do not know how many species share our world. Each year 7000 to 10,000 new species are named. The total number of named species is currently around 1.4 million. However, many scientists believe that 7 million to 10 million species may exist, and estimates range as high as 30 million. Of all of the species that have been identified, about 5 percent are in the Monera and Protista kingdoms. An additional 22 percent are plants and fungi, and the rest are animals. This distribution has little to do with the actual abundance of these organisms and a lot to do with the size of the organisms, how easy they are to classify, and the number of scientists studying them.

The kingdom Protista, defined as comprising all single-celled eukaryotic organisms, is not a natural grouping and scientists disagree about which organisms it should include. Plants, animals, and fungi all have close protistan relatives and the separation of single-celled organisms from multicellular organisms is sometimes problematic. It is especially so for the algae, which have both single-celled and multicellular representatives within most smaller taxonomic groupings. Can closely related organisms be placed into separate kingdoms, Protista and Plantae, simply on the basis of multicellularity? If you look at different textbooks, you will see that the algae, photosynthetic organisms with simple reproduction, are sometimes placed entirely into Protista, and

sometimes they are split between Protista and Plantae depending on whether they are single-celled or multicellular. Some taxonomists split the multicellular algae into two kingdoms, placing the multicellular brown and red algae with the protists and the multicellular green algae into the plant kingdom. These different attempts to classify closely related organisms are good examples of how difficult it is to develop standard criteria for grouping organisms, even at the kingdom level.

One approach to this problem, enthusiastically endorsed by Lynn Margulis, a biologist at the University of Massachusetts, is the creation of the kingdom Protocista. This taxonomic category would include single-celled organisms and their close descendants (for example, the multicellular algae but not the animals, fungi, and plants). Margulis describes the kingdom Protocista as "the entire motley and unruly group of nonplant, nonanimal, nonfungal organisms representative of lineages of the earliest descendants of the eukaryotes."

It is conceptually difficult to group one of the largest multicellular organisms in the world, the brown algae called giant kelps, with simple microscopic single-celled organisms. Kelps, some of which are up to 60 meters long, possess a tissue-like level of organization that is relatively complex and can transport materials over long distances, as can the tissues of higher plants. The cells in kelps and some other algae are specialized and show division of labor. However, kelps reproduce like other algae and differently from plants. Thomas Cavalier-Smith of the University of British Columbia has proposed that brown algae merit their own kingdom (kingdom Chromista) based on ultrastructural features and molecular comparisons of all algae. So, even among the algae, there are clear differences that some scientists believe are sufficient to justify the status of separate kingdom.

As we learn more about the relationships between organisms and refine the criteria used to classify them, classification schemes will change. As the superficially simple question "In which kingdom should we place the algae?" illustrates, the taxonomic categories in textbooks are tentative and subject to revision as we continue to discover more about life on Earth.

ANIMALS / SPECIES / ZOOLOGY

02 – 3 The Origins of Cetaceans

It should be obvious that cetaceans-whales, porpoises, and dolphins-are mammals. They breathe through lungs, not through gills, and give birth to live young. Their streamlined bodies, the absence of hind legs, and the presence of a fluke and blowhole cannot disguise their affinities with land dwelling mammals. However, unlike the cases of sea otters and pinnipeds (seals, sea lions, and walruses, whose limbs are functional both on land and at sea), it is not easy to envision what the first whales looked like. Extinct but already fully marine cetaceans are known from the fossil record. How was the gap between a walking mammal and a swimming whale bridged? Missing until recently were fossils clearly intermediate, or transitional, between land mammals and cetaceans.

Very exciting discoveries have finally allowed scientists to reconstruct the most likely origins of cetaceans. In 1979, a team looking for fossils in northern Pakistan found what proved to be the oldest fossil whale. The fossil was officially named *Pakicetus* in honor of the country where the discovery was made. *Pakicetus* was found embedded in rocks formed from river deposits that were 52 million years old. The river that formed these deposits was actually not far from an ancient ocean known as the Tethys Sea.

The fossil consists of a complete skull of an archaeocyte, an extinct group of ancestors of modern cetaceans. Although limited to a skull, the *Pakicetus* fossil provides precious details on the origins of cetaceans. The skull is cetacean-like but its jawbones lack the enlarged space that is filled with fat or oil and used for receiving underwater sound in modern whales. *Pakicetus* probably detected sound through the ear opening as in land mammals. The skull also lacks a blowhole, another cetacean adaptation for diving. Other features, however, show experts that *Pakicetus* is a transitional form between a group of extinct flesh-eating mammals, the mesonychids, and cetaceans. It has been suggested that *Pakicetus* fed on fish in shallow water and was not yet adapted for life in the open ocean. It probably bred and gave birth on land.

Another major discovery was made in Egypt in 1989. Several skeletons of another early whale, *Basilosaurus*, were found in sediments left by the Tethys Sea and now exposed in the Sahara desert.

This whale lived around 40 million years ago, 12 million years after *Pakicetus*. Many incomplete skeletons were found but they included, for the first time in an archaeocyte, a complete hind leg that features a foot with three tiny toes. Such legs would have been far too small to have supported the 50-foot-long *Basilosaurus* on land. *Basilosaurus* was undoubtedly a fully marine whale with possibly nonfunctional, or vestigial, hind legs.

An even more exciting find was reported in 1994, also from Pakistan. The now extinct whale *Ambulocetus natans* ("the walking whale that swam") lived in the Tethys Sea 49 million years ago. It lived around 3 million years after *Pakicetus* but 9 million before *Basilosaurus*. The fossil luckily includes a good portion of the hind legs. The legs were strong and ended in long feet very much like those of a modern pinniped. The legs were certainly functional both on land and at sea. The whale retained a tail and lacked a fluke, the major means of locomotion in modern cetaceans. The structure of the backbone shows, however, that *Ambulocetus* swam like modern whales by moving the rear portion of its body up and down, even though a fluke was missing. The large hind legs were used for propulsion in water. On land, where it probably bred and gave birth, *Ambulocetus* may have moved around very much like a modern sea lion. It was undoubtedly a whale that linked life on land with life at sea.

04 – 1 Deer Populations of the Puget Sound

Two species of deer have been prevalent in the Puget Sound area of Washington State in the Pacific Northwest of the United States. The black-tailed deer, a lowland, west-side cousin of the mule deer of eastern Washington, is now the most common. The other species, the Columbian white-tailed deer, in earlier times was common in the open prairie country; it is now restricted to the low, marshy islands and flood plains along the lower Columbia River.

Nearly any kind of plant of the forest understory can be part of a deer's diet. Where the forest inhibits the growth of grass and other meadow plants, the black-tailed deer browses on huckleberry, salal, dogwood, and almost any other shrub or herb. But this is fair-weather feeding. What keeps the black-tailed deer alive in the harsher seasons of plant decay and dormancy? One compensation for not hibernating is the built-in urge to migrate. Deer may move from high-elevation browse areas in summer down to the lowland areas in late fall. Even with snow on the ground, the high bushy understory is exposed; also snow and wind bring down leafy branches of cedar, hemlock, red alder, and other arboreal fodder.

The numbers of deer have fluctuated markedly since the entry of Europeans into Puget Sound country. The early explorers and settlers told of abundant deer in the early 1800s and yet almost in the same breath bemoaned the lack of this succulent game animal. Famous explorers of the north American frontier, Lewis and Clark arrived at the mouth of the Columbia River on November 14, 1805, in nearly starved circumstances. They had experienced great difficulty finding game west of the Rockies and not until the second of December did they kill their first elk. To keep 40 people alive that winter, they consumed approximately 150 elk and 20 deer. And when game moved out of the lowlands in early spring, the expedition decided to return east rather than face possible starvation. Later on in the early years of the nineteenth century, when Fort Vancouver became the headquarters of the Hudson's Bay Company, deer populations continued to fluctuate. David Douglas, Scottish botanical explorer of the 1830s, found a disturbing change in the animal life around the fort during the period between his first visit in 1825 and his final contact with the fort in 1832. A recent Douglas biographer states: "The deer which once picturesquely dotted the meadows around the fort

were gone [in 1832], hunted to extermination in order to protect the crops."

Reduction in numbers of game should have boded ill for their survival in later times. A worsening of the plight of deer was to be expected as settlers encroached on the land, logging, burning, and clearing, eventually replacing a wilderness landscape with roads, cities, towns, and factories. No doubt the numbers of deer declined still further. Recall the fate of the Columbian white-tailed deer, now in a protected status. But for the black-tailed deer, human pressure has had just the opposite effect. Wildlife zoologist Helmut Buechner (1953), in reviewing the nature of biotic changes in Washington through recorded time, says that "since the early 1940s, the state has had more deer than at any other time in its history, the winter population fluctuating around approximately 320,000 deer (mule and black-tailed deer), which will yield about 65,000 of either sex and any age annually for an indefinite period."

The causes of this population rebound are consequences of other human actions. First, the major predators of deer—wolves, cougar, and lynx—have been greatly reduced in numbers. Second, conservation has been insured by limiting times for and types of hunting. But the most profound reason for the restoration of high population numbers has been the fate of the forests. Great tracts of lowland country deforested by logging, fire, or both have become ideal feeding grounds of deer. In addition to finding an increase of suitable browse, like huckleberry and vine maple, Arthur Einarsen, longtime game biologist in the Pacific Northwest, found quality of browse in the open areas to be substantially more nutritive. The protein content of shade-grown vegetation, for example, was much lower than that for plants grown in clearings.

05 – 3 The Cambrian Explosion

The geologic timescale is marked by significant geologic and biological events, including the origin of Earth about 4.6 billion years ago, the origin of life about 3.5 billion years ago, the origin of eukaryotic life-forms (living things that have cells with true nuclei) about 1.5 billion years ago, and the origin of animals about 0.6 billion years ago. The last event marks the beginning of the Cambrian period. Animals originated relatively late in the history of Earth—in only the last 10 percent of Earth's history. During a geologically brief 100-million-year period, all modern animal groups (along with other animals that are now extinct) evolved. This rapid origin and diversification of animals is often referred to as "the Cambrian explosion."

Scientists have asked important questions about this explosion for more than a century. Why did it occur so late in the history of Earth? The origin of multicellular forms of life seems a relatively simple step compared to the origin of life itself. Why does the fossil record not document the series of evolutionary changes during the evolution of animals? Why did animal life evolve so quickly? Paleontologists continue to search the fossil record for answers to these questions.

One interpretation regarding the absence of fossils during this important 100-million-year period is that early animals were soft bodied and simply did not fossilize. Fossilization of soft-bodied animals is less likely than fossilization of hard-bodied animals, but it does occur. Conditions that promote fossilization of soft-bodied animals include very rapid covering by sediments that create an environment that discourages decomposition. In fact, fossil beds containing soft-bodied animals have been known for many years.

The Ediacara fossil formation, which contains the oldest known animal fossils, consists exclusively of soft-bodied forms. Although named after a site in Australia, the Ediacara formation is worldwide in distribution and dates to Precambrian times. This 700-million-year-old formation gives few clues to the origins of modern animals, however, because paleontologists believe it represents an evolutionary experiment that failed. It contains no ancestors of modern animal groups.

A slightly younger fossil formation containing animal remains is the Tommotian formation, named after a locale in Russia. It dates to the very early Cambrian period, and it also contains only soft-

bodied forms. At one time, the animals present in these fossil beds were assigned to various modern animal groups, but most paleontologists now agree that all Tommotian fossils represent unique body forms that arose in the early Cambrian period and disappeared before the end of the period, leaving no descendants in modern animal groups.

A third fossil formation containing both soft-bodied and hard-bodied animals provides evidence of the result of the Cambrian explosion. This fossil formation, called the Burgess Shale, is in Yoho National Park in the Canadian Rocky Mountains of British Columbia. Shortly after the Cambrian explosion, mud slides rapidly buried thousands of marine animals under conditions that favored fossilization. These fossil beds provide evidence of about 32 modern animal groups, plus about 20 other animal body forms that are so different from any modern animals that they cannot be assigned to any one of the modern groups. These unassignable animals include a large swimming predator called *Anomalocaris* and a soft-bodied animal called *Wiwaxia*, which ate detritus or algae. The Burgess Shale formation also has fossils of many extinct representatives of modern animal groups. For example, a well-known Burgess Shale animal called *Sidneyia* is a representative of a previously unknown group of arthropods (a category of animals that includes insects, spiders, mites, and crabs).

Fossil formations like the Burgess Shale show that evolution cannot always be thought of as a slow progression. The Cambrian explosion involved rapid evolutionary diversification, followed by the extinction of many unique animals. Why was this evolution so rapid? No one really knows. Many zoologists believe that it was because so many ecological niches were available with virtually no competition from existing species. Will zoologists ever know the evolutionary sequences in the Cambrian explosion? Perhaps another ancient fossil bed of soft-bodied animals from 600-million-year-old seas is awaiting discovery.

08 – 2 Extinction of the Dinosaurs

Paleontologists have argued for a long time that the demise of the dinosaurs was caused by climatic alterations associated with slow changes in the positions of continents and seas resulting from plate tectonics. Off and on throughout the Cretaceous (the last period of the Mesozoic era, during which dinosaurs flourished), large shallow seas covered extensive areas of the continents. Data from diverse sources, including geochemical evidence preserved in seafloor sediments, indicate that the Late Cretaceous climate was milder than today's. The days were not too hot, nor the nights too cold. The summers were not too warm, nor the winters too frigid. The shallow seas on the continents probably buffered the temperature of the nearby air, keeping it relatively constant.

At the end of the Cretaceous, the geological record shows that these seaways retreated from the continents back into the major ocean basins. No one knows why. Over a period of about 100,000 years, while the seas pulled back, climates around the world became dramatically more extreme: warmer days, cooler nights; hotter summers, colder winters. Perhaps dinosaurs could not tolerate these extreme temperature changes and became extinct.

If true, though, why did cold-blooded animals such as snakes, lizards, turtles, and crocodiles survive the freezing winters and torrid summers? These animals are at the mercy of the climate to maintain a livable body temperature. It's hard to understand why they would not be affected, whereas dinosaurs were left too crippled to cope, especially if, as some scientists believe, dinosaurs were warm-blooded. Critics also point out that the shallow seaways had retreated from and advanced on the continents numerous times during the Mesozoic, so why did the dinosaurs survive the climatic changes associated with the earlier fluctuations but not with this one? Although initially appealing, the hypothesis of a simple climatic change related to sea levels is insufficient to explain all the data.

Dissatisfaction with conventional explanations for dinosaur extinctions led to a surprising observation that, in turn, has suggested a new hypothesis. Many plants and animals disappear abruptly from the fossil record as one moves from layers of rock documenting the end of the Cretaceous up into rocks representing the beginning of the Cenozoic (the era after the Mesozoic). Between the last layer of Cretaceous rock and the first layer of Cenozoic

rock, there is often a thin layer of clay. Scientists felt that they could get an idea of how long the extinctions took by determining how long it took to deposit this one centimeter of clay and they thought they could determine the time it took to deposit the clay by determining the amount of the element iridium (Ir) it contained.

It has not been common at Earth's since the very beginning of the planet's history. Because it usually exists in a metallic state, it was preferentially incorporated in Earth's core as the planet cooled and consolidated. Ir is found in high concentrations in some meteorites, in which the solar system's original chemical composition is preserved. Even today, microscopic meteorites continually bombard Earth, falling on both land and sea. By measuring how many of these meteorites fall to Earth over a given period of time, scientists can estimate how long it might have taken to deposit the observed amount of Ir in the boundary clay. These calculations suggest that a period of about one million years would have been required. However, other reliable evidence suggests that the deposition of the boundary clay could not have taken one million years. So the unusually high concentration of Ir seems to require a special explanation.

In view of these facts, scientists hypothesized that a single large asteroid, about 10 to 15 kilometers across, collided with Earth, and the resulting fallout created the boundary clay. Their calculations show that the impact kicked up a dust cloud that cut off sunlight for several months, inhibiting photosynthesis in plants; decreased surface temperatures on continents to below freezing; caused extreme episodes of acid rain; and significantly raised long-term global temperatures through the greenhouse effect. This disruption of food chain and climate would have eradicated the dinosaurs and other organisms in less than fifty years.

15 – 2 Mass Extinctions

Cases in which many species become extinct within a geologically short interval of time are called mass extinctions. There was one such event at the end of the Cretaceous period (around 70 million years ago). There was another, even larger, mass extinction at the end of the Permian period (around 250 million years ago). The Permian event has attracted much less attention than other mass extinctions because mostly unfamiliar species perished at that time.

The fossil record shows at least five mass extinctions in which many families of marine organisms died out. The rates of extinction happening today are as great as the rates during these mass extinctions. Many scientists have therefore concluded that a sixth great mass extinction is currently in progress.

What could cause such high rates of extinction? There are several hypotheses, including warming or cooling of Earth, changes in seasonal fluctuations or ocean currents, and changing positions of the continents. Biological hypotheses include ecological changes brought about by the evolution of cooperation between insects and flowering plants or of bottom-feeding predators in the oceans. Some of the proposed mechanisms required a very brief period during which all extinctions suddenly took place; other mechanisms would be more likely to have taken place more gradually, over an extended period, or at different times on different continents. Some hypotheses fail to account for simultaneous extinctions on land and in the seas. Each mass extinction may have had a different cause. Evidence points to hunting by humans and habitat destruction as the likely causes for the current mass extinction.

American paleontologists David Raup and John Sepkoski, who have studied extinction rates in a number of fossil groups, suggest that episodes of increased extinction have recurred periodically, approximately every 26 million years since the mid-Cretaceous period. The late Cretaceous extinction of the dinosaurs and ammonoids was just one of the more drastic in a whole series of such recurrent extinction episodes. The possibility that mass extinctions may recur periodically has given rise to such hypotheses as that of a companion star with a long-period orbit deflecting other bodies from their normal orbits, making some of them fall to Earth as

meteors and causing widespread devastation upon impact.

Of the various hypotheses attempting to account for the late Cretaceous extinctions, the one that has attracted the most attention in recent years is the asteroid-impact hypothesis first suggested by Luis and Walter Alvarez. According to this hypothesis, Earth collided with an asteroid with an estimated diameter of 10 kilometers, or with several asteroids, the combined mass of which was comparable. The force of collision spewed large amounts of debris into the atmosphere, darkening the skies for several years before the finer particles settled. The reduced level of photosynthesis led to a massive decline in plant life of all kinds, and this caused massive starvation first of herbivores and subsequently of carnivores. The mass extinction would have occurred very suddenly under this hypothesis.

One interesting test of the Alvarez hypothesis is based on the presence of the rare-earth element iridium (Ir). Earth's crust contains very little of this element, but most asteroids contain a lot more. Debris thrown into the atmosphere by an asteroid collision would presumably contain large amounts of iridium, and atmospheric currents would carry this material all over the globe. A search of sedimentary deposits that span the boundary between the Cretaceous and Tertiary periods shows that there is a dramatic increase in the abundance of iridium briefly and precisely at this boundary. This iridium anomaly offers strong support for the Alvarez hypothesis even though no asteroid itself has ever been recovered.

An asteroid of this size would be expected to leave an immense crater, even if the asteroid itself was disintegrated by the impact. The intense heat of the impact would produce heat-shocked quartz in many types of rock. Also, large blocks thrown aside by the impact would form secondary craters surrounding the main crater. To date, several such secondary craters have been found along Mexico's Yucatan Peninsula, and heat-shocked quartz has been found both in Mexico and in Haiti. A location called Chicxulub, along the Yucatan coast, has been suggested as the primary impact site.

11 – 2 Orientation and Navigation

To South Americans, robins are birds that fly north every spring. To North Americans, the robins simply vacation in the south each winter. Furthermore, they fly to very specific places in South America and will often come back to the same trees in North American yards the following spring. The question is not why they would leave the cold of winter so much as how they find their way around. The question perplexed people for years, until, in the 1950s, a German scientist named Gustave Kramer provided some answers and, in the process, raised new questions.

Kramer initiated important new kinds of research regarding how animals orient and navigate. Orientation is simply facing in the right direction; navigation involves finding one's way from point A to point B.

Early in his research, Kramer found that caged migratory birds became very restless at about the time they would normally have begun migration in the wild. Furthermore, he noticed that as they fluttered around in the cage, they often launched themselves in the direction of their normal migratory route. He then set up experiments with caged starlings and found that their orientation was, in fact, in the proper migratory direction except when the sky was overcast, at which times there was no clear direction to their restless movements. Kramer surmised, therefore, that they were orienting according to the position of the Sun. To test this idea, he blocked their view of the Sun and used mirrors to change its apparent position. He found that under these circumstances, the birds oriented with respect to the new "Sun." They seemed to be using the Sun as a compass to determine direction. At the time, this idea seemed preposterous. How could a bird navigate by the Sun when some of us lose our way with road maps? Obviously, more testing was in order.

So, in another set of experiments, Kramer put identical food boxes around the cage, with food in only one of the boxes. The boxes were stationary, and the one containing food was always at the same point of the compass. However, its position with respect to the surroundings could be changed by revolving either the inner cage containing the birds or the outer walls, which served as the background. As long as the birds could see the Sun, no matter how their surroundings were altered, they went directly to the correct food box.

Whether the box appeared in front of the right wall or the left wall, they showed no signs of confusion. On overcast days, however, the birds were disoriented and had trouble locating their food box.

In experimenting with artificial suns, Kramer made another interesting discovery. If the artificial Sun remained stationary, the birds would shift their direction with respect to it at a rate of about 15 degrees per hour, the Sun's rate of movement across the sky. Apparently, the birds were assuming that the "Sun" they saw was moving at that rate. When the real Sun was visible, however, the birds maintained a constant direction as it moved across the sky. In other words, they were able to compensate for the Sun's movement. This meant that some sort of biological clock was operating—and a very precise clock at that.

What about birds that migrate at night? Perhaps they navigate by the night sky. To test the idea, caged night-migrating birds were placed on the floor of a planetarium during their migratory period. A planetarium is essentially a theater with a domelike ceiling onto which a night sky can be projected for any night of the year. When the planetarium sky matched the sky outside, the birds fluttered in the direction of their normal migration. But when the dome was rotated, the birds changed their direction to match the artificial sky. The results clearly indicated that the birds were orienting according to the stars.

There is accumulating evidence indicating that birds navigate by using a wide variety of environmental cues. Other areas under investigation include magnetism, landmarks, coastlines, sonar, and even smells. The studies are complicated by the fact that the data are sometimes contradictory and the mechanisms apparently change from time to time. Furthermore, one sensory ability may back up another.

11 – 3 Begging by Nestlings

Many signals that animals make seem to impose on the signalers costs that are overly damaging. A classic example is noisy begging by nestling songbirds when a parent returns to the nest with food. These loud cheeps and peeps might give the location of the nest away to a listening hawk or raccoon, resulting in the death of the defenseless nestlings. In fact, when tapes of begging tree swallows were played at an artificial swallow nest containing an egg, the egg in that “noisy” nest was taken or destroyed by predators before the egg in a nearby quiet nest in 29 of 37 trials.

Further evidence for the costs of begging comes from a study of differences in the begging calls of warbler species that nest on the ground versus those that nest in the relative safety of trees. The young of ground-nesting warblers produce begging cheeps of higher frequencies than do their tree-nesting relatives. These higher-frequency sounds do not travel as far, and so may better conceal the individuals producing them, who are especially vulnerable to predators in their ground nests. David Haskell created artificial nests with clay eggs and placed them on the ground beside a tape recorder that played the begging calls of either tree-nesting or of ground-nesting warblers. The eggs “advertised” by the tree-nesters’ begging calls were found bitten significantly more often than the eggs associated with the ground-nesters’ calls.

The hypothesis that begging calls have evolved properties that reduce their potential for attracting predators yields a prediction: baby birds of species that experience high rates of nest predation should produce softer begging signals of higher frequency than nestlings of other species less often victimized by nest predators. This prediction was supported by data collected in one survey of 24 species from an Arizona forest, and more evidence that predator pressure favors the evolution of begging calls that are hard to detect and pinpoint.

Given that predators can make it costly to beg for food, what benefit do begging nestlings derive from their communications? One possibility is that a noisy baby bird provides accurate signals of its real hunger and good health, making it worthwhile for the listening parent to give it food in a nest where several other offspring are usually available to be fed. If this hypothesis is true, then it follows that nestlings should adjust the intensity of their signals in relation to the signals produced by their

nestmates, who are competing for parental attention. When experimentally deprived baby robins are placed in a nest with normally fed siblings, the hungry nestlings beg more loudly than usual—but so do their better-fed siblings, though not as loudly as the hungrier birds.

If parent birds use begging intensity to direct food to healthy offspring capable of vigorous begging, then parents should make food delivery decisions on the basis of their offsprings’ calls. Indeed, if you take baby tree swallows out of a nest for an hour feeding half the set and starving the other half, when the birds are replaced in the nest, the starved youngsters beg more loudly than the fed birds, and the parent birds feed the active beggars more than those who beg less vigorously.

As these experiments show, begging apparently provides a signal of need that parents use to make judgments about which offspring can benefit most from a feeding. But the question arises, why don’t nestlings beg loudly when they aren’t all that hungry? By doing so, they could possibly secure more food, which should result in more rapid growth or larger size, either of which is advantageous. The answer lies apparently not in the increased energy costs of exaggerated begging—such energy costs are small relative to the potential gain in calories—but rather in the damage that any successful cheater would do to its siblings, which share genes with one another. An individual’s success in propagating his or her genes can be affected by more than just his or her own personal reproductive success. Because close relatives have many of the same genes, animals that harm their close relatives may in effect be destroying some of their own genes. Therefore, a begging nestling that secures food at the expense of its siblings might actually leave behind fewer copies of its genes overall than it might otherwise.

13 – 2 Biological Clocks

Survival and successful reproduction usually require the activities of animals to be coordinated with predictable events around them. Consequently, the timing and rhythms of biological functions must closely match periodic events like the solar day, the tides, the lunar cycle, and the seasons. The relations between animal activity and these periods, particularly for the daily rhythms, have been of such interest and importance that a huge amount of work has been done on them and the special research field of chronobiology has emerged. Normally, the constantly changing levels of an animal's activity – sleeping, feeding, moving, reproducing, metabolizing, and producing enzymes and hormones, for example – are well coordinated with environmental rhythms, but the key question is whether the animal's schedule is driven by external cues, such as sunrise or sunset, or is instead dependent somehow on internal timers that themselves generate the observed biological rhythms. Almost universally, biologists accept the idea that all eukaryotes (a category that includes most organisms except bacteria and certain algae) have internal clocks. By isolating organisms completely from external periodic cues, biologists learned that organisms have internal clocks. For instance, apparently normal daily periods of biological activity were maintained for about a week by the fungus *Neurospora* when it was intentionally isolated from all geophysical timing cues while orbiting in a space shuttle. The continuation of biological rhythms in an organism without external cues attests to its having an internal clock.

When crayfish are kept continuously in the dark, even for four to five months, their compound eyes continue to adjust on a daily schedule for daytime and nighttime vision. Horseshoe crabs kept in the dark continuously for a year were found to maintain a persistent rhythm of brain activity that similarly adapts their eyes on a daily schedule for bright or for weak light. Like almost all daily cycles of animals deprived of environmental cues, those measured for the horseshoe crabs in these conditions were not exactly 24 hours. Such a rhythm whose period is approximately—but not exactly – a day is called circadian. For different individual horseshoe crabs, the circadian period ranged from 22.2 to 25.5 hours. A particular animal typically maintains its own characteristic cycle duration with great precision for many days. Indeed, stability of the biological clock's period is

one of its major features, even when the organism's environment is subjected to considerable changes in factors, such as temperature, that would be expected to affect biological activity strongly. Further evidence for persistent internal rhythms appears when the usual external cycles are shifted – either experimentally or by rapid east-west travel over great distances. Typically, the animal's daily internally generated cycle of activity continues without change. As a result, its activities are shifted relative to the external cycle of the new environment. The disorienting effects of this mismatch between external time cues and internal schedules may persist, like our jet lag, for several days or weeks until certain cues such as the daylight/darkness cycle reset the organism's clock to synchronize with the daily rhythm of the new environment.

Animals need natural periodic signals like sunrise to maintain a cycle whose period is precisely 24 hours. Such an external cue not only coordinates an animal's daily rhythms with particular features of the local solar day but also—because it normally does so day after day—seems to keep the internal clock's period close to that of Earth's rotation. Yet despite this synchronization of the period of the internal cycle, the animal's timer itself continues to have its own genetically built-in period close to, but different from, 24 hours. Without the external cue, the difference accumulates and so the internally regulated activities of the biological day drift continuously, like the tides, in relation to the solar day. This drift has been studied extensively in many animals and in biological activities ranging from the hatching of fruit fly eggs to wheel running by squirrels. Light has a predominating influence in setting the clock. Even a fifteen-minute burst of light in otherwise sustained darkness can reset an animal's circadian rhythm. Normally, internal rhythms are kept in step by regular environmental cycles. For instance, if a homing pigeon is to navigate with its Sun compass, its clock must be properly set by cues provided by the daylight/darkness cycle.

15 – 1 A Warm-Blooded Turtle

When it comes to physiology, the leatherback turtle is, in some ways, more like a reptilian whale than a turtle. It swims farther into the cold of the northern and southern oceans than any other sea turtle, and it deals with the chilly waters in a way unique among reptiles.

A warm-blooded turtle may seem to be a contradiction in terms. Nonetheless, an adult leatherback can maintain a body temperature of between 25 and 26°C (77-79°F) in seawater that is only 8°C (46.4°F). Accomplishing this feat requires adaptations both to generate heat in the turtle's body and to keep it from escaping into the surrounding waters. Leatherbacks apparently do not generate internal heat the way we do, or the way birds do, as a by-product of cellular metabolism. A leatherback may be able to pick up some body heat by basking at the surface; its dark, almost black body color may help it to absorb solar radiation. However, most of its internal heat comes from the action of its muscles.

Leatherbacks keep their body heat in three different ways. The first, and simplest, is size. The bigger the animal is, the lower its surface-to-volume ratio; for every ounce of body mass, there is proportionately less surface through which heat can escape. An adult leather back is twice the size of the biggest cheloniid sea turtles and will therefore take longer to cool off. Maintaining a high body temperature through sheer bulk is called gigantothermy. It works for elephants, for whales, and, perhaps, it worked for many of the larger dinosaurs. It apparently works, in a smaller way, for some other sea turtles. Large loggerhead and green turtles can maintain their body temperature at a degree or two above that of the surrounding water, and gigantothermy is probably the way they do it. Muscular activity helps, too, and an actively swimming green turtle may be 7°C (12.6°F) warmer than the waters it swims through.

Gigantothermy, though, would not be enough to keep a leatherback warm in cold northern waters. It is not enough for whales, which supplement it with a thick layer of insulating blubber (fat). Leatherbacks do not have blubber, but they do have a reptilian equivalent: thick, oil-saturated skin, with a layer of fibrous, fatty tissue just beneath it. Insulation protects the leatherback everywhere but on its head and flippers. Because the flippers are comparatively thin and blade-like, they are the one

part of the leatherback that is likely to become chilled. There is not much that the turtle can do about this without compromising the aerodynamic shape of the flipper. The problem is that as blood flows through the turtle's flippers, it risks losing enough heat to lower the animal's central body temperature when it returns. The solution is to allow the flippers to cool down without drawing heat away from the rest of the turtle's body. The leatherback accomplishes this by arranging the blood vessels in the base of its flipper into a countercurrent exchange system.

In a countercurrent exchange system, the blood vessels carrying cooled blood from the flippers run close enough to the blood vessels carrying warm blood from the body to pick up some heat from the warmer blood vessels; thus, the heat is transferred from the outgoing to the ingoing vessels before it reaches the flipper itself. This is the same arrangement found in an old-fashioned steam radiator, in which the coiled pipes pass heat back and forth as water courses through them. The leatherback is certainly not the only animal with such an arrangement; gulls have a countercurrent exchange in their legs. That is why a gull can stand on an ice floe without freezing.

All this applies, of course, only to an adult leatherback. Hatchlings are simply too small to conserve body heat, even with insulation and countercurrent exchange systems. We do not know how old, or how large, a leatherback has to be before it can switch from a cold-blooded to a warm-blooded mode of life. Leatherbacks reach their immense size in a much shorter time than it takes other sea turtles to grow. Perhaps their rush to adulthood is driven by a simple need to keep warm.

17 – 2 Animal Signals in the Rain Forest

The daytime quality of light in forests varies with the density of the vegetation, the angle of the Sun, and the amount of cloud in the sky. Both animals and plants have different appearances in these various lighting conditions. A color or pattern that is relatively indistinct in one kind of light may be quite conspicuous in another.

In the varied and constantly changing light environment of the forest, an animal must be able to send visual signals to members of its own species and at the same time avoid being detected by predators. An animal can hide from predators by choosing the light environment in which its pattern is least visible. This may require moving to different parts of the forest at different times of the day or under different weather conditions, or it may be achieved by changing color according to the changing light conditions. Many species of amphibians (frogs and toads) and reptiles (lizards and snakes) are able to change their color patterns to camouflage themselves. Some also signal by changing color. The chameleon lizard has the most striking ability to do this. Some chameleon species can change from a rather dull appearance to a full riot of carnival colors in seconds. By this means, they signal their level of aggression or readiness to mate.

Other species take into account the changing conditions of light by performing their visual displays only when the light is favorable. A male bird of paradise may put himself in the limelight by displaying his spectacular plumage in the best stage setting to attract a female. Certain butterflies move into spots of sunlight that have penetrated to the forest floor and display by opening and closing their beautifully patterned wings in the bright spotlights. They also compete with each other for the best spot of sunlight.

Very little light filters through the canopy of leaves and branches in a rain forest to reach ground level—or close to the ground—and at those levels the yellow-to-green wavelengths predominate. A signal might be most easily seen if it is maximally bright. In the green-to-yellow lighting conditions of the lowest levels of the forest, yellow and green would be the brightest colors, but when an animal is signaling, these colors would not be very visible if the animal was sitting in an area with a yellowish or greenish background. The best signal depends not only on its brightness but also on how well it

contrasts with the background against which it must be seen. In this part of the rain forest, therefore, red and orange are the best colors for signaling, and they are the colors used in signals by the ground-walking Australian brush turkey. This species, which lives in the rain forests and scrublands of the east coast of Australia, has a brown-to-black plumage with bare, bright-red skin on the head and neck and a neck collar of orange-yellow loosely hanging skin. During courtship and aggressive displays, the turkey enlarges its colored neck collar by inflating sacs in the neck region and then flings about a pendulous part of the colored signaling apparatus as it utters calls designed to attract or repel. This impressive display is clearly visible in the light spectrum illuminating the forest floor.

Less colorful birds and animals that inhabit the rain forest tend to rely on other forms of signaling other than the visual, particularly over long distances. The piercing cries of the rhinoceros and hornbill characterize the Southeast Asian rain forest, as do the unmistakable calls of the gibbons. In densely wooded environments, sound is the best means of communication over distance because in comparison with light, it travels with little impediment from trees and other vegetation. In forests, visual signals can be seen only at short distances, where they are not obstructed by trees. The male riflebird exploits both of these modes of signaling simultaneously in his courtship display. The sounds made as each wing is opened carry extremely well over distance and advertise his presence widely. The ritualized visual display communicates in close quarters when a female has approached.

17 – 3 Symbiotic Relationships

A symbiotic relationship is an interaction between two or more species in which one species lives in or on another species. There are three main types of symbiotic relationships: parasitism, commensalism, and mutualism. The first and the third can be key factors in the structure of a biological community; that is, all the populations of organisms living together and potentially interacting in a particular area.

Parasitism is a kind of predator-prey relationship in which one organism, the parasite, derives its food at the expense of its symbiotic associate, the host. Parasites are usually smaller than their hosts. An example of a parasite is a tapeworm that lives inside the intestines of a larger animal and absorbs nutrients from its host. Natural selection favors the parasites that are best able to find and feed on hosts. At the same time, defensive abilities of hosts are also selected for. As an example, plants make chemicals toxic to fungal and bacterial parasites, along with ones toxic to predatory animals (sometimes they are the same chemicals). In vertebrates, the immune system provides a multiple defense against internal parasites.

At times, it is actually possible to watch the effects of natural selection in host-parasite relationships. For example, Australia during the 1940 s was overrun by hundreds of millions of European rabbits. The rabbits destroyed huge expanses of Australia and threatened the sheep and cattle industries. In 1950, myxoma virus, a parasite that affects rabbits, was deliberately introduced into Australia to control the rabbit population. Spread rapidly by mosquitoes, the virus devastated the rabbit population. The virus was less deadly to the offspring of surviving rabbits, however, and it caused less and less harm over the years. Apparently, genotypes (the genetic make-up of an organism) in the rabbit population were selected that were better able to resist the parasite. Meanwhile, the deadliest strains of the virus perished with their hosts as natural selection favored strains that could infect hosts but not kill them. Thus, natural selection stabilized this host-parasite relationship.

In contrast to parasitism, in commensalism, one partner benefits without significantly affecting the other. Few cases of absolute commensalism probably exist, because it is unlikely that one of the partners will be completely unaffected. Commensal

associations sometimes involve one species' obtaining food that is inadvertently exposed by another. For instance, several kinds of birds feed on insects flushed out of the grass by grazing cattle. It is difficult to imagine how this could affect the cattle, but the relationship may help or hinder them in some way not yet recognized.

The third type of symbiosis, mutualism, benefits both partners in the relationship. Legume plants and their nitrogen-fixing bacteria, and the interactions between flowering plants and their pollinators, are examples of mutualistic association. In the first case, the plants provide the bacteria with carbohydrates and other organic compounds, and the bacteria have enzymes that act as catalysts that eventually add nitrogen to the soil, enriching it. In the second case, pollinators (insects, birds) obtain food from the flowering plant, and the plant has its pollen distributed and seeds dispersed much more efficiently than they would be if they were carried by the wind only. Another example of mutualism would be the bull's horn acacia tree, which grows in Central and South America. The tree provides a place to live for ants of the genus *Pseudomyrmex*. The ants live in large, hollow thorns and eat sugar secreted by the tree. The ants also eat yellow structures at the tip of leaflets: these are protein rich and seem to have no function for the tree except to attract ants. The ants benefit the host tree by attacking virtually anything that touches it. They sting other insects and large herbivores (animals that eat only plants) and even clip surrounding vegetation that grows near the tree. When the ants are removed, the trees usually die, probably because herbivores damage them so much that they are unable to compete with surrounding vegetation for light and growing space.

The complex interplay of species in symbiotic relationships highlights an important point about communities: Their structure depends on a web of diverse connections among organisms.

20 – 3 Fossil Preservation

When one considers the many ways by which organisms are completely destroyed after death, it is remarkable that fossils are as common as they are. Attack by scavengers and bacteria, chemical decay, and destruction by erosion and other geologic agencies make the odds against preservation very high. However, the chances of escaping complete destruction are vastly improved if the organism happens to have a mineralized skeleton and dies in a place where it can be quickly buried by sediment. Both of these conditions are often found on the ocean floors, where shelled invertebrates (organisms without spines) flourish and are covered by the continuous rain of sedimentary particles. Although most fossils are found in marine sedimentary rocks, they also are found in terrestrial deposits left by streams and lakes. On occasion, animals and plants have been preserved after becoming immersed in tar or quicksand, trapped in ice or lava flows, or engulfed by rapid falls of volcanic ash.

The term "fossil" often implies petrification, literally a transformation into stone. After the death of an organism, the soft tissue is ordinarily consumed by scavengers and bacteria. The empty shell of a snail or clam may be left behind, and if it is sufficiently durable and resistant to dissolution, it may remain basically unchanged for a long period of time. Indeed, unaltered shells of marine invertebrates are known from deposits over 100 million years old. In many marine creatures, however, the skeleton is composed of a mineral variety of calcium carbonate called aragonite. Although aragonite has the same composition as the more familiar mineral known as calcite, it has a different crystal form, is relatively unstable, and in time changes to the more stable calcite.

Many other processes may alter the shell of a clam or snail and enhance its chances for preservation. Water containing dissolved silica, calcium carbonate, or iron may circulate through the enclosing sediment and be deposited in cavities such as marrow cavities and canals in bone once occupied by blood vessels and nerves. In such cases, the original composition of the bone or shell remains, but the fossil is made harder and more durable. This addition of a chemically precipitated substance into pore spaces is termed "permineralization."

Petrification may also involve a simultaneous exchange of the original substance of a dead plant or animal with mineral matter of a different composition. This process is termed "replacement" because solutions have dissolved the original material and replaced it with an equal volume of the new substance. Replacement can be a marvelously precise process, so that details of shell ornamentation, tree rings in wood, and delicate structures in bone are accurately preserved.

Another type of fossilization, known as carbonization, occurs when soft tissues are preserved as thin films of carbon. Leaves and tissue of soft-bodied organisms such as jellyfish or worms may accumulate, become buried and compressed, and lose their volatile constituents. The carbon often remains behind as a blackened silhouette.

Although it is certainly true that the possession of hard parts enhances the prospect of preservation, organisms having soft tissues and organs are also occasionally preserved. Insects and even small invertebrates have been found preserved in the hardened resins of conifers and certain other trees. X-ray examination of thin slabs of rock sometimes reveals the ghostly outlines of tentacles, digestive tracts, and visual organs of a variety of marine creatures. Soft parts, including skin, hair, and viscera of ice age mammoths, have been preserved in frozen soil or in the oozing tar of oil seeps.

The probability that actual remains of soft tissue will be preserved is improved if the organism dies in an environment of rapid deposition and oxygen deprivation. Under such conditions, the destructive effects of bacteria are diminished. The Middle Eocene Messel Shale (from about 48 million years ago) of Germany accumulated in such an environment. The shale was deposited in an oxygen-deficient lake where lethal gases sometimes bubbled up and killed animals. Their remains accumulated on the floor of the lake and were then covered by clay and silt. Among the superbly preserved Messel fossils are insects with iridescent exoskeletons (hard outer coverings), frogs with skin and blood vessels intact, and even entire small mammals with preserved fur and soft tissue.

26 – 2 Survival of Plants and Animals in Desert Conditions

The harsh conditions in deserts are intolerable for most plants and animals. Despite these conditions, however, many varieties of plants and animals have adapted to deserts in a number of ways. Most plant tissues die if their water content falls too low: the nutrients that feed plants are transmitted by water; water is a raw material in the vital process of photosynthesis; and water regulates the temperature of a plant by its ability to absorb heat and because water vapor lost to the atmosphere through the leaves helps to lower plant temperatures. Water controls the volume of plant matter produced. The distribution of plants within different areas of desert is also controlled by water. Some areas, because of their soil texture, topographical position, or distance from rivers or groundwater, have virtually no water available to plants, whereas others do.

The nature of plant life in deserts is also highly dependent on the fact that they have to adapt to the prevailing aridity. There are two general classes of vegetation: long-lived perennials, which may be succulent (water-storing) and are often dwarfed and woody, and annuals or ephemerals, which have a short life cycle and may form a fairly dense stand immediately after rain.

The ephemeral plants evade drought. Given a year of favorable precipitation, such plants will develop vigorously and produce large numbers of flowers and fruit. This replenishes the seed content of the desert soil. The seeds then lie dormant until the next wet year, when the desert blooms again.

The perennial vegetation adjusts to the aridity by means of various avoidance mechanisms. Most desert plants are probably best classified as xerophytes. They possess drought-resisting adaptations: loss of water through the leaves is reduced by means of dense hairs covering waxy leaf surfaces, by the closure of pores during the hottest times to reduce water loss, and by the rolling up or shedding of leaves at the beginning of the dry season. Some xerophytes, the succulents (including cacti), store water in their structures. Another way of countering drought is to have a limited amount of mass above ground and to have extensive root networks below ground. It is not unusual for the roots of some desert perennials to extend downward more than ten meters. Some plants are woody in type — an adaptation designed

to prevent collapse of the plant tissue when water stress produces wilting. Another class of desert plant is the phreatophyte. These have adapted to the environment by the development of long taproots that penetrate downward until they approach the assured water supply provided by groundwater. Among these plants are the date palm, tamarisk, and mesquite. They commonly grow near stream channels, springs, or on the margins of lakes.

Animals also have to adapt to desert conditions, and they may do it through two forms of behavioral adaptation: they either escape or retreat. Escape involves such actions as aestivation, a condition of prolonged dormancy, or torpor, during which animals reduce their metabolic rate and body temperature during the hot season or during very dry spells.

Seasonal migration is another form of escape, especially for large mammals or birds. The term retreat is applied to the short-term escape behavior of desert animals, and it usually assumes the pattern of a daily rhythm. Birds shelter in nests, rock overhangs, trees, and dense shrubs to avoid the hottest hours of the day, while mammals like the kangaroo rat burrow underground.

Some animals have behavioral, physiological, and morphological (structural) adaptations that enable them to withstand extreme conditions. For example, the ostrich has plumage that is so constructed that the feathers are long but not too dense. When conditions are hot, the ostrich erects them on its back, thus increasing the thickness of the barrier between solar radiation and the skin. The sparse distribution of the feathers, however, also allows considerable lateral air movement over the skin surface, thereby permitting further heat loss by convection. Furthermore, the birds orient themselves carefully with regard to the Sun and gently flap their wings to increase convection cooling.

27 – 3 Predator-Prey Cycles

How do predators affect populations of the prey animals? The answer is not as simple as might be thought. Moose reached Isle Royale in Lake Superior by crossing over winter ice and multiplied freely there in isolation without predators. When wolves later reached the island, naturalists widely assumed that the wolves would play a key role in controlling the moose population. Careful studies have demonstrated, however, that this is not the case. The wolves eat mostly old or diseased animals that would not survive long anyway. In general, the moose population is controlled by food availability, disease and other factors rather than by wolves.

When experimental populations are set up under simple laboratory conditions, the predator often exterminates its prey and then becomes extinct itself, having nothing left to eat. However, if safe areas like those prey animals have in the wild are provided, the prey population drops to low level but not extinction. Low prey population levels then provide inadequate food for the predators, causing the predator population to decrease. When this occurs, the prey population can rebound. In this situation the predator and prey population may continue in this cyclical pattern for some time.

Population cycles are characteristic of small mammals, and they sometimes appear to be brought about by predators. Ecologists studying hare populations have found that the North American snow shoe hare follows a roughly ten-year cycle. Its numbers fall tenfold to thirty in a typical cycle, and a hundredfold change can occur. Two factors appear to be generating the cycle: food plants and predators.

The preferred foods of snowshoe hares are willow and birch twigs. As hare density increases, the quantity of these twigs decreases, forcing the hares to feed on low-quality high-fiber food. Lower birth rates, low juvenile survivorship, and low growth rates follow, so there is a corresponding decline in hare abundance. Once the hare population has declined, it takes two to three year for the quantity of twigs to recover.

A key predator of the snowshoe hare is the Canada lynx. The Canada lynx shows a ten-year cycle of abundance that parallels the abundance cycle of hares. As hare numbers fall, so do lynx numbers, as their food supply depleted.

What causes the predator-prey oscillations? Do increasing number of hares lead to overharvesting of plants, which in turn results in reduced hare populations, or do increasing numbers of lynx lead to overharvesting hares? Field experiments carried out by Charles Krebs and coworkers in 1992 provide an answer. Krebs investigated experimental plots in Canada's Yukon territory that contained hare populations. When food was added to those plots (no food effect) and predators were excluded (no predator effect) from an experimental area, hare numbers increased tenfold and stayed there - the cycle was lost. However, the cycle was retained if either of the factors was allowed to operate alone: if predators were excluded but food was not added (food effect alone), or if food was added in the presence of predators (predator effect alone). Thus both factors can affect the cycle, which, in practice, seems to be generated by conjunction of the two factors.

Predators are an essential factor in maintaining communities that are rich and diverse in species. Without predators, the species that is the best competitor for food, shelter, nesting sites, and other environmental resources tends to dominate and exclude the species with which it competes. This phenomenon is known as "competitor exclusion". However, if the community contains a predator of the strongest competitor species, then the population of that competitor is controlled. Thus even the less competitive species are able to survive. For example, sea stars prey on a variety of bivalve mollusks and prevent these bivalves from monopolizing habitats on the sea floor. This opens up space for many other organisms. When sea stars are removed, species diversity falls sharply. Therefore, from the stand point of diversity, it is usually a mistake to eliminate a major predator from a community.

28 – 3 Buck Rubs and Buck Scrapes

A conspicuous sign indicating the presence of white-tailed deer in a woodlot is a buck rub. A male deer makes a buck rub by stripping the bark (outer layer) of a small tree with its antlers. When completed, the buck rub is an obvious visual signal to us and presumable to other deer in the area. A rub is usually located at the shoulder height of a deer (one meter or less above the ground) on a smooth-barked, small-diameter (16-25 millimeters) tree. The smooth bark of small red maples makes this species ideal for buck rubs in the forests of the mid-eastern United States.

Adult male deer usually produce rubs in late summer or early autumn when the outer velvet layer is being shed from their antlers. Rubs are created about one to two months before the breeding season (the rut). Hence for a long time biologists believed that male deer used buck rubs not only to clean and polish antlers but also to provide practice for the ensuing male-to-male combat during the rut. However, biologists also noted deer sniff and lick an unfamiliar rub, which suggests that this visual mark on a small tree plays an important communication purpose in the social life of deer.

Buck rubs also have a scent produced by glands in the foreheads of deer that is transferred to the tree when the rub is made. These odors make buck rubs an important means of olfactory communication between deer. The importance of olfactory communication (using odors to communicate) in the way of life of deer was documented by a study of captive adult male deer a few decades ago, which noted that males rubbed their foreheads on branches and twigs, especially as autumn approached. A decade later another study reported that adult male white-tailed deer exhibited forehead rubbing just before and during the rut. It was found that when a white-tailed buck makes a rub, it moves both antlers and forehead glands along the small tree in a vertical direction. This forehead rubbing behavior coincides with a high level of glandular activity in the modified scent glands found on the foreheads of male deer; the glandular activity causes the forehead pelage (hairy covering) of adult males to be distinctly darker than in females or younger males.

Forehead rubbing by male deer on buck rubs presumably sends a great deal of information to other members of the same species. First, the

chemicals deposited on the rub provide information on the individual identity of an animal; no two mammals produce the same scent. For instance, as we all know, dogs recognize each other via smell. Second, because only male deer rub, the buck rub and its associated chemicals indicate the sex of the deer producing the rub. Third, older, more dominant bucks produce more buck rubs and probably deposit more glandular secretions on a given rub. Thus the presence of many well-marked rubs is indicative of older, higher-status males being in the general vicinity rather than simply being a crude measure of relative deer abundance in a given area. The information conveyed by the olfactory signals on a buck rub make it the social equivalent of some auditory signals in other deer species, such as trumpeting by bull elk.

Because both sexes of white-tailed respond to buck rubs by smelling and licking them, rubs may serve a very important additional function. Fresher buck rubs (less than two days old), in particular, are visited more frequently by adult females than older rubs. In view of this behavior it has been suggested that chemicals present in fresh buck rubs may help physiologically induce and synchronize fertility in females that visit these rubs. This would be an obvious advantage to wide-ranging deer, especially to a socially dominant buck when courting several adult females during the autumn rut. Another visual signal produced by white-tailed deer is termed a buck scrape. Scrapes consist of a clearing (about 0.5 meter in diameter) and shallow depression made by pushing aside the leaves covering the ground; after making the scrape, the deer typically urinates in the depression. Thus, like a buck rub, a scrape is both a visual and an olfactory signal. Buck scrapes are generally created after leaf-fall in autumn, which is just before or during the rut. Scrapes are usually placed in open or conspicuous places, such as along a deer trail. Most are made by older males, although females and younger males (2.5 years old or less) occasionally make scrapes.

29 – 2 Competition

When several individuals of the same species or of several different species depend on the same limited resource, a situation may arise that is referred to as competition. The existence of competition has been long known to naturalists; its effects were described by Darwin in considerable detail. Competition among individuals of the same species (intraspecies competition), one of the major mechanisms of natural selection, is the concern of evolutionary biology. Competition among the individuals of different species (interspecies competition) is a major concern of ecology. It is one of the factors controlling the size of competing populations, and extreme cases it may lead to the extinction of one of the competing species. This was described by Darwin for indigenous New Zealand species of animals and plants, which died out when competing species from Europe were introduced.

No serious competition exists when the major needed resource is in superabundant supply, as in most cases of the coexistence of herbivores (plant eaters). Furthermore, most species do not depend entirely on a single resource, if the major resource for a species becomes scarce, the species can usually shift to alternative resources. If more than one species is competing for a scarce resource, the competing species usually switch to different alternative resources. Competition is usually most severe among close relatives with similar demands on the environment. But it may also occur among totally unrelated forms that compete for the same resource, such as seed-eating rodents and ants. The effects of such competition are graphically demonstrated when all the animals or all the plants in an ecosystem come into competition, as happened 2 million years ago at the end of Pliocene, when North and South America became joined by the Isthmus of Panama. North and South American species migrating across the Isthmus now came into competition with each other. The result was the extermination of a large fraction of the South American mammals, which were apparently unable to withstand the competition from invading North American species - although added predation was also an important factor.

To what extent competition determines the composition of a community and the density of particular species has been the source of considerable controversy. The problem is that competition ordinarily cannot be observed directly

but must be inferred from the spread or increase of one species and the concurrent reduction or disappearance of another species. The Russian biologist G. F. Gause performed numerous two-species experiments in the laboratory, in which one of the species became extinct when only a single kind of resource was available. On the basis of these experiments and of field observations, the so-called law of competitive exclusion was formulated, according to which no two species can occupy the same niche. Numerous seeming exceptions to this law have since been found, but they can usually be explained as cases in which the two species, even though competing for a major joint resource, did not really occupy exactly the same niche.

Competition among species is of considerable evolutionary importance. The physical structure of species competing for resources in the same ecological niche tends to gradually evolve in ways that allow them to occupy different niches. Competing species also tend to change their ranges so that their territories no longer overlap. The evolutionary effect of competition on species has been referred to as "species selection"; however, this description is potentially misleading. Only the individuals of a species are subject to the pressures of natural selection. The effect on the well-being and existence of a species is just the result of the effects of selection on all the individuals of the species. Thus species selection is actually a result of individual selection.

Competition may occur for any needed resource. In the case of animals it is usually food; in the case of forest plants it may be light; in the case of substrate inhabitants it may be space, as in many shallow-water bottom-dwelling marine organisms. Indeed, it may be for any of the factors, physical as well as biotic, that are essential for organisms. Competition is usually the more severe the denser the population. Together with predation, it is the most important density-dependent factor in regulating population growth.

30 – 2 The Pace of Evolutionary Change

A heated debate has enlivened recent studies of evolution. Darwin's original thesis, and the viewpoint supported by evolutionary gradualists, is that species change continuously but slowly and in small increments. Such changes are all but invisible over the short time scale of modern observations, and, it is argued, they are usually obscured by innumerable gaps in the imperfect fossil record. Gradualism, with its stress on the slow pace of change, is a comforting position, repeated over and over again in generations of textbooks. By the early twentieth century, the question about the rate of evolution had been answered in favor of gradualism to most biologists' satisfaction.

Sometimes a closed question must be reopened as new evidence or new arguments based on old evidence come to light. In 1972 paleontologists Stephen Jay Gould and Niles Eldredge challenged conventional wisdom with an opposing viewpoint, the punctuated equilibrium hypothesis, which posits that species give rise to new species in relatively sudden bursts, without a lengthy transition period. These episodes of rapid evolution are separated by relatively long static spans during which a species may hardly change at all.

The punctuated equilibrium hypothesis attempts to explain a curious feature of the fossil record — one that has been familiar to paleontologists for more than a century but has usually been ignored. Many species appear to remain unchanged in the fossil record for millions of years — a situation that seems to be at odds with Darwin's model of continuous change. Intermediate fossil forms, predicted by gradualism, are typically lacking. In most localities a given species of clam or coral persists essentially unchanged throughout a thick formation of rock, only to be replaced suddenly by a new and different species.

The evolution of North American horse, which was once presented as a classic textbook example of gradual evolution, is now providing equally compelling evidence for punctuated equilibrium. A convincing 50-million-year sequence of modern horse ancestors — each slightly larger, with more complex teeth, a longer face, and a more prominent central toe — seemed to provide strong support for Darwin's contention that species evolve gradually. But close examination of those fossil deposits now reveals a somewhat different story.

Horses evolved in discrete steps, each of which persisted almost unchanged for millions of years and was eventually replaced by a distinctive newer model. The four-toed *Eohippus* preceded the three-toed *Miohippus*, for example, but North American fossil evidence suggests a jerky, uneven transition between the two. If evolution had been a continuous, gradual process, one might expect that almost every fossil specimen would be slightly different from every year.

If it seems difficult to conceive how major changes could occur rapidly, consider this: an alteration of a single gene in flies is enough to turn a normal fly with a single pair of wings into one that has two pairs of wings.

The question about the rate of evolution must now be turned around: does evolution ever proceed gradually, or does it always occur in short bursts? Detailed field studies of thick rock formations containing fossils provide the best potential tests of the competing theories.

Occasionally, a sequence of fossil-rich layers of rock permits a comprehensive look at one type of organism over a long period of time. For example, Peter Sheldon's studies of trilobites, a now extinct marine animal with a segmented body, offer a detailed glimpse into three million years of evolution in one marine environment. In that study, each of eight different trilobite species was observed to undergo a gradual change in the number of segments — typically an increase of one or two segments over the whole time interval. No significant discontinuities were observed, leading Sheldon to conclude that environmental conditions were quite stable during the period he examined.

Similar exhaustive studies are required for many different kinds of organisms from many different periods. Most researchers expect to find that both modes of transition from one species to another are at work in evolution. Slow, continuous change may be the norm during periods of environmental stability, while rapid evolution of new species occurs during periods of environment stress. But a lot more studies like Sheldon's are needed before we can say for sure.

32 – 3 Distributions of Tropical Bee Colonies

In 1977 ecologists Stephen Hubbell and Leslie Johnson recorded a dramatic example of how social interactions can produce and enforce regular spacing in a population. They studied competition and nest spacing in populations of stingless bees in tropical dry forests in Costa Rica. Though these bees do not sting, rival colonies of some species fight fiercely over potential nesting sites.

Stingless bees are abundant in tropical and subtropical environments, where they gather nectar and pollen from a wide variety of flowers. They generally nest in trees and live in colonies made up of hundreds to thousands of workers. Hubbell and Johnson observed that some species of stingless bees are highly aggressive to members of their species from other colonies, while other species are not. Aggressive species usually forage in groups and feed mainly on flowers that occur in high-density clumps. Nonaggressive species feed singly or in small groups and on more widely distributed flowers.

Hubbell and Johnson studied several species of stingless bees to determine whether there is a relationship between aggressiveness and patterns of colony distribution. They predicted that the colonies of aggressive species would show regular distributions, while those of nonaggressive species would show random or closely grouped (clumped) distribution. They concentrated their studies on a thirteen-hectare tract of tropical dry forest that contained numerous nests of nine species of stingless bees.

Though Hubbell and Johnson were interested in how bee behavior might affect colony distributions, they recognized that the availability of potential nest sites for colonies could also affect distributions. So as one of the first steps in their study, they mapped the distributions of trees suitable for nesting. They found that potential nest trees were distributed randomly through the study area. They also found that the number of potential nest sites was much greater than the number of bee colonies. What did these measurements show the researchers? The number of colonies in the study area was not limited by availability of suitable trees, and a clumped or regular distribution of colonies was not due to an underlying clumped or regular distribution of potential nest sites.

Hubbell and Johnson mapped the nests of five of the nine species of stingless bees accurately, and

the nests of four of these species were distributed regularly. All four species with regular nest distributions were highly aggressive to bees from other colonies of their own species. The fifth species was not aggressive, and its nests were randomly distributed over the study area.

The researchers also studied the process by which the aggressive species establish new colonies. Their observations provide insights into the mechanisms that establish and maintain the regular nest distribution of these species. Aggressive species apparently mark prospective nest sites with pheromones, chemical substances secreted by some animals for communication with other members of their species. The pheromone secreted by these stingless bees attracts and aggregates members of their colony to the prospective nest site; however, it also attracts workers from other nests.

If workers from two different colonies arrive at the prospective nest at the same time, they may fight for possession. Fights may be escalated into protracted battles. The researchers observed battles over a nest tree that lasted for two weeks. Each dawn, fifteen to thirty workers from two competing colonies arrived at the contested nest site. The workers from the two colonies faced off in two swarms and displayed and fought with each other. In the displays, pairs of bees faced each other, slowly flew vertically to a height of about three meters, and then grappled each other to the ground. When the two bees hit the ground, they separated, faced off, and performed another aerial display. Bees did not appear to be injured in these fights, which were apparently ritualized. The two swarms abandoned the battle at about 8 or 9 A.M. each morning, only to re-form and begin again the next day just after dawn. While this contest over an unoccupied nest site produced no obvious mortality, fights over occupied nests sometimes kill over 1,000 bees in a single battle.

33 – 3 Extinction Episodes of the Past

It was not until the Cambrian period, beginning about 600 million years ago, that a great proliferation of macroscopic species occurred on Earth and produced a fossil record that allows us to track the rise and fall of biodiversity. Since the Cambrian period, biodiversity has generally risen, but there have been some notable exceptions. Biodiversity collapsed dramatically during at least five periods because of mass extinctions around the globe. The five major mass extinctions receive most of the attention, but they are only one end of a spectrum of extinction events. Collectively, more species went extinct during smaller events that were less dramatic but more frequent. The best known of the five major extinction events, the one that saw the demise of the dinosaurs, is the Cretaceous-Tertiary extinction.

Starting about 280 million years ago, reptiles were the dominant large animals in terrestrial environments. In popular language this was the era "when dinosaurs ruled Earth", when a wide variety of reptile species occupying many ecological niches. However, no group or species can maintain its dominance indefinitely, and when, after over 200 million years, the age of dinosaurs came to a dramatic end about 65 million years ago, mammals began to flourish, evolving from relatively few types of small terrestrial animals into the myriad of diverse species, including bats and whales, that we know today. Paleontologists label this point in Earth's history as the end of the Cretaceous period and the beginning of the Tertiary period, often abbreviated as the K-T boundary. This time was also marked by changes in many other types of organisms. Overall, about 38 percent of the families of marine animals were lost, with percentages much higher in some groups. Ammonoid mollusks went from being very diverse and abundant to being extinct. An extremely abundant set of planktonic marine animals called foraminifera largely disappeared, although they rebounded later. Among plants, the K-T boundary saw a sharp but brief rise in the abundance of primitive vascular plants such as ferns, club mosses, horsetails, and conifers and other gymnosperms. The number of flowering plants (angiosperms) was reduced at this time, but they then began to increase dramatically.

What caused these changes? For many years scientists assumed that a cooling of the climate was responsible, with dinosaurs being particularly vulnerable because, like modern reptiles, they were

ectothermic (dependent on environmental heat, or cold-blooded). It is now widely believed that at least some species of dinosaurs had a metabolic rate high enough for them to be endotherms (animals that maintain a relatively consistent body temperature by generating heat internally). Nevertheless, climatic explanations for the K-T extinction are not really challenged by the ideas that dinosaurs may have been endothermic, because even endotherms can be affected by a significant change in the climate.

Explanations for the K-T extinction were revolutionized in 1980 when a group of physical scientists led by Luis Alvarez proposed that 65 million years ago Earth was struck by a 10-kilometer-wide meteorite traveling at 90,000 kilometers per hour. They believed that this impact generated a thick cloud of dust that enveloped Earth, shutting out much of the incoming solar radiation and reducing plant photosynthesis to very low levels. Short-term effects might have included huge tidal waves and extensive fires. In other words, a series of events arising from a single cataclysmic event caused the massive extinctions. Initially, the meteorite theory was based on a single line of evidence. At locations around the globe, geologists had found an unusually high concentration of iridium in the layer of sedimentary rocks that was formed about 65 million years ago. Iridium is an element that is usually uncommon near Earth's surface, but it is abundant in some meteorites. Therefore, Alvarez and his colleagues concluded that it was likely that the iridium in sedimentary rocks deposited at the K-T boundary had originated in a giant meteorite or asteroid. Most scientists came to accept the meteorite theory after evidence came to light that a circular formation, 180 kilometers in diameter and centered on the north coast of the Yucatan Peninsula, was created by a meteorite impact about 65 million years ago.

35 – 3 Seasonal Succession in Phytoplankton

Phytoplankton are minute, free-floating aquatic plants. In addition to the marked changes in abundance observed in phytoplankton over the course of a year, there is also a marked change in species composition. This change in the dominant species from season to season is called seasonal succession, and it occurs in a wide variety of locations. Under seasonal succession, one or more species dominate the phytoplankton for a shorter or longer period of time and then are replaced by another set of species. This pattern is repeated yearly. This succession is different from typical terrestrial ecological succession in which various plants replace one another until finally a so-called climax community develops, which persists for many years.

What are the factors causing this phenomenon? Considering that seasonal succession is most often and clearly seen in temperate seas, which have a marked change in temperature during a year, temperature has been suggested as a cause. This may be one of the factors, but it is unlikely to be the sole cause because there are species that become dominant species at various temperatures. Furthermore, temperature changes rather slowly in seawater, and the replacement of dominant species often is much more rapid.

Another suggested reason is the change in nutrient level over the year, with differing concentrations favoring different phytoplankton species. While this factor may also contribute, observations suggest that phytoplankton populations rise and fall much more quickly than nutrient concentrations change.

Yet another explanation is that species succession is a consequence of changes in seawater brought about by the phytoplankton living in it. Each species of phytoplankton secretes or excretes organic molecules into the seawater. These metabolites can have an effect on the organisms living in the seawater, either inhibiting or promoting their growth. For any individual organism, the amount of metabolite secreted is small. But the effect of secretions by all the individuals of the dominant species can be significant both for themselves and for other species.

These organic metabolites could, and probably do, include a number of different classes of organic compounds. Some are likely toxins, such as those released by the dinoflagellates (a species of plankton) during red tides, which inhibit growth of

other photosynthetic organisms. In such cases, the population explosion of dinoflagellates is so great that the water becomes brownish red in color from the billions of dinoflagellate cells. Although each cell secretes a minute amount of toxin, the massive dinoflagellate numbers cause the toxin to reach concentrations that kill many creatures. This toxin can be concentrated in such filter-feeding organisms as clams and mussels, rendering them toxic to humans.

Another class of metabolite is the vitamins. It is now known that certain phytoplankton species have requirements for certain vitamins, and that there are considerable differences among species as to requirements. The B vitamins, especially vitamin B12, thiamine and biotin, seem to be the most generally required. Some species may be unable to thrive until a particular vitamin, or group of vitamins, is present in the water. These vitamins are produced only by another species: hence, a succession of species could occur whereby first the vitamin-producing species is present and then the vitamin-requiring species follows.

Other organic compounds that may inhibit or promote various species include amino acids, carbohydrates, and fatty acids. Although it is suspected that these organic metabolites may have an important role in species succession and it has been demonstrated in the laboratory that phytoplankton species vary both in their ability to produce necessary vitamins and in their requirements for such in order to grow, evidence is still inadequate as to their real role in the sea.

There is also evidence to suggest that grazers (animals that feed on plants or stationary animals), particularly selective grazers, can influence the phytoplankton species composition. Many copepods (small, herbivorous crustaceans) and invertebrate larvae pick out selected phytoplankton species from mixed groups, changing the species composition.

A growing body of evidence now suggests that all of the factors considered here are operating simultaneously to produce species succession. The importance of any factor will vary with the particular phytoplankton species and the environmental conditions.

37 – 2 Direct Species Translocation

It is becoming increasingly common for conservationists to move individual animals or entire species from one site to another. This may be either to establish a new population where a population of conspecifics (animals or plants belonging to the same species) has become extinct or to add individuals to an existing population. The former is termed reintroduction and the latter reinforcement. In both cases, wild individuals are captured in one location and translocated directly to another.

Direct translocation has been used a wide range of plants and animals and was carried out to maintain populations as a source of food long before conservation was a familiar term. The number of translocations carried out under the banner of conservation has increased rapidly, and this has led to criticism of the technique because of the lack of evaluation of its efficacy and because of its potential disadvantages. The nature of translocation ranges from highly organized and researched national or international programs to ad hoc releases of rescued animals by well-intentioned animal lovers. In a fragmented landscape where many populations and habitats are isolated from others, translocations can play an effective role in conservation strategies; they can increase the number of existing populations or increase the size, genetic diversity, and demographic balance of a small population, consequently increasing its chances of survival.

Translocation clearly has a role in the recovery of species that have substantially declined and is the most likely method by which many sedentary species can recover all or part of their former range. However, against this is the potential for reinforcement translocations to spread disease from one population to another or to introduce deleterious or maladaptive genes to a population. Additionally, translocation of predators or competitors may have negative impacts on other species, resulting in an overall loss of diversity. Last but not least of these considerations is the effort and resources required in this type of action, which need to be justified by evidence of the likely benefits.

Despite the large number of translocations that have taken place, there is surprisingly little evidence of the efficacy of such actions. This is partly because many translocations have not been

strictly for conservation; neither have they been official nor legal, let alone scientific in their approach. Successful translocations inevitably get recorded and gain attention, whereas failures may never be recorded at all. This makes appraisal of the method very difficult. One key problem is a definition of success. Is translocation successful if the individuals survive the first week or a year, or do they need to reproduce for one or several generations? Whatever the answer, it is clear that a general framework is required to ensure that any translocation is justified, has a realistic chance of success, and will be properly monitored and evaluated for the benefit of future efforts.

An example of apparent translocation success involves the threatened Seychelles warbler. This species was once confined to Cousin Island, one of the Seychelles islands, and reduced to 26 individuals. Careful habitat management increased this number to over 300 birds, but the single population remained vulnerable to local catastrophic events. The decision was taken to translocate individuals to two nearby islands to reduce this risk. The translocations took place in 1988 and 1990, and both have resulted in healthy breeding populations. A successful translocation exercise also appears to have been achieved with red howler monkeys in French Guiana. A howler population was translocated from a site due to be flooded for hydroelectric power generation. The release site was an area where local hunting had reduced the density of the resident howler population. Released troops of monkeys were kept under visual observation and followed by radio tracking of 16 females. Although the troops appeared to undergo initial problems, causing them to split up, all the tracked females settled into normal behavioral patterns.

Unfortunately, the success stories are at least matched by accounts of failure. Reviewing translocation of amphibians and reptiles, researchers C.Kenneth Dodd and Richard A. Siegel concluded that most projects have not demonstrated success as conservation techniques and should not be advocated as though they were acceptable management and mitigation practices.

38 – 2 The Raccoons' Success

Raccoons have a vast transcontinental distribution, occurring throughout most of North America and Central America. They are found from southern Canada all the way to Panama, as well as on islands near coastal areas. They occur in each of the 49 states of the continental United States. Although raccoons are native only to the Western Hemisphere, they have been successfully transplanted to other parts of the globe.

Following a decline to a relatively low population level in the 1930s, raccoons began to prosper following their 1943 breeding season. A rapid population surge continued throughout the 1940s, and high numbers have been sustained ever since. By the late 1980s, the number of raccoons in North America was estimated to be at least 15 to 20 times the number that existed during the 1930s. By now, their numbers have undoubtedly grown even more, as they have continued to expand into new habitats where they were once either rare or absent, such as sandy prairies, deserts, coastal marshes, and mountains. Their spread throughout the Rocky Mountain West is indicative of the fast pace at which they can exploit new environments. Despite significant numbers being harvested and having suffered occasional declines, typically because of disease, the raccoon has consistently maintained high population levels.

Several factors explain the raccoon's dramatic increase in abundance and distribution. First, their success has been partially attributed to the growth of cities, as they often thrive in suburban and even urban settings. Furthermore, they have been deliberately introduced throughout the continent. Within the United States, they are commonly taken from one area to another, both legally and illegally, to restock hunting areas and, presumably, because people simply want them to be part of their local fauna. Their appearance and subsequent flourishing in Utah's Great Salt Lake valley within the last 40 years appears to be from such an introduction. As an example of the ease with which transplanted individuals can succeed, raccoons from Indiana (midwestern United States) have reportedly been able to flourish on islands off the coast of Alaska.

The raccoon's expansion in various areas may also be due to the spread of agriculture. Raccoons have been able to exploit crops, especially corn but also cereal grains, which have become dependable

food sources for them. The expansion of agriculture, however, does not necessarily lead to rapid increases in their abundance. Farming in Kansas and eastern Colorado (central and western United States) proceeded rapidly in the 1870s and 1880s, but this was about 50 years before raccoons started to spread out from their major habitat, the wooded river bottomlands. They have also expanded into many areas lacking any agriculture other than grazing and into places without forests or permanent streams.

Prior to Europeans settling and farming the Great Plains Region, raccoons probably were just found along its rivers and streams and in the wooded areas of its southeastern section. With the possible exception of the southern part of the province of Manitoba, their absence was notable throughout Canada. They first became more widely distributed in the southern part of Manitoba, and by the 1940s were abundant throughout its southeastern portion. In the 1950s their population swelled in Canada. The control of coyotes in the prairie region in the 1950s may have been a factor in raccoon expansion. If their numbers are sufficient coyotes might be able to suppress raccoon populations (though little direct evidence supports this notion). By the 1960s the raccoon had become a major predator of the canvasback ducks nesting in southwestern Manitoba.

The extermination of the wolf from most of the contiguous United States may have been a critical factor in the raccoon's expansion and numerical increase. In the eighteenth century, when the wolfs range included almost all of North America, raccoons apparently were abundant only in the deciduous forests of the East, Gulf Coast, and Great Lakes regions, though they also extended into the wooded bottomlands of the Midwest's major rivers. In such areas, their arboreal habits and the presence of hollow den trees should have offered some protection from wolves and other large predators. Even though raccoons may not have been a significant part of their diet, wolves surely would have tried to prey on those exposed in relatively treeless areas.

39 – 2 The Extinction of Moa

Between 80 and 85 million years ago, Gondwanaland, a giant continent made up of what today is Africa, Antarctica, Australia, and South America, broke up, thus causing what is now New Zealand to become separated from the larger landmass. After the separation, any creature unable to cross a considerable distance of ocean could not migrate to New Zealand. Snakes and most mammals evolved after the separation. Thus, there are no New Zealand snakes, and bats, which flew there, and seals, which swam there, were the only mammals on New Zealand when Polynesian settlers (the Maori) arrived there about a thousand years ago.

When the Maori arrived in New Zealand, they encountered birds that had been evolving for 80 million years without the presence of mammalian predators. The most striking of these animals must have been moa. Now extinct, moa were gigantic wingless birds that stood as much as 10 feet (3 meters) tall and weighed as much as 550 pounds (250 kilograms). They are known from a diverse array of remains including eggshells, eggs, a few mummified carcasses, vast numbers of bones, and some older fossilized bone. The species of moa that are currently recognized occupied ecological niches customarily filled elsewhere by large mammalian browsing herbivores. They may have had relatively low reproductive rates; apparently, they usually laid only one egg at a time.

It seems possible that when Captain James Cook first visited New Zealand in 1769, moa (or at least one of the moa species) may have still survived in the remote areas in the western part of New Zealand's South Island. If so, these individuals would have been the last of their kind. Climatic conditions in New Zealand appear to have been relatively stable over the period during which moa became extinct. Different factors could have worked in concert to account for their abrupt disappearance.

Vegetation was considerably altered by the Maori occupation of New Zealand, a change not easily explained by climate variation or other possible factors. Forest and shrubland burning appears to have reduced the prime habitat of many moa species. However, the main forest burning started around 700 years ago, after what current archaeological evidence indicates was the most intensive stage of moa hunting. While there

appears to have been extensive burning on the east side of New Zealand's South Island, large forest tracts remained in the most southern part of the island. Because major habitat destruction seems to have occurred after moa populations already were depleted, and because some habitat that could have sheltered moa populations remained, it would seem that other factors were also at work in the extinction of these birds.

For South Island, human predation appears to have been a significant factor in the depletion of the population of moa. At one excavated Maori site, moa remains filled six railway cars. The density of Maori settlements and artifacts increased substantially at the time of the most intensive moa hunting (900 to 600 years ago). This period was followed by a time of decline in the Maori population and a societal transition to smaller, less numerous settlements. The apparent decline fits the pattern expected as a consequence of the Maori's overexploitation of moa.

Finally, the Maori introduced the Polynesian rat and the dog to New Zealand. The actions of these potential nest predators could have reduced moa populations without leaving much direct evidence. The Maori may have also inadvertently brought pests and disease organisms in fowls, which could have crossed over to eradicate moa populations. The possibility of analyzing ancient DNA to identify past diseases of extinct animals is being explored. However, evidence of such diseases is difficult to determine directly from paleoecological or archaeological remains. For these reasons, it is hard to determine the likelihood that introduced disease organisms were a cause of the decline of moa, but they are potentially significant.

While the last of these possible causes remains speculative, definite clues exist for the action of the first two causes. The story of moa species and their demise raises ecological issues on the vulnerability of species to human-caused changes—including altered vegetative cover of the landscape, change in the physical environment, and modification of the flora and fauna of a region by eliminating some species and introducing others.

74 – 1 Early Horses

The earliest-known horse fossils come from the Eocene epoch (from 57 to 34 million years ago), and they are so different from the modern version that it was not initially realized that there was any relation. Dawn horse, or Eohippus, as this animal has been called, has been found in both Europe and North America. Eohippus was tiny, about the size of a small dog, and apparently lived in wooded areas. These animals had hooves but, in contrast to modern horses, they had four hoofed toes on the front feet, three on the rear, and the hooves were padded. Eohippus was also pug-nosed by comparison to modern horses, and its teeth reveal that it was a browser that fed on a variety of plants. In fact, although it was completely herbivorous (plant eating), this little horse had canine teeth—a reminder that many of its predecessors were carnivores (meat eaters).

Throughout the Eocene and Oligocene epochs (from 57 to 23 million years ago), the descendants of Eohippus evolved in a fairly straightforward way that is well documented by their fossils. They became gradually larger; the middle toe, eventually to become the single hoof of the modern horses, became stronger and more prominent, and the grinding surfaces of the teeth became larger, with complex, resistant ridges. But the resemblance to Eohippus remained clear. It was only in the Miocene (from 23 to 5 million years ago), coincident with the spread of grassy prairies, that abrupt changes took place, resulting in several different lineages of horse evolution, only one of which is still extant: the modern horse.

Several of the physical characteristics of today's animals are ultimately linked to their diet of grasses. Foremost among the Miocene modifications that led toward present-day horses were changes in the teeth and the shape of the head. Compared to the succulent leaves of tropical trees that were the fodder for some of the horse's ancestors, grasses are abrasive and much more difficult to chew and grind. The response of the Miocene horses was to develop teeth with much more elaborate and resistant grinding surfaces and with much larger crowns, at least part of which could grow out of the gums as they were worn down. These changes meant that the head had to be much deeper, and the muzzle longer, to make room for the long rows of grinding teeth along the horse's cheeks. At about the same time, the legs and feet of the ancestors of today's horses became

better adapted to rapid running across the spreading grasslands. This occurred through fusion of several of the independent bones in the lower parts of the legs, making them stiffer, and through further emphasis of the central hoofed toe, which by now bore the entire weight of the animal. In place of a foot, the horse has a single toe at the end of its leg.

By the middle or late Miocene, many of the extant horses were at least superficially similar to modern horses. Based on the fossil record, much of their development seems to have occurred in North America, but by the Pleistocene epoch (from 1.6 million to 10,000 years ago), the modern horse genus, *Equus*, had spread over much of the world. Then, inexplicably, only 8,000 to 10,000 years ago, horses disappeared from North America. The reason for this extinction is unknown. Whatever the cause, it is a fact that the plains of North America were without these animals for thousands of years, until horses brought from Europe by the early Spanish explorers escaped and began to repopulate the vast grasslands.

It is clear that many of the familiar features of the modern horse are directly or indirectly related to its diet and its preferred environment, grasslands. But how and why did the grasslands develop when they did? There are competing theories on this question, but only a few are consistent with the evidence. Most of these invoke a change in global climate as an important factor, perhaps the overriding factor. In particular, grasslands expanded rapidly as the climate in continental interiors became cooler and drier.

40 – 3 Amphibian Thermoregulation

In contrast to mammals and birds, amphibians are unable to produce thermal energy through their metabolic activity, which would allow them to regulate their body temperature independent of the surrounding or ambient temperature. However, the idea that amphibians have no control whatsoever over their body temperature has been proven false because their body temperature does not always correspond to the surrounding temperature. While amphibians are poor thermoregulators, they do exercise control over their body temperature to a limited degree.

Physiological adaptations can assist amphibians in colonizing habitats where extreme conditions prevail. The tolerance range in body temperature represents the range of temperatures within which a species can survive. One species of North American newt is still active when temperatures drop to -2°C while one South American frog feels comfortable even when temperatures rise to 41°C —the highest body temperature measured in a free-ranging amphibian. Recently it has been shown that some North American frog and toad species can survive up to five days with a body temperature of -6°C with approximately one-third of their body fluids frozen. The other tissues are protected because they contain the frost-protective agents glycerin or glucose. Additionally, in many species the tolerance boundaries are flexible and can change as a result of acclimatization (long-term exposure to particular conditions).

Frog species that remain exposed to the sun despite high diurnal (daytime) temperatures exhibit some fascinating modifications in the skin structure that function as morphological adaptations. Most amphibian skin is fully water permeable and is therefore not a barrier against evaporation or solar radiation. The African savanna frog *Hyperolius viridiflavus* stores guanine crystals in its skin, which enable it to better reflect solar radiation, thus providing protection against overheating. The tree frog *Phyllomedusa sauvagii* responds to evaporative losses with gland secretions that provide a greasy film over its entire body that helps prevent desiccation (dehydration).

However, behavior is by far the most important factor in thermoregulation. The principal elements in behavioral thermoregulation are basking (heliothermy), heat exchange with substrates such as rock or earth (thigmothermy), and diurnal and

annual avoidance behaviors, which include moving to shelter during the day for cooling and hibernating or estivating (reducing activity during cold or hot weather, respectively). Heliothermy is especially common among frogs and toads: it allows them to increase their body temperature by more than 10°C . The Andean toad *Bufo spinulosus* exposes itself immediately after sunrise on moist ground and attains its preferred body temperature by this means, long before either ground or air is correspondingly warmed. A positive side effect of this approach is that it accelerates the digestion of the prey consumed overnight, thus also accelerating growth. Thigmothermy is a behavior present in most amphibians, although pressing against the ground serves a dual purpose: heat absorption by conductivity and water absorption through the skin. The effect of thigmothermy is especially evident in the Andean toad during rainfall: its body temperature corresponds to the temperature of the warm earth and not to the much cooler air temperature.

Avoidance behavior occurs whenever physiological and morphological adaptations are insufficient to maintain body temperature within the vital range. Nocturnal activity in amphibians with low tolerance for high ambient temperatures is a typical thermoregulatory behavior of avoidance. Seasonal avoidance behavior is extremely important in many amphibians. Species whose habitat lies in the temperate latitudes are confronted by lethal low temperatures in winter, while species dwelling in semi- and regions are exposed to long dry, hot periods in summer.

In amphibians hibernation occurs in mud or deep holes away from frost. North of the Pyrenees Mountains, the natterjack toad offers a good example of hibernation, passing the winter dug deep into sandy ground. Conversely, natterjacks in southern Spain remain active during the mild winters common to the region and are instead forced into inactivity during the dry, hot summer season. Summer estivation also occurs by burrowing into the ground or hiding in cool, deep rock crevasses to avoid desiccation and lethal ambient temperature. Amphibians are therefore hardly at mercy of ambient temperature, since by means of the mechanisms described above they are more than exercising some control over their body temperature.

42 – 1 Geographic Isolation of Species

Biologist Ernst Mayr defined a species as “an actually or potentially interbreeding population that does not interbreed with other such populations when there is opportunity to do so.” A key event in the origin of many species is the separation of a population with its gene pool (all of the genes in a population at any one time) from other populations of the same species, thereby preventing population interbreeding. With its gene pool isolated, a separate population can follow its own evolutionary course. In the formation of many species, the initial isolation of a population seems to have been a geographic barrier. This mode of evolving new species is called allopatric speciation.

Many factors can isolate a population geographically. A mountain range may emerge and gradually split a population of organisms that can inhabit only lowland lakes, certain fish populations might become isolated in this way. Similarly, a creeping glacier may gradually divide a population, or a land bridge such as the Isthmus of Panama may form and separate the marine life in the ocean waters on either side.

How formidable must a geographic barrier be to keep populations apart? It depends on the ability of the organisms to move across barriers. Birds and coyotes can easily cross mountains and rivers. The passage of wind-blown tree pollen is also not hindered by such barriers, and the seeds of many plants may be earned back and forth on animals. In contrast, small rodents may find a deep canyon or a wide river an effective barrier. For example, the Grand Canyon, in the southwestern United States, separate the range of the white-tailed antelope squirrel from that of the closely related Harris' antelope squirrel. Smaller, with a shorter tail that is white underneath, the white-tailed antelope squirrel inhabits deserts north of the canyon and west of the Colorado River in southern California. Harris' antelope squirrel has a more limited range in deserts south of the Grand Canyon.

Geographic isolation creates opportunities for new species to develop, but it does not necessarily lead to new species because speciation occurs only when the gene pool undergoes enough changes to establish reproductive barriers between the isolated population and its parent population. The likelihood of allopatric speciation increases when a population is small as well as isolated, making it more likely than a large population to have its gene

pool changed substantially. For example, in less than two million years, small populations of stray animals and plants from the South American mainland that managed to colonize the Galapagos Islands gave rise to all the species that now inhabit the islands.

When oceanic islands are far enough apart to permit populations to evolve in isolation, but close enough to allow occasional dispersions to occur, they are effectively outdoor laboratories of evolution. The Galapagos island chain is one of the world's greatest showcases of evolution. Each island was born from underwater volcanoes and was gradually covered by organisms derived from strays that rode the ocean currents and winds from other islands and continents. Organisms can also be carried to islands by other organisms, such as sea birds that travel long distances with seeds clinging to their feathers.

The species on the Galapagos Islands today, most of which occur nowhere else, descended from organisms that floated, flew, or were blown over the sea from the South American mainland. For instance, the Galapagos island chain has a total of thirteen species of closely related birds called Galapagos finches. These birds have many similarities but differ in their feeding habits and their beak type, which is correlated with what they eat. Accumulated evidence indicates that all thirteen finch species evolved from a single small population of ancestral birds that colonized one of the islands. Completely isolated on the island after migrating from the mainland, the founder population may have undergone significant changes in its gene pool and become a new species. Later, a few individuals of this new species may have been blown by storms to a neighboring island. Isolated on this second island, the second founder population could have evolved into a second new species, which could later recolonize the island from which its founding population emigrated. Today each Galapagos island has multiple species of finches, with as many as ten on some islands.

42 – 2 Explaining Dinosaur Extinction

Dinosaurs rapidly became extinct about 65 million years ago as part of a mass extinction known as the K-T event, because it is associated with a geological signature known as the K-T boundary, usually a thin band of sedimentation found in various parts of the world (K is the traditional abbreviation for the Cretaceous, derived from the German name Kreidezeit). Many explanations have been proposed for why dinosaurs became extinct. For example, some have blamed dinosaur extinction on the development of flowering plants, which were supposedly more difficult to digest and could have caused constipation or indigestion—except that flowering plants first evolved in the Early Cretaceous, about 60 million years before the dinosaurs died out. In fact, several scientists have suggested that the duckbill dinosaurs and horned dinosaurs, with their complex battery of grinding teeth, evolved to exploit this new resource of rapidly growing flowering plants. Others have blamed extinction on competition from the mammals, which allegedly ate all the dinosaur eggs—except that mammals and dinosaurs appeared at the same time in the Late Triassic, about 190 million years ago, and there is no reason to believe that mammals suddenly acquired a taste for dinosaur eggs after 120 million years of coexistence. Some explanations (such as the one stating that dinosaurs all died of diseases) fail because there is no way to scientifically test them, and they cannot move beyond the realm of speculation and guesswork.

This focus on explaining dinosaur extinction misses an important point: the extinction at the end of the Cretaceous was a global event that killed off organisms up and down the food chain. It wiped out many kinds of plankton in the ocean and many marine organisms that lived on the plankton at the base of the food chain. These included a variety of clams and snails, and especially the ammonites, a group of shelled squidlike creatures that dominated the Mesozoic seas and had survived many previous mass extinctions. The K-T event marked the end of the marine reptiles, such as the mosasaurs and the plesiosaurs, which were the largest creatures that had ever lived in the seas and which ruled the seas long before whales evolved. On land, there was also a crisis among the land plants, in addition to the disappearance of dinosaurs. So any event that can explain the destruction of the base of the food chain (plankton

in the ocean, plants on land) can better explain what happened to organisms at the top of the food chain, such as the dinosaurs. By contrast, any explanation that focuses strictly on the dinosaurs completely misses the point. The Cretaceous extinctions were a global phenomenon, and dinosaurs were just a part of a bigger picture.

According to one theory, the Age of Dinosaurs ended suddenly 65 million years ago when a giant rock from space plummeted to Earth. Estimated to be ten to fifteen kilometers in diameter, this bolide (either a comet or an asteroid) was traveling at cosmic speeds of 20-70 kilometers per second, or 45,000-156,000 miles per hour. Such a huge mass traveling at such tremendous speeds carries an enormous amount of energy. When the bolide struck, this energy was released and generated a huge shock wave that leveled everything for thousands of kilometers around the impact and caused most of the landscape to burst into flames. The bolide struck an area of the Yucatan Peninsula of Mexico known as Chicxulub, excavating a crater 15-20 kilometers deep and at least 170 kilometers in diameter. The impact displaced huge volumes of seawater, causing much flood damage in the Caribbean. Meanwhile, the bolide itself excavated 100 cubic kilometers of rock and debris from the site, which rose to an altitude of 100 kilometers. Most of it fell back immediately, but some of it remained as dust in the atmosphere for months. This material, along with the smoke from the fires, shrouded Earth, creating a form of nuclear winter. According to computerized climate models, global temperatures fell to near the freezing point, photosynthesis halted, and most plants on land and in the sea died. With the bottom of the food chain destroyed, dinosaurs could not survive.

45 – 3 Feeding Strategies in the Ocean

In the open sea, animals can often find food reliably available in particular regions or seasons (e.g., in coastal areas in springtime). In these circumstances, animals are neither constrained to get the last calorie out of their diet nor is energy conservation a high priority. In contrast, the food levels in the deeper layers of the ocean are greatly reduced, and the energy constraints on the animals are much more severe. To survive at those levels, animals must maximize their energy input, finding and eating whatever potential food source may be present.

In the near-surface layers, there are many large, fast carnivores as well as an immense variety of planktonic animals, which feed on plankton (small, free-floating plants or animals) by filtering them from currents of water that pass through a specialized anatomical structure. These filter-feeders thrive in the well-illuminated surface waters because oceans have so many very small organisms, from bacteria to large algae to larval crustaceans. Even fishes can become successful filter-feeders in some circumstances. Although the vast majority of marine fishes are carnivores, in near-surface regions of high productivity the concentrations of larger phytoplankton (the plant component of plankton) are sufficient to support huge populations of filter-feeding sardines and anchovies. These small fishes use their gill filaments to strain out the algae that dominate such areas. Sardines and anchovies provide the basis for huge commercial fisheries as well as a food resource for large numbers of local carnivores, particularly seabirds. At a much larger scale, baleen whales and whale sharks are also efficient filter-feeders in productive coastal or polar waters, although their filtered particles comprise small animals such as copepods and krill rather than phytoplankton.

Filtering seawater for its particulate nutritional content can be an energetically demanding method of feeding, particularly when the current of water to be filtered has to be generated by the organism itself, as is the case for all planktonic animals. Particulate organic matter of at least 2.5 micrograms per cubic liter is required to provide a filter-feeding planktonic organism with a net energy gain. This value is easily exceeded in most coastal waters, but in the deep sea, the levels of organic matter range from next to nothing to around 7

micrograms per cubic liter. Even though mean levels may mask much higher local concentrations, it is still the case that many deep-sea animals are exposed to conditions in which a normal filter-feeder would starve.

There are, therefore, fewer successful filter-feeders in deep water, and some of those that are there have larger filtering systems to cope with the scarcity of particles. Another solution for such animals is to forage in particular layers of water where the particles may be more concentrated. Many of the groups of animals that typify the filter-feeding lifestyle in shallow water have deep-sea representatives that have become predatory. Their filtering systems, which reach such a high degree of development in shallow-water species, are greatly reduced. Alternative methods of active or passive prey capture have been evolved, including trapping and seizing prey, entangling prey, and sticky tentacles.

In the deeper waters of the oceans, there is a much greater tendency for animals to await the arrival of food particles or prey rather than to search them out actively (thus minimizing energy expenditure). This has resulted in a more stealthy style of feeding, with the consequent emphasis on lures and/or the evolution of elongated appendages that increase the active volume of water controlled or monitored by the animal. Another consequence of the limited availability of prey is that many animals have developed ways of coping with much larger food particles, relative to their own body size, than the equivalent shallower species can process. Among the fishes there is a tendency for the teeth and jaws to become appreciably enlarged. In such creatures, not only are the teeth hugely enlarged and/or the jaws elongated but the size of the mouth opening may be greatly increased by making the jaw articulations so flexible that they can be effectively dislocated. Very large or long teeth provide almost no room for cutting the prey into a convenient size for swallowing, the fish must gulp the prey down whole.

44 – 1 From Fish to Terrestrial Vertebrates

One of the most significant evolutionary events that occurred on Earth was the transition of water-dwelling fish to terrestrial tetrapods (four-limbed organisms with backbones). Fish probably originated in the oceans, and our first records of them are in marine rocks. However, by the Devonian Period (408 million to 362 million years ago), they had radiated into almost all available aquatic habitats, including freshwater settings. One of the groups whose fossils are especially common in rocks deposited in fresh water is the lobe-finned fish.

The freshwater Devonian lobe-finned fish rhipidistian crossopterygian is of particular interest to biologists studying tetrapod evolution. These fish lived in river channels and lakes on large deltas. The delta rocks in which these fossils are found are commonly red due to oxidized iron minerals, indicating that the deltas formed in a climate that had alternate wet and dry periods. If there were periods of drought, any adaptations allowing the fish to survive the dry conditions would have been advantageous. In these rhipidistians, several such adaptations existed. It is known that they had lungs as well as gills for breathing. Cross sections cut through some of the fossils reveal that the mud filling the interior of the carcass differed in consistency and texture depending on its location inside the fish. These differences suggest a saclike cavity below the front end of the gut that can only be interpreted as a lung. Gills were undoubtedly the main source of oxygen for these fish, but the lungs served as an auxiliary breathing device for gulping air when the water became oxygen depleted, such as during extended periods of drought. So, these fish had already evolved one of the prime requisites for living on land, the ability to use air as a source of oxygen.

A second adaptation of these fish was in the structure of the lobe fins. The fins were thick, fleshy, and quite sturdy, with a median axis of bone down the center. They could have been used as feeble locomotor devices on land, perhaps good enough to allow a fish to flop its way from one pool of water that was almost dry to an adjacent pond that had enough water and oxygen for survival. These fins eventually changed into short, stubby legs. The bones of the fins of a Devonian rhipidistian exactly match in number and position the limb bones of the earliest known tetrapods, the amphibians. It should be emphasized that the

evolution of lungs and limbs was in no sense an anticipation of future life on land. These adaptations developed because they helped fish to survive in their existing aquatic environment.

What ecological pressures might have caused fishes to gradually abandon their watery habitat and become increasingly land-dwelling creatures? Changes in climate during the Devonian may have had something to do with this if freshwater areas became progressively more restricted. Another impetus may have been new sources of food. The edges of ponds and streams surely had scattered dead fish and other water-dwelling creatures. In addition, plants had emerged into terrestrial habitats in areas near streams and ponds, and crabs and other arthropods were also members of this earliest terrestrial community. Thus, by the Devonian the land habitat marginal to freshwater was probably a rich source of protein that could be exploited by an animal that could easily climb out of water. Evidence from teeth suggests that these earliest tetrapods did not utilize land plants as food; they were presumably carnivorous and had not developed the ability to feed on plants.

How did the first tetrapods make the transition to a terrestrial habitat? Like early land plants such as rhyniophytes, they made only a partial transition, they were still quite tied to water. However, many problems that faced early land plants were not applicable to the first tetrapods. The ancestors of these animals already had a circulation system, and they were mobile, so that they could move to water to drink. Furthermore, they already had lungs, which rhipidistians presumably used for auxiliary breathing. The principal changes for the earliest tetrapods were in the skeletal system—changes in the bones of the fins, the vertebral column, pelvic girdle, and pectoral girdle.

46 – 4 Crown of Thorns Starfish and Coral Reefs

The crown of thorns starfish, *Acanthaster Tlanci*, is large, twenty-five to thirty-five centimeters in diameter, and has seven to twenty-one arms that are covered in spines. It feeds primarily on coral and is found from the Indian Ocean to the west coast of Central America, usually at quite low population densities. Since the mid-1950s, population outbreaks at densities four to six times greater than normal have occurred at the same time in places such as Hawaii, Tahiti, Panama, and the Great Barrier Reef. The result has often been the loss of a fifty percent to nearly one hundred percent of the coral cover over large areas.

A single *Acanthaster* can consume five to six square meters of coral polyps per year, and dense populations can destroy up to six square kilometers per year and move on rapidly. *Acanthasters* show a preference for branching corals, especially Acroporids. After an outbreak in a particular area, it is common to find that Acroporids have been selectively removed, leaving a mosaic of living and dead corals. In places where Acroporids previously dominated the community devastation can be almost complete, and local areas of reefs have collapsed.

Areas of dead coral are usually colonized rapidly by algae and often are later colonized by sponges and soft corals. Increases in abundance of plant-eating fish and decreases in abundance of coral-feeding fish accompany these changes. Coral larvae settle among the algae and eventually establish flourishing coral colonies. In ten to fifteen years the reefs often return to about the same percentage of coral cover as before. Development of a four-species diversity takes about twenty years.

Two schools of thought exist concerning the cause of these outbreaks. One group holds that they are natural phenomena that have occurred many times in the past, citing old men's recollections of earlier outbreaks and evidence from traditional cultures. The other group maintains that recent human activities ranging from physical coral destruction through pollution to predator removal have triggered these events.

One theory, the adult aggregation hypothesis, maintains that most species is more abundant than we realize when a storm destroys coral and causes a food shortage. The adult *Acanthasters* converge

on remaining portions of healthy coral and feed hungrily. Certainly there have been outbreaks of *Acanthaster* following large storms, but there is little evidence that the storms have caused the enough reef damage to create a food shortage for these starfish.

Two other hypotheses attempt to explain the increased abundance of *Acanthaster* after episodes of high terrestrial runoff following storms. The first hypothesis is that low salinity and high temperatures favor the survival of the starfish larvae. The second hypothesis emphasizes the food web aspect, suggesting that strong fresh water runoff brings additional nutrients to the coastal waters, stimulating phytoplankton production and promoting more rapid development and better survival of the starfish larvae.

Those favoring anthropogenic (human influenced) causes have pointed to the large proportion of outbreaks that have been near centers of human populations. It has been suggested that coral polyps are the main predators of the starfish larvae. Destruction of coral by blasting and other bad land use practices would reduce predation on the starfish larvae and cause a feedback in which increases in *Acanthaster* populations cause still further coral destruction. Unfortunately, there are too few documented instances of physical destruction of coral being followed by outbreaks of *Acanthaster* for these hypotheses to be fully supported.

Another group of hypothesis focuses on removal of *Acanthaster's* predators. Some have suggested that the predators might have been killed off by pollution whereas others have suggested that the harvesting of vertebrate and invertebrate predators of *Acanthaster* could have reduced mortality and caused increased abundance of adults. The problem with this group of hypothesis is that it is difficult to understand how reduced predation would lead to sudden increases in *Acanthaster* numbers in several places at the same time in specific years. It seems probable that there is no single explanation but that there are elements of the truth in several of the hypotheses. That is there are natural processes that have led to outbreaks in the past, but human impact has increased the frequency and severity of the outbreaks.

45 – 1 Microscopes The Beringia Landscape

During the peak of the last ice age, northeast Asia (Siberia) and Alaska were connected by a broad land mass called the Bering Land Bridge. This land bridge existed because so much of Earth's water was frozen in the great ice sheets that sea levels were over 100 meters lower than they are today. Between 25,000 and 10,000 years ago, Siberia, the Bering Land Bndge, and Alaska shared many environmental characteristics. These included a common mammalian fauna of large mammals, a common flora composed of broad grasslands as well as wind-swept dunes and tundra, and a common climate with cold, dry winters and somewhat wanner summers. The recognition that many aspects of the modern flora and fauna were present on both sides of the Bering Sea as remnants of the ice-age landscape led to this region being named Beringia.

It is through Beringia that small groups of large mammal hunters, slowly expanding their hunting territories, eventually colonized North and South America. On this archaeologists generally agree, but that is where the agreement stops. One broad area of disagreement in explaining the peopling of the Americas is the domain of paleoecologists, but it is critical to understanding human history: what was Beringia like?

The Beringian landscape was very different from what it is today. Broad, windswept valleys; glaciated mountains; sparse vegetation; and less moisture created a rather forbidding land mass. This land mass supported herds of now-extinct species of mammoth, bison, and horse and somewhat modern versions of caribou, musk ox, elk, and saiga antelope. These grazers supported in turn a number of impressive carnivores, including the giant short-faced bear, the saber-tooth cat, and a large species of lion.

The presence of mammal species that require grassland vegetation has led Arctic biologist Dale Guthrie to argue that while cold and dry, there must have been broad areas of dense vegetation to support herds of mammoth, horse, and bison. Further, nearly all of the ice-age fauna had teeth that indicate an adaptation to grasses and sedges; they could not have been supported by a modern flora of mosses and lichens. Guthrie has also demonstrated that the landscape must have been subject to intense and continuous winds, especially in winter. He makes this argument based on the

anatomy of horse and bison, which do not have the ability to search for food through deep snow cover. They need landscapes with strong winds that remove the winter snows, exposing the dry grasses beneath. Guthrie applied the term "mammoth steppe" to characterize this landscape.

In contrast, Paul Colinvaux has offered a counterargument based on the analysis of pollen in lake sediments dating to the last ice age. He found that the amount of pollen recovered in these sediments is so low that the Beringian landscape during the peak of the last glaciation was more likely to have been what he termed a "polar desert," with little or only sparse vegetation, in no way was it possible that this region could have supported large herds of mammals and thus, human hunters. Guthrie has argued against this view by pointing out that radiocarbon analysis of mammoth, horse, and bison bones from Beringian deposits revealed that the bones date to the period of most intense glaciation.

The argument seemed to be at a standstill until a number of recent studies resulted in a spectacular suite of new finds. The first was the discovery of a 1,000-square-kilometer preserved patch of Beringian vegetation dating to just over 17,000 years ago—the peak of the last ice age. The plants were preserved under a thick ash fall from a volcanic eruption. Investigations of the plants found grasses, sedges, mosses, and many other varieties in a nearly continuous cover, as was predicted by Guthrie. But this vegetation had a thin root mat with no soil formation, demonstrating that there was little long-term stability in plant cover, a finding supporting some of the arguments of Colinvaux. A mixture of continuous but thin vegetation supporting herds of large mammals is one that seems plausible and realistic with the available data.

47 – 2 Termite Ingenuity

Termites, social insects which live in colonies that, in some species, contain 2 million individuals or more, are often incorrectly referred to as white ants. But they are certainly not ants. Termites, unlike ants, have gradual metamorphosis with only three life stage: egg, nymph, and adult. Ants and the other social members of their order, certain bees and wasps, have complete metamorphosis in four life stages; egg, larva, pupa, and adult. The worker and soldier castes of social ants, bees, and wasps consist of only females, all daughters of a single queen that mated soon after she matured and thereafter never mated again. The worker and soldier castes of termites consist of both males and females, and the queen lives permanently with a male consort.

Since termites are small and soft-bodied, they easily become desiccated and must live in moist places with a high relative humidity. They do best when the relative humidity in their nest is above 96 percent and the temperature is fairly high, an optimum of about 79°F for temperate zone species and about 86°F for tropical species. Subterranean termites, the destructive species that occurs commonly throughout the eastern United States, attain these conditions by nesting in moist soil that is in contact with wood, their only food. The surrounding soil keeps the nest moist and tends to keep the temperature at a more or less favorable level. When it is cold in winter, subterranean termites move to burrows below the frost line.

Some tropical termites are more ingenious engineers, constructing huge above-ground nests with built-in “air conditioning” that keeps the nest moist, at a constant temperature, and well supplied with oxygen. Among the most architecturally advanced of these termites is an African species, *Macrotermes natalensis*. Renowned Swiss entomologist Martin Luscher described the mounds of this fungus-growing species as being as much as 16 feet tall, 16 feet in diameter at their base, and with a cement-like wall of soil mixed with termite saliva that is from 16 to 23 inches thick. The thick and dense wall of the mound insulates the interior microclimate from the variations in humidity and temperature of the outside atmosphere. Several narrow and relatively thin-walled ridges on the outside of the mound extend from near its base almost to its top.

According to Luscher, a medium-sized nest of *Macrotermes* has a population of about 2 million individuals. The metabolism of so many termites and of the fungus that they grow in their gardens as food helps keep the interior of the nest warm and supplies some moisture to the air in the nest. The termites saturate the atmosphere of the nest, bringing it to about 100 percent relative humidity, by carrying water up from the soil.

But how is this well-insulated nest ventilated? Its many occupants require over 250 quarts of oxygen (more than 1,200 quarts of air) per day. How can so much oxygen diffuse through the thick walls of the mound? Even the pores in the wall are filled with water, which almost stops the diffusion of gases. The answer lies in the construction of the nest. The interior consists of a large central core in which the fungus is grown, below it is “cellar” of empty space, above it is an “attic” of empty space, and within the ridges on the outer wall of the nest, there are many small tunnels that connect the cellar and the attic. The warm air in the fungus gardens rises through the nest up to the attic. From the attic, the air passes into the tunnels in the ridges and flows back down to the cellar. Gases, mainly oxygen coming in and carbon dioxide going out, easily diffuse into or out of the ridges, since their walls are thin and their surface area is large because they protrude far out from the wall of the mound. Thus air that flows down into the cellar through the ridges is relatively rich in oxygen, and has lost much of its carbon dioxide. It supplies the nest’s inhabitants with fresh oxygen as it rises through the fungus-growing area back up to the attic.

48 – 2 Determining Dinosaur Diet

Determining what extinct dinosaurs ate is difficult, but we can infer some aspects of their dietary preferences. Traditionally, this information has been derived from direct evidence, such as stomach contents, and indirect evidence, such as establishing a correlation between particular body characteristics and diets of living animals and then inferring habits for dinosaurs.

Animals such as house cats and dogs have large, stabbing canine teeth at the front of the mouth and smaller, equally sharp teeth farther back in their jaws. Many of these animals are also armed with sharp claws. The advantage of teeth and claws as predatory tools is obvious. Now consider animals like cows, horses, rabbits, and mice. These animals have flat teeth at the back of the jaw that are analogous to and have the same function as grindstones. Unlike the meat-slicing and stabbing teeth of carnivores, the teeth of these animals grind and shred plant material before digestion.

More clues exist in other parts of the skull. The jaw joint of carnivores such as dogs and cats has the mechanical advantage of being at the same level as the tooth row, allowing the jaws to close with tremendous speed and forcing the upper teeth to occlude against the lower teeth with great precision. In herbivorous animals, rapid jaw closure is less important. Because the flat teeth of herbivores work like grindstones, however, the jaws must move both side to side and front to back. The jaw joints of many advanced herbivores, such as cows, lie at a different level than the tooth row, allowing transverse tearing, shredding, and compression of plant material. If we extend such observations to extinct dinosaurs, we can infer dietary preferences (such as carnivory and herbivory), even though we cannot determine the exact diet. The duck-billed dinosaurs known as hadrosaurs are a good example of a group whose jaw joint is below the level of the tooth row, which probably helped them grind up tough, fibrous vegetation.

Paleontologists would like to be much more specific about a dinosaur's diet than simply differentiating carnivore from herbivore. This finer level of resolution requires direct fossil evidence of dinosaur meals. Stomach contents are only rarely preserved, but when present, allow us to determine exactly what these animals were eating.

In the stomach contents of specimens of *Coelophysis* (a small, long-necked dinosaur) are bones from juvenile animals of the same species. At one time, these were thought to represent embryonic animals, suggesting that this small dinosaur gave birth to live young rather than laying eggs. Further research indicated that the small dinosaurs were too large and too well developed to be prehatchling young. In addition, the juveniles inside the body cavity were of different sizes. All the evidence points to the conclusion that these are the remains of prey items and that, as an adult, *Coelophysis* was at least in part a cannibal.

Fossilized stomach contents are not restricted to carnivorous dinosaurs. In a few rare cases, most of them "mummies" (unusually well preserved specimens), fossilized plant remains have been found inside the body cavity of hadrosaurs. Some paleontologists have argued that these represent stream accumulations rather than final meals. The best known of these cases is the second *Edmontosaurus* mummy collected by the Sternbergs. In the chest cavity of this specimen, which is housed in the Senckenberg Museum in Germany, are the fossil remains of conifer needles, twigs, seeds, and fruits. Similar finds in *Corythosaurus* specimens from Alberta, Canada, have also been reported, indicating that at least two kinds of Late Cretaceous hadrosaurs fed on the sorts of tress that are common in today's boreal woodlands.

A second form of direct evidence comes from coprolites (fossilized bodily waste). Several dinosaur fossil localities preserve coprolites. Coprolites yield unequivocal evidence about the dietary habits of dinosaurs. Many parts of plants and animals are extremely resistant to the digestive systems of animals and pass completely through the body with little or no alteration. Study of coprolites has indicated that the diets of some herbivorous dinosaurs were relatively diverse, while other dinosaurs appear to have been specialists, feeding on particular types of plants. The problem with inferring diets from coprolites is the difficulty in accurately associating a particular coprolite with a specific dinosaur.

54 – 2 Overkill of the North American Megafauna

Thousands of years ago, in North America's past, all of its megafauna—large mammals such as mammoths and giant bears—disappeared. One proposed explanation for this event is that when the first Americans migrated over from Asia, they hunted the megafauna to extinction. These people, known as the Clovis society after a site where their distinctive spear points were first found, would have been able to use this food source to expand their population and fill the continent rapidly. Yet many scientists argue against this "Pleistocene overkill" hypothesis. Modern humans have certainly been capable of such drastic effects on animals, but could ancient people with little more than stone spears similarly have caused the extinction of numerous species of animals? Thirty-five genera or groups of species (and many individual species) suffered extinction in North America around 11,000 B.C., soon after the appearance and expansion of Paleo-Indians throughout the Americas (27 genera disappeared completely, and another 8 became locally extinct, surviving only outside North America).

Although the climate changed at the end of the Pleistocene, warming trends had happened before. A period of massive extinction of large mammals like that seen about 11,000 years ago had not occurred during the previous 400,000 years, despite these changes. The only apparently significant difference in the Americas 11,000 years ago was the presence of human hunters of these large mammals. Was this coincidence or cause-and-effect?

We do not know. Ecologist Paul S. Martin has championed the model that associates the extinction of large mammals at the end of the Pleistocene with human predation. With researcher J. E. Mosimann, he has co-authored a work in which a computer model showed that in around 300 years, given the right conditions, a small influx of hunters into eastern Beringia 12,000 years ago could have spread across North America in a wave and wiped out game animals to feed their burgeoning population.

The researchers ran the model several ways, always beginning with a population of 100 humans in Edmonton, in Alberta, Canada, at 11,500 years ago. Assuming different initial North American big-game-animal populations (75-150 million animals)

and different population growth rates for the human settlers (0.65%-3.5%), and varying kill rates, Mosimann and Martin derived figures of between 279 and 1,157 years from initial contact to big-game extinction.

Many scholars continue to support this scenario. For example, geologist Larry Agenbroad has mapped the locations of dated Clovis sites alongside the distribution of dated sites where the remains of woolly mammoths have been found in both archaeological and purely paleontological contexts. These distributions show remarkable synchronicity (occurrence at the same time).

There are, however, many problems with this model. Significantly, though a few sites are quite impressive, there really is very little archaeological evidence to support it. Writing in 1982, Martin himself admitted the paucity of evidence; for example, at that point, the remains of only 38 individual mammoths had been found at Clovis sites. In the years since, few additional mammoths have been added to the list; there are still fewer than 20 Clovis sites where the remains of one or more mammoths have been recovered, a minuscule proportion of the millions that necessarily would have had to have been slaughtered within the overkill scenario.

Though Martin claims the lack of evidence actually supports his model—the evidence is sparse because the spread of humans and the extinction of animals occurred so quickly—this argument seems weak. And how could we ever disprove it? As archaeologist Donald Grayson points out, in other cases where extinction resulted from the quick spread of human hunters—for example, the extinction of the moa, the large flightless bird of New Zealand—archaeological evidence in the form of remains is abundant. Grayson has also shown that the evidence is not so clear that all or even most of the large herbivores in late Pleistocene America became extinct after the appearance of Clovis. Of the 35 extinct genera, only 8 can be confidently assigned an extinction date of between 12,000 and 10,000 years ago. Many of the older genera, Grayson argues, may have succumbed before 12,000 B.C., at least half a century before the Clovis showed up in the American West.

55 – 3 How Herding Can Provide Safety

In open grasslands, there is no place for a large animal to hide. Thus a watchful grazing animal will see the slight movement that betrays the presence of a predator long before it is close enough to launch an attack. It sounds as though the hunters (predators) stand no chance at all. Unfortunately for the grazers, life is not so simple, however. A grazing animal must lower its head and look at the ground to feed. Its attention may be occupied for only a few seconds before it raises its head and resumes its watch while chewing the food it took: but hunters are patient and skillful and are concentrating intensely. Those few seconds provide time enough to advance a few steps and then freeze, body flattened against the ground. It may take hours: but eventually, these repeated small advances will put the hunter within range—close enough to outrun its prey—and the long time the hunt has taken will have been worthwhile because the resulting feast will be highly nutritious.

Clearly, the grazers are at a disadvantage because while they eat they are vulnerable to attack. The hunters also have a weakness, however: and it is one that allows the grazers to survive. Hunters can attack only one prey animal at a time. This applies even to the predators that hunt as a team, such as lionesses, wolves, and hunting dogs. Their hunt involves running down or ambushing an individual. Teamwork allows them to hunt animals much bigger and stronger than themselves and to hunt more successfully, but it does not allow them to attack more than one individual at a time.

The grazers exploit this weakness by making it as difficult as they can for the predators to choose an individual as a target. They do not graze alone: scattered widely across the landscape, but together as a herd. The approaching hunter sees not a solitary animal, but a crowd of animals, all of them moving, so they are constantly crossing and recrossing each other's paths. No sooner does the hunter choose an individual than another animal has crossed in front of it and the target has disappeared into the herd. From the hunter's point of view this is highly confusing behavior—as, indeed, it is meant to be.

There is another advantage to the grazers: A herd is much more alert than a solitary animal. An animal has to relax its guard while it is taking food; but in a herd there are at any time some animals

with their heads down, biting, and others, with their heads up, watching. What is more, those with their heads up are looking in different directions so that together they are alert to any movement anywhere on the landscape around them. There is no way for a hunter to approach a herd unobserved. When a member of the herd spots trouble, it starts to move away. Other members of the herd move with it and the entire herd starts to move. If the trouble is serious and close, the herd will run. The individual raising the alarm is simply protecting itself but in doing so it is warning all of the others.

Herding is highly successful, provided members of the herd stay together in a tight bunch. The hunter moves with the herd: watching for an individual to wander away from the others. When that happens, it tries to move between that individual and the rest of the herd, preventing it from rejoining. Once it has done that the hunter has a good chance of making a kill. If the herd starts to run, a solitary hunter may abandon the chase: but a pack of wolves or hunting dogs will regard the running herd of animals as an opportunity and set off in pursuit. As the herd runs, one or two old or sick animals, or young animals that become separated from their mothers, may fall behind.

As soon as the hunter or hunters seize their prey, they lose interest in all other grazers since then they, too, must concentrate on eating, at which point the herd stops running, those who were left behind rejoin the group: and they all resume grazing.

56 – 4 Habitats and Chipmunk Species

There are eight chipmunk species in the Sierra Nevada mountain range, and most of them look pretty much alike. But eight different species of chipmunks scurrying around a picnic area will not be found. Nowhere in the Sierra do all eight species occur together. Each species tends strongly to occupy a specific habitat type, within an elevational range, and the overlap among them is minimal.

The eight chipmunk species of the Sierra Nevada represent but a few of the 15 species found in western North America, yet the whole of eastern North America makes do with but one species: the Eastern chipmunk. Why are there so many very similar chipmunks in the West? The presence of tall mountains interspersed with vast areas of arid desert and grassland makes the West ecologically far different from the East. The West affords much more opportunity for chipmunk populations to become geographically isolated from one another, a condition of species formation. Also, there are more extremes in western habitats. In the Sierra Nevada, high elevations are close to low elevations, at least in terms of mileage, but ecologically they are very different.

Most ecologists believe that ancient populations of chipmunks diverged genetically when isolated from one another by mountains and unfavorable ecological habitat. These scattered populations first evolved into races—adapted to the local ecological conditions—and then into species, reproductively isolated from one another. This period of evolution was relatively recent, as evidenced by the similar appearance of all the western chipmunk species.

Ecologists have studied the four chipmunk species that occur on the eastern slope of the Sierra and have learned just how these species interact while remaining separate, each occupying its own elevational zone. The sagebrush chipmunk is found at the lowest elevation, among the sagebrush. The yellow pine chipmunk is common in low to mid-elevations and open conifer forests, including pinon and ponderosa and Jeffrey pine forests. The lodgepole chipmunk is found at higher elevations, among the lodgepoles, firs, and high-elevation pines. The alpine chipmunk is higher still, venturing among the talus slopes, alpine meadows, and high-elevation pines and junipers. Obviously, the ranges of each species overlap. Why don't sagebrush

chipmunks move into the pine zones? Why don't alpine chipmunks move to lower elevations and share the conifer forests with lodgepole chipmunks?

The answer, in one word, is aggression. Chipmunk species actively defend their ecological zones from encroachment by neighboring species. The yellow pine chipmunk is more aggressive than the sagebrush chipmunk, possibly because it is a bit larger. It successfully bullies its smaller evolutionary cousin, excluding it from the pine forests. Experiments have shown that the sagebrush chipmunk is physiologically able to live anywhere in the Sierra Nevada, from high alpine zones to the desert. The little creature is apparently restricted to the desert not because it is specialized to live only there but because that is the only habitat where none of the other chipmunk species can live. The fact that sagebrush chipmunks tolerate very warm temperatures makes them, and only them, able to live where they do. The sagebrush chipmunk essentially occupies its habitat by default. In one study, ecologists established that yellow pine chipmunks actively exclude sagebrush chipmunks from pine forests; the ecologists simply trapped all the yellow pine chipmunks in a section of forest and moved them out. Sagebrush chipmunks immediately moved in, but yellow pine chipmunks did not enter sagebrush desert when sagebrush chipmunks were removed.

The most aggressive of the four eastern-slope species is the lodgepole chipmunk, a feisty rodent indeed. It actively prevents alpine chipmunks from moving downslope, and yellow pine chipmunks from moving upslope. There is logic behind the lodge-pole's aggressive demeanor. It lives in the cool, shaded conifer forests, and of the four species, it is the least able to tolerate heat stress. It is, in other words, the species of the strictest habitat needs: it simply must be in those shaded forests. However, if it shared its habitat with alpine and yellow pine chipmunks, either or both of these species might outcompete it, taking most of the available food. Such a competition could effectively eliminate lodgepole chipmunks from the habitat. Lodgepoles survive only by virtue of their aggression.

57 – 1 Pests and Pesticides

Around 1870, a little fruit-eating insect arrived in San Jose, California, on some nursery stock shipped from Asia. The pest, which became known as the San Jose scale, quickly spread through the United States and Canada, killing orchard trees as it went. Farmers found that the best way to control the scale was to spray their orchards with a mixture of sulfur and lime. Within a few weeks of spraying a tree, the insect vanished completely.

Around the turn of the century, however, farmers began to notice that the sulfur-lime mixture was not working all that well. A handful of scales would survive a spraying and eventually rebound to their former numbers. In Clarkston Valley in Washington State, orchard growers became convinced that manufacturers were adulterating their pesticide. They built their own factory to guarantee a pure poison, which they drenched over their trees, yet the scale kept spreading uncontrollably. An entomologist named A. L. Melander inspected the trees and found scales living happily under a thick crust of dried spray.

Melander began to suspect that adulteration was not to blame. In 1912, he compared how effective the sprays were in different parts of Washington. In Yakima and Sunnyside, he found that sulfur-lime could wipe out every last scale on a tree, while in Clarkston between 4 and 13 percent survived. On the other hand, the Clarkston scales were annihilated by a different pesticide made from fuel oil, just as the insects in other parts of Washington were. In other words, the scales of Clarkston had a peculiar resistance to sulfur-lime.

Melander wondered why. He knew that if individuals eat small amounts of certain poisons, such as arsenic, they can build up an immunity. But San Jose scales bred so quickly that no single scale experienced more than a single spray of sulfur-lime, giving them no chance to develop immunity.

A radical idea occurred to Melander. Perhaps mutations made a few scales resistant to sulfur-lime. When farmers sprayed their trees, these resistant scales survived, as did a few nonresistant ones that hadn't received a fatal dose. The surviving scales would then breed, and the resistant genes would become more common in the following generations. Depending on the proportions of the survivors, the trees might become covered by resistant or nonresistant scales. In the Clarkston Valley region, farmers had

been using sulfur-lime longer than anywhere else in the Northwest and were desperately soaking their trees with the stuff. In the process, they were driving the evolution of more resistant scales.

Melander offered his ideas in 1914, but no one paid much attention to him; they were too busy discovering even more powerful pesticides. In 1939 the Swiss chemist Paul Muller found that a compound of chlorine and hydrocarbons called DDT could kill insects more effectively than any previous pesticide had. DDT was cheap and easy to make, it could kill many species of insects, and it was stable enough to be stored for years. It could be used in small doses, and it didn't seem to pose any health risks to humans. Between 1941 and 1976, 4.5 million tons of DDT were produced. DDT was so powerful and cheap that farmers gave up old-fashioned ways of controlling pests, such as draining standing water or breeding resistant strains of crops.

DDT and similar pesticides created the delusion that pests could be not merely controlled but eradicated, so farmers began spraying pesticides on their crops as a matter of course, rather than to control outbreaks. Meanwhile, public health workers saw in DDT the hope of controlling mosquitoes, which spread diseases such as malaria.

DDT certainly saved a great many lives and crops, but even in its early days, some scientists saw signs of its doom. In 1946 Swedish scientists discovered houseflies that could no longer be killed with DDT. Houseflies in other countries became resistant as well in later years, and soon other species could withstand it. Melander's warning was becoming a reality. By 1992 more than 500 species were resistant to DDT, and the number is still climbing. As DDT began to fail, farmers at first just applied more of it; when more no longer worked, they switched to newer pesticides.

57 – 2 The Debate over Spontaneous Generation

Until the second half of the nineteenth century, many scientists and philosophers believed that some forms of life could arise spontaneously from nonliving matter; they called this hypothetical process spontaneous generation. Not much more than 100 years ago, people commonly believed that toads, snakes, and mice could be born of moist soil; that flies could emerge from manure; and that maggots, the larvae of flies, could arise from decaying corpses.

A strong opponent of spontaneous generation, the Italian physician Francesco Redi, set out in 1668 to demonstrate that maggots did not arise spontaneously from decaying meat. Redi filled three jars with decaying meat and sealed them tightly. Then he arranged three other jars similarly but left them open. Maggots appeared in the open vessels after flies entered the jars and laid their eggs, but the sealed containers showed no signs of maggots. Still, Redi's antagonists were not convinced; they claimed that fresh air was needed for spontaneous generation. So Redi set up a second experiment, in which three jars were covered with a fine net instead of being sealed. No larvae appeared in the net-covered jars, even though air was present. Maggots appeared only when flies were allowed to leave their eggs on the meat.

Redi's results were a serious blow to the long-held belief that large forms of life could arise from nonlife. However, many scientists still believed that tiny microorganisms were simple enough to be generated from nonliving materials.

The case for spontaneous generation of microorganisms seemed to be strengthened in 1745, when John Needham, an Englishman, found that even after he heated nutrient fluids (chicken broth and corn broth) before pouring them into covered flasks, the cooled solutions were soon teeming with microorganisms. Needham claimed that microbes developed spontaneously from the fluids. Twenty years later, Lazzaro Spallanzani an Italian scientist, suggested that microorganisms from the air probably had entered Needham's solutions after they were boiled. Spallanzani showed that nutrient fluids heated after being sealed in a flask did not develop microbial growth. Needham responded by claiming the "vital force" necessary for spontaneous generation had been

destroyed by the heat and was kept out of the flasks by the seals.

This intangible "vital force" was given all the more credence shortly after Spallanzani's experiment, when Laurent Lavoisier showed the importance of oxygen to life. Spallanzani's observations were criticized on the grounds that there was not enough oxygen in the sealed flasks to support microbial life.

The issue was still unresolved in 1858, when the German scientist Rudolf Virchow challenged spontaneous generation with the concept of biogenesis, the claim that living cells can arise only from preexisting living cells. Arguments about spontaneous generation continued until 1861, when the work of the French scientist Louis Pasteur ended the debate.

With a series of ingenious and persuasive experiments, Pasteur demonstrated that microorganisms are present in the air and can contaminate sterile solutions, but air itself does not create microbes. He filled several short-necked flasks with beef broth and then boiled their contents. Some were then left open and allowed to cool. In a few days, these flasks were found to be contaminated with microbes. The other flasks, sealed after boiling, were free of microorganisms. From these results, Pasteur reasoned that microbes in the air were the agents responsible for contaminating nonliving matter such as the broths in Needham's flasks.

Pasteur next placed broth in open-ended long-necked flasks and bent the necks into S-shaped curves. The contents of these flasks were then boiled and cooled. The broth in the flasks did not decay and showed no signs of life, even after months. Pasteur's unique design allowed air to pass into the flask, but the curved neck trapped any airborne microorganisms that might have contaminated the broth.

Pasteur showed that microorganisms can be present in nonliving matter—on solids, in liquids, and in the air. His work provided evidence that microorganisms cannot originate from mysterious forces present in nonliving materials. Rather, any appearance of "spontaneous" life in nonliving solutions can be attributed to microorganisms that were already present in the air or in the fluids themselves.

58 – 2 Pinyon Pines and Pinyon Jays

Organisms of different species frequently evolve adaptations that make them utterly and specifically dependent on each other for resources. For example, many trees must have mycorrhizal fungi living in their root systems, fungi that are indispensable to the tree because they facilitate mineral uptake from the soil. In turn, the fungi rely on the trees to photosynthesize: they use some of the tree's chemical food as their only energy source. Such intimate relationships are examples of coevolution through which two species become so interdependent that they can thrive only in each other's presence.

At first glance, the relationship between pinyon pine trees and the pinyon jay does not appear to be one of coevolution. Pinyon pines produce cones and seeds that attract seed predators, especially the pinyon jay. A seed, of course, contains an embryonic plant for the next generation. Why sacrifice it to an animal? Some plants have poisonous seeds, an obvious adaptation to reduce loss to animals. Pinyons, however, have an array of characteristics that combine to encourage jays to visit the pines and help themselves to the seeds. Cones are positioned upward and outward on the tree, so the seeds inside are in plain sight of the jays, essentially inviting them to partake. Pinyon seeds are unusually large, and each seed is high in energy. The seed coat is thin, meaning that birds such as pinyon jays can not only ingest the seeds but also digest them. In many plants, an indigestible seed coat permits the seed to pass unharmed through the bird's alimentary system. Pinyon seed coats differ in color between edible and nonviable seeds, signaling the jays as to which they should select.

Attracting seed predators would not seem to be a successful survival strategy for pinyon pines. However, pinyon jays are behaviorally adapted to bury any seeds in excess of their immediate survival needs. This is a useful behavior for the jays, providing they can retrieve some of the buried seeds during winter, and good for the trees, as the unretrieved seeds are ready to germinate. Still, for the jays to bury any seeds, there must be an abundance of seeds far beyond the jays' immediate needs. One tree could never produce so many seeds, but if all the pinyons in a region produced heavy seed crops at once, they would indeed "flood the market" with vastly more pinyon seeds than the local population of jays could

consume. In fact, that is exactly what the pinyon pines do.

It requires a great deal of energy to make so many seeds, so much energy that it is unlikely that a pinyon population could produce such a bumper seed crop every year. More important, however, it would be to the severe disadvantage of the trees to produce large seed crops annually, even if they could. Doing so would make the resource not only abundant but also predictable. Seed predators could, over the years, steadily build their populations, eventually increasing so much that they could, indeed, consume virtually all of the seeds. It is much more adaptive for the plants to produce seed cornucopias intermittently. Doing so has several major advantages. First, energy can be stored for some years and then devoted to cone and seed production, ensuring adequate energy to produce a large seed crop. Second, seed-predator populations will decline in years of low seed production, either through starvation, reduced reproduction, or emigration. Pinyons in most areas have a roughly six-year interval between heavy seed crops.

Pinyon jays have so completely adapted to the cycle of the pinyons that their reproduction is tied to it. Most species of birds mature sexually in response to changes in day length. In pinyon jays, however, day length is only one stimulus for reproduction. The other cue is availability of pinyon seeds. When seeds are abundant, jays can breed very early in spring, continue breeding through summer, and reenter breeding condition as early as the winter solstice. So the jay population temporarily enlarges in response to one of the intermittent large crops of pinyon seeds. But the increase in the population of jays is less than the increase in the size of the seed crop. Thus many of the seeds of a bumper crop are buried and subsequently germinate, many more than would germinate without planting by the jays.

58 – 4 Cetacean Intelligence

We often hear that whales, dolphins, and porpoises are as intelligent as humans maybe even more so. Are they really that smart? There is no question that cetaceans are among the most intelligent of animals. Dolphins, killer whales, and pilot whales in captivity quickly learn tricks. The military has trained bottlenose dolphins to find bombs and missile heads and to work as underwater spies.

This type of learning, however, is called conditioning. The animal simply learns that when it performs a particular behavior, it gets a reward, usually a fish. Many animals, including rats, birds, and even invertebrates, can be conditioned to perform tricks. We certainly don't think of these animals as our mental rivals. Unlike most other animals, however, dolphins quickly learn by observations and may spontaneously imitate human activities. One tame dolphin watched a diver cleaning an underwater viewing window, seized a feather in its beak, and began imitating the diver—complete with sound effects! Dolphins have also been seen imitating seals, turtles, and even water-skiers.

Given the seeming intelligence of cetaceans, people are always tempted to compare them with humans and other animals. Studies on discrimination and problem-solving skills in the bottlenose dolphin, for instance, have concluded that its intelligence lies "somewhere between that of a dog and a chimpanzee." Such comparisons are unfair. It is important to realize that intelligence is a very human concept and that we evaluate it in human terms. After all, not many people would consider themselves stupid because they couldn't locate and identify a fish by its echo. Why should we judge cetaceans by their ability to solve human problems?

Both humans and cetaceans have large brains with an expanded and distinctively folded surface, the cortex. The cortex is the dominant association center of the brain, where abilities such as memory and sensory perception are centered. Cetaceans have larger brains than ours, but the ratio of brain to body weight is higher in humans. Again, direct comparisons are misleading. In cetaceans it is mainly the portions of the brain associated with hearing and the processing of sound information that are expanded. The enlarged portions of our brain deal largely with vision and hand-eye coordination. Cetaceans and humans almost

certainly perceive the world in very different ways. Their world is largely one of sounds, ours one of sights.

Contrary to what is depicted in movies and on television, the notion of "talking" to dolphins is also misleading. Although they produce a rich repertoire of complex sounds, they lack vocal cords and their brains probably process sound differently from ours. Bottlenose dolphins have been trained to make sounds through the blowhole that sound something like human sounds, but this is a far cry from human speech. By the same token, humans cannot make whale sounds. We will probably never be able to carry on an unaided conversation with cetaceans.

As in chimpanzees, captive bottlenose dolphins have been taught American Sign Language. These dolphins have learned to communicate with trainers who use sign language to ask simple questions. Dolphins answer back by pushing a "yes" or "no" paddle. They have even been known to give spontaneous responses not taught by the trainers. Evidence also indicates that these dolphins can distinguish between commands that differ from each other only by their word order, a truly remarkable achievement. Nevertheless, dolphins do not seem to have a real language like ours. Unlike humans, dolphins probably cannot convey very complex messages.

Observations of cetaceans in the wild have provided some insights on their learning abilities. Several bottlenose dolphins off western Australia, for instance, have been observed carrying large cone-shaped sponges over their beaks. They supposedly use the sponges for protection against stingrays and other hazards on the bottom as they search for fish to eat. This is the first record of the use of tools among wild cetaceans.

Instead of "intelligence," some people prefer to speak of "awareness." In any case, cetaceans probably have a very different awareness and perception of their environment than do humans. Maybe one day we will come to understand cetaceans on their terms instead of ours, and perhaps we will discover a mental sophistication rivaling our own.

64 – 1 Characteristics of Pterosaurs

The extinct flying reptiles called pterosaurs were the second group of animals (after insects) to evolve flight. Most pterosaurs were about the size of modern seagulls. A few were as small as sparrows, but some of the later species were the largest flying animals that have ever lived. In 1817, Theodore Von Soemmerring published the first description of a pterosaur fossil, and thinking that it was that of an unusual bat species, he drew his reconstruction with a very batlike posture and wing. His early reconstruction of a pterosaur has haunted the public and scientific perception of pterosaurs ever since. Soemmerring's reconstruction is understandable given that he was the first to try to describe a pterosaur, that few naturalists of the time accepted the idea of major groups of extinct animals, and that both pterosaurs' and bats' wings consist of a membrane supported by enormously elongated finger bones. Soemmerring showed his pterosaur with the laterally directed legs and reoriented feet of bats and with the wing membrane stretching from the arm and finger along the sides of the body and legs all the way to the ankle. The reconstruction also included a membrane stretching between the legs, similar to that in bats. Even though other scientists developed less batlike descriptions of pterosaurs in the late 1800s, the popular literature, and even some scientific literature, continued to describe pterosaurs as batlike into the 1980s.

Bats perch by hanging upside down from tree limbs and roofs of caves. Though many are surprisingly agile climbers, bats are generally awkward when crawling on level surfaces. Did pterosaurs also hang upside down and avoid landing on the ground? Until recently, some paleontologists thought they did, but most scientists now agree that pterosaurs got around on the ground reasonably well. What is still uncertain is whether pterosaurs walked on all fours or just on their hindlegs. Pterosaurs' ancestors were bipedal (two-footed) and used their tails to balance their forward-tilted trunks and heads. Early pterosaurs also had long tails and probably could have run on their hindlegs, certainly handy for an animal with wings for forelimbs. These early pterosaurs, however, could have used their forelimbs for walking because their arm and hand bones were only slightly enlarged—most of the wing was supported by the gigantic fourth finger.

Later pterosaurs are more enigmatic: their arms and especially their hands seem too long to be used comfortably for walking, but their tails were too short to counterbalance their bodies if they walked just on their hindlegs. Birds also have short, stubby tail skeletons, but they manage to walk quite well on their hindlegs. Birds manage this by angling the thighs forward to get their feet under the body's center of gravity. They hold their thighs at this unstable angle with extensive hip and thigh muscles. Some researchers have suggested that pterosaurs' hipbones were too small to anchor extensive thigh-positioning muscles, but others have responded that pterosaurs' leg and foot bones are so strikingly birdlike that pterosaurs must surely have walked like birds. Recently, however, some pterosaur experts have concluded that a number of fossil trackways—trails of preserved footprints—were made by pterydactyloid pterosaurs, and these animals clearly walked on all four limbs. Perhaps some early pterosaurs walked on their hindlegs, but according to current evidence, most species probably walked on all fours. In any case, large pterosaurs, with eight or ten-meter wingspans and weighing as much as an adult human, do not seem likely candidates for a batlike existence confined to clambering about in trees and hanging upside down from branches.

Pterosaurs also possessed some obvious adaptations for powered flight. They had large sternums (breastbones) for attaching powerful flight muscles, well-developed shoulder bones to carry the body's weight in flight, and air-filled bones to lighten the skeleton. Some even had a furcula (a fused breastbone also found in birds): perhaps to flex like a spring and help raise the wings during the upstroke. How competent were they at flying? The original batlike reconstructions, along with their classification as reptiles, suggested to many earlier biologists that pterosaurs were only gliders. Biologists now, however, generally agree that pterosaurs were capable of powered, flapping flight. Indeed, the shoulder joint is clearly specialized for the down-and-forward, up-and-back movement of normal flapping.

66 – 3 What is Coral?

Corals belong to a group of animals called cnidarians. This group encompasses hard and soft corals, sea fans, gorgonians, hydroids, jellyfish, and sea anemones. Although the group is remarkably diverse, there are a few features shared by all. They have a free-swimming larval stage and a simple body plan: a central mouth, through which material passes in and out of the body, and a ring of tentacles. The other distinguishing feature of the cnidarians is the presence of nematocysts (stinging cells), which are used to catch prey.

Individual corals are known as coral polyps: each is a single animal. In isolation a coral polyp looks very similar to a sea anemone, but unlike sea anemones, which live separated from each other, corals form colonies. A few species remain solitary, but in most cases, new polyps bud off the initial founding polyp, and gradually colonies of thousands or even millions of polyps will grow, each connected to its neighbors by living tissue. Freed from the limitations of living alone, colonies can grow to immense sizes and live a very long time.

Corals come in all shapes and sizes, but the basic plan is the same, polyps live as a surface layer on some sort of structure, be it hard and inflexible or rubbery. The hard, or true corals, for which reefs are most famous, build limestone skeletons beneath the living tissue. Soft corals, as their name implies, do not form solid skeletons; instead they secrete limestone crystal structures called sclerites, which are embedded within a jellylike matrix beneath the polyps. When they die, little is left of most soft corals as the limestone element of their makeup is so small and easily breaks up.

By contrast, some sea whips and fans and, notably, the black corals, have such dense skeletons of limestone, protein, and minerals that they are highly durable and are collected and polished for jewelry. The skeleton of black corals is more dense the slower it grows, and it grows more slowly at depth. So it is the very deep, older colonies that are most highly prized.

Coral reefs would not exist if it were not for the ability of coral polyps to secrete limestone or calcium carbonate. Sea water surrounding a coral is very rich in dissolved calcium carbonate, but the fluid inside the polyp cannot retain a large quantity of calcium carbonate, so it is laid down as

microscopic needle-shaped crystals beneath and around the polyp. This process occurs in two stages. As the polyp expands to feed with its tentacles at night, it lifts off the skeleton, rather like a glove coming part way off a hand. At this stage, the calcium carbonate crystals form ridges. During the following day (when the coral polyp is retracted and lying on its new structure), the valleys between the ridges fill in with more calcium carbonate and the skeleton takes on a smoother appearance. Because the skeleton of hard corals is made of limestone or calcium carbonate, it is pure white.

Each coral species lays down its skeleton in a different way. This gives rise to the extraordinary range of shapes and forms that hard corals take. Some form large boulders, where the polyps live in small, isolated depressions or grooves in the skeleton. Some grow in branches, which can be small and stubby, while others are spreading, treelike structures. Still others grow delicate, leaflike plates or flat tables. The range is huge, and to complicate the matter further, the same species will grow in a different way depending on the physical characteristics of the place in which it finds itself. Although a coral's genetic blueprint is fundamental in determining what it looks like, its appearance is also affected by waves, currents, light, and competition for space on the reef. Such variations in coral form have complicated the matter of identification over the years.

65 – 1 The Pit Organs of Snakes

It is thought that the facial pits or pit organs on the head of some snakes are specialized infrared (heat) receptors. When a rattlesnake strikes, the direction of the strike seems to be guided by the infrared radiation from its prey. A rattlesnake strikes only at warm-blooded prey, and when the prey is dead and at room temperature, the snake will not strike. However, a blindfolded snake strikes correctly at a dead rat that is pulled across the cage, provided the rat is warmer than the surroundings. Blindfolded, the snake cannot be guided by vision; nor is it guided by the sense of smell, for it will strike correctly even at a moving, cloth-wrapped electric bulb. The pit organs are evidently involved in sensing the location of warm objects. All snakes that have pit organs feed preferentially on warm-blooded prey, and this further supports the view that these organs are infrared sensors. In the rattlesnake the pit organs are located, one on each side, between the nostril and the eye: they are connected to many nerves, and this in itself suggests a sensory role for the organ.

The sensitivity of the facial pit has been examined by recording the activity in the nerve leading from the organ. A variety of stimuli, such as sound, vibration, or light of moderate intensity (with the infrared part of the spectrum filtered out), has no detectable effect on the activity in the nerve. However, if objects of a temperature different from the surroundings are brought into the receptive field around the head, there is a striking change in nerve activity, regardless of the temperature of the intervening air.

How is the infrared radiation sensed? The pit is covered by a thin transparent membrane, and it has been suggested that a rise in temperature in the pit behind the membrane could cause an expansion of the gas with a consequent deformation of the membrane. This in turn could be sensed by a suitable receptor. This hypothesis is highly improbable, for a cut in the membrane that opens the pit to the outside air causes no loss in responsiveness, a result that is incompatible with the hypothesis that a pressure change is sensed.

We are left with two other possibilities to consider: Either the effect is photochemical, which means that the infrared radiation is absorbed by a specific compound, analogous to the light-sensitive pigments in the eye, or the pit organs are sensitive

to the slight temperature rise caused when infrared radiation reaches it. The infrared radiation emitted from a mammalian body has low quantum energy, which makes any photochemical effect on a pigment extremely unlikely. Pure infrared radiation can be produced by a laser, and experiments with such radiation of known wavelength provide strong evidence that the mode of reception in the facial pit organ is entirely thermal.

Can the pit organs be used for perception of the infrared source in the way our two eyes are used for stereoscopic vision? This seems likely, not only from observations of the precision with which a snake can strike, but also from studies of its brain activity. When infrared radiation falls on the facial pit organ, electric activity can be recorded from the optic tectum, the part of the brain with which the optic nerve is connected. This in itself is suggestive; although the nerves from the pit organ are completely separate from the optic nerves, the same part of the brain seems to handle visual information, which is known to be stereoscopic, and infrared information. The optic tectum has left and right parts, located on either side of the brain. Each part receives input from each eye via the optic chiasma, the point at which the optic nerves from the left and right eye meet. Many of the neurons in the tectum respond to stimulation of the pit organ on the opposite side of the head. This is reminiscent of the way information from the eyes is handled; the crossover of the optic nerve in the optic chiasma is essential for stereoscopic vision and interpretation of distance. Information from the two pit organs is apparently coordinated and interpreted in a similar way, a conclusion in agreement with recorded changes in the neural activity in the tectum when the infrared source is in a position to irradiate both pits at once. It therefore appears that the facial pits indeed provide stereoscopic perception and substantially aid in the precision of estimating the location of prey.

69 – 1 Why Snakes Have Forked Tongues

The forked tongue of snakes has intrigued people for millennia, inspiring many hypotheses. In many cultures and religions, the forked tongue symbolizes malevolence and deceit. The first person known to inquire about the functional significance of the forked tongue was Aristotle; he suggested that it would double the pleasure of sensations of taste. By the beginning of the twentieth century, the consensus was that the snake's tongue is a tactile organ, that is, the snake uses it to tap the ground much as a blind person uses a cane.

In 1920, Broman suggested what seemed to be a winning hypothesis: When the snake retracts its tongue, the tips (or tines) of the forked tongue are inserted into openings on both sides of the roof of the mouth, through these openings chemical stimuli reach special organs that help snakes detect smells – the vomeronasal organs (VNO). These organs are highly developed in snakes, lizards, and many mammals. They are a second system for detecting smells that appear to have evolved specifically to detect pheromones, the chemical signals that animals secrete as messages to other animals of their species. Broman suggested that forked tongue flicks out, picking up chemical signals, and then delivers these to the VNO. This hypothesis was widely accepted into the 1980s. Then X-ray movie studies of tongue flicks in snakes and lizards with forked tongue disproved the hypothesis; they showed that when the tongue is withdrawn into the mouth, it enters a sheath and the tips do not go into the openings to the VNO. Instead, the chemical molecules are deposited on pads at the bottom of the mouth, and closing the mouth presses the pads and molecules against the opening.

If the tongue is not forked to fit into the VNO, then what function could the forked shape serve? Schwenk proposes a solution that encompasses observations from several fields – animal behavior, ecology, sensory physiology, and neuroanatomy. He hypothesizes that the forked tongue allows the snake to sense chemical stimuli at two points simultaneously, thereby giving it the ability to detect differences in an odor trail. Obtaining two simultaneous readings enhances the ability of the snake to detect the edges of odor trails, and thus to follow pheromone trails accurately. This ability is important in seeking both prey and mates.

This spatial chemical perception is like other systems for spatial perception that are based on simultaneous stimulation of two separated sense organs – for example, auditory localization, which depends on differential stimulation at the two ears. Similarly, the use of two eyes permits stereovision.

Several kinds of evidence support the hypothesis that forked tongues evolved as chemosensory edge detectors to enhance the ability to follow odor trails: Snakes and lizards spread the tines of their tongue apart when they retrieve odor molecules, then draw the tines together when retracting the tongue. The greater the distance between sampling points, the better the animals sample differences within an odor trail. Lizards that forage widely have forked tongues, whereas lizard species without forked tongues tend not to forage widely. Forked tongues have evolved independently at least twice in different families of reptiles, indicating their value as an adaptation. In the snake nervous system, each tine of the tongue is linked to a nucleus in the other side of the brain, and the two nuclei are linked across the two hemispheres. This arrangement is similar to the anatomy of auditory centers in mammals and birds that permits the computation of differences between what one ear hears and what the other ear hears and thus mediates auditory localization.

Species in other orders have also evolved paired chemical receptors to guide individuals to mates or prey. For example, male gypsy moths have large, elaborate, odor-detecting antennae with which they track potential mates over large distances, and the ant nest beetle has spoon-shaped antennae extending from each side of the head with which it detects and follows the pheromones of the ants that are its food.

PHOTOGRAPHY / THEATER / CINAMA

01 – 2 The Origins of Theater

In seeking to describe the origins of theater, one must rely primarily on speculation, since there is little concrete evidence on which to draw. The most widely accepted theory, championed by anthropologists in the late nineteenth and early twentieth centuries, envisions theater as emerging out of myth and ritual. The process perceived by these anthropologists may be summarized briefly. During the early stages of its development, a society becomes aware of forces that appear to influence or control its food supply and well-being. Having little understanding of natural causes, it attributes both desirable and undesirable occurrences to supernatural or magical forces, and it searches for means to win the favor of these forces. Perceiving an apparent connection between certain actions performed by the group and the result it desires, the group repeats, refines and formalizes those actions into fixed ceremonies, or rituals.

Stories (myths) may then grow up around a ritual. Frequently the myths include representatives of those supernatural forces that the rites celebrate or hope to influence. Performers may wear costumes and masks to represent the mythical characters or supernatural forces in the rituals or in accompanying celebrations. As a person becomes more sophisticated, its conceptions of supernatural forces and causal relationships may change. As a result, it may abandon or modify some rites. But the myths that have grown up around the rites may continue as part of the group's oral tradition and may even come to be acted out under conditions divorced from these rites. When this occurs, the first step has been taken toward theater as an autonomous activity, and thereafter entertainment and aesthetic values may gradually replace the former mystical and socially efficacious concerns.

Although origin in ritual has long been the most popular, it is by no means the only theory about how the theater came into being. Storytelling has been proposed as one alternative. Under this theory, relating and listening to stories are seen as fundamental human pleasures. Thus, the recalling of an event (a hunt, battle, or other feat) is elaborated through the narrator's pantomime and impersonation and eventually through each role being assumed by a different person.

A closely related theory sees theater as evolving out of dances that are primarily pantomimic, rhythmical or gymnastic, or from imitations of animal noises and sounds. Admiration for the performer's skill, virtuosity, and grace are seen as motivation for elaborating the activities into fully realized theatrical performances.

In addition to exploring the possible antecedents of theater, scholars have also theorized about the motives that led people to develop theater. Why did theater develop, and why was it valued after it ceased to fulfill the function of ritual? Most answers fall back on the theories about the human mind and basic human needs. One, set forth by Aristotle in the fourth century B.C., sees humans as naturally imitative-as taking pleasure in imitating persons, things, and actions and in seeing such imitations. Another, advanced in the twentieth century, suggests that humans have a gift for fantasy, through which they seek to reshape reality into more satisfying forms than those encountered in daily life. Thus, fantasy or fiction (of which drama is one form) permits people to objectify their anxieties and fears, confront them, and fulfill their hopes in fiction if not fact. The theater, then, is one tool whereby people define and understand their world or escape from unpleasant realities.

But neither the human imitative instinct nor a penchant for fantasy by itself leads to an autonomous theater. Therefore, additional explanations are needed. One necessary condition seems to be a somewhat detached view of human problems. For example, one sign of this condition is the appearance of the comic vision, since comedy requires sufficient detachment to view some deviations from social norms as ridiculous rather than as serious threats to the welfare of the entire group. Another condition that contributes to the development of autonomous theater is the emergence of the aesthetic sense. For example, some early societies ceased to consider certain rites essential to their well-being and abandoned them, nevertheless, they retained as parts of their oral tradition the myths that had grown up around the rites and admired them for their artistic qualities rather than for their religious usefulness.

02 – 3 Early Cinema

The cinema did not emerge as a form of mass consumption until its technology evolved from the initial "peepshow" format to the point where images were projected on a screen in a darkened theater. In the peepshow format, a film was viewed through a small opening in a machine that was created for that purpose. Thomas Edison's peepshow device, the Kinetoscope, was introduced to the public in 1894. It was designed for use in Kinetoscope parlors, or arcades, which contained only a few individual machines and permitted only one customer to view a short, 50-foot film at any one time. The first Kinetoscope parlors contained five machines. For the price of 25 cents (or 5 cents per machine), customers moved from machine to machine to watch five different films (or, in the case of famous prizefights, successive rounds of a single fight).

These Kinetoscope arcades were modeled on phonograph parlors, which had proven successful for Edison several years earlier. In the phonograph parlors, customers listened to recordings through individual ear tubes, moving from one machine to the next to hear different recorded speeches or pieces of music. The Kinetoscope parlors functioned in a similar way. Edison was more interested in the sale of Kinetoscopes (for roughly \$1,000 apiece) to these parlors than in the films that would be run in them (which cost approximately \$10 to \$15 each). He refused to develop projection technology, reasoning that if he made and sold projectors, then exhibitors would purchase only one machine—a projector—from him instead of several.

Exhibitors, however, wanted to maximize their profits, which they could do more readily by projecting a handful of films to hundreds of customers at a time (rather than one at a time) and by charging 25 to 50 cents admission. About a year after the opening of the first Kinetoscope parlor in 1894, showmen such as Louis and Auguste Lumiere, Thomas Armat and Charles Francis Jenkins, and Orville and Woodville Latham (with the assistance of Edison's former assistant, William Dickson) perfected projection devices. These early projection devices were used in vaudeville theaters, legitimate theaters, local town halls, makeshift storefront theaters, fairgrounds, and amusement parks to show films to a mass audience.

With the advent of projection in 1895-1896, motion pictures became the ultimate form of mass consumption. Previously, large audiences had viewed spectacles at the theater, where vaudeville, popular dramas, musical and minstrel shows, classical plays, lectures, and slide-and-lantern shows had been presented to several hundred spectators at a time. But the movies differed significantly from these other forms of entertainment, which depended on either live performance or (in the case of the slide-and-lantern shows) the active involvement of a master of ceremonies who assembled the final program.

Although early exhibitors regularly accompanied movies with live acts, the substance of the movies themselves is mass-produced, prerecorded material that can easily be reproduced by theaters with little or no active participation by the exhibitor. Even though early exhibitors shaped their film programs by mixing films and other entertainments together in whichever way they thought would be most attractive to audiences or by accompanying them with lectures, their creative control remained limited. What audiences came to see was the technological marvel of the movies: the lifelike reproduction of the commonplace motion of trains, of waves striking the shore, and of people walking in the street, and the magic made possible by trick photography and the manipulation of the camera.

With the advent of projection, the viewer's relationship with the image was no longer private, as it had been with earlier peepshow devices such as the Kinetoscope and the Mutoscope, which was a similar machine that reproduced motion by means of successive images on individual photographic cards instead of on strips of celluloid. It suddenly became public—an experience that the viewer shared with dozens, scores, and even hundreds of others. At the same time, the image that the spectator looked at expanded from the minuscule peepshow dimensions of 1 or 2 inches (in height) to the life-size proportions of 6 or 9 feet.

12 – 2 Transition to Sound in Film

The shift from silent to sound film at the end of the 1920s marks, so far, the most important transformation in motion picture history. Despite all the highly visible technological developments in theatrical and home delivery of the moving image that have occurred over the decades since then, no single innovation has come close to being regarded as a similar kind of watershed. In nearly every language, however, the words are phrased, the most basic division in cinema history lies between films that are mute and films that speak.

Yet this most fundamental standard of historical periodization conceals a host of paradoxes. Nearly every movie theater, however modest, had a piano or organ to provide musical accompaniment to silent pictures. In many instances, spectators in the era before recorded sound experienced elaborate aural presentations alongside movies' visual images, from the Japanese benshi (narrators) crafting multivoiced dialogue narratives to original musical compositions performed by symphony-size orchestras in Europe and the United States. In Berlin, for the premiere performance outside the Soviet Union of *The Battleship Potemkin*, film director Sergei Eisenstein worked with Austrian composer Edmund Meisel (1874-1930) on a musical score matching sound to image; the Berlin screenings with live music helped to bring the film its wide international fame.

Beyond that, the triumph of recorded sound has overshadowed the rich diversity of technological and aesthetic experiments with the visual image that were going forward simultaneously in the 1920s. New color processes, larger or differently shaped screen sizes, multiple-screen projections, even television, were among the developments invented or tried out during the period, sometimes with startling success. The high costs of converting to sound and the early limitations of sound technology were among the factors that suppressed innovations or retarded advancement in these other areas. The introduction of new screen formats was put off for a quarter century, and color, though utilized over the next two decades for special productions, also did not become a norm until the 1950s.

Though it may be difficult to imagine from a later perspective, a strain of critical opinion in the 1920s predicted that sound film would be a technical novelty that would soon fade from sight, just as had

many previous attempts, dating well back before the First World War, to link images with recorded sound. These critics were making a common assumption—that the technological inadequacies of earlier efforts (poor synchronization, weak sound amplification, fragile sound recordings) would invariably occur again. To be sure, their evaluation of the technical flaws in 1920s sound experiments was not so far off the mark, yet they neglected to take into account important new forces in the motion picture field that, in a sense, would not take no for an answer.

These forces were the rapidly expanding electronics and telecommunications companies that were developing and linking telephone and wireless technologies in the 1920s. In the United States, they included such firms as American Telephone and Telegraph, General Electric, and Westinghouse. They were interested in all forms of sound technology and all potential avenues for commercial exploitation. Their competition and collaboration were creating the broadcasting industry in the United States, beginning with the introduction of commercial radio programming in the early 1920s. With financial assets considerably greater than those in the motion picture industry, and perhaps a wider vision of the relationships among entertainment and communications media, they revitalized research into recording sound for motion pictures.

In 1929 the United States motion picture industry released more than 300 sound films—a rough figure, since a number were silent films with music tracks, or films prepared in dual versions, to take account of the many cinemas not yet wired for sound. At the production level, in the United States the conversion was virtually complete by 1930. In Europe it took a little longer, mainly because there were more small producers for whom the costs of sound were prohibitive, and in other parts of the world problems with rights or access to equipment delayed the shift to sound production for a few more years (though cinemas in major cities may have been wired in order to play foreign sound films). The triumph of sound cinema was swift, complete, and enormously popular.

22 – 2 The Birth of Photography

Perceptions of the visible world were greatly altered by the invention of photography in the middle of the nineteenth century. In particular, and quite logically, the art of painting was forever changed, though not always in the ways one might have expected. The realistic and naturalistic painters of the mid- and late-nineteenth century were all intently aware of photography — as a thing to use, to learn from, and react to.

Unlike most major inventions, photography had been long and impatiently awaited. The images produced by the camera obscura, a boxlike device that used a pinhole or lens to throw an image onto a ground-glass screen or a piece of white paper, were already familiar — the device had been much employed by topographical artists like the Italian painter Canaletto in his detailed views of the city of Venice. What was lacking was a way of giving such images permanent form. This was finally achieved by Louis Daguerre (1787-1851), who perfected a way of fixing them on a silvered copper plate. His discovery, the “daguerreotype”, was announced in 1839.

A second and very different process was patented by the British inventor William Henry Talbot (1800-1877) in 1841. Talbot’s “calotype” was the first negative-to-positive process and the direct ancestor of the modern photograph. The calotype was revolutionary in its use of chemically treated paper in which areas hit by light became dark in tone, producing a negative image. This “negative”, as Talbot called it, could then be used to print multiple positive images on another piece of treated paper.

The two processes produced very different results. The daguerreotype was a unique image that reproduced what was in front of the camera lens in minute, unselective detail and could not be duplicated. The calotype could be made in series, and was thus the equivalent of an etching or an engraving. Its general effect was soft edged and tonal.

One of the things that most impressed the original audience for photography was the idea of authenticity. Nature now seemed able to speak for itself, with a minimum of interference. The title Talbot chose for his book, *The Pencil of Nature* (the first part of which was published in 1844), reflected this feeling. Artists were fascinated by photography because it offered a way of examining the world in

much greater detail. They were also afraid of it, because it seemed likely to make their own efforts unnecessary.

Photography did indeed make certain kinds of painting obsolete — the daguerreotype virtually did away with the portrait miniature. It also made the whole business of making and owning images democratic. Portraiture, once a luxury for the privileged few, was suddenly well within the reach of many more people.

In the long term, photography’s impact on the visual arts was far from simple. Because the medium was so prolific, in the sense that it was possible to produce a multitude of images very cheaply, it was soon treated as the poor relation of fine art, rather than its destined successor. Even those artists who were most dependent on photography became reluctant to admit that they made use of it, in case this compromised their professional standing.

The rapid technical development of photography — the introduction of lighter and simpler equipment, and of new emulsions that coated photographic plates, film, and paper and enabled images to be made at much faster speeds — had some unanticipated consequences. Scientific experiments made by photographers such as Eadweard Muybridge (1830-1904) and Etienne-Jules Marey (1830-1904) demonstrated that the movements of both humans and animals differed widely from the way they had been traditionally represented in art. Artists, often reluctantly, were forced to accept the evidence provided by the camera. The new candid photography — unposed pictures that were made when the subjects were unaware that their pictures were being taken — confirmed these scientific results, and at the same time, thanks to the radical cropping (trimming) of images that the camera often imposed, suggested new compositional formats. The accidental effects obtained by candid photographers were soon being copied by artists such as the French painter Degas.

44 – 2 The Use of the Camera Obscura

The precursor of the modern camera, the camera obscura is a darkened enclosure into which light is admitted through a lens in a small hole. The image of the illuminated area outside the enclosure is thrown upside down as if by magic onto a surface in the darkened enclosure. This technique was known as long ago as the fifth century B.C. in China. Aristotle also experimented with it in the fourth century B.C., and Leonardo da Vinci described it in his notebooks in 1490. In 1558 Giovanni Battista Della Porta wrote in his twenty-volume work *Magia naturalis* (meaning "natural magic") instructions for adding a convex lens to improve the quality of the image thrown against a canvas or panel in the darkened area where its outlines could be traced. Later, portable camera obscuras were developed, with interior mirrors and drawing tables on which the artist could trace the image. For the artist, this technique allows forms and linear perspective to be drawn precisely as they would be seen from a single viewpoint. Mirrors were also used to reverse projected images to their original positions.

Did some of the masters of painting, then, trace their images using a camera obscura? Some art historians are now looking for clues of artists' use of such devices. One of the artists whose paintings are being analyzed from this point of view is the great Dutch master, Jan Vermeer, who lived from 1632 to 1675 during the flowering of art and science in the Netherlands, including the science of optics. Vermeer produced only about 30 known paintings, including his famous *The Art of Painting*. The room shown in it closely resembles the room in other Vermeer paintings, with lighting coming from a window on the left, the same roof beams, and similar floor tiles, suggesting that the room was fitted with a camera obscura on the side in the foreground. The map hung on the opposite wall was a real map in Vermeer's possession, reproduced in such faithful detail that some kind of tracery is suspected. When one of Vermeer's paintings was X-rayed, it did not have any preliminary sketches on the canvas beneath the paint, but rather the complete image drawn in black and white without any trial sketches. Vermeer did not have any students, did not keep any records, and did not encourage anyone to visit his studio, facts that can be interpreted as protecting his secret use of a camera obscura.

In recent times the British artist David Hockney has published his investigations into the secret use of the camera obscura, claiming that for up to 400 years, many of Western art's great masters probably used the device to produce almost photographically realistic details in their paintings. He includes in this group Caravaggio, Hans Holbein, Leonardo da Vinci, Diego Velazquez, Jean-Auguste-Dominique Ingres, Agnolo Bronzino, and Jan van Eyck. From an artist's point of view, Hockney observed that a camera obscura compresses the complicated forms of a three-dimensional scene into two-dimensional shapes that can easily be traced and also increases the contrast between light and dark, leading to the chiaroscuro effect seen in many of these paintings. In Jan van Eyck's *The Marriage of Giovanni Arnolfini and Giovanna Cenami*, the complicated foreshortening in the chandelier and the intricate detail in the bride's garments are among the clues that Hockney thinks point to the use of the camera obscura.

So what are we to conclude? If these artists did use a camera obscura, does that diminish their stature? Hockney argues that the camera obscura does not replace artistic skill in drawing and painting. In experimenting with it, he found that it is actually quite difficult to use for drawing, and he speculates that the artists probably combined their observations from life with tracing of shapes.

57 – 3 The Problem of Narrative Clarity in Silent Films

Beginning in 1904, American commercial filmmaking became increasingly oriented toward storytelling. Moreover, with the new emphasis on one-reel films, narratives became longer and necessitated a series of camera shots. Filmmakers faced the challenge of making story films that would be comprehensible to audiences. How could techniques of editing, camerawork, acting, and lighting be combined so as to clarify what was happening in a film? How could the spectator grasp where and when the action was occurring?

Over the span of several years, filmmakers solved such problems. Sometimes they influenced each other, while at other times two filmmakers might happen on the same technique independently. Some devices were tried and abandoned. By 1917, filmmakers had worked out a system of formal principles that were standard in American filmmaking. That system has come to be called the classical Hollywood cinema. Despite that name, many of the basic principles of the system were being worked out before filmmaking was centered in Hollywood and, indeed, many of those principles were first tried in other countries. In the years before the First World War, film style was still largely international, since films circulated widely outside their country of origin.

The basic problem that confronted filmmakers early in the silent-movie era was that audiences could not understand the causal, spatial, and temporal relations in many films. If the editing abruptly changed locales, the spectator might not grasp where the new action was occurring. An actor's elaborate pantomime might fail to convey the meaning of a crucial action. A review of a 1906 film lays out the problem: "Regardless of the fact that there are a number of good motion pictures brought out, it is true that there are some which, although photographically good, are poor because the manufacturer, being familiar with the picture and the plot, does not take into consideration that the film was not made for him but for the audience. A movie recently seen was very good photographically, but the story could not be understood by the audience." In a few theaters, a lecturer might explain the plot as the film unrolled, but producers could not rely on such aids.

Filmmakers came to assume that a film should guide the spectator's attention, making every

aspect of the story on the screen as clear as possible. In particular, films increasingly set up a chain of narrative causes and effects. One event would plainly lead to an effect, which would in turn cause another effect, and so on. Moreover, an event was typically caused by a character's beliefs or desires. Character psychology had not been particularly important in early films. Comical chases or brief melodramas depended more on physical action or familiar situations than on character traits. Increasingly after 1907, however, character psychology motivated actions. By following a series of characters' goals and resulting conflicts, the spectator could comprehend the action.

Every aspect of the silent-film style came to be used to enhance narrative clarity. Staging or framing action in depth could show the spatial relationships among elements. Intertitles could add narrative information beyond what the images conveyed. Closer views of the actors could suggest their emotions more precisely. Color, set design, and lighting could imply time of day, the milieu of the action, and so on.

Some of the most important innovations of this period involved the ways in which camera shots were put together, or edited, to create a story. In one sense, editing was a boon to the filmmaker, permitting instant movement from one space to another or cuts to closer views to reveal details. But if the spectator could not keep track of the temporal or spatial relations between one shot and the next, editing might also lead to confusion. In some cases, intertitles could help. Editing also came to emphasize continuity among shots. Certain visual cues indicated that time was flowing uninterrupted across cuts. Between scenes, other cues might suggest how much time had been skipped over. When a cut moved from one space to another, the director found ways to orient the viewer.

66 – 1 The Actor and the Audience

Actors, even when they are well rehearsed, can never fully anticipate how well they will perform before an actual audience. The actor who has been brilliant in rehearsal can crumble before an audience and completely lose the "edge" of his or her performance in the face of stage fright and apprehension. The presence of an audience can affect performance in other ways as well. Or-and this is more likely-an actor who seemed fairly unexciting at rehearsal can suddenly take fire and dazzle the audience with unexpected energy, subtlety, and depth. One celebrated example of this phenomenon was achieved by Lee J Cobb in the original production of Arthur Miller's *Death of a Salesman*, in which Cobb had the title role. The presence of an audience can affect performance in other ways as well. Roles rehearsed in all solemnity can suddenly turn comical in performance; conversely, roles developed for comic potential in rehearsal may be received soberly by an audience and lose their comedic aspect entirely.

Sudden and dramatic change, however, is not the norm as the performance phase replaces rehearsal: most actors cross over from final dress rehearsal to opening night with only the slightest shift: indeed, this is generally thought to be the goal of a disciplined and professional rehearsal schedule. Holding back until opening night, the once-popular acting practice of restraining emotional display until opening night, is universally disavowed today, and opening night recklessness is viewed as a sure sign of the amateur, who relies primarily on guts and adrenaline to get through the performance. Deliberate revision of a role in performance, in response to the first waves of laughter or applause, is similarly frowned upon in all but the most inartistic of theaters today.

Nevertheless, a fundamental shift does occur in the actor's awareness between rehearsal and performance, and this cannot and should not be denied: indeed, it is essential to the creation of theater art. This shift is set up by an elementary feedback: the actor is inevitably aware, with at least a portion of his or her mind, of the audience's reaction to his or her own performance and that of the other players; there is always, in any acting performance, a subtle adjustment to the audience that sees it. The outward manifestations of this adjustment are usually all but imperceptible: the split-second hold for a laugh to die down, the slight

special projection of a certain line to make sure that it reaches the back row, the quick turn of a head to make a characterization or plot transition extra clear.

In addition, the best actors consistently radiate a quality known to the theater world as presence. It is a quality difficult to describe, but it has the effect of making both the character whom the actor portrays and the self of the actor who represents that character especially vibrant and in the present for the audience: it is the quality of an actor who takes the stage and acknowledges, in some inexplicable yet indelible manner, that he or she is there to be seen. Performance is not a one-way statement given from the stage to the house; it is a two-way participatory communication between the actors and the audience members in which the former employ text and movement and the latter employ applause, laughter, silence, and attention.

Even when the audience is silent and invisible-and, owing to the brightness of the stage lights, the audience is frequently invisible to the actor-the performer feels its presence. There is nothing extrasensory about this: the absence of sound is itself a signal, for when several hundred people sit without shuffling, coughing, or muttering, their silence betokens a level of attention for which the actor customarily strives. Laughter, gasps, sighs, and applause similarly feed back into the actor's consciousness-and unconsciousness-and spur (or sometimes, alas, distract) the actor's efforts. The veteran actor can determine quickly how to ride the crest of audience laughter and how to hold the line just long enough that it will pierce the lingering chuckles but not be overridden by them, he or she also knows how to vary the pace and/or redouble his or her energy when sensing restlessness or boredom on the other side of the curtain line. Performance technique, or the art of reading an audience, is more instinctual than learned. The timing it requires is of such complexity that no actor could master it rationally; he or she can develop it only out of experience.

ART / ARCHITECTURE / CULTURE / ART HISTORY / ARCHEOLOGY / SCULPTURE / VISUAL ARTS

03 – 1 Architecture

Architecture is the art and science of designing structures that organize and enclose space for practical and symbolic purposes. Because architecture grows out of human needs and aspirations, it clearly communicates cultural values. Of all the visual arts, architecture affects our lives most directly for it determines the character of the human environment in major ways.

Architecture is a three-dimensional form. It utilizes space, mass, texture, line, light, and color. To be architecture, a building must achieve a working harmony with a variety of elements. Humans instinctively seek structures that will shelter and enhance their way of life. It is the work of architects to create buildings that are not simply constructions but also offer inspiration and delight. Buildings contribute to human life when they provide shelter, enrich space, complement their site, suit the climate, and are economically feasible. The client who pays for the building and defines its function is an important member of the architectural team. The mediocre design of many contemporary buildings can be traced to both clients and architects.

In order for the structure to achieve the size and strength necessary to meet its purpose, architecture employs methods of support that, because they are based on physical laws, have changed little since people first discovered them—even while building materials have changed dramatically. The world's architectural structures have also been devised in relation to the objective limitations of materials. Structures can be analyzed in terms of how they deal with downward forces created by gravity. They are designed to withstand the forces of compression (pushing together), tension (pulling apart), bending, or a combination of these in different parts of the structure.

Even development in architecture has been the result of major technological changes. Materials and methods of construction are integral parts of the design of architecture structures. In earlier times it was necessary to design structural systems suitable for the materials that were available, such as wood, stone, brick. Today technology has progressed to the point where it is possible to invent new building materials to suit the type of

structure desired. Enormous changes in materials and techniques of construction within the last few generations have made it possible to enclose space with much greater ease and speed and with a minimum of material. Progress in this area can be measured by the difference in weight between buildings built now and those of comparable size built one hundred years ago.

Modern architectural forms generally have three separate components comparable to elements of the human body: a supporting skeleton or frame, an outer skin enclosing the interior spaces, and equipment, similar to the body's vital organs and systems. The equipment includes plumbing, electrical wiring, hot water, and air-conditioning. Of course in early architecture—such as igloos and adobe structures—there was no such equipment, and the skeleton and skin were often one.

Much of the world's great architecture has been constructed of stone because of its beauty, permanence, and availability. In the past, whole cities grew from the arduous task of cutting and piling stone upon. Some of the world's finest stone architecture can be seen in the ruins of the ancient Inca city of Machu Picchu high in the eastern Andes Mountains of Peru. The doorways and windows are made possible by placing over the open spaces thick stone beams that support the weight from above. A structural invention had to be made before the physical limitations of stone could be overcome and new architectural forms could be created. That invention was the arch, a curved structure originally made of separate stone or brick segments. The arch was used by the early cultures of the Mediterranean area chiefly for underground drains, but it was the Romans who first developed and used the arch extensively in aboveground structures. Roman builders perfected the semicircular arch made of separate blocks of stone. As a method of spanning space, the arch can support greater weight than a horizontal beam. It works in compression to divert the weight above it out to the sides, where the weight is borne by the vertical elements on either side of the arch. The arch is among the many important structural breakthroughs that have characterized architecture throughout the centuries.

04 – 2 Cave Art in Europe

The earliest discovered traces of art are beads and carvings, and then paintings, from sites dating back to the Upper Paleolithic period. We might expect that early artistic efforts would be crude, but the cave paintings of Spain and southern France show a marked degree of skill. So do the naturalistic paintings on slabs of stone excavated in southern Africa. Some of those slabs appear to have been painted as much as 28,000 years ago, which suggests that painting in Africa is as old as painting in Europe. But painting may be even older than that. The early Australians may have painted on the walls of rock shelters and cliff faces at least 30,000 years ago, and maybe as much as 60,000 years ago.

The researchers Peter Ucko and Andree Rosenfeld identified three principal locations of paintings in the caves of western Europe : (1) in obviously inhabited rock shelters and cave entrances; (2) in galleries immediately off the inhabited areas of caves; and (3) in the inner reaches of caves, whose difficulty of access has been interpreted by some as a sign that magical-religious activities were performed there.

The subjects of the paintings are mostly animals. The paintings rest on bare walls, with no backdrops or environmental trappings. Perhaps, like many contemporary peoples, Upper Paleolithic men and women believed that the drawing of a human image could cause death or injury, and if that were indeed their belief, it might explain why human figures are rarely depicted in cave art. Another explanation for the focus on animals might be that these people sought to improve their luck at hunting. This theory is suggested by evidence of chips in the painted figures, perhaps made by spears thrown at the drawings. But if improving their hunting luck was the chief motivation for the paintings, it is difficult to explain why only a few show signs of having been speared. Perhaps the paintings were inspired by the need to increase the supply of animals. Cave art seems to have reached a peak toward the end of the Upper Paleolithic period, when the herds of game were decreasing.

The particular symbolic significance of the cave paintings in southwestern France is more explicitly revealed, perhaps, by the results of a study conducted by researchers Patricia Rice and Ann Paterson. The data they present suggest that the animals portrayed in the cave paintings were

mostly the ones that the painters preferred for meat and for materials such as hides. For example, wild cattle (bovines) and horses are portrayed more often than we would expect by chance, probably because they were larger and heavier (meatier) than other animals in the environment. In addition, the paintings mostly portray animals that the painters may have feared the most because of their size, speed, natural weapons such as tusks and horns, and the unpredictability of their behavior. That is, mammoths, bovines, and horses are portrayed more often than deer and reindeer. Thus, the paintings are consistent with the idea that the art is related to the importance of hunting in the economy of Upper Paleolithic people. Consistent with this idea, according to the investigators, is the fact that the art of the cultural period that followed the Upper Paleolithic also seems to reflect how people got their food. But in that period, when getting food no longer depended on hunting large game animals (because they were becoming extinct), the art ceased to focus on portrayals of animals.

Upper Paleolithic art was not confined to cave paintings. Many shafts of spears and similar objects were decorated with figures of animals. The anthropologist Alexander Marshack has an interesting interpretation of some of the engravings made during the Upper Paleolithic. He believes that as far back as 30,000 B.C., hunters may have used a system of notation, engraved on bone and stone, to mark phases of the Moon. If this is true, it would mean that Upper Paleolithic people were capable of complex thought and were consciously aware of their environment. In addition to other artworks, figurines representing the human female in exaggerated form have also been found at Upper Paleolithic sites. It has been suggested that these figurines were an ideal type or an expression of a desire for fertility.

10 – 1 Chinese Pottery

China has one of the world's oldest continuous civilizations—despite invasions and occasional foreign rule. A country as vast as China with so long-lasting a civilization has a complex social and visual history, within which pottery and porcelain play a major role.

The function and status of ceramics in China varied from dynasty to dynasty, so they may be utilitarian, burial, trade-collectors', or even ritual objects, according to their quality and the era in which they were made. The ceramics fall into three broad types—earthenware, stoneware, and porcelain—for vessels, architectural items such as roof tiles, and modeled objects and figures. In addition, there was an important group of sculptures made for religious use, the majority of which were produced in earthenware.

The earliest ceramics were fired to earthenware temperatures, but as early as the fifteenth century B.C., high-temperature stonewares were being made with glazed surfaces. During the Six Dynasties period (A.D. 265-589), kilns in north China were producing high-fired ceramics of good quality. Whitewares produced in Hebei and Henan provinces from the seventh to the tenth centuries evolved into the highly prized porcelains of the Song dynasty (A.D. 960-1279), long regarded as one of the high points in the history of China's ceramic industry. The tradition of religious sculpture extends over most historical periods but is less clearly delineated than that of stonewares or porcelains, for it embraces the old custom of earthenware burial ceramics with later religious images and architectural ornament. Ceramic products also include lead-glazed tomb models of the Han dynasty, three-color lead-glazed vessels and figures of the Tang dynasty, and Ming three-color temple ornaments, in which the motifs were outlined in a raised trail of slip—as well as the many burial ceramics produced in imitation of vessels made in materials of higher intrinsic value.

Trade between the West and the settled and prosperous Chinese dynasties introduced new forms and different technologies. One of the most far-reaching examples is the impact of the fine ninth-century A.D. Chinese porcelain wares imported into the Arab world. So admired were these pieces that they encouraged the development of earthenware made in imitation of porcelain and instigated research into the method

of their manufacture. From the Middle East the Chinese acquired a blue pigment—a purified form of cobalt oxide unobtainable at that time in China—that contained only a low level of manganese. Cobalt ores found in China have a high manganese content, which produces a more muted blue-gray color. In the seventeenth century, the trading activities of the Dutch East India Company resulted in vast quantities of decorated Chinese porcelain being brought to Europe, which stimulated and influenced the work of a wide variety of wares, notably Delft. The Chinese themselves adapted many specific vessel forms from the West, such as bottles with long spouts, and designed a range of decorative patterns especially for the European market.

Just as painted designs on Greek pots may seem today to be purely decorative, whereas in fact they were carefully and precisely worked out so that at the time, their meaning was clear, so it is with Chinese pots. To twentieth-century eyes, Chinese pottery may appear merely decorative, yet to the Chinese the form of each object and its adornment had meaning and significance. The dragon represented the emperor, and the phoenix, the empress; the pomegranate indicated fertility, and a pair of fish, happiness; mandarin ducks stood for wedded bliss; the pine tree, peach, and crane are emblems of long life; and fish leaping from waves indicated success in the civil service examinations. Only when European decorative themes were introduced did these meanings become obscured or even lost.

From early times pots were used in both religious and secular contexts. The imperial court commissioned work and in the Yuan dynasty (A.D. 1279-1368) an imperial ceramic factory was established at Jingdezhen. Pots played an important part in some religious ceremonies. Long and often lyrical descriptions of the different types of ware exist that assist in classifying pots, although these sometimes confuse an already large and complicated picture.

11 – 1 Ancient Egyptian Sculpture

In order to understand ancient Egyptian art, it is vital to know as much as possible of the elite Egyptians' view of the world and the functions and contexts of the art produced for them. Without this knowledge we can appreciate only the formal content of Egyptian art, and we will fail to understand why it was produced or the concepts that shaped it and caused it to adopt its distinctive forms. In fact, a lack of understanding concerning the purposes of Egyptian art has often led it to be compared unfavorably with the art of other cultures: Why did the Egyptians not develop sculpture in which the body turned and twisted through space like classical Greek statuary? Why do the artists seem to get left and right confused? And why did they not discover the geometric perspective as European artists did in the Renaissance? The answer to such questions has nothing to do with a lack of skill or imagination on the part of Egyptian artists and everything to do with the purposes for which they were producing their art.

The majority of three-dimensional representations, whether standing, seated, or kneeling, exhibit what is called frontality: they face straight ahead, neither twisting nor turning. When such statues are viewed in isolation, out of their original context and without knowledge of their function, it is easy to criticize them for their rigid attitudes that remained unchanged for three thousand years. Frontality is, however, directly related to the functions of Egyptian statuary and the contexts in which the statues were set up. Statues were created not for their decorative effect but to play a primary role in the cults of the gods, the king, and the dead. They were designed to be put in places where these beings could manifest themselves in order to be the recipients of ritual actions. Thus it made sense to show the statue looking ahead at what was happening in front of it, so that the living performer of the ritual could interact with the divine or deceased recipient. Very often such statues were enclosed in rectangular shrines or wall niches whose only opening was at the front, making it natural for the statue to display frontality. Other statues were designed to be placed within an architectural setting, for instance, in front of the monumental entrance gateways to temples known as pylons, or in pillared courts, where they would be placed against or between pillars: their frontality worked perfectly within the architectural context.

Statues were normally made of stone, wood, or metal. Stone statues were worked from single rectangular blocks of material and retained the compactness of the original shape. The stone between the arms and the body and between the legs in standing figures or the legs and the seat in seated ones was not normally cut away. From a practical aspect this protected the figures against breakage and psychologically gives the images a sense of strength and power, usually enhanced by a supporting back pillar. By contrast, wooden statues were carved from several pieces of wood that were pegged together to form the finished work, and metal statues were either made by wrapping sheet metal around a wooden core or cast by the lost wax process. The arms could be held away from the body and carry separate items in their hands; there is no back pillar. The effect is altogether lighter and freer than that achieved in stone, but because both perform the same function, formal wooden and metal statues still display frontality.

Apart from statues representing deities, kings, and named members of the elite that can be called formal, there is another group of three-dimensional representations that depicts generic figures, frequently servants, from the nonelite population. The function of these is quite different. Many are made to be put in the tombs of the elite in order to serve the tomb owners in the afterlife. Unlike formal statues that are limited to static poses of standing, sitting, and kneeling, these figures depict a wide range of actions, such as grinding grain, baking bread, producing pots, and making music, and they are shown in appropriate poses, bending and squatting as they carry out their tasks.

23 – 3 Rock Art of the Australia Aborigines

Ever since European first explored Australia, people have been trying to understand the ancient rock drawings and carvings created by the Aborigines, the original inhabitants of the continent. Early in the nineteenth century, encounters with Aboriginal rock art tended to be infrequent and open to speculative interpretation, but since the late nineteenth century, awareness of the extent and variety of Australian rock art has been growing. In the latter decades of the twentieth century there were intensified efforts to understand and record the abundance of Australian rock art.

The systematic study of this art is a relatively new discipline in Australia. Over the past four decades new discoveries have steadily added to the body of knowledge. The most significant data have come from a concentration on three major questions. First, what is the age of Australian rock art? Second, what is its stylistic organization and is it possible to discern a sequence or a pattern of development between styles? Third, is it possible to interpret accurately the subject matter of ancient rock art, bringing to bear all available archaeological techniques and the knowledge of present-day Aboriginal informants?

The age of Australia's rock art is constantly being revised, and earlier datings have been proposed as the result of new discoveries. Currently, reliable scientific evidence dates the earliest creation of art on rock surfaces in Australia to somewhere between 30,000 and 50,000 years ago. This in itself is an almost incomprehensible span of generations, and one that makes Australia's rock art the oldest continuous art tradition in the world.

Although the remarkable antiquity of Australia's rock art is now established, the sequences and meanings of its images have been widely debated. Since the mid-1970s, a reasonably stable picture has been formed of the organization of Australian rock art. In order to create a sense of structure to this picture, researchers have relied on a distinction that still underlies the forms of much indigenous visual culture — a distinction between geometric and figurative elements. Simple geometric repeated patterns — circles, concentric circles, and lines — constitute the iconography (characteristic images) of the earliest rock-art sites found across Australia. The frequency with which certain simple motifs appear in these oldest sites has led rock-art researchers to adopt a descriptive term — the

Panaramitee style — a label which takes its name from the extensive rock pavements at Panaramitee North in desert South Australia, which are covered with motifs pecked into the surface. Certain features of these engravings lead to the conclusion that they are of great age — geological changes had clearly happened after the designs had been made and local Aboriginal informants, when first questioned about them, seemed to know nothing of their origins. Furthermore, the designs were covered with “desert varnish”, a glaze that develops on rock surfaces over thousands of years of exposure to the elements. The simple motifs found at Panaramitee are common to many rock-art sites across Australia. Indeed, sites with engravings of geometric shapes are also to be found on the island of Tasmania, which was separated from the mainland of the continent some 10,000 years ago.

In the 1970s when the study of Australian archaeology was in an exciting phase of development, with the great antiquity of rock art becoming clear, Lesley Maynard, the archaeologist who coined the phrase “Panaramitee style”, suggested that a sequence could be determined for Australian rock art, in which a geometric style gave way to a simple figurative style (outlines of figures and animals), followed by a range of complex figurative styles that, unlike the pan-Australian geometric tradition, tended to much greater regional diversity. While accepting that this sequence fits the archaeological profile of those sites, which were occupied continuously over many thousands of years a number of writers have warned that the underlying assumption of such a sequence — a development from the simple and the geometric to the complex and naturalistic — obscures the cultural continuities in Aboriginal Australia, in which geometric symbolism remains fundamentally important. In this context the simplicity of a geometric motif may be more apparent than real. Motifs of seeming simplicity can encode complex meanings in Aboriginal Australia. And has not twentieth-century art shown that naturalism does not necessarily follow abstraction in some kind of predetermined sequence?

24 – 3 Moving into Pueblos

In the Mesa Verde area of the ancient North American Southwest, living patterns changed in the thirteenth century, with large numbers of people moving into large communal dwellings called pueblos, often constructed at the edges of canyons, especially on the sides of cliffs. Abandoning small extended-family households to move into these large pueblos with dozens if not hundreds of other people was probably traumatic. Few of the cultural traditions and rules that today allow us to deal with dense populations existed for these people accustomed to household autonomy and the ability to move around the landscape almost at will. And besides the awkwardness of having to share walls with neighbors, living in aggregated pueblos introduced other problems. For people in cliff dwellings, hauling water, wood, and food to their homes was a major chore. The stress on local resources, especially in the firewood needed for daily cooking and warmth, was particularly intense, and conditions in aggregated pueblos were not very hygienic.

Given all the disadvantages of living in aggregated towns, why did people in the thirteenth century move into these closely packed quarters? For transitions of such suddenness, archaeologists consider either pull factors (benefits that drew families together) or push factors (some external threat or crisis that forced people to aggregate). In this case, push explanations dominate.

Population growth is considered a particularly influential push. After several generations of population growth, people packed the landscape in densities so high that communal pueblos may have been a necessary outcome. Around Sand Canyon, for example, populations grew from 5 -12 people per square kilometer in the tenth century to as many as 30 - 50 by the 1200s. As densities increased, domestic architecture became larger, culminating in crowded pueblos. Some scholars expand on this idea by emphasizing a corresponding need for arable land to feed growing numbers of people: construction of small dams, reservoirs, terraces, and field houses indicates that farmers were intensifying their efforts during the 1200s. Competition for good farmland may also have prompted people to bond together to assert rights over the best fields.

Another important push was the onset of the Little Ice Age, a climatic phenomenon that led to cooler

temperatures in the Northern Hemisphere. Although the height of the Little Ice Age was still around the corner, some evidence suggests that temperatures were falling during the thirteenth century. The environmental changes associated with this transition are not fully understood, but people living closest to the San Juan Mountains, to the northeast of Mesa Verde, were affected first. Growing food at these elevations is always difficult because of the short growing season. As the Little Ice Age progressed, farmers probably moved their fields to lower elevations, infringing on the lands of other farmers and pushing people together, thus contributing to the aggregations. Archaeologists identify a corresponding shift in populations toward the south and west toward Mesa Verde and away from higher elevations.

In the face of all these pushes, people in the Mesa Verde area had yet another reason to move into communal villages: the need for greater cooperation. Sharing and cooperation were almost certainly part of early Puebloan life, even for people living in largely independent single-household residences scattered across the landscape. Archaeologists find that even the most isolated residences during the eleventh and twelfth centuries obtained some pottery, and probably food, from some distance away, while major ceremonial events were opportunities for sharing food and crafts. Scholars believe that this cooperation allowed people to contend with a patchy environment in which precipitation and other resources varied across the landscape: if you produce a lot of food one year, you might trade it for pottery made by a distant ally who is having difficulty with crops — and the next year, the flow of goods might go in the opposite direction. But all of this appears to have changed thirteenth century. Although the climate remained as unpredictable as ever between one year and the next, it became much less locally diverse. In a bad year for farming, everyone was equally affected. No longer was it helpful to share widely. Instead, the most sensible thing would be for neighbors to combine efforts to produce as much food as possible, and thus aggregated towns were a sensible arrangement.

27 – 1 Crafts in the Ancient Near East

Some of the earliest human civilizations arose in southern Mesopotamia, in what is now southern Iraq, in the fourth millennium B.C.E. In the second half of the millennium, in the south around the city of Uruk, there was an enormous escalation in the area occupied by permanent settlements. A large part of that increase took place in Uruk itself, which became a real urban center surrounded by a set of secondary settlements. While population estimates are notoriously unreliable, scholars assume that Uruk inhabitants were able to support themselves from the agricultural production of the field surrounding the city, which could be reached with a daily commute. But Uruk's dominant size in the entire region, far surpassing that of other settlements, indicates that it was a regional center and a true city. Indeed, it was the first city in human history.

The vast majority of its population remained active in agriculture, even those people living within the city itself. But a small segment of the urban society started to specialize in nonagricultural tasks as a result of the city's role as a regional center. Within the productive sector, there was a growth of a variety of specialist craftspeople. Early in the Uruk period, the use of undecorated utilitarian pottery was probably the result of specialized mass production. In an early fourth-millennium level of the Eanna archaeological site at Uruk, a pottery style appears that is most characteristic of this process, the so-called beveled-rim bowl. It is a rather shallow bowl that was crudely made in a mold; hence, in only a limited number of standard sizes. For some unknown reason, many were discarded, often still intact, and thousands have been found all over the Near East. The beveled-rim bowl is one of the most telling diagnostic finds for identifying an Uruk-period site. Of importance is the fact that it was produced rapidly in large amounts, most likely by specialists in a central location.

A variety of documentation indicates that certain goods, once made by a family member as one of many duties, were later made by skilled artisans. Certain images depict groups of people, most likely women, involved in weaving textiles, an activity we know from later third-millennium texts to have been vital in the economy and to have been centrally administered. Also, a specialized metal-producing workshop may have been excavated in a small area at Uruk. It contained a number of channels lined by a sequence of holes, about 50 centimeters

deep, all showing burn marks and filled with ashes. This has been interpreted as the remains of a workshop where molten metal was scooped up from the channel and poured into molds in the holes. Some type of mass production by specialists were involved here.

Objects themselves suggest that they were the work of skilled professionals. In the late Uruk period (3500-3100 B.C.E.), there first appeared a type of object that remained characteristic for Mesopotamia throughout its entire history: the cylinder seal. This was a small cylinder, usually no more than 3 centimeters high and 2 centimeters in diameter, of shell, bone, faience (a glassy type of stoneware), or various types of stones, on which a scene was carved into the surface. When rolled over a soft material - primarily the clay of bullae (round seals), tablets, or clay lumps attached to boxes, jars, or door bolts - the scene would appear in relief, easily legible. The technological knowledge needed to carve it was far superior to that for stamp seals, which had happened in the early Neolithic period (approximately 10,000-5000 B.C.E.). From the first appearance of cylinder seals, the carved scenes could be highly elaborate and refined, indicating the work of specialist stone-cutters. Similarly, the late Uruk period shows the first monumental art, relief, and statuary in the round, made with a degree of mastery that only a professional could have produced.

29 – 1 Characteristics of Roman Pottery

The pottery of ancient Romans is remarkable in several ways. The high quality of Roman pottery is very easy to appreciate when handling actual pieces of tableware or indeed kitchenware and amphorae (the large jars used throughout the Mediterranean for the transport and storage of liquids, such as wine and oil). However, it is impossible to do justice to Roman wares on the page, even when words can be backed up by photographs and drawing. Most Roman pottery is light and smooth to touch and very tough, although, like all pottery, it shatters if dropped on a hard surface. It is generally made with carefully selected and purified clay, worked to thin-walled and standardized shapes on a fast wheel and fired in a kiln (pottery oven) capable of ensuring a consistent finish. With handmade pottery, inevitably there are slight differences between individual vessels of the same design and occasional minor blemishes (flaws). But what strikes the eye and the touch most immediately and most powerfully with Roman pottery is its consistent high quality.

This is not just an aesthetic consideration but also a practical one. These vessels are solid (brittle, but not fragile), they are pleasant and easy to handle (being light and smooth), and, with their hard and sometimes glossy (smooth and shiny) surfaces, they hold liquids well and are easy to wash. Furthermore, their regular and standardized shapes would have made them simple to stack and store. When people today are shown a very ordinary Roman pot and, in particular, are allowed to handle it, they often comment on how modern it looks and feels, and they need to be convinced of its true age.

As impressive as the quality of Roman pottery is its sheer massive quantity. When considering quantities, we would ideally like to have some estimates for overall production from particular sites of pottery manufacture and for overall consumption at specific settlements. Unfortunately, it is in the nature of the archaeological evidence, which is almost invariably only a sample of what once existed, that such figures will always be elusive. However, no one who has ever worked in the field would question the abundance of Roman pottery, particularly in the Mediterranean region. This abundance is notable in Roman settlements (especially urban sites) where the labor that archaeologists have to put into the washing and sorting of potsherds (fragments of pottery)

constitutes a high proportion of the total work during the initial phases of excavation.

Only rarely can we derive any "real" quantities from deposits of broken pots. However, there is one exceptional dump, which does represent a very large part of the site's total history of consumption and for which an estimate of quantity has been produced. On the left bank of the Tiber River in Rome, by one of the river ports of the ancient city, is a substantial hill some 50 meters high called Monte Testaccio. It is made up entirely of broken oil amphorae, mainly of the second and third centuries A.D. It has been estimated that Monte Testaccio contains the remains of some 53 million amphorae, in which around 6,000 million liters of oil were imported into the city from overseas. Imports into imperial Rome were supported by the full might of the state and were therefore quite exceptional - but the size of the operations at Monte Testaccio, and the productivity and complexity that lay behind them, nonetheless cannot fail to impress. This was a society with similarities to modern one - moving goods on a gigantic scale, manufacturing high-quality containers to do so, and occasionally, as here, even discarding them on delivery.

Roman pottery was transported not only in large quantities but also over substantial distances. Many Roman pots, in particular amphorae and the fine wares designed for use at tables, could travel hundreds of miles - all over the Mediterranean and also further afield. But maps that show the various spots where Roman pottery of a particular type has been found tell only part of the story. What is more significant than any geographical spread is the access that different levels of society had to good-quality products. In all but the remotest regions of the empire, Roman pottery of a high standard is common at the sites of humble villages and isolated farmsteads.

34 – 1 Islamic Art and the Book

The arts of the Islamic book, such as calligraphy and decorative drawing, developed during A.D. 900 to 1500, and luxury books are some of the most characteristic examples of Islamic art produced in this period. This came about from two major developments: paper became common, replacing parchment as the major medium for writing, and rounded scripts were regularized and perfected so that they replaced the angular scripts of the previous period, which because of their angularity were uneven in height. Books became major vehicles for artistic expression, and the artists who produced them, notably calligraphers and painters, enjoyed high status, and their workshops were often sponsored by princes and their courts. Before A.D. 900, manuscripts of the Koran (the book containing the teachings of the Islamic religion) seem to have been the most common type of book produced and decorated, but after that date a wide range of books were produced for a broad spectrum of patrons. These continued to include, of course, manuscripts of the Koran, which every Muslim wanted to read, but scientific works, histories, romances, and epic and lyric poetry were also copied in fine handwriting and decorated with beautiful illustrations. Most were made for sale on the open market, and cities boasted special souks (markets) where books were bought and sold. The mosque of Marrakech in Morocco is known as the Kutubiyya, or Booksellers' Mosque, after the adjacent market. Some of the most luxurious books were specific commissions made at the order of a particular prince and signed by the calligrapher and decorator.

Papermaking had been introduced to the Islamic lands from China in the eighth century. It has been said that Chinese papermakers were among the prisoners captured in a battle fought near Samarkand between the Chinese and the Muslims in 751, and the technique of papermaking—in which cellulose pulp extracted from any of several plants is first suspended in water, caught on a fine screen, and then dried into flexible sheets—slowly spread westward. Within fifty years, the government in Baghdad was using paper for documents. Writing in ink on paper, unlike parchment, could not easily be erased, and therefore paper had the advantage that it was difficult to alter what was written on it. Papermaking spread quickly to Egypt—and eventually to Sicily and Spain—but it was several centuries before

paper supplanted parchment for copies of the Koran, probably because of the conservative nature of religious art and its practitioners. In western Islamic lands, parchment continued to be used for manuscripts of the Koran throughout this period.

The introduction of paper spurred a conceptual revolution whose consequences have barely been explored. Although paper was never as cheap as it has become today, it was far less expensive than parchment, and therefore more people could afford to buy books. Paper is thinner than parchment, so more pages could be enclosed within a single volume. At first, paper was made in relatively small sheets that were pasted together, but by the beginning of the fourteenth century, very large sheets—as much as a meter across—were available. These large sheets meant that calligraphers and artists had more space on which to work. Paintings became more complicated, giving the artist greater opportunities to depict space or emotion. The increased availability of paper, particularly after 1250, encouraged people to develop systems of representation, such as architectural plans and drawings. This in turn allowed the easy transfer of artistic ideas and motifs over great distances from one medium to another, and in a different scale in ways that had been difficult, if not impossible, in the previous period.

Rounded styles of Arabic handwriting had long been used for correspondence and documents alongside the formal angular scripts used for inscriptions and manuscripts of the Koran. Around the year 900, Ibn Muqla, who was a secretary and vizier at the Abbasid court in Baghdad, developed a system of proportioned writing. He standardized the length of alif, the first letter of the Arabic alphabet, and then determined what the size and shape of all other letters should be, based on the alif. Eventually, six round forms of handwriting, composed of three pairs of big and little scripts known collectively as the Six Pens, became the standard repertory of every calligrapher.

37 – 3 Modern Architecture in the United States

At the end of the nineteenth century, there were basically two kinds of buildings in the United States. On one hand were the buildings produced for the wealthy or for civic purposes, which tended to echo the architecture of the past and to use traditional styles of ornamentation. On the other hand were purely utilitarian structures, such as factories and grain elevators, which employed modern materials such as steel girders and plate glass in an undisguised and unadorned manner. Such buildings, however, were viewed in a category separate from "fine" architecture, and in fact were often designed by engineers and builders rather than architects. The development of modern architecture might in large part be seen as an adaptation of this sort of functional building and its pervasive application for daily use. Indeed, in this influential book *Toward a New Architecture*, the Swiss architect Le Corbusier illustrated his text with photographs of American factories and grain storage silos, as well as ships, airplanes, and other industrial objects. Nonetheless, modern architects did not simply employ these new materials in a strictly practical fashion—they consciously exploited their aesthetic possibilities. For example, glass could be used to open up walls and eliminate their stone and brick masonry because large spaces could now be spanned with steel beams.

The fundamental premise of modern architecture was that the appearance of the building should exhibit the nature of its materials and forms of physical support. This often led to effects that looked odd from a traditional standpoint but that became hallmarks of modern architecture for precisely this reason. For example, in traditional architecture, stone or brick walls served a structural role, but in a steel-beam building the walls were essentially hung from the internal skeleton of steel beams, which meant that walls and corners no longer needed to be solid but could be opened up in unexpected ways. At the Fagus shoe factory in Germany, for example, German architect Walter Gropius placed glass walls in the corners, effectively breaking open the box of traditional architecture and creating a new sense of light and openness. Similarly, steel beams could be used to construct balconies that projected out from the building without any support beneath them. These dramatic balconies quickly became a signature of modern architects such as Frank Lloyd Wright. Wright's most dramatic residence, Fallingwater,

has balconies that thrust far out over a stream in a way that seems to defy gravity.

The ways in which new technology transformed architectural design are dramatically illustrated through the evolution of the high-rise office building. After ten or twelve stories, masonry construction reaches a maximum possible height, since it runs into difficulties of compression and of inadequate lateral strength to combat wind shear. Steel construction, on the other hand, can support a building of 50 or 100 stories without difficulty. Such buildings were so different from any previous form of architecture that they quickly acquired a new name—the skyscraper.

From the standpoint of real estate developers, the purpose of skyscrapers was to increase rental space in valuable urban locations. But to create usable high-rise buildings, a number of technical challenges needed to be solved. One problem was getting people to the upper floors, since after five or six stories it becomes exhausting to climb stairs. Updated and electrified versions of the freight elevator that had been introduced by Elisha Graves Otis in 1853 (several decades before skyscraper construction) solved this problem. Another issue was fire safety. The metal supporting buildings became soft when exposed to fire and collapsed relatively quickly. (They could melt at 2700 Fahrenheit, whereas major fires achieve temperatures of 3000 degrees). However, when the metal is encased in fire-retardant materials, its vulnerability to fire is much decreased. In Chicago, a system was developed for surrounding the metal components with hollow tiles made from brick-like terra-cotta. Such tiles are impervious to fire. The terra-cotta tiles were used both to encase the supporting members and as flooring. A structure built with steel beams protected by terra-cotta tiles was still three times lighter than a comparably sized building that used masonry construction, so the weight of the tiles was not a problem.

41 – 1 Navajo Art

The Navajo, a Native American people living in the southwestern United States, live in small scattered settlements. In many respects, such as education, occupation, and leisure activities, their life is like that of other groups that contribute to the diverse social fabric of North American culture in the twenty-first century. At the same time, they have retained some traditional cultural practices that are associated with particular art forms. For example, the most important traditional Navajo rituals include the production of large floor paintings. These are actually made by pouring thin, finely controlled streams of colored sands or pulverized vegetable and mineral substances, pollen, and flowers in precise patterns on the ground. The largest of these paintings may be up to 5.5 meters in diameter and cover the entire floor of a room. Working from the inside of the design outward, the Navajo artist and his assistants will sift the black, white, bluish-gray, orange, and red materials through their fingers to create the finely detailed imagery. The paintings and chants used in the ceremonies are directed by well-trained artists and singers who enlist the aid of spirits who are impersonated by masked performers. The twenty-four known Navajo chants can be represented by up to 500 sand paintings. These complex paintings serve as memory aids to guide the singers during the performance of the ritual songs, which can last up to nine days.

The purpose and meaning of the sand paintings can be explained by examining one of the most basic ideals of Navajo society, embodied in their word *hozho* (beauty or harmony, goodness, and happiness). It coexists with *hochxo* ("ugliness," or "evil," and "disorder") in a world where opposing forces of dynamism and stability create constant change. When the world, which was created in beauty, becomes ugly and disorderly, the Navajo gather to perform rituals with songs and make sand paintings to restore beauty and harmony to the world. Some illness is itself regarded as a type of disharmony. Thus, the restoration of harmony through a ceremony can be part of a curing process.

Men make sand paintings that are accurate copies of paintings from the past. The songs sung over the paintings are also faithful renditions of songs from the past. By re-creating these arts, which reflect the original beauty of creation, the Navajo bring beauty to the present world. As relative newcomers

to the Southwest, a place where their climate, neighbors, and rulers could be equally inhospitable, the Navajo created these art forms to affect the world around them, not just through the recounting of the actions symbolized, but through the beauty and harmony of the artworks themselves. The paintings generally illustrate ideas and events from the life of a mythical hero, who, after being healed by the gods, gave gifts of songs and paintings. Working from memory, the artists re-create the traditional form of the image as accurately as possible.

The Navajo are also world-famous for the designs on their woven blankets. Navajo women own the family flocks, control the shearing of the sheep, the carding, the spinning, and dying of the thread, and the weaving of the fabrics. While the men who make faithful copies of sand paintings from the past represent the principle of stability in Navajo thought, women embody dynamism and create new designs for every weaving they make. Weaving is a paradigm of the creativity of a mythic ancestor named Spider Woman who wove the universe as a cosmic web that united earth and sky. It was she who, according to legend, taught Navajo women how to weave. As they prepare their materials and weave, Navajo women imitate the transformations that originally created the world. Working on their looms, Navajo weavers create images through which they experience harmony with nature. It is their means of creating beauty and thereby contributing to the beauty, harmony, and healing of the world. Thus, weaving is a way of seeing the world and being part of it.

47 – 1 Roman Cultural Influence on Britain

After the Roman Empire's conquest of Britain in the first century A.D., the presence of administrators, merchants, and troops on British soil, along with the natural flow of ideas and goods from the rest of the empire, had an enormous influence on life in the British Isles. Cultural influences were of three types: the bringing of objects, the transfer of craft workers, and the introduction of massive civil architecture. Many objects were not art in even the broadest sense and comprised utilitarian items of clothing, utensils, and equipment. We should not underestimate the social status associated with such mundane possessions which had not previously been available. The flooding of Britain with red-gloss pottery from Gaul (modern-day France), decorated with scenes from Classical mythology, probably brought many into contact with the styles and artistic concepts of the Greco-Roman world for the first time, whether or not the symbolism was understood. Mass-produced goods were accompanied by fewer more aesthetically impressive objects such as statuettes. Such pieces perhaps first came with officials for their own religious worship; others were then acquired by native leaders as diplomatic gifts or by purchase. Once seen by the natives, such objects created a fashion which rapidly spread through the province.

In the most extreme instances, natives literally bought the whole package of Roman culture. The Fishbourne villa, built in the third quarter of the first century A.D., probably for the native client king Cogidubnus, amply illustrates his Roman pretensions. It was constructed in the latest Italian style with imported marbles and stylish mosaics. It was lavishly furnished with imported sculptures and other Classical objects. A visitor from Rome would have recognized its owner as a participant in the contemporary culture of the empire, not at all provincial in taste. Even if those from the traditional families looked down on him, they would have been unable to dismiss him as uncultured. Although exceptional, this demonstrates how new cultural symbols bound provincials to the identity of the Roman world.

Such examples established a standard to be copied. One result was an influx of craft worker, particularly those skilled in artistic media like stone-carving which had not existed before the conquest. Civilian workers came mostly from Gaul and Germany. The magnificent temple built beside the sacred spring at Bath was constructed only about

twenty years after the conquest. Its detail shows that it was carved by artists from northeast Gaul. In the absence of a tradition of Classical stone-carving and building, the desire to develop Roman amenities would have been difficult to fulfill. Administrators thus used their personal contacts to put the Britons in touch with architects and masons. As many of the officials in Britain had strong links with Gaul, it is not surprising that early Roman Britain owes much to craft workers from that area. Local workshops did develop and stylistically similar groups of sculpture show how skills in this new medium became widespread. Likewise skills in the use of mosaic, wall painting, ceramic decoration, and metal-working developed throughout the province with the eventual emergence of characteristically Romano-British styles.

This art had a major impact on the native peoples, and one of the most importance factors was a change in the scale of buildings. Pre-Roman Britain was highly localized, with people rarely traveling beyond their own region. On occasion large groups amassed for war or religious festivals, but society remained centered on small communities. Architecture of this era reflected this with even the largest of the fortified towns and hill forts containing no more than clusters of medium-sized structures. The spaces inside even the largest roundhouses were modest, and the use of rounded shapes and organic building materials gave buildings a human scale. But the effect of Roman civil architecture was significant. The sheer size of space enclosed within buildings like the basilica of London must have been astonishing. This was an architecture of dominance in which subject peoples were literally made to feel small by buildings that epitomized imperial power. Supremacy was accentuated by the unyielding straight lines of both individual buildings and planned settlements since these too provided a marked contrast with the natural curvilinear shapes dominant in the native realm.

52 – 2 Natufian Culture

In the archaeological record of the Natufian period, from about 12,500 to 10,200 years ago, in the part of the Middle East known as the Levant—roughly east of the Mediterranean and north of the Arabian Peninsula—we see clear evidence of agricultural origins. The stone tools of the Natufians included many sickle-shaped cutting blades that show a pattern of wear characteristic of cereal harvesting. Also, querns (hand mills) and other stone tools used for processing grain occur in abundance at Natufian sites, and many such tools show signs of long, intensive use. Along with the sickle blades are many grinding stones, primarily mortars and pestles of limestone or basalt. There is also evidence that these heavy grinding stones were transported over long distances, more than 30 kilometers in some cases, and this is not something known to have been done by people of preceding periods. Fishhooks and weights for sinking fishing nets attest to the growing importance of fish in the diet in some areas. Stone vessels indicate an increased need for containers, but there is no evidence of Natufian clay working or pottery. Studies of the teeth of Natufians also strongly suggest that these people specialized in collecting cereals and may have been cultivating them and in the process of domesticating them, but they were also still hunter—foragers who intensively hunted gazelle and deer in more lush areas and wild goats and equids in more arid zones.

The Natufians had a different settlement pattern from that of their predecessors. Some of their base camps were far larger (over 1,000 square meters) than any of those belonging to earlier periods, and they may have lived in some of these camps for half the year or even more. In some of the camps, people made foundations and other architectural elements out of limestone blocks. Trade in shell, obsidian, and other commodities seems to have been on the rise, and anthropologists suspect that the exchange of perishables (such as skins, foodstuffs) and salt was also on the increase. With the growing importance of wild cereals in the diet, salt probably became for the first time a near necessity: people who eat a lot of meat get many essential salts from this diet, but diets based on cereals can be deficient in salts. Salt was probably also important as a food preservative in early villages.

As always, there is more to a major cultural change than simply a shift in economics. The Natufians

made (and presumably wore) beads and pendants in many materials, including gemstones and marine shells that had to be imported, and it is possible that this ornamentation actually reflects a growing sense of ethnic identity and perhaps some differences in personal and group status. Cleverly carved figurines of animals, women, and other subjects occur in many sites, and Natufian period cave paintings have been found in Anatolia, Syria, and Iran. More than 400 Natufian burials have been found, most of them simple graves set in house floors. As archaeologist Belfer-Cohen notes, these burials may reflect an ancestor cult and a growing sense of community emotional ties and attachment to a particular place, and toward the end of the Natufian period, people in this area were making a strict separation between living quarters and burial grounds. In contrast with the Pleistocene cultures of the Levant, Natufian culture appears to have experienced considerable social change.

The question of why the Natufians differed from their predecessors in these and other ways and why they made these first steps toward farming as a way of life remains unclear. There were climate changes, of course, and growing aridity and rising population densities may have forced them to intensify the exploitation of cereals, which in turn might have stimulated the development of sickles and other tools and the permanent communities that make agriculture efficient. But precisely how these factors interacted with others at play is poorly understood.

53 – 3 Paleolithic Cave Paintings

In any investigation of the origins of art, attention focuses on the cave paintings created in Europe during the Paleolithic era (c. 40,000-10,000 years ago) such as those depicting bulls and other animals in the Lascaux cave in France. Accepting that they are the best preserved and most visible signs of what was a global creative explosion, how do we start to explain their appearance? Instinctively, we may want to update the earliest human artists by assuming that they painted for the sheer joy of painting. The philosophers of Classical Greece recognized it as a defining trait of humans to "delight in works of imitation"—to enjoy the very act and triumph of representation. If we were close to a real lion or snake, we might feel frightened. But a well-executed picture of a lion or snake will give us pleasure. Why suppose that our Paleolithic ancestors were any different?

This simple acceptance of art for art's sake has a certain appeal. To think of Lascaux as a gallery allows it to be a sort of special viewing place where the handiwork of accomplished artists might be displayed. Plausibly, daily existence in parts of Paleolithic Europe may not have been so hard, with an abundance of ready food and therefore the leisure time for art. The problems with this explanation, however, are various. In the first place, the proliferation of archaeological discoveries—and this includes some of the world's innumerable rock art sites that cannot be dated—has served to emphasize a remarkably limited repertoire of subjects. The images that recur are those of animals. Human figures are unusual, and when they do make an appearance, they are rarely done with the same attention to form accorded to the animals. If Paleolithic artists were simply seeking to represent the beauty of the world around them, would they not have left a far greater range of pictures—of trees, flowers, of the Sun and the stars?

A further question to the theory of art for art's sake is posed by the high incidence of Paleolithic images that appear not to be imitative of any reality whatsoever. These are geometrical shapes or patterns consisting of dots or lines. Such marks may be found isolated or repeated over a particular surface but also scattered across more recognizable forms. A good example of this may be seen in the geologically spectacular grotto of Pêche Merle, in the Lot region of France. Here we encounter some favorite animals from the

Paleolithic repertoire—a pair of stout-bellied horses. But over and around the horses' outlines are multiple dark spots, daubed in disregard for the otherwise naturalistic representation of animals. What does such patterning imitate? There is also the factor of location. The caves of Lascaux might conceivably qualify as underground galleries, but many other paintings have been found in recesses totally unsuitable for any kind of viewing—tight nooks and crannies that must have been awkward even for the artists to penetrate, let alone for anyone else wanting to see the art.

Finally, we may doubt the notion that the Upper Paleolithic period was a paradise in which food came readily, leaving humans ample time to amuse themselves with art. For Europe it was still the Ice Age. An estimate of the basic level of sustenance then necessary for human survival has been judged at 2200 calories per day. This consideration, combined with the stark emphasis upon animals in the cave art, has persuaded some archaeologists that the primary motive behind Paleolithic images must lie with the primary activity of Paleolithic people: hunting.

Hunting is a skill. Tracking, stalking, chasing, and killing the prey are difficult, sometimes dangerous activities. What if the process could be made easier—by art? In the early decades of the twentieth century, Abbé Henri Breuil argued that the cave paintings were all about "sympathetic magic". The artists strived diligently to make their animal images evocative and realistic because they were attempting to capture the spirit of their prey. What could have prompted their studious attention to making such naturalistic, recognizable images? According to Breuil, the artists may have believed that if a hunter were able to make a true likeness of some animal, then that animal was virtually trapped. Images, therefore, may have had the magical capacity to confer success or luck in the hunt.

59 – 3 Ancient Greek Pottery

An extremely important and long-standing industry in ancient Greece was the ceramics industry, giving rise to the huge number of pieces of pottery archaeologists and historians pore over. In reality, pottery production can almost be seen as a combination of a natural resources-based economy and a production-based economy, as the primary material-good clay-was a result of the natural geology. However, not all city-states made equal use of their clay resources, and only two-Corinth and Athens-were really active in pottery export in the Archaic Age (750 -500 BC) and following.

As with weaving, pottery no doubt began as, and in many places remained, a household industry, probably practiced exclusively in the summer when the clay and kindling wood were drier and easier to transport and burn. But evidence of a more concentrated effort at production for export appeared already in the Bronze Age (3000-1100 B c.) and continued in the Dark Ages (1100-800 B c), as with the copious export of Euboean scyphi (drinking cups) to such locations in the Mediterranean as Cyprus, the Levant, and even Italy. These were followed by the Corinthian cups and scyphi, and finally the Athenian kylix (a shallow type of wine goblet). The demand for Greek vessels to drink out of appears to have stemmed from a desire for Greek things to drink, and amphorae and pithoi (storage vessels), especially from Attica, also proliferated in the Mediterranean.

This need for commercial production led to the rise of workshops and factories. Such workshops were usually family-owned businesses run by the head of the household, his sons (and possibly daughters), and, depending on the size of the industry, additional servants and/or slaves. The general assumption concerning many professional potters is that they were not citizens of the cities in which they worked. In Athens, these were the metics, or resident aliens, as well as slaves. The identity of the workers in the primary ceramic working area in Corinth-the Potters' Quarter-is uncertain. But an inscribed pottery shard naming the Phoenician goddess Astarte suggests that there were Phoenician immigrants (from what is now Lebanon) working there, or at least a heavy Phoenician influence.

Happily for modern researchers, many artisans signed their works. This gives an idea not only of who these ancient ceramicists were, but also of

how labor was divided. Modern scholars have the names of a few potter-families from sixth-century B c Athens: Nearkhos, who worked with his sons Tleson and Ergoteles; Ergotimos, who was succeeded by his son Eukheiros (literally, "Good Hand"); and Amasis and his son Kleophrades, who produced pottery from the mid-sixth century B c to the dawn of the fifth. The signatures indicate, however, that although these men fashioned and fired the ceramics, others were employed to paint the more elaborate figural vases. Such artists traveled around to different workshops, like the sixth-century Epictetus, whose name as a painter accompanied about six different names of potters.

Starting around 550 B.C., numbers were scratched onto the bottoms of some vases, often appearing after the name of the vase in full or abbreviated form. These may have been batch numbers, indicating groups of pots fired together and intended for specific merchants or markets. The merchants were also indicated by markings on some vases, whereby a signature sign or abbreviation indicated to which exporter or importer in the receiving city the pottery was destined. In the sixth century B c ., such merchants were frequently Iomans (Greeks from cities on the Mediterranean coast of Asia Minor), who were responsible for much of the import-export business at Naucratis in the south and Gravisca in the west.

But the most important markings from the point of view of economics are the actual "price tags" etched into some pots. An important difference in the price occurred between painted and undecorated vases, with a markup of anywhere from 25 to 50 percent for the figural vases. Clearly, this was because the pottery manufacturer had to pay for the additional labor.

58 – 1 The Development of Instrumental Music

Until the sixteenth century, almost all music was written for the voice rather than for musical instruments. Even during the Renaissance (from the fourteenth to the sixteenth century), instrumental music was, for the most part, the result of substituting an instrument for a voice in music written for singing or dancing. The seventeenth century marked the rise of music that lacked extramusical meaning. Like a mathematical equation or geometric formula, the instrumental music of the early modern era carried no explicit narrative content—it was neither a vehicle of religious expression nor a means of supporting a secular (nonreligious) vocalized text. Such music was written without consideration for the associational content traditionally provided by a set of sung lyrics. The idea of music as an aesthetic exercise, composed for its own sake rather than to serve a religious or communal purpose, was a notable feature of the seventeenth century and one that has distinguished modern Western European music from the musical traditions of Asia and Africa.

Not surprisingly, the rise of instrumental music was accompanied by improvements in instruments and refinements in tuning. Indeed, instrumental music came to dominate musical composition at the very moment that Western musicians were perfecting such stringed instruments as the violin, viola, and cello and such keyboard instruments as the organ and harpsichord. By the early eighteenth century, musicians were adopting the system of tuning known as equal temperament, whereby the octave was divided into twelve half-steps of equal size. Johann Sebastian Bach's *Well-Tempered Clavier* (1722) was an attempt to popularize this system to a skeptical musical public. The new attention paid to improving instruments and systematizing tuning mirrored the efforts of scientists and philosophers to bring precision and uniformity to the tools and methods for scientific inquiry.

In the seventeenth century, northern Italy was the world center for the manufacture of violins. The Amati, Guarneri, and Stradivari families of Cremona, Italy, established the techniques of making high-quality violins that were sought in all of the great courts of Europe. Transmitted from father to son, the construction techniques used to produce these instruments were guarded so secretly that modern violinmakers have never successfully imitated them. Elsewhere, around

1650, earlier instruments were standardized and refined. Also during this period amateur music making was widespread, and professional performance also took a great leap forward, as a new breed of virtuosi inspired the writing of treatises on performance techniques.

Three main types of composition—the sonata, the suite, and the concerto—dominated seventeenth-century instrumental music. All three reflect the baroque taste for dramatic contrasts in tempo and texture. The sonata (from the Italian word for “sounded,” that is, music played and not sung) is a piece written for a few instruments—often no more than one or two. It usually consisted of three movements of contrasting tempo—fast/slow/fast—each based on a song or dance form of the time. The suite, written for any combination of instruments, is a sequence or series of movements derived from various European court or folk dances—for example, the sarabande, the pavane, the minuet, and gigue, or jig. Henry Purcell (1659–1695) in England, François Couperin (1668–1733) in France, and Johann Sebastian Bach (1685–1750) in Germany all contributed to the development of the suite as a musical genre. Finally, the concerto (from the same root as *concertato*, which describes opposing or contrasting bodies of sound) is a composition consisting of two groups of instruments, one small and the other large, playing in dialogue. The typical baroque concerto, the *concerto grosso* (“large concerto”) featured several movements whose number and kind varied considerably.

The leading Italian instrumental composer of the baroque era was Antonio Vivaldi (1678–1741). Vivaldi wrote some 450 concertos. He systematized the *concerto grosso* into a three-movement form (fast/slow/fast) and increased the distinctions between solo and ensemble groups in each movement. Of the many exciting compositions Vivaldi wrote for solo violin and ensemble, the most glorious is *The Four Seasons*, a group of four violin concertos, each of which musically describes a single season.

62 – 2 The Use of Jade and Bronze in China

Jade and bronze were chosen early in China's history as the materials for society's most precious objects. The main reason for this is very straightforward. Both materials are beautiful to the eye and in the case of jade, also to the touch. Jade is a dense stone that can be ground to smooth, soft surfaces and glows in subtle greens, grays, and browns. Bronze, an alloy of tin and copper, is a light, bright gold color when polished.

Jade was employed (c. 4500 B c) for exceptionally elegant versions of utilitarian stone tools; bronze was the choice (c. 1650 B c) for the highest-quality cooking pots. Made in jade, the tools were not for daily use but for displays of status and power, and the bronze cooking pots were not for ordinary meals but were reserved for offerings of food and wine to ancestors. Jade and bronze were thus used for special ritual or ceremonial versions of standard everyday items.

The materials themselves were scarcer and required more labor to work than ordinary stone and ceramic. That this was the case must have been as clear in the past as it is today and must have marked the objects as in some way exceptional. Not only the materials but also the ways in which they were worked demonstrated their exalted functions. Jade scepters (ceremonial objects of various shapes that were carried by influential people as symbols of authority) were ground more thinly than the stone tools they copied, such as axes or reaping knives. Had they been used to chop down a tree, they would have broken. Bronze cooking pots were made in intricate forms, with extra knobs and handles and dense decoration all of which would have been impracticable on everyday ceramics.

There would be no point in using these scarce and labor-intensive materials in place of common ones if they could not be immediately recognized as outstanding. Craftsmanship was therefore directed toward exploiting and displaying the particular qualities of jade and bronze that make them recognizably different from ordinary stone and ceramic. These qualities had to be made visually noticeable. Other qualities, such as weight or texture, can only be appreciated by holding the object, and fewer people will have the opportunity to pick it up and feel its features with their hands than will be able to glance at it quickly.

Visual distinctions deployed to separate the ceremonial from the everyday can also be used to refer to smaller differences in the ranking of ritual items, separating those owned by kings from those owned by nobles. Such differences can be in size or mass—one piece being larger or thicker than another—or in skill, with one being better crafted than another. We make such judgments when looking at ancient objects today, and it seems likely that some sort of similar assessment was made at the time.

Two hierarchies, both visible to the eye, were thus achieved. First and most fundamental was a hierarchy of function, the ritual object standing above the everyday object. Second was the hierarchy within each category, with the more elaborately worked and often larger object higher up the scale. The rich and powerful owned the jade scepters and bronze vessels. What is more, these scepters and vessels not only expressed or reflected the status of their owners but they also probably reinforced it. In this way, jades and bronzes became bound up with the religious and political structures of the early Chinese state, and they remained so linked up to the twentieth century.

It is therefore not surprising that at times when rulers and their courts wished to assert their authority, they commissioned large numbers of conspicuous bronzes. Further, when they wished to distract attention from weaknesses in society, they emphasized their power even more by increasing expenditure on ritual objects. Thus changes in the aesthetic qualities of jades and bronzes can inform us about the society that produced them.

Later generations of Chinese scholars of the Song Dynasty (A.D. 960-1279) quickly recognized the link between the bronzes and jades and political power. In their search for political legitimacy, they collected and studied ancient artifacts, pursuits that in themselves reinforced the high status of jade and bronze in society. From the idea that political power resided in ancient objects, including calligraphy and painting, grew the first art collections.

69 – 3 Ancient Southwestern Cultures

Around 2,500 years ago, Amerindian peoples began to live in settled communities in what is now the southwestern United States, growing maize – a crop unique to the New World – together with squash and, later, beans. These cultures developed advanced technologies that allowed them to manipulate their environment and thrive in its arid conditions, though ultimately environmental factors may also have led to their disappearance.

One of these cultures, the Hohokam, practiced irrigation-based agriculture. Knowledge of Hohokam irrigation systems is based on archaeological remains found in southern Arizona as well as on extrapolation from technologies currently in use. Because the land was desert with less than 10 inches of rainfall a year, agriculture required catchment dams, irrigation canals, and water conservation. Between A.D. 300 and 900, the Hohokam manipulated the arid environment by building large canals, some as wide as 30 feet across, 7 feet deep, and 8 miles long, that could fertilize as many as 8,000 acres at a time with water from the Salt and Gila river systems.

The Hohokam could grow crops and store them throughout the year, producing a continual supply of food. An environmental consequence of irrigated agriculture, however, is that water from the rivers leaches out salts, drawing them up from the subsoil and leaving salinated (salty) topsoil. Just as the Mesopotamian settlements of the Near East were abandoned owing to salinization of the land, the Hohokam may have been forced to abandon their villages. Alternatively, they may have remained evolving and adapting to changed conditions. By the time the Spanish arrived in the region in the sixteenth century, they met Amerindians called Akimel O'odham, who were still practicing a form of irrigation, but not on the enormous scale developed by the Hohokam. Other explanations for the disappearance of Hohokam culture include flooding, drought, earthquakes, and deforestation.

A second group of southwestern peoples, the Anasazi, also developed advanced technologies for manipulating their environment. They constructed large multistoried communal houses in villages: or pueblos, along the edges of a river in the bottom of New Mexico's Chaco Canyon where they planted crops. The communities consisted of many small villages with larger central cities containing four-story apartments and ramrod-straight roads that

linked Chaco Canyon with other communities in the Southwest.

One Anasazi pueblo, Mesa Verde, illustrates the sophistication with which these peoples transformed their environment. Mesa Verde is situated 7,000 to 8,000 feet above sea level on a vast forested plateau into which erosion has cut numerous steep-walled canyons. Mesa Verde culture and agriculture developed gradually over time. The people of the earliest culture, from around A.D. 500 to 750, were basket weavers who lived underground in pit houses. Around A.D. 750 to 1,100, they added ceremonial chambers called kivas and began building houses above ground out of bricks. Between A.D. 1,100 and 1,300, they reached the peak of their complex culture, building ladders ascending to the cliff areas as well as pathways descending to agricultural fields in the valleys. The vast agricultural network covered some 80 square miles of development, supporting numerous pueblos located about 1,800 to 2,000 feet above the river.

Not only did the Anasazi build roads and pueblos, they also put their understanding of astronomy to practical use. In particular, the Anasazi attained the ability to predict the advent of growing and harvesting seasons based on arrival of solstices (longest and shortest days) around June 21 and December 21 and equinoxes (days with equal periods of light and darkness) around March 21 and September 21. They did so by building an observatory on Fajada Butte in Chaco Canyon, a stone-slab structure assembled so that on solstices and equinoxes, a sliver of sunlight entered between two stones and fell on a petroglyph (an image carved onto a rock) in a characteristic way. With this knowledge they gained a measure of control over the hazards of an environment.

The Anasazi disappeared from their homeland at the height of their cultural development around the end of the thirteenth century. Deforestation, social disruption, and the onset of an 11-year drought around A.D. 1280 (as evidenced by the narrowness of tree-ring growth) may have been factors leading to the abandonment of the pueblos.

WRITING / PRINTING

39 – 1 Early Writing Systems

Scholars agree that writing originated somewhere in the Middle East, probably Mesopotamia, around the fourth millennium B.C.E. It is from the great libraries and word-hoards of these ancient lands that the first texts emerged. They were written on damp clay tablets with a wedged (or V-shaped) stick; since the Latin word for wedge is *cunea*, the texts are called cuneiform. The clay tablets usually were not fired; sun drying was probably reckoned enough to preserve the text for as long as it was being used. Fortunately, however, many tablets survived because they were accidentally fired when the buildings they were stored in burned.

Cuneiform writing lasted for some 3,000 years, in a vast line of succession that ran through Sumer, Akkad, Assyria, Nineveh, and Babylon, and preserved for us fifteen languages in an area represented by modern-day Iraq, Syria, and western Iran. The oldest cuneiform texts recorded the transactions of tax collectors and merchants, the receipts and bills of sale of an urban society. They had to do with things like grain, goats, and real estate. Later, Babylonian scribes recorded the laws and kept other kinds of records. Knowledge conferred power. As a result, the scribes were assigned their own goddess, Nisaba, later replaced by the god Nabu of Borsippa, whose symbol is neither weapon nor dragon but something far more fearsome, the cuneiform stick.

Cuneiform texts on science, astronomy, medicine, and mathematics abound, some offering astoundingly precise data. One tablet records the speed of the Moon over 248 days; another documents an early sighting of Halley's Comet, from September 22 to September 28, 164 B.C.E. More esoteric texts attempt to explain old Babylonian customs, such as the procedure for curing someone who is ill, which included rubbing tar and gypsum on the sick person's door and drawing a design at the foot of the person's bed. What is clear from the vast body of texts (some 20,000 tablets were found in King Ashurbanipal's library at Nineveh) is that scribes took pride in their writing and knowledge.

The foremost cuneiform text, the Babylonian Epic of Gilgamesh, deals with humankind's attempts to conquer time. In it, Gilgamesh, king and warrior, is crushed by the death of his best friend and so sets out on adventures that prefigure mythical heroes of

ancient Greek legends such as Hercules. His goal is not just to survive his ordeals but to make sense of this life. Remarkably, versions of Gilgamesh span 1,500 years, between 2100 B.C.E. and 600 B.C.E., making the story the epic of an entire civilization.

The ancient Egyptians invented a different way of writing and a new substance to write on—papyrus, a precursor of paper, made from a wetland plant. The Greeks had a special name for this writing: hieroglyphic, literally "sacred writing". This, they thought, was language fit for the gods, which explains why it was carved on walls of pyramids and other religious structures. Perhaps hieroglyphics are Egypt's great contribution to the history of writing: hieroglyphic writing, in use from 3100 B.C.E. until 394 C.E., resulted in the creation of texts that were fine art as well as communication. Egypt gave us the tradition of the scribe not just as educated person but as artist and calligrapher.

Scholars have detected some 6,000 separate hieroglyphic characters in use over the history of Egyptian writing, but it appears that never more than a thousand were in use during any one period. It still seems a lot to recall, but what was lost in efficiency was more than made up for in the beauty and richness of the texts. Writing was meant to impress the eye with the vastness of creating itself. Each symbol or glyph - the flowering reed (pronounced like V), the owl ("m"), the quail chick ("w"), etcetera - was a tiny work of art. Manuscripts were compiled with an eye to the overall design. Egyptologists have noticed that the glyphs that constitute individual words were sometimes shuffled to make the text more pleasing to the eye with little regard for sound or sense.

73 – 2 Writing History

Historians throughout the ages have been beset by many pitfalls in their attempts to write history. Modern historians writing the history of Africa not only face problems similar to those experienced by their colleagues writing the history of other cultures. Most medieval African societies south of the Sahara lacked an alphabet and a written language. African written languages were confined mainly to Ethiopia, to Sudan, and to the maritime cities of the east coast. Early Greek, Roman, Amharic, Fulani, and Swahili texts still survive to give us some idea of bygone days. Future collectors will almost certainly find additional records when they search for documents in North Africa and in faraway countries like Iran and Saudi Arabia. In tropical countries, however, old papers are preserved only with great difficulty, for white ants and atmospheric humidity pose serious problems even to modern archivists. Early record keepers were even less equal to their task, and much material has been lost or destroyed.

The lack of written material means that we shall never be able to read, for example, the memoirs of a king of the Luba people in the eighteenth century. Worse still, we cannot draw on administrative archives. Dull as they look, administrative files have a peculiar value denied to other written records. They are composed of documents drawn up or used in the course of an administrative or executive transaction of which they had formed a part. This does not necessarily make official memorandums, communiques, and such more reliable, for officials often enough were dishonest with their inferiors and superiors alike. But unlike the authors of so many biographies—they hardly ever wrote in order to deceive posterity; hence, archival evidence has an unusual degree of authenticity. The lack of such material makes the historian's job hard indeed. For where there is no writing, there are large areas of human behavior that cannot be described accurately except by speculative extrapolations from literate cultures or by conjectures determined to a large extent by our own cultural preconceptions.

Some historians, therefore, try to supplement their sources with information from indigenous traditions and folklore, heroic legends, genealogical tables, and personal reminiscences. The use of oral history, however, again presents many difficulties. Historians can look at the same document again and again; so can their

successors. An artifact can be preserved in a museum, and different conclusions can be arrived at by going over the same evidence again and again. But two or three people cannot interview the same informant twice; even should they do so, they will not necessarily elicit the same information, for the interviewers themselves help impose a subjective pattern on the information they obtain. Human witnesses, moreover, will remember previous interviews; their memory can never be erased like a magnetic tape in such a way that they can be questioned over and over again without changing their answers at least to some degree.

Admittedly, it is easy to underestimate the value of material elicited by oral historians who investigate preliterate societies. Professional poets and singers may faithfully record the same lay or the same genealogical table for many generations, for their memory is beyond question. Nevertheless, oral tradition is always subject to change. The Lunda people, who dwell on the Luapula River, are, for instance, very interested in their own past. But as Ian Cunnison, a British anthropologist, observes, "this is not history as we know it. It is the handing on of what is already known or believed, and it does not involve the asking of questions about the past, and the search for the answer to them."

This does not mean that oral traditions are without value to historians. They enlarge their imagination, they help them pose new questions, and they tell them a great deal concerning a people's culture; they also provide a substratum of fact. They are, however, difficult to interpret, and the conventional historian's intellectual tools may often prove inadequate for the task. Oral traditions, moreover, like all other forms of evidence, always stand in need of corroboration and can never be accepted on the basis of their own authority.

46 – 1 The Origins of Writing

It was in Egypt and Mesopotamia (modern-day Iraq) that civilization arose, and it is there that we find the earliest examples of that key feature of civilization, writing. These examples, in the form of inscribed clay tablets that date to shortly before 3000 B.C.E., have been discovered among the archaeological remains of the Sumerians, a gifted people settled in southern Mesopotamia.

The Egyptians were not far behind in developing writing, but we cannot follow the history of their writing in detail because they used a perishable writing material. In ancient times the banks of the Nile were lined with papyrus plants, and from the papyrus reeds the Egyptians made a form of paper; it was excellent in quality but, like any paper, fragile. Mesopotamia's rivers boasted no such useful reeds, but its land did provide good clay, and as a consequence the clay tablet became the standard material. Though clumsy and bulky it has a virtue dear to archaeologists: it is durable. Fire, for example, which is death to papyrus paper or other writing materials such as leather and wood, simply bakes it hard, thereby making it even more durable. So when a conqueror set a Mesopotamian palace ablaze, he helped ensure the survival of any clay tablets in it. Clay, moreover, is cheap, and forming it into tablets is easy, factors that helped the clay tablet become the preferred writing material not only throughout Mesopotamia but far outside it as well, in Syria, Asia Minor, Persia, and even for a while in Crete and Greece. Excavators have unearthed clay tablets in all these lands. In the Near East they remained in use for more than two and a half millennia, and in certain areas they lasted down to the beginning of the common era until finally yielding, once and for all, to more convenient alternatives.

The Sumerians perfected a style of writing suited to clay. This script consists of simple shapes, basically just wedge shapes and lines that could easily be incised in soft clay with a reed or wooden stylus; scholars have dubbed it cuneiform from the wedge-shaped marks (*cunei* in Latin) that are its hallmark. Although the ingredients are merely wedges and lines, there are hundreds of combinations of these basic forms that stand for different sounds or words. Learning these complex signs required long training and much practice; inevitably, literacy was largely limited to a small professional class, the scribes.

The Akkadians conquered the Sumerians around the middle of the third millennium B.C.E., and they took over the various cuneiform signs used for writing Sumerian and gave them sound and word values that fit their own language. The Babylonians and Assyrians did the same, and so did peoples in Syria and Asia Minor. The literature of the Sumerians was treasured throughout the Near East, and long after Sumerian ceased to be spoken, the Babylonians and Assyrians and others kept it alive as a literary language, the way Europeans kept Latin alive after the fall of Rome. For the scribes of these non-Sumerian languages, training was doubly demanding since they had to know the values of the various cuneiform signs for Sumerian as well as for their own language.

The contents of the earliest clay tablets are simple notations of numbers of commodities—animals, jars, baskets, etc. Writing, it would appear, started as a primitive form of bookkeeping. Its use soon widened to document the multitudinous things and acts that are involved in daily life, from simple inventories of commodities to complicated governmental rules and regulations.

Archaeologists frequently find clay tablets in batches. The batches, some of which contain thousands of tablets, consist for the most part of documents of the types just mentioned: bills, deliveries, receipts, inventories, loans, marriage contracts, divorce settlements, court judgments, and so on. These records of factual matters were kept in storage to be available for reference—they were, in effect, files, or, to use the term preferred by specialists in the ancient Near East, archives. Now and then these files include pieces of writing that are of a distinctly different order, writings that do not merely record some matter of fact but involve creative intellectual activity. They range from simple textbook material to literature—and they make an appearance very early, even from the third millennium B C E.

49 – 2 Movable Type

Nothing divided the medieval world in Europe more decisively from the Early Modern period than printing with movable type. It was a German invention and the culmination of a complex process. The world of antiquity had recorded its writings mainly on papyrus. Between 200 B.C and A.D 300, this was supplemented by vellum, calf skin treated and then smoothed by pumice stone. To this in late Roman times was added parchment, similarly made from the smoothed skin of sheep or goats. In the early Middle Ages, Europe imported an industrial process from China, which turned almost any kind of fibrous material into pulp that was then spread in sheets. This was known as cloth parchment. By about 1150, the Spanish had developed the first mill for making cheap paper (a word contracted from "papyrus", which became the standard term). One of the most important phenomena of the later Middle Ages was the growing availability of cheap paper. Even in England, where technology lagged far behind, a sheet of paper, or eight octavo pages, cost only a penny by the fifteenth century.

In the years 1446-1448, two German goldsmiths, Johannes Gutenberg and Johann Fust, made use of cheap paper to introduce a critical improvement in the way written pages were reproduced. Printing from wooden blocks was the old method; what the Germans did was to invent movable type for the letterpress. It had three merits: it could be used repeatedly until worn out; it was cast in metal from a mold and so could be renewed without difficulty; and it made lettering uniform. In 1450, Gutenberg began work on his Bible, the first printed book, known as the Gutenberg. It was completed in 1455 and is a marvel. As Gutenberg, apart from getting the key idea, had to solve a lot of practical problems, including imposing paper and ink into the process and the actual printing itself, for which he adapted the screw press used by winemakers, it is amazing that his first product does not look at all rudimentary. Those who handle it are struck by its clarity and quality.

Printing was one of those technical revolutions that developed its own momentum at extraordinary speed. Europe in the fifteenth century was a place where intermediate technology - that is, workshops with skilled craftspeople - was well established and spreading fast, especially in Germany and Italy. Such workshops were able to take on printing easily, and it thus became Europe's first true

industry. The process was aided by two factors: the new demand for cheap classical texts and the translation of the Latin Bible into "modern" languages. Works of reference were also in demand. Presses sprang up in several German cities, and by 1470, Nuremberg, Germany had established itself as the center of the international publishing trade, printing books from 24 presses and distributing them at trade fairs all over western and central Europe. The old monastic scriptoria-monastery workshops where monks copied texts by hand-worked closely alongside the new presses, continuing to produce the luxury goods that movable-type printing could not yet supply. Printing, however, was primarily aimed at a cheap mass sale.

Although there was no competition between the technologies, there was rivalry between nations. The Italians made energetic and successful efforts to catch up with Germany. Their most successful scriptorium quickly imported two leading German printers to set up presses in their book-producing shop. German printers had the disadvantage of working with the complex typeface that the Italians sneeringly referred to as "Gothic" and that later became known as black letter. Outside Germany, readers found this typeface disagreeable. The Italians, on the other hand, had a clear typeface known as roman that became the type of the future.

Hence, although the Germans made use of the paper revolution to introduce movable type, the Italians went far to regain the initiative by their artistry. By 1500 there were printing firms in 60 German cities, but there were 150 presses in Venice alone. However, since many nations and governments wanted their own presses, the trade quickly became international. The cumulative impact of this industrial spread was spectacular. Before printing, only the very largest libraries, of which there were a dozen in Europe, had as many as 600 books. The total number of books on the entire Continent was well under 100,000. But by 1500, after only 45 years of the printed book, there were 9 million in circulation.

53 – 1 Evidence of the Earliest Writing

Although literacy appeared independently in several parts of the prehistoric world, the earliest evidence of writing is the cuneiform Sumerian script on the clay tablets of ancient Mesopotamia, which, archaeological detective work has revealed, had its origins in the accounting practices of commercial activity. Researchers demonstrated that preliterate people, to keep track of the goods they produced and exchanged, created a system of accounting using clay tokens as symbolic representations of their products. Over many thousands of years, the symbols evolved through several stages of abstraction until they became wedge-shaped (cuneiform) signs on clay tablets, recognizable as writing.

The original tokens (circa 8500 B.C.E.) were three-dimensional solid shapes—tiny spheres, cones, disks, and cylinders. A debt of six units of grain and eight head of livestock, for example might have been represented by six conical and eight cylindrical tokens. To keep batches of tokens together, an innovation was introduced (circa 3250 B. C. E.) whereby they were sealed inside clay envelopes that could be broken open and counted when it came time for a debt to be repaid. But because the contents of the envelopes could easily be forgotten, two-dimensional representations of the three-dimensional tokens were impressed into the surface of the envelopes before they were sealed. Eventually, having two sets of equivalent symbols—the internal tokens and external markings—came to seem redundant, so the tokens were eliminated (circa 3250-3100 B.C.E.), and only solid clay tablets with two-dimensional symbols were retained. Over time, the symbols became more numerous, varied, and abstract and came to represent more than trade commodities, evolving eventually into cuneiform writing.

The evolution of the symbolism is reflected in the archaeological record first of all by the increasing complexity of the tokens themselves. The earliest tokens, dating from about 10,000 to 6,000 years ago, were of only the simplest geometric shapes. But about 3500 B.C.E., more complex tokens came into common usage, including many naturalistic forms shaped like miniature tools, furniture, fruit, and humans. The earlier, plain tokens were counters for agricultural products, whereas the complex ones stood for finished products, such as bread, oil, perfume, wool, and rope, and for items produced in workshops, such as metal, bracelets, types of cloth, garments, mats, pieces of furniture, tools, and a variety of stone and pottery vessels. The signs marked on clay tablets likewise evolved from simple

wedges, circles, ovals, and triangles based on the plain tokens to pictographs derived from the complex tokens.

Before this evidence came to light, the inventors of writing were assumed by researchers to have been an intellectual elite. Some, for example, hypothesized that writing emerged when members of the priestly caste agreed among themselves on written signs. But the association of the plain tokens with the first farmers and of the complex tokens with the first artisans—and the fact that the token-and-envelope accounting system invariably represented only small-scale transactions—testifies to the relatively modest social status of the creators of writing.

And not only of literacy, but numeracy (the representation of quantitative concepts) as well. The evidence of the tokens provides further confirmation that mathematics originated in people's desire to keep records of flocks and other goods. Another immensely significant step occurred around 3100 B.C.E., when Sumerian accountants extended the token-based signs to include the first real numerals. Previously, units of grain had been represented by direct one-to-one correspondence —by repeating the token or symbol for a unit of grain the required number of times. The accountants, however, devised numeral signs distinct from commodity signs, so that eighteen units of grain could be indicated by preceding a single grain symbol with a symbol denoting "18." Their invention of abstract numerals and abstract counting was one of the most revolutionary advances in the history of mathematics.

What was the social status of the anonymous accountants who produced this breakthrough? The immense volume of clay tablets unearthed in the ruins of the Sumerian temples where the accounts were kept suggests a social differentiation within the scribal class, with a virtual army of lower-ranking tabulators performing the monotonous job of tallying commodities. We can only speculate as to how high or low the inventors of true numerals were in the scribal hierarchy, but it stands to reason that this labor-saving innovation would have been the brainchild of the lower-ranking types whose drudgery is eased.

60 – 2 The Revolution of Cheap Print

The first half of the nineteenth century saw dramatic changes in the economics of the printed word in both the United States and Europe, though the changes generally happened earlier and on a wider basis in America. In the 1830s and 1840s, sharp reductions in prices for newspapers and books in America highlighted the advent of an era of cheap print. Now there were daily newspapers that instead of 6 cents per copy sold for a penny or two. Now there were novels that instead of an earlier price of \$2 sold for 25 cents or less, when the same books in Britain cost the equivalent of more than \$7. So steep were the declines in the price of print over so short a period that they amounted to an information-price revolution, the first of several such episodes of declining prices that have profoundly affected information and culture during the past two centuries. Two mid-nineteenth-century American cultural innovations, the "penny press" and the "dime novel," were actually named for their low price. These were criticized for being cheap in both senses of that word low in price and low in taste. But low price did not necessarily mean lowbrow increasingly, book publishers issued even the most esteemed works in cheap as well as expensive editions to reach as wide a public as possible. The information-price revolution also affected religious and political publishing, as reading became a basis of mass persuasion for the first time in history.

Cheap print was not entirely unprecedented. In seventeenth- and eighteenth-century England and France, cheap collections of stories, ballads, and other miscellany had circulated among the lower classes. But since only a minority of the poor could read, most listened while a few read aloud; thus cheap print reached not so much a reading as a listening public. The expansion of cheap print in the nineteenth century in America and Europe was on a much larger scale, and it took place during a great increase in popular literacy. Together these amounted to a cultural watershed. Traditionally, even in literate homes, books and other publications had been relatively rare and treasured objects; reading meant returning to a few texts, especially religious works. But with the explosion of print, reading became more varied, and readers scanned newspapers, magazines, and cheap books that they soon passed on or discarded. Intensive reading of religious and other works did

not disappear, but reading became an increasingly common form of diversion as well as devotion.

The usual explanation for the rise of cheap print emphasizes new technology. Unquestionably, the full development of cheap print could not have happened without technological change. Print, however, had already become cheaper in America before technological advances played a significant role; new technology arrived once the process was under way, not at the beginning. This was no accident: the continuing expansion of print created an incentive for technological innovation. To conceive of technology as the causal force is to understate the prior importance of politics, culture, and markets in creating the conditions that allowed investments in new technology to pay off.

Cheap print was public policy in America. While European governments taxed newspapers and other publications, the United States let them go tax free and even subsidized them, to a degree, through the postal system. The rise of cheap books and other forms of cheap print in the United States also reflected distinctive patterns of nineteenth-century American consumer markets. As the economic historian Nathan Rosenberg remarks, citing the cases of cutlery, guns, boots, and clothing, "Americans readily accepted products which had been deliberately designed for low cost, mass production methods" at a time when handmade goods persisted in Britain. Books fit this pattern. Americans had not been primarily responsible for introducing new manufacturing technology to the production of books. On the contrary, most of the key advances in printing and papermaking before 1850 had traveled west across the Atlantic rather than the reverse. But the industrialization of book production proceeded more rapidly in the United States, where the market by the middle decades of the century was not only larger than in Britain but also apparently more sensitive to price than to quality, perhaps because elite readers constituted a smaller proportion of book buyers.

HISTORY / SOCIETIES / ECONOMY / BIOGRAPHY / ANCIENT SOCIETIES

07 – 2 Ancient Rome and Greece

There is a quality of cohesiveness about the Roman world that applied neither to Greece nor perhaps to any other civilization, ancient or modern. Like the stone of Roman wall, which were held together both by the regularity of the design and by that peculiarly powerful Roman cement, so the various parts of the Roman realm were bonded into a massive, monolithic entity by physical, organizational, and psychological controls. The physical bonds included the network of military garrisons, which were stationed in every province, and the network of stone-built roads that linked the provinces with Rome. The organizational bonds were based on the common principles of law and administration and on the universal army of officials who enforced common standards of conduct. The psychological controls were built on fear and punishment—on the absolute certainty that anyone or anything that threatened the authority of Rome would be utterly destroyed.

The source of Roman obsession with unity and cohesion may well have lain in the pattern of Rome's early development. Whereas Greece had grown from scores of scattered cities, Rome grew from one single organism. While the Greek world had expanded along the Mediterranean seas lanes, the Roman world was assembled by territorial conquest. Of course, the contrast is not quite so stark: in Alexander the Great the Greeks had found the greatest territorial conqueror of all time; and the Romans, once they moved outside Italy, did not fail to learn the lessons of sea power. Yet the essential difference is undeniable. The key to the Greek world lay in its high-powered ships; the key to Roman power lay in its marching legions. The Greeks were wedded to the sea; the Romans, to the land. The Greek was a sailor at heart; the Roman, a landsman.

Certainly, in trying to explain the Roman phenomenon, one would have to place great emphasis on this almost instinct for the territorial imperative. Roman priorities lay in the organization, exploitation, and defense of their territory. In all probability it was the fertile plain of Latium, where the Latins who founded Rome originated, that created the habits and skills of landed settlement, landed property, landed economy, landed administration, and a land-based society. From this

arose the Roman genius for military organization and orderly government. In turn, a deep attachment to the land, and to the stability which rural life engenders, fostered the Roman virtues: *gravitas*, a sense of responsibility, *pietas*, a sense of devotion to family and country, and *iustitia*, a sense of the natural order.

Modern attitudes to Roman civilization range from the infinitely impressed to the thoroughly disgusted. As always, there are the power worshippers, especially among historians, who are predisposed to admire whatever is strong, who feel more attracted to the might of Rome than to the subtlety of Greece. At the same time, there is a solid body of opinion that dislikes Rome. For many, Rome is at best the imitator and the continuator of Greece on a larger scale. Greek civilization had quality; Rome, mere quantity. Greece was original; Rome, derivative. Greece had style; Rome had money. Greece was the inventor; Rome, the research and development division. Such indeed was the opinion of some of the more intellectual Romans. "Had the Greeks held novelty in such disdain as we," asked Horace in his epistle, "what work of ancient date would now exist?"

Rome's debt to Greece was enormous. The Romans adopted Greek religion and moral philosophy. In literature, Greek writers were consciously used as models by their Latin successors. It was absolutely accepted that an educated Roman should be fluent in Greek. In speculative philosophy and the sciences, the Romans made virtually no advance on early achievements.

Yet it would be wrong to suggest that Rome was somehow a junior partner in Greco-Roman civilization. The Roman genius was projected into new spheres—especially into those of law, military organization, administration, and engineering. Moreover, the tensions that arose within the Roman state produced literary and artistic sensibilities of the highest order. It was no accident that many leading Roman soldiers and statesmen were writers of high caliber.

07 – 3 Agriculture, Iron, and the Bantu Peoples

There is evidence of agriculture in Africa prior to 3000 B.C.. It may have developed independently, but many scholars believe that the spread of agriculture and iron throughout Africa linked it to the major centers of the Near East and Mediterranean world. The drying up of what is now the Sahara desert had pushed many peoples to the south into sub-Saharan Africa. These peoples settled at first in scattered hunting-and-gathering bands, although in some places near lakes and rivers, people who fished, with a more secure food supply, lived in larger population concentrations. Agriculture seems to have reached these people from the Near East, since the first domesticated crops were millets and sorghums whose origins are not African but west Asian. Once the idea of planting diffused, Africans began to develop their own crops, such as certain varieties of rice, and they demonstrated a continued receptiveness to new imports. The proposed areas of the domestication of African crops lie in a band that extends from Ethiopia across southern Sudan to West Africa. Subsequently, other crops, such as bananas, were introduced from Southeast Asia.

Livestock also came from outside Africa. Cattle were introduced from Asia, as probably were domestic sheep and goats. Horses were apparently introduced by the Hyksos invaders of Egypt (1780-1560 B.C.) and then spread across the Sudan to West Africa. Rock paintings in the Sahara indicate that horses and chariots were used to traverse the desert and that by 300-200 B.C., there were trade routes across the Sahara. Horses were adopted by peoples of the West African savannah, and later their powerful cavalry forces allowed them to carve out large empires. Finally, the camel was introduced around the first century A.D. This was an important innovation, because the camel's abilities to thrive in harsh desert conditions and to carry large loads cheaply made it an effective and efficient means of transportation. The camel transformed the desert from a barrier into a still difficult, but more accessible, route of trade and communication.

Iron came from West Asia, although its routes of diffusion were somewhat different than those of agriculture. Most of Africa presents a curious case in which societies moved directly from a technology of stone to iron without passing through the intermediate stage of copper or bronze metallurgy, although some early copper-working sites have

been found in West Africa. Knowledge of iron making penetrated into the forest and savannahs of West Africa at roughly the same time that iron making was reaching Europe. Evidence of iron making has been found in Nigeria, Ghana, and Mali.

This technological shift caused profound changes in the complexity of African societies. Iron represented power. In West Africa the blacksmith who made tools and weapons had an important place in society, often with special religious powers and functions. Iron hoes, which made the land more productive, and iron weapons, which made the warrior more powerful, had symbolic meaning in a number of West African societies. Those who knew the secrets of making iron gained ritual and sometimes political power.

Unlike in the Americas, where metallurgy was a very late and limited development, Africans had iron from a relatively early date, developing ingenious furnaces to produce the high heat needed for production and to control the amount of air that reached the carbon and iron ore necessary for making iron. Much of Africa moved right into the Iron Age, taking the basic technology and adapting it to local conditions and resources.

The diffusion of agriculture and later of iron was accompanied by a great movement of people who may have carried these innovations. These people probably originated in eastern Nigeria. Their migration may have been set in motion by an increase in population caused by a movement of peoples fleeing the desiccation, or drying up, of the Sahara. They spoke a language, proto-Bantu ("Bantu" means "the people"), which is the parent tongue of a language of a large number of Bantu languages still spoken throughout sub-Saharan Africa. Why and how these people spread out into central and southern Africa remains a mystery, but archaeologists believe that their iron weapons allowed them to conquer their hunting-gathering opponents, who still used stone implements. Still, the process is uncertain, and peaceful migration—or simply rapid demographic growth—may have also caused the Bantu explosion.

08 – 1 The Rise of Teotihuacán

The city of Teotihuacán, which lay about 50 kilometers northeast of modern-day Mexico City, began its growth by 200-100 B.C. At its height, between about A.D. 150 and 700, it probably had a population of more than 125,000 people and covered at least 20 square kilometers. It had over 2,000 apartment complexes, a great market, a large number of industrial workshops, an administrative center, a number of massive religious edifices, and a regular grid pattern of streets and buildings. Clearly, much planning and central control were involved in the expansion and ordering of this great metropolis. Moreover, the city had economic and perhaps religious contacts with most parts of Mesoamerica (modern Central America and Mexico).

How did this tremendous development take place, and why did it happen in the Teotihuacán Valley? Among the main factors are Teotihuacán's geographic location on a natural trade route to the south and east of the Valley of Mexico, the obsidian resources in the Teotihuacán Valley itself, and the valley's potential for extensive irrigation. The exact role of other factors is much more difficult to pinpoint—for instance, Teotihuacán's religious significance as a shrine, the historical situation in and around the Valley of Mexico toward the end of the first millennium B.C., the ingenuity and foresightedness of Teotihuacán's elite, and, finally, the impact of natural disasters, such as the volcanic eruptions of the late first millennium B.C.

This last factor is at least circumstantially implicated in Teotihuacán's rise. Prior to 200 B.C., a number of relatively small centers coexisted in and near the Valley of Mexico. Around this time, the largest of these centers, Cuicuilco, was seriously affected by a volcanic eruption, with much of its agricultural land covered by lava. With Cuicuilco eliminated as a potential rival, any one of a number of relatively modest towns might have emerged as a leading economic and political power in Central Mexico. The archaeological evidence clearly indicates, though, that Teotihuacán was the center that did arise as the predominant force in the area by the first century A.D.

It seems likely that Teotihuacán's natural resources, along with the city elite's ability to recognize their potential, gave the city a competitive edge over its neighbors. The valley, like many other places in Mexican and Guatemalan

highlands, was rich in obsidian. The hard volcanic stone was a resource that had been in great demand for many years, at least since the rise of the Olmecs (a people who flourished between 1200 and 400 B.C.), and it apparently had a secure market. Moreover, recent research on obsidian tools found at Olmec sites has shown that some of the obsidian obtained by the Olmecs originated near Teotihuacán. Teotihuacán obsidian must have been recognized as a valuable commodity for many centuries before the great city arose.

Long-distance trade in obsidian probably gave the elite residents of Teotihuacán access to a wide variety of exotic good, as well as a relatively prosperous life. Such success may have attracted immigrants to Teotihuacán. In addition, Teotihuacán's elite may have consciously attempted to attract new inhabitants. It is also probable that as early as 200 B.C. Teotihuacán may have achieved some religious significance and its shrine (or shrines) may have served as an additional population magnet. Finally, the growing population was probably fed by increasing the number and size of irrigated fields.

The picture of Teotihuacán that emerges is a classic picture of positive feedback among obsidian mining and working, trade, population growth, irrigation, and religious tourism. The thriving obsidian operation, for example, would necessitate more miners, additional manufacturers of obsidian tools, and additional traders to carry the goods to new markets. All this led to increased wealth, which in turn would attract more immigrants to Teotihuacán. The growing power of the elite, who controlled the economy, would give them the means to physically coerce people to move to Teotihuacán and serve as additions to the labor force. More irrigation works would have to be built to feed the growing population, and this resulted in more power and wealth for the elite.

09 – 1 Colonizing the Americas via the Northwest Coast

It has long been accepted that the Americas were colonized by a migration of peoples from Asia, slowly traveling across a land bridge called Beringia (now the Bering Strait between northeastern Asia and Alaska) during the last Ice Age. The first water craft theory about this migration was that around 11,000-12,000 years ago there was an ice-free corridor stretching from eastern Beringia to the areas of North America south of the great northern glaciers. It was this midcontinental corridor between two massive ice sheets – the Laurentide to the east and the Cordilleran to the west – that enabled the southward migration. But belief in this ice-free corridor began to crumble when paleoecologist Glen MacDonald demonstrated that some of the most important radiocarbon dates used to support the existence of an ice-free corridor were incorrect. He persuasively argued that such an ice-free corridor did not exist until much later, when the continental ice began its final retreat.

Support is growing for the alternative theory that people using watercraft, possibly skin boats, moved southward from Beringia along the Gulf of Alaska and then southward along the Northwest coast of North America possibly as early as 16,000 years ago. This route would have enabled humans to enter southern areas of the Americas prior to the melting of the continental glaciers. Until the early 1970s, most archaeologists did not consider the coast a possible migration route into the Americas because geologists originally believed that during the last Ice Age the entire Northwest Coast was covered by glacial ice. It had been assumed that the ice extended westward from the Alaskan/Canadian Mountains to the very edge of the continental shelf, the flat, submerged part of the continent that extends into the ocean. This would have created a barrier of ice extending from the Alaska Peninsula, through the Gulf of Alaska and southward along the Northwest Coast of north America to what is today the state of Washington.

The most influential proponent of the coastal migration route has been Canadian archaeologist Knut Fladmark. He theorized that with the use of watercraft, people gradually colonized unglaciated refuges and areas along the continental shelf exposed by the lower sea level. Fladmark's hypothesis received additional support from the fact that the greatest diversity in native American

languages occurs along the west coast of the Americas, suggesting that this region has been settled the longest.

More recent geologic studies documented deglaciation and the existence of ice-free areas throughout major coastal areas of British Columbia, Canada, by 13,000 years ago. Research now indicates that sizable areas of southeastern Alaska along the inner continental shelf were not covered by ice toward the end of the last Ice Age. One study suggests that except for a 250-mile coastal area between southwestern British Columbia and Washington State, the Northwest Coast of North America was largely free of ice by approximately 16,000 years ago. Vast areas along the coast may have been deglaciated beginning around 16,000 years ago, possibly providing a coastal corridor for the movement of plants, animals, and humans sometime between 13,000 and 14,000 years ago.

The coastal hypothesis has gained increasing support in recent years because the remains of large land animals, such as caribou and brown bears, have been found in southeastern Alaska dating between 10,000 and 12,500 years ago. This is the time period in which most scientists formerly believed the area to be inhospitable for humans. It has been suggested that if the environment were capable of supporting breeding populations of bears, there would have been enough food resources to support humans. Fladmark and others believe that the first human colonization of America occurred by boat along the Northwest Coast during the very late Ice Age, possibly as early as 14,000 years ago. The most recent geologic evidence indicates that it may have been possible for people to colonize ice-free regions along the continental shelf that were still exposed by the lower sea level between 13,000 and 14,000 years ago.

The coastal hypothesis suggests an economy based on marine mammal hunting, saltwater fishing, shellfish gathering, and the use of watercraft. Because of the barrier of ice to the east, the Pacific Ocean to the west, and populated areas to the north, there may have been a greater impetus for people to move in a southerly direction.

10 – 3 Seventeenth-Century European Economic Growth

In the late sixteenth century and into the seventeenth, Europe continued the growth that had lifted it out of the relatively less prosperous medieval period (from the mid 400s to the late 1400s). Among the key factors behind this growth were increased agricultural productivity and an expansion of trade.

Populations cannot grow unless the rural economy can produce enough additional food to feed more people. During the sixteenth century, farmers brought more land into cultivation at the expense of forests and fens (low-lying wetlands). Dutch land reclamation in the Netherlands in the sixteenth and seventeenth centuries provides the most spectacular example of the expansion of farmland: the Dutch reclaimed more than 36,000 acres from 1590 to 1615 alone.

Much of the potential for European economic development lay in what at first glance would seem to have been only sleepy villages. Such villages, however, generally lay in regions of relatively advanced agricultural production, permitting not only the survival of peasants but also the accumulation of an agricultural surplus for investment. They had access to urban merchants, markets, and trade routes.

Increased agricultural production in turn facilitated rural industry, an intrinsic part of the expansion of industry. Woolens and textile manufacturers, in particular, utilized rural cottage (in-home) production, which took advantage of cheap and plentiful rural labor. In the German states, the ravages of the Thirty Years' War (1618-1648) further moved textile production into the countryside. Members of poor peasant families spun or wove cloth and linens at home for scant remuneration in an attempt to supplement meager family income.

More extended trading networks also helped develop Europe's economy in this period. English and Dutch ships carrying rye from the Baltic states reached Spain and Portugal. Population growth generated an expansion of small-scale manufacturing, particularly of handicrafts, textiles, and metal production in England, Flanders, parts of northern Italy, the southwestern German states, and parts of Spain. Only iron smelting and mining required marshaling a significant amount of capital (wealth invested to create more wealth).

The development of banking and other financial services contributed to the expansion of trade. By the middle of the sixteenth century, financiers and traders commonly accepted bills of exchange in place of gold or silver for other goods. Bills of exchange, which had their origins in medieval Italy, were promissory notes (written promises to pay a specified amount of money by a certain date) that could be sold to third parties. In this way, they provided credit. At mid-century, an Antwerp financier only slightly exaggerated when he claimed, "One can no more trade without bills of exchange than sail without water." Merchants no longer had to carry gold and silver over long, dangerous journeys. An Amsterdam merchant purchasing soap from a merchant in Marseille could go to an exchanger and pay the exchanger the equivalent sum in guilders, the Dutch currency. The exchanger would then send a bill of exchange to a colleague in Marseille, authorizing the colleague to pay the Marseille merchant in the merchant's own currency after the actual exchange of goods had taken place.

Bills of exchange contributed to the development of banks, as exchangers began to provide loans. Not until the eighteenth century, however, did such banks as the Bank of Amsterdam and the Bank of England begin to provide capital for business investment. Their principal function was to provide funds for the state.

The rapid expansion in international trade also benefitted from an infusion of capital, stemming largely from gold and silver brought by Spanish vessels from the Americas. This capital financed the production of goods, storage, trade, and even credit across Europe and overseas. Moreover, an increased credit supply was generated by investments and loans by bankers and wealthy merchants to states and by joint-stock partnerships—an English innovation (the first major company began in 1600). Unlike short-term financial cooperation between investors for a single commercial undertaking, joint-stock companies provided permanent funding of capital by drawing on the investments of merchants and other investors who purchased shares in the company.

14 – 2 Maya Water Problems

To understand the ancient Mayan people who lived in the area that is today southern Mexico and Central America and the ecological difficulties they faced, one must first consider their environment, which we think of as “jungle” or “tropical rainforest”. This view is inaccurate, and the reason proves to be important. Properly speaking, tropical rainforests grow in high-rainfall equatorial areas that remain wet or humid all year round. But the Maya homeland lies more than sixteen hundred kilometers from the equator, at latitudes 17 to 22 degrees north, in a habitat termed a “seasonal tropical forest”. That is, while there does tend to be a rainy season from May to October, there is also a dry season from January through April. If one focuses on the wet months, one calls the Maya homeland a “seasonal tropical forest”; if one focuses on the dry months, one could instead describe it as a “seasonal desert”.

From north to south in the Yucatan Peninsula, where the Maya lived, rainfall ranges from 18 to 100 inches (457 to 2,540 millimeters) per year, and the soils become thicker, so that the southern peninsula was agriculturally more productive and supported denser populations. But rainfall in the Maya homeland is unpredictably variable between years; some recent years have had three or four times more rain than other years. As a result, modern farmers attempting to grow corn in the ancient Maya homelands have faced frequent crop failures, especially in the north. The ancient Maya were presumably more experienced and did better, but nevertheless they too must have faced risks of crop failures from droughts and hurricanes.

Although southern Maya areas received more rainfall than northern areas, problems of water were paradoxically more severe in the wet south. While that made things hard for ancient Maya living in the south, it has also made things hard for modern archaeologists who have difficulty understanding why ancient droughts caused bigger problems in the wet south than in the dry north. The likely explanation is that an area of underground freshwater underlies the Yucatan Peninsula, but surface elevation increases from north to south, so that as one moves south the land surface lies increasingly higher above the water table. In the northern peninsula the elevation is sufficiently low that the ancient Maya were able to reach the water table at deep sinkholes called cenotes, or at deep caves. In low-elevation north

coastal areas without sinkholes, the Maya would have been able to get down to the water table by digging wells up to 75 feet (22 meters) deep. But much of the south lies too high above the water table for cenotes or wells to reach down to it. Making matters worse, most of the Yucatan Peninsula consists of karst, a porous sponge-like limestone terrain where rain runs straight into the ground and where little or no surface water remains available.

How did those dense southern Maya populations deal with the resulting water problem? It initially surprises us that many of their cities were not built next to the rivers but instead on high terrain in rolling uplands. The explanation is that the Maya excavated depressions, or modified natural depressions, and then plugged up leaks in the karst by plastering the bottoms of the depressions in order to create reservoirs, which collected rain from large plastered catchment basins and stored it for use in the dry season. For example, reservoirs at the Maya city of Tikal held enough water to meet the drinking water needs of about 10,000 people for a period of 18 months. At the city of Coba the Maya built dikes around a lake in order to raise its level and make their water supply more reliable. But the inhabitants of Tikal and other cities dependent on reservoirs for drinking water would still have been in deep trouble if 18 months passed without rain in a prolonged drought. A shorter drought in which they exhausted their stored food supplies might already have gotten them in deep trouble, because growing crops required rain rather than reservoirs.

14 – 3 Pastoralism in Ancient Inner Eurasia

Pastoralism is a lifestyle in which economic activity is based primarily on livestock. Archaeological evidence suggests that by 3000 B.C., and perhaps even earlier, there had emerged on the steppes of Inner Eurasia the distinctive types of pastoralism that were to dominate the region's history for several millennia. Here, the horse was already becoming the animal of prestige in many regions, though sheep, goats, and cattle could also play a vital role. It is the use of horses for transportation and warfare that explains why Inner Eurasian pastoralism proved the most mobile and the most militaristic of all major forms of pastoralism. The emergence and spread of pastoralism had a profound impact on the history of Inner Eurasia, and also, indirectly, on the parts of Asia and Europe just outside this area. In particular, pastoralism favors a mobile lifestyle, and this mobility helps to explain the impact of pastoralist societies on this part of the world.

The mobility of pastoralist societies reflects their dependence on animal-based foods. While agriculturalists rely on domesticated plants, pastoralists rely on domesticated animals. As a result, pastoralists, like carnivores in general, occupy a higher position on the food chain. All else being equal, this means they must exploit larger areas of land than do agriculturalists to secure the same amount of food, clothing, and other necessities. So pastoralism is a more extensive lifeway than farming is. However, the larger the terrain used to support a group, the harder it is to exploit that terrain while remaining in one place. So, basic ecological principles imply a strong tendency within pastoralist lifeways toward nomadism (a mobile lifestyle). As the archaeologist Roger Cribb puts it, "The greater the degree of pastoralism, the stronger the tendency toward nomadism." A modern Turkic nomad interviewed by Cribb commented: "The more animals you have, the farther you have to move."

Nomadism has further consequences. It means that pastoralist societies occupy and can influence very large territories. This is particularly true of the horse pastoralism that emerged in the Inner Eurasian steppes, for this was the most mobile of all major forms of pastoralism. So, it is no accident that with the appearance of pastoralist societies there appear large areas that share similar cultural, ecological, and even linguistic features. By the late fourth millennium B.C., there is already evidence of

large culture zones reaching from Eastern Europe to the western borders of Mongolia. Perhaps the most striking sign of mobility is the fact that by the third millennium B.C., most pastoralists in this huge region spoke related languages ancestral to the modern Indo-European languages. The remarkable mobility and range of pastoral societies explain, in part, why so many linguists have argued that the Indo-European languages began their astonishing expansionist career not among farmers in Anatolia (present-day Turkey), but among early pastoralists from Inner Eurasia. Such theories imply that the Indo-European languages evolved not in Neolithic (10,000 to 3,000 B.C.) Anatolia, but among the foraging communities of the cultures in the region of the Don and Dnieper rivers, which took up stock breeding and began to exploit the neighboring steppes.

Nomadism also subjects pastoralist communities to strict rules of portability. If you are constantly on the move, you cannot afford to accumulate large material surpluses. Such rules limit variations in accumulated material goods between pastoralist households (though they may also encourage a taste for portable goods of high value such as silks or jewelry). So, by and large, nomadism implies a high degree of self-sufficiency and inhibits the appearance of an extensive division of labor. Inequalities of wealth and rank certainly exist, and have probably existed in most pastoralist societies, but except in periods of military conquest, they are normally too slight to generate the stable, hereditary hierarchies that are usually implied by the use of the term class. Inequalities of gender have also existed in pastoralist societies, but they seem to have been softened by the absence of steep hierarchies of wealth in most communities, and also by the requirement that women acquire most of the skills of men, including, often, their military skills.

05 – 2 The Origin of the Pacific Island People

The greater Pacific region, traditionally called Oceania, consists of three cultural areas: Melanesia, Micronesia, and Polynesia. Melanesia, in the southwest Pacific, contains the large islands of New Guinea, the Solomons, Vanuatu, and New Caledonia. Micronesia, the area north of Melanesia, consists primarily of small scattered islands. Polynesia is the central Pacific area in the great triangle defined by Hawaii, Easter Island, and New Zealand. Before the arrival of Europeans, the islands in the two largest cultural areas, Polynesia and Micronesia, together contained a population estimated at 700,000.

Speculation on the origin of these Pacific islanders began as soon as outsiders encountered them, in the absence of solid linguistic, archaeological, and biological data, many fanciful and mutually exclusive theories were devised. Pacific islanders are variously thought to have come from North America, South America, Egypt, Israel, and India, as well as Southeast Asia. Many older theories implicitly deprecated the navigational abilities and overall cultural creativity of the Pacific islanders. For example, British anthropologists G. Elliot Smith and W. J. Perry assumed that only Egyptians would have been skilled enough to navigate and colonize the Pacific. They inferred that the Egyptians even crossed the Pacific to found the great civilizations of the New World (North and South America). In 1947 Norwegian adventurer Thor Heyerdahl drifted on a balsa-log raft westward with the winds and currents across the Pacific from South America to prove his theory that Pacific islanders were Native Americans (also called American Indians). Later Heyerdahl suggested that the Pacific was peopled by three migrations: by Native Americans from the Pacific Northwest of North America drifting to Hawaii, by Peruvians drifting to Easter Island, and by Melanesians. In 1969 he crossed the Atlantic in an Egyptian-style reed boat to prove Egyptian influences in the Americas. Contrary to these theorists, the overwhelming evidence of physical anthropology, linguistics, and archaeology shows that the Pacific islanders came from Southeast Asia and were skilled enough as navigators to sail against the prevailing winds and currents.

The basic cultural requirements for the successful colonization of the Pacific islands include the appropriate boat-building, sailing, and navigation skills to get to the islands in the first place, domesticated plants and gardening skills suited to

often marginal conditions, and a varied inventory of fishing implements and techniques. It is now generally believed that these prerequisites originated with peoples speaking Austronesian languages (a group of several hundred related languages) and began to emerge in Southeast Asia by about 5,000 B.C.E. The culture of that time, based on archaeology and linguistic reconstruction, is assumed to have had a broad inventory of cultivated plants including taro, yams, banana, sugarcane, breadfruit, coconut, sago, and rice. Just as important, the culture also possessed the basic foundation for an effective maritime adaptation, including outrigger canoes and a variety of fishing techniques that could be effective for overseas voyaging.

Contrary to the arguments of some that much of the Pacific was settled by Polynesians accidentally marooned after being lost and adrift, it seems reasonable that this feat was accomplished by deliberate colonization expeditions that set out fully stocked with food and domesticated plants and animals. Detailed studies of the winds and currents using computer simulations suggest that drifting canoes would have been a most unlikely means of colonizing the Pacific. These expeditions were likely driven by population growth and political dynamics on the home islands, as well as the challenge and excitement of exploring unknown waters. Because all Polynesians, Micronesians, and many Melanesians speak Austronesian languages and grow crops derived from Southeast Asia, all these peoples most certainly derived from that region and not the New World or elsewhere. The undisputed pre-Columbian presence in Oceania of the sweet potato, which is a New World domesticate, has sometimes been used to support Heyerdahl's "American Indians in the Pacific" theories. However, this is one plant out of a long list of Southeast Asian domesticates. As Patrick Kirch, an American anthropologist, points out, rather than being brought by rafting South Americans, sweet potatoes might just have easily been brought back by returning Polynesian navigators who could have reached the west coast of South America.

06 – 2 William Smith

In 1769 in a little town in Oxfordshire, England, a child with the very ordinary name of William Smith was born into the poor family of a village blacksmith. He received rudimentary village schooling, but mostly he roamed his uncle's farm collecting the fossils that were so abundant in the rocks of the Cotswold hills. When he grew older, William Smith taught himself surveying from books he bought with his small savings, and at the age of eighteen he was apprenticed to a surveyor of the local parish. He then proceeded to teach himself geology, and when he was twenty-four, he went to work for the company that was excavating the Somerset Coal Canal in the south of England.

This was before the steam locomotive, and canal building was at its height. The companies building the canals to transport coal needed surveyors to help them find the coal deposits worth mining as well as to determine the best courses for the canals. This job gave Smith an opportunity to study the fresh rock outcrops created by the newly dug canal. He later worked on similar jobs across the length and breadth of England, all the while studying the newly revealed strata and collecting all the fossils he could find. Smith used mail coaches to travel as much as 10,000 miles per year. In 1815 he published the first modern geological map, "A Map of the Strata of England and Wales with a Part of Scotland," a map so meticulously researched that it can still be used today.

In 1831 when Smith was finally recognized by the Geological Society of London as the "father of English geology," it was not only for his maps but also for something even more important. Ever since people had begun to catalog the strata in particular outcrops, there had been the hope that these could somehow be used to calculate geological time. But as more and more accumulations of strata were cataloged in more and more places, it became clear that the sequences of rocks sometimes differed from region to region and that no rock type was ever going to become a reliable time marker throughout the world. Even without the problem of regional differences, rocks present a difficulty as unique time markers. Quartz is quartz—a silicon ion surrounded by four oxygen ions—there's no difference at all between two-million-year-old Pleistocene quartz and Cambrian quartz created over 500 million years ago.

As he collected fossils from strata throughout England, Smith began to see that the fossils told a different story from the rocks. Particularly in the younger strata, the rocks were often so similar that he had trouble distinguishing the strata, but he never had trouble telling the fossils apart. While rock between two consistent strata might in one place be shale and in another sandstone, the fossils in that shale or sandstone were always the same. Some fossils endured through so many millions of years that they appear in many strata, but others occur only in a few strata, and a few species had their births and extinctions within one particular stratum. Fossils are thus identifying markers for particular periods in Earth's history.

Not only could Smith identify rock strata by the fossils they contained, he could also see a pattern emerging: certain fossils always appear in more ancient sediments, while others begin to be seen as the strata become more recent. By following the fossils, Smith was able to put all the strata of England's earth into relative temporal sequence. About the same time, Georges Cuvier made the same discovery while studying the rocks around Paris. Soon it was realized that this principle of faunal (animal) succession was valid not only in England or France but virtually everywhere. It was actually a principle of floral succession as well, because plants showed the same transformation through time as did fauna. Limestone may be found in the Cambrian or—300 million years later—in the Jurassic strata, but a trilobite—the ubiquitous marine arthropod that had its birth in the Cambrian—will never be found in Jurassic strata, nor a dinosaur in the Cambrian.

16 – 1 Trade and the Ancient Middle East

Trade was the mainstay of the urban economy in the Middle East, as caravans negotiated the surrounding desert, restricted only by access to water and by mountain ranges. This has been so since ancient times, partly due to the geology of the area, which is mostly limestone and sandstone, with few deposits of metallic ore and other useful materials. Ancient demands for obsidian (a black volcanic rock useful for making mirrors and tools) led to trade with Armenia to the north, while jade for cutting tools was brought from Turkistan, and the precious stone lapis lazuli was imported from Afghanistan. One can trace such expeditions back to ancient Sumeria, the earliest known Middle Eastern civilization. Records show merchant caravans and trading posts were set up by the Sumerians in the surrounding mountains and deserts of Persia and Arabia, where they traded grain for raw materials, such as timber and stones, as well as for metals and gems.

Reliance on trade had several important consequences. Production was generally in the hands of skilled individual artisans doing piecework under the tutelage of a master who was also the shop owner. In these shops differences of rank were blurred as artisans and masters labored side by side in the same modest establishment, were usually members of the same guild and religious sect, lived in the same neighborhoods, and often had assumed (or real) kinship relationships. The worker was bound to the master by a mutual contract that either one could repudiate, and the relationship was conceptualized as one of partnership.

This mode of craft production favored the growth of self-governing and ideologically egalitarian craft guilds everywhere in the Middle Eastern city. These were essentially professional associations that provided for the mutual aid and protection of their members, and allowed for the maintenance of professional standards. The growth of independent guilds was furthered by the fact that surplus was not a result of domestic craft production but resulted primarily from international trading; the government left working people to govern themselves, much as shepherds of tribal confederacies were left alone by their leaders. In the multiplicity of small-scale local egalitarian or quasi-egalitarian organizations for fellowship, worship, and production that flourished in this laissez-faire environment, individuals could interact

with one another within a community of harmony and ideological equality, following their own popularly elected leaders and governing themselves by shared consensus while minimizing distinctions of wealth and power.

The mercantile economy was also characterized by a peculiar moral stance that is typical of people who live by trade—an attitude that is individualistic, calculating, risk taking, and adaptive to circumstances. As among tribespeople, personal relationships and a careful weighing of character have always been crucial in a mercantile economy with little regulation, where one's word is one's bond and where informal ties of trust cement together an international trade network. Nor have merchants and artisans ever had much tolerance for aristocratic professions of moral superiority, favoring instead an egalitarian ethic of the open market, where steady hard work, the loyalty of one's fellows, and entrepreneurial skill make all the difference. And, like the pastoralists, Middle Eastern merchants and artisans unhappy with their environment could simply pack up and leave for greener pastures—an act of self-assertion wholly impossible in most other civilizations throughout history.

Dependence on long-distance trade also meant that the great empires of the Middle East were built both literally and figuratively on shifting sand. The central state, though often very rich and very populous, was intrinsically fragile, since the development of new international trade routes could undermine the monetary base and erode state power, as occurred when European seafarers circumvented Middle Eastern merchants after Vasco da Gama's voyage around Africa in the late fifteenth century opened up a southern route. The ecology of the region also permitted armed predators to prowl the surrounding barrens, which were almost impossible for a state to control. Peripheral peoples therefore had a great advantage in their dealings with the center, making government authority insecure and anxious.

17 – 1 Europe's Early Sea Trade with Asia

In the fourteenth century, a number of political developments cut Europe's overland trade routes to southern and eastern Asia, with which Europe had had important and highly profitable commercial ties since the twelfth century. This development, coming as it did when the bottom had fallen out of the European economy, provided an impetus to a long-held desire to secure direct relations with the East by establishing a sea trade. Widely reported, if somewhat distrusted, accounts by figures like the famous traveler from Venice, Marco Polo, of the willingness of people in China to trade with Europeans and of the immensity of the wealth to be gained by such contact made the idea irresistible. Possibilities for trade seemed promising, but no hope existed for maintaining the traditional routes over land. A new way had to be found.

The chief problem was technological: How were the Europeans to reach the East? Europe's maritime tradition had developed in the context of easily navigable seas—the Mediterranean, the Baltic, and, to a lesser extent, the North Sea between England and the Continent—not of vast oceans. New types of ships were needed, new methods of finding one's way, new techniques for financing so vast a scheme. The sheer scale of the investment it took to begin commercial expansion at sea reflects the immensity of the profits that such East-West trade could create. Spices were the most sought-after commodities. Spices not only dramatically improved the taste of the European diet but also were used to manufacture perfumes and certain medicines. But even high-priced commodities like spices had to be transported in large bulk in order to justify the expense and trouble of sailing around the African continent all the way to India and China.

The principal seagoing ship used throughout the Middle Ages was the galley, a long, low ship fitted with sails but driven primarily by oars. The largest galleys had as many as 50 oarsmen. Since they had relatively shallow hulls, they were unstable when driven by sail or when on rough water: hence they were unsuitable for the voyage to the East. Even if they hugged the African coastline, they had little chance of surviving a crossing of the Indian Ocean. Shortly after 1400, shipbuilders began developing a new type of vessel properly designed to operate in rough, open water: the caravel. It had a wider and deeper hull than the galley and hence could carry more cargo. Increased stability made it possible to add multiple masts and sails. In the

largest caravels, two main masts held large square sails that provided the bulk of the thrust driving the ship forward, while a smaller forward mast held a triangular-shaped sail, called a lateen sail, which could be moved into a variety of positions to maneuver the ship.

The astrolabe had long been the primary instrument for navigation, having been introduced in the eleventh century. It operated by measuring the height of the Sun and the fixed stars: by calculating the angles created by these points, it determined the degree of latitude at which one stood (the problem of determining longitude, though, was not solved until the eighteenth century). By the early thirteenth century, Western Europeans had also developed and put into use the magnetic compass, which helped when clouds obliterated both the Sun and the stars. Also beginning in the thirteenth century, there were new maps refined by precise calculations and the reports of sailors that made it possible to trace one's path with reasonable accuracy. Certain institutional and practical norms had become established as well. A maritime code known as the Consulate of the Sea, which originated in the western Mediterranean region in the fourteenth century, won acceptance by a majority of sea goers as the normative code for maritime conduct; it defined such matters as the authority of a ship's officers, protocols of command, pay structures, the rights of sailors, and the rules of engagement when ships met one another on the sea-lanes. Thus by about 1400 the key elements were in place to enable Europe to begin its seaward adventure.

19 – 1 The Roman Army's Impact on Britain

In the wake of the Roman Empire's conquest of Britain in the first century A.D., a large number of troops stayed in the new province, and these troops had a considerable impact on Britain with their camps, fortifications, and participation in the local economy. Assessing the impact of the army on the civilian population starts from the realization that the soldiers were always unevenly distributed across the country. Areas rapidly incorporated into the empire were not long affected by the military. Where the army remained stationed, its presence was much more influential. The imposition of a military base involved the requisition of native lands for both the fort and the territory needed to feed and exercise the soldiers' animals. The imposition of military rule also robbed local leaders of opportunities to participate in local government, so social development was stunted and the seeds of disaffection sown. This then meant that the military had to remain to suppress rebellion and organize government.

Economic exchange was clearly very important as the Roman army brought with it very substantial spending power. Locally with respect to a particular place or situation, a fort had two kinds of impact. Its large population needed food and other supplies. Some of these were certainly brought from long distances, but demands were inevitably placed on the local area. Although goods could be requisitioned, they were usually paid for, and this probably stimulated changes in the local economy. When not campaigning, soldiers needed to be occupied; otherwise they represented a potentially dangerous source of friction and disloyalty. Hence a writing tablet dated 25 April tells of 343 men at one fort engaged on tasks like shoemaking, building a bathhouse, operating kilns, digging clay, and working lead. Such activities had a major effect on the local area, in particular with the construction of infrastructure such as roads, which improved access to remote areas.

Each soldier received his pay, but in regions without a developed economy there was initially little on which it could be spent. The pool of excess cash rapidly stimulated a thriving economy outside fort gates. Some of the demand for the services and goods was no doubt fulfilled by people drawn from far afield, but some local people certainly became entwined in this new economy. There was informal marriage with soldiers, who until AD 197 were not legally entitled to wed, and whole new

communities grew up near the forts. These settlements acted like small towns, becoming centers for the artisan and trading populations.

The army also provided a mean of personal advancement for auxiliary soldiers recruited from the native peoples, as a man obtained hereditary Roman citizenship on retirement after service in an auxiliary regiment. Such units recruited on an ad hoc (as needed) basis from the area in which they were stationed, and there was evidently large-scale recruitment within Britain. The total numbers were at least 12,500 men up to the reign of the emperor Hadrian (A.D. 117-138), with a peak around A.D. 80. Although a small proportion of the total population, this perhaps had a massive local impact when a large proportion of the young men were removed from an area. Newly raised regiments were normally transferred to another province from whence it was unlikely that individual recruits would ever return. Most units raised in Britain went elsewhere on the European continent, although one is recorded in Morocco. The reverse process brought young men to Britain, where many continued to live after their 20 to 25 years of service, and this added to the cosmopolitan Roman character of the frontier population. By the later Roman period, frontier garrisons (groups of soldiers) were only rarely transferred, service in units became effectively hereditary, and forts were no longer populated or maintained at full strength.

This process of settling in as a community over several generations, combined with local recruitment, presumably accounts for the apparent stability of the British northern frontier in the later Roman period. It also explains why some of the forts continued in occupation long after Rome ceased to have any formal authority in Britain, at the beginning of the fifth century A.D. The circumstances that had allowed natives to become Romanized also led the self-sustaining military community of the frontier area to become effectively British.

20 – 1 Westward Migration

The story of the westward movement of population in the United States is, in the main, the story of the expansion of American agriculture — of the development of new areas for the raising of livestock and the cultivation of wheat, corn, tobacco, and cotton. After 1815 improved transportation enabled more and more western farmers to escape a self-sufficient way of life and enter a national market economy. During periods when commodity prices were high, the rate of westward migration increased spectacularly. "Old America seemed to be breaking up and moving westward," observed an English visitor in 1817, during the first great wave of migration. Emigration to the West reached a peak in the 1830's. Whereas in 1810 only a seventh of the American people lived west of the Appalachian Mountains, by 1840 more than a third lived there.

Why were these hundreds of thousands of settlers — most of them farmers, some of them artisans — drawn away from the cleared fields and established cities and villages of the East? Certain characteristics of American society help to explain this remarkable migration. The European ancestors of some Americans had for centuries lived rooted to the same village or piece of land until some religious, political, or economic crisis uprooted them and drove them across the Atlantic. Many of those who experienced this sharp break thereafter lacked the ties that had bound them and their ancestors to a single place. Moreover, European society was relatively stratified; occupation and social status were inherited. In American society, however, the class structure was less rigid; some people changed occupations easily and believed it was their duty to improve their social and economic position. As a result, many Americans were an inveterately restless, rootless, and ambitious people. Therefore, these social traits helped to produce the nomadic and daring settlers who kept pushing westward beyond the fringes of settlement. In addition, there were other immigrants who migrated west in search of new homes, material success, and better lives.

The West had plenty of attractions: the alluvial river bottoms, the fecund soils of the rolling forest lands, the black loams of the prairies were tempting to New England farmers working their rocky, sterile land and to southeastern farmers plagued with soil depletion and erosion. In 1820 under a new land law, a farm could be bought for \$100. The

continued proliferation of banks made it easier for those without cash to negotiate loans in paper money. Western Farmers borrowed with the confident expectation that the expanding economy would keep farm prices high, thus making it easy to repay loans when they fell due.

Transportation was becoming less of a problem for those who wished to move west and for those who hand farm surpluses to send to market. Prior to 1815, western farmers who did not live on navigable waterways were connected to them only by dirt roads and mountain trails. Livestock could be driven across the mountains, but the cost of transporting bulky grains in this fashion was several times greater than their value in eastern markets. The first step toward an improvement of western transportation was the construction of turnpikes. These roads made possible a reduction in transportation costs and thus stimulated the commercialization of agriculture along their routes.

Two other developments presaged the end of the era of turnpikes and started a transportation revolution that resulted in increased regional specialization and the growth of a national market economy. First came the steamboat; although flatboats and keelboats continued to be important until the 1850's steamboats eventually superseded all other craft in the carrying of passengers and freight. Steamboats were not only faster but also transported upriver freight for about one tenth of what it had previously cost on hand-propelled keelboats. Next came the Erie Canal, an enormous project in its day, spanning about 350 miles. After the canal went into operation, the cost per mile of transporting a ton of freight from Buffalo to New York City declined from nearly 20 cents to less than 1 cent. Eventually, the western states diverted much of their produce from the rivers to the Erie Canal, a shorter route to eastern markets.

20 – 2 Early Settlements in the Southwest Asia

The universal global warming at the end of the Ice Age had dramatic effects on temperate regions of Asia, Europe, and North America. Ice sheets retreated and sea levels rose. The climatic changes in southwestern Asia were more subtle, in that they involved shifts in mountain snow lines, rainfall patterns, and vegetation cover. However, these same cycles of change had momentous impacts on the sparse human populations of the region. At the end of the Ice Age, no more than a few thousand foragers lived along the eastern Mediterranean coast, in the Jordan and Euphrates valleys. Within 2,000 years, the human population of the region numbered in the tens of thousands, all as a result of village life and farming. Thanks to new environmental and archaeological discoveries, we now know something about this remarkable change in local life.

Pollen samples from freshwater lakes in Syria and elsewhere tell us forest cover expanded rapidly at the end of the Ice Age, for the southwestern Asian climate was still cooler and considerably wetter than today. Many areas were richer in animal and plant species than they are now, making them highly favorable for human occupation. About 9000 B.C., most human settlements lay in the area along the Mediterranean coast and in the Zagros Mountains of Iran and their foothills. Some local areas, like the Jordan River valley, the middle Euphrates valley, and some Zagros valleys, were more densely populated than elsewhere. Here more sedentary and more complex societies flourished. These people exploited the landscape intensively, foraging on hill slopes for wild cereal grasses and nuts, while hunting gazelle and other game on grassy lowlands and in river valleys. Their settlements contain exotic objects such as seashells, stone bowls, and artifacts made of obsidian (volcanic glass), all traded from afar. This considerable volume of intercommunity exchange brought a degree of social complexity in its wake.

Thanks to extremely fine-grained excavation and extensive use of flotation methods (through which seeds are recovered from soil samples), we know a great deal about the foraging practices of the inhabitants of Abu Hureyra in Syria's Euphrates valley. Abu Hureyra was founded about 9500 B.C., a small village settlement of cramped pit dwellings (houses dug partially in the soil) with reed roofs supported by wooden uprights. For the next 1,500 years, its inhabitants enjoyed a somewhat warmer

and damper climate than today, living in a well-wooded steppe area where wild cereal grasses were abundant. They subsisted on offspring migrations of Persian gazelles from the south. With such a favorable location, about 300 to 400 people lived in a sizable, permanent settlement. They were no longer a series of small bands but lived in a large community with more elaborate social organization, probably grouped into clans of people of common descent.

The flotation samples from the excavations allowed botanists to study shifts in plant-collecting habits as if they were looking through a telescope at a changing landscape. Hundreds of tiny plant remains show how the inhabitants exploited nut harvests in nearby pistachio and oak forests. However, as the climate dried up, the forests retreated from the vicinity of the settlement. The inhabitants turned to wild cereal grasses instead, collecting them by the thousands, while the percentage of nuts in the diet fell. By 8200 B.C., drought conditions were so severe that the people abandoned their long-established settlement, perhaps dispersing into smaller camps.

Five centuries later, about 7700 B.C., a new village rose on the mound. At first the inhabitants still hunted gazelle intensively. Then, about 7000 B.C., within the space of a few generations, they switched abruptly to herding domesticated goats and sheep and to growing einkorn, pulses, and other cereal grasses. Abu Hureyra grew rapidly until it covered nearly 30 acres. It was a close-knit community of rectangular, one-story mud-brick houses, joined by narrow lanes and courtyards, finally abandoned about 5000 B.C.. Many complex factors led to the adoption of the new economies, not only at Abu Hureyra, but at many other locations such as 'Ain Ghazal, also in Syria, where goat toe bones showing the telltale marks of abrasion caused by foot tethering (binding) testify to early herding of domestic stock.

25 – 2 The Decline of Venetian Shipping

In the late thirteenth century, northern Italian cities such as Genoa, Florence, and Venice began an economic resurgence that made them into the most important economic centers of Europe. By the seventeenth century, however, other European powers had taken over, as the Italian cities lost much of their economic might.

This decline can be seen clearly in the changes that affected Venetian shipping and trade. First, Venice's intermediary functions in the Adriatic Sea, where it had dominated the business of shipping for other parties, were lost to direct trading. In the fifteenth century there was little problem recruiting sailors to row the galleys (large ships propelled by oars): guilds (business associations) were required to provide rowers, and through a draft system free citizens served compulsorily when called for. In the early sixteenth century the shortage of rowers was not serious because the demand for galleys was limited by a move to round ships (round-hulled ships with more cargo space), with required fewer rowers. But the shortage of crews proved to be a greater and greater problem, despite continuous appeal to Venice's tradition of maritime greatness. Even though sailors' wages doubled among the northern Italian cities from 1550 to 1590, this did not elicit an increased supply.

The problem in shipping extended to the Arsenale, Venice's huge and powerful shipyard. Timber ran short, and it was necessary to procure it from farther and farther away. In ancient Roman times, the Italian peninsula had great forest of fir preferred for warships, but scarcity was apparent as early as the early fourteenth century. Arsenale officers first brought timber from the foothills of the Alps, then from north toward Trieste, and finally from across the Adriatic. Private shipbuilders were required to buy their oak abroad. As the costs of shipbuilding rose, Venice clung to its outdated standard while the Dutch were innovation in the lighter and more easily handled ships.

The step from buying foreign timber to buying foreign ships was regarded as a short one, especially when complaints were heard in the latter sixteenth century that the standards and traditions of the Arsenale were running down. Work was stretched out and done poorly. Older workers had been allowed to stop work a half hour before the regular time, and in 1601 younger workers left with them. Merchants complained that the privileges

reserved for Venetian-built and owned ships were first extended to those Venetians who bought ships from abroad and then to foreign-built and owned vessels. Historian Frederic Lane observes that after the loss of ships in battle in the late sixteenth century, the shipbuilding industry no longer had the capacity to recover that it had displayed at the start of the century.

The conventional explanation for the loss of Venetian dominance in trade is establishment of the Portuguese direct sea route to the East, replacing the overland Silk Road from the Black Sea and the highly profitable Indian Ocean-caravan-eastern Mediterranean route to Venice. The Portuguese Vasco da Gama's Voyage around southern Africa to India took place at the end of the fifteenth century, and by 1502 the trans-Arabian caravan route had been cut off by political unrest.

The Venetian Council finally allowed round ships to enter the trade that was previously reserved for merchant galleys, thus reducing transport cost by one third. Prices of spices delivered by ship from the eastern Mediterranean came to equal those of spices transported by Portuguese vessels, but the increase in quantity with both routes in operation drove the price far down. Gradually, Venice's role as a storage and distribution center for spices and silk, dyes cotton, and gold decayed, and by the early seventeenth century Venice had lost its monopoly in markets such as France and southern Germany.

Venetian shipping had started to decline from about 1530-before the entry into the Mediterranean of large volumes of Dutch and British shipping-and was clearly outclassed by the end of the century. A contemporary of Shakespeare (1564-1616) observed that the productivity of Italian shipping had declined, compared with that of the British, because of conservatism and loss of expertise. Moreover, Italian sailors were deserting and emigrating, and captains, no longer recruited from the ranks of nobles, were weak on navigations.

26 – 3 Sumer and the First Cities of the Ancient Near East

The earliest of the city states of the ancient Near East appeared at the southern end of the Mesopotamian plain, the area between the Tigris and Euphrates rivers in what is now Iraq. It was here that the civilization known as Sumer emerged in its earliest form in the fifth millennium. At first sight, the plain did not appear to be a likely home for a civilization. There were few natural resources, no timber, stone, or metals. Rainfall was limited, and what water there was rushed across the plain in the annual flood of melted snow. As the plain fell only 20 meters in 500 kilometers, the beds of the rivers shifted constantly. It was this that made the organization of irrigation, particularly the building of canals to channel and preserve the water, essential. Once this was done and the silt carried down by the rivers was planted, the rewards were rich: four to five times what rain-fed earth would produce. It was these conditions that allowed an elite to emerge, probably as an organizing class, and to sustain itself through the control of surplus crops.

It is difficult to isolate the factors that led to the next development — the emergence of urban settlements. The earliest, that of Eridu, about 4500 B.C.E., and Uruk, a thousand years later, center on impressive temple complexes built of mud brick. In some way, the elite had associated themselves with the power of the gods. Uruk, for instance, had two patron gods — Anu, the god of the sky and sovereign of all other gods, and Inanna, a goddess of love and war — and there were others, patrons of different cities. Human beings were at their mercy. The biblical story of the Flood may originate in Sumer. In the earliest version, the gods destroy the human race because its clamor had been so disturbing to them.

It used to be believed that before 3000 B.C.E. the political and economic life of the cities was centered on their temples, but it now seems probable that the cities had secular rulers from earliest times. Within the city lived administrators, craftspeople, and merchants. (Trading was important, as so many raw materials, the semiprecious stones for the decoration of the temples, timbers for roofs, and all metals, had to be imported.) An increasingly sophisticated system of administration led in about 3300 B.C.E. to the appearance of writing. The earliest script was based on logograms, with a symbol being used to

express a whole word. The logograms were incised on damp clay tablets with a stylus with a wedge shape at its end. (The Romans called the shape *cuneus* and this gives the script its name of *cuneiform*.) Two thousand logograms have been recorded from these early centuries of writing. A more economical approach was to use a sign to express not a whole word but a single syllable. (To take an example: the Sumerian word for “head” was “sag.” Whenever a word including a syllable in which the sound “sag” was to be written, the sign for “sag” could be used to express that syllable with the remaining syllables of the word expressed by other signs.) By 2300 B.C.E. the number of signs required had been reduced to 600, and the range of words that could be expressed had widened. Texts dealing with economic matters predominated, as they always had done; but at this point works of theology, literature, history, and law also appeared.

Other innovations of the late fourth millennium include the wheel, probably developed first as a more efficient way of making pottery and then transferred to transport. A tablet engraved about 3000 B.C.E. provides the earliest known example from Sumer, a roofed boxlike sledge mounted on four solid wheels. A major development was the discovery, again about 3000 B.C.E., that if copper, which had been known in Mesopotamia since about 3500 B.C.E., was mixed with tin, a much harder metal, bronze, would result. Although copper and stone tools continued to be used, bronze was far more successful in creating sharp edges that could be used as anything from saws and scythes to weapons. The period from 3000 to 1000 B.C.E., when the use of bronze became widespread, is normally referred to as the Bronze Age.

28 – 2 Early Saharan Pastoralists

The Sahara is a highly diverse, albeit dry, region that has undergone major climatic changes since 10,000 B.C. As recently as 6,000 B.C., the southern frontier of the desert was far to the north of where it is now, while semiarid grassland and shallow freshwater lakes covered much of what are now arid plains. This was a landscape where antelope of all kinds abounded — along with *Bosprimigenius*, a kind of oxen that has become extinct. The areas that are now desert were, like all arid regions, very susceptible to cycles of higher and lower levels of rainfall, resulting in major, sudden changes in distributions of plants and animals. The people who hunted the sparse desert animals responded to drought by managing the wild resources they hunted and gathered, especially wild oxen, which had to have regular water supplies to survive.

Even before the drought, the Sahara was never well watered. Both humans and animals were constantly on the move, in search of food and reliable water supplies. Under these circumstances, archaeologist Andrew Smith believes, the small herds of *Bosprimigenius* in the desert became smaller, more closely knit breeding units as the drought took hold. The beasts were more disciplined, so that it was easier for hunters to predict their habits, and capture animals at will. At the same time, both cattle and humans were more confined in their movements, staying much closer to permanent water supplies for long periods of time. As a result, cattle and humans came into close association.

Smith believes that the hunters were well aware of the more disciplined ways in which their prey behaved. Instead of following the cattle on their annual migrations, the hunters began to prevent the herd from moving from one spot to another. At first, they controlled the movement of the herd while ensuring continuance of their meat diet. But soon they also gained genetic control of the animals, which led to rapid physical changes in the herd. South African farmers who maintain herds of wild eland (large African antelopes with short, twisted horns) report that the offspring soon diminish in size, unless wild bulls are introduced constantly from outside. The same effects of inbreeding may have occurred in controlled cattle populations, with some additional, and perhaps unrecognized, advantages. The newly domesticated animals behaved better, were easier

to control, and may have enjoyed a higher birth rate, which in turn yielded greater milk supplies. We know from rock paintings deep in the Sahara that the herders were soon selecting breeding animals to produce offspring with different horn shapes and hide colors.

It is still unclear whether domesticated cattle were tamed independently in northern Africa or introduced to the continent from Southwest Asia. Whatever the source of the original tamed herds might have been, it seems entirely likely that much the same process of juxtaposition (living side by side) and control occurred in both southwest Asia and northern Africa, and even in Europe, among peoples who had an intimate knowledge of the behavior of wild cattle. The experiments with domestication probably occurred in many places, as people living in ever-drier environments cast around for more predictable food supplies.

The cattle herders had only a few possessions: unsophisticated pots and polished adzes. They also hunted with bow and arrow. The Saharan people left a remarkable record of their lives painted on the walls of caves deep in the desert. Their artistic endeavors have been preserved in paintings of wild animals, cattle, goats, humans, and scenes of daily life that extend back perhaps to 5,000 B.C.. The widespread distribution of pastoral sites of this period suggests that the Saharans ranged their herds over widely separated summer and winter grazing grounds.

About 3,500 B.C., climatic conditions again deteriorated. The Sahara slowly became drier and lakes vanished. On the other hand, rainfall increased in the interior of western Africa, and the northern limit of the tsetse fly, an insect fatal to cattle, moved south. So the herders shifted south, following the major river systems into savanna regions. By this time, the Saharan people were probably using domestic crops, experimenting with such summer rainfall crops as sorghum and millet as they move out of areas where they could grow wheat, barley, and other Mediterranean crops.

32 – 2 Siam, 1851-1910

In the late nineteenth century, political and social changes were occurring rapidly in Siam (now Thailand). The old ruling families were being displaced by an evolving centralized government. These families were pensioned off (given a sum of money to live on) or simply had their revenues taken away or restricted; their sons were enticed away to schools for district officers, later to be posted in some faraway province; and the old patron-client relations that had bound together local societies simply disintegrated. Local rulers could no longer protect their relatives and attendants in legal cases, and with the ending in 1905 of the practice of forcing peasant farmers to work part-time for local rulers, the rulers no longer had a regular base for relations with rural populations. The old local ruling families, then, were severed from their traditional social context.

The same situation viewed from the perspective of the rural population is even more complex. According to the government's first census of the rural population, taken in 1905, there were about thirty thousand villages in Siam. This was probably a large increase over the figure even two or three decades earlier, during the late 1800s. It is difficult to imagine it now, but Siam's Central Plain in the late 1800s was nowhere near as densely settled as it is today. There were still forests closely surrounding Bangkok into the last half of the nineteenth century, and even at century's end there were wild elephants and tigers roaming the countryside only twenty or thirty miles away.

Much population movement involved the opening up of new lands for rice cultivation. Two things made this possible and encouraged it to happen. First, the opening of the kingdom to the full force of international trade by the Bowring Treaty (1855) rapidly encouraged economic specialization in the growing of rice, mainly to feed the rice-deficient portions of Asia (India and China in particular). The average annual volume of rice exported from Siam grew from under 60 million kilograms per year in the late 1850s to more than 660 million kilograms per year at the turn of the century; and over the same period the average price per kilogram doubled. During the same period, the area planted in rice increased from about 230, 000 acres to more than 350, 000 acres. This growth was achieved as the result of the collective decisions of thousands of peasant families to expand the amount of land they cultivated, clear and plant new

land, or adopt more intensive methods of agriculture.

They were able to do so because of our second consideration. They were relatively freer than they had been half a century earlier. Over the course of the Fifth Reign (1868 – 1910), the ties that bound rural people to the aristocracy and local ruling elites were greatly reduced. Peasants now paid a tax on individuals instead of being required to render labor service to the government. Under these conditions, it made good sense to thousands of peasant families to in effect work full-time at what they had been able to do only part-time previously because of the requirement to work for the government: grow rice for the marketplace.

Numerous changes accompanied these developments. The rural population both dispersed and grew, and was probably less homogeneous and more mobile than it had been a generation earlier. The villages became more vulnerable to arbitrary treatment by government bureaucrats as local elites now had less control over them. By the early twentieth century, as government modernization in a sense caught up with what had been happening in the countryside since the 1870s, the government bureaucracy intruded more and more into village life. Provincial police began to appear, along with district officers and cattle registration and land deeds and registration for compulsory military service. Village handicrafts diminished or died out completely as people bought imported consumer goods, like cloth and tools, instead of making them themselves. More economic variation took shape in rural villages, as some grew prosperous from farming while others did not. As well as can be measured, rural standards of living improved in the Fifth Reign. But the statistical averages mean little when measured against the harsh realities of peasant life.

33 – 1 The First Civilizations

Evidence suggests that an important stimulus behind the rise of early civilizations was the development of settled agriculture, which unleashed a series of changes in the organization of human communities that culminated in the rise of large ancient empires.

The exact time and place that crops were first cultivated successfully is uncertain. Many prehistorians believe that farming may have emerged independently in several different areas of the world when small communities, driven by increasing population and a decline in available food resources, began to plant seeds in the ground in an effort to guarantee their survival. The first farmers, who may have lived as long as 10,000 years ago, undoubtedly used simple techniques and still relied primarily on other forms of food production, such as hunting, foraging, or pastoralism. The real breakthrough took place when farmers began to cultivate crops along the floodplains of river systems. The advantage was that crops grown in such areas were not as dependent on rainfall and therefore produced a more reliable harvest. An additional benefit was that the sediment carried by the river waters deposited nutrients in the soil, thus enabling the farmer to cultivate a single plot of ground for many years without moving to a new location. Thus, the first truly sedentary (that is, nonmigratory) societies were born. As time went on, such communities gradually learned how to direct the flow of water to enhance the productive capacity of the land, while the introduction of the iron plow eventually led to the cultivation of heavy soils not previously susceptible to agriculture.

The spread of this river valley agriculture in various parts of Asia and Africa was the decisive factor in the rise of the first civilizations. The increase in food production in these regions led to a significant growth in population, while efforts to control the flow of water to maximize the irrigation of cultivated areas and to protect the local inhabitants from hostile forces outside the community provoked the first steps toward cooperative activities on a large scale. The need to oversee the entire process brought about the emergence of an elite that was eventually transformed into a government.

The first clear steps in the rise of the first civilizations took place in the fourth and third millennia B.C. in Mesopotamia, northern Africa,

India, and China. How the first governments took shape in these areas is not certain, but anthropologists studying the evolution of human communities in various parts of the world have discovered that one common stage in the process is the emergence of what are called "big men" within a single village or a collection of villages. By means of their military prowess, dominant personalities, or political talents, these people gradually emerge as the leaders of that community. In time, the "big men" become formal symbols of authority and pass on that authority to others within their own family. As the communities continue to grow in size and material wealth, the "big men" assume hereditary status, and their allies and family members are transformed into a hereditary monarchy.

The appearance of these sedentary societies had a major impact on the social organizations, religious beliefs, and way of life of the peoples living within their boundaries. With the increase in population and the development of centralized authority came the emergence of the cities. While some of these urban centers were identified with a particular economic function, such as proximity to gold or iron deposits or a strategic location on a major trade route, others served primarily as administrative centers or the site of temples for the official cult or other ritual observances. Within these cities, new forms of livelihood appeared to satisfy the growing need for social services and consumer goods. Some people became artisans or merchants, while others became warriors, scholars, or priests. In some cases, the physical division within the first cities reflected the strict hierarchical character of the society as a whole, with a royal palace surrounded by an imposing wall and separate from the remainder of the urban population. In other instances, such as the Indus River Valley, the cities lacked a royal precinct and the ostentatious palaces that marked their contemporaries elsewhere.

72 – 3 Cities and Kingdoms in Alexander's Empire

As Alexander the Great spread his empire (the Hellenistic empire) from Macedon and Greece over a large swath of southwestern Asia and northern Africa, he established a number of Kingdoms that became subjects of the empire. One of the major developments of these new kingdoms was the resurgence of monarchy, which had many repercussions. For most Greeks, monarchs were something out of the heroic past, something found in Homer's mythical epic the Iliad but not in daily life. Furthermore, most Hellenistic kingdoms embraced numerous different peoples who had little in common. Hellenistic kings thus needed a new political concept to unite them. One solution was the creation of a ruler cult that linked the king's authority with that of the gods. Thus, royal power had divine approval and was meant to create a political and religious bond between the kings and their subjects. These deified kings were not considered gods as mighty as Zeus or Apollo, and the new ruler cults probably made little religious impact on those ruled. Nonetheless, the ruler cult was an easily understandable symbol of unity within the kingdom.

Monarchy also included royal women, who began to play an active part in political and diplomatic life. Some of them did so in their own right, others in collaboration with their husbands. For the most part they served as examples that women too were capable of shouldering vast responsibilities and performing them successfully. Although Alexander's generals created huge kingdoms, the concept of monarchy, even when combined with the ruler cult, never replaced the ideal of the polis (the ancient Greek city-state).

Consequently, the monarchies never won the deep emotional loyalty that Greeks had once felt for the polis. Hellenistic kings needed large numbers of Greeks to run their kingdoms. Otherwise royal business would grind to a halt, and the conquerors would soon be swallowed up by the far more numerous conquered population. Obviously, then, the kings had to encourage Greeks to immigrate and build new homes. The Hellenistic kings thus confronted the problem of making life in the new monarchies resemble the traditional Greek way of life. Since Greek civilization was urban, the kings continued Alexander's policy of establishing cities throughout their kingdoms in order to entice Greeks to immigrate. Yet the creation of these cities posed

a serious political problem that the Hellenistic kings failed to solve.

To the Greeks, civilized life was unthinkable without the polis, which was far more than a mere city. The Greek polis was by definition autonomous state run by its citizens, free of any outside power or restraint. Hellenistic kings, however, refused to grant sovereignty to their cities. In effect, these kings willingly built cities but refused to build a polis.

Hellenistic monarchs gave their cities all the external trappings of a polis. Each had an assembly of citizens, a council to prepare legislation, and a board of magistrates to conduct the city's political business. Yet, however similar to the Greek polis they appeared, these cities could not engage in diplomatic dealings, make treaties, pursue their own foreign policy, or wage their own wars. None could govern its own affairs without interference from the king, who, even if he stood in the background, was the real sovereign. In the eyes of the king, the cities were important parts of the kingdom, but the welfare of the whole kingdom came first. The cities had to follow royal orders, and the king often placed his own officials in the cities to see that his decrees were followed.

A new Hellenistic city differed from a Greek polis in other ways as well. The Greek polis had enjoyed political and social unity even though it was normally composed of citizens, non-citizens and resident aliens. The polis had one body of law and one set of customs. In the Hellenistic city, Greeks represented an elite citizen class. Natives and non-Greek foreigners who lived in Hellenistic cities usually possessed lesser rights than Greeks and often had their own laws. In some instances this disparity spurred natives to assimilate Greek culture in order to rise politically and socially. The Hellenistic city was not homogeneous and could not spark the intensity of feeling that marked the polis.

71 – 2 Minoan Palaces

The Minoan culture on the island of Crete in the Aegean Sea flourished from about 3000 to 1100 B.C. In what is known as the Palace Period (ca. 2000-1450 B.C.), power was centralized in palaces and, later, in villas. According to one authority, the five primary economic functions of Minoan palaces during much of the Palace era were (1) production of manufactured goods, (2) consumption of food and manufactured goods, (3) regulation of local and internal exchange, (4) regulation of international and external exchange, and (5) use as depositories (storage facilities)

The production and storage of manufactured goods are evident from the archaeological digs at Phaistos and Mallia, important palace sites. Excavations of the earliest phases at Phaistos (those from the so-called First Palace Period, which ended about 1700 B.C.) revealed two areas dedicated to economic activity. Unit A contained several large storage vessels originally filled with foodstuffs, including liquids such as wine and oil. Nearby Unit B was the palace workshop. Here, excavators found tools used for stoneworking (a lapidary workshop), several loom weights (a weaving workshop), and two potter's wheels (a ceramics workshop). Clay sealings from a smaller room in Unit B may indicate where finished products were processed for storage or export. By the Second Palace Period (roughly 1700-1450 B.C.), there were even more food storage vessels present and an archive room (a room for storing documents).

Similar finds appeared at Mallia. In the northwest quarter of the palace, excavators discovered obsidian, soapstone, and a reddish marble called rosso antico, all evidently part of the lapidary (stonecutting) workshop. A potter's workshop was also present within the palace walls. During part of this period, the workshop of a bronze smithy was located just outside the palace walls. It is actually surprising that such an industry would be so close to any residential quarters, considering the unpleasant fumes given off by the work and the rather high potential for fires. Nevertheless, at a somewhat later date, the palace walls were extended so that the smithy was located within the palace itself. Clearly, this was an industry over which the palace wanted to keep very close control. The role of Minoan palaces as depositories and regulators of local distribution and trade may be seen in the koulouras—large, stone-lined pits located

at Knossos, Phaistos, and in a slightly altered guise, at Mallia. There is continued debate as to the purpose of these huge storage bins. It was originally suggested that they were rubbish pits. Some modern scholars believe that they were giant tree planters. But the usual interpretation is that they were for grain storage, with the koulouras at Knossos being able to hold enough grain to feed 1,000 people and the koulouras at Phaistos being able to hold enough for 300 people. In such a case, the palace would have received a substantial portion of the agricultural produce of the surrounding farms, stored it, and then distributed it to the more specialized, nonagricultural populace of the palace region.

Palatial control over foreign trade is more difficult to prove archaeologically, as there is often no way to determine where on Crete any specific item found abroad was made. One argument often brought to the fore is that only the palaces would have the capital (to use a modern term) to finance the goods and shipping for long trade journeys, not to mention to handle the risks of possible sea wrecks. Another argument, however, lies in the nature of the Minoan goods found abroad. For example, Kamares ware pottery from Minoan Crete has come to light on the coasts of Cyprus and in areas of the Near East such as Egypt and Syria. This Kamares ware is clearly a product of palatial manufacture. The ceramics from before and after the classical Kamares wares are clearly local creations—Knossian ware being distinct from Malian ware. By contrast, the Kamares ware made in the palaces is similar from palace to palace but is utterly distinct from the provincial wares. The number of foreign goods stored in the palaces, especially Zakro, also gives evidence for the palatial control of international exchange.

41 – 3 Trade and Early State Formation

Bartering was a basic trade mechanism for many thousands of years; often sporadic and usually based on notions of reciprocity, it involved the mutual exchange of commodities or objects between individuals or groups. Redistribution of these goods through society lay in the hands of chiefs, religious leaders, or kin groups. Such redistribution was a basic element in chiefdoms. The change from redistribution to formal trade—often based on regulated commerce that perhaps involved fixed prices and even currency—was closely tied to growing political and social complexity and hence to the development of the state in the ancient world.

In the 1970s, a number of archaeologists gave trade a primary role in the rise of ancient states. British archaeologist Colin Renfrew attributed the dramatic flowering of the Minoan civilization on Crete and through the Aegean to intensified trading contacts and to the impact of olive and vine cultivation on local communities. As agricultural economies became more diversified and local food supplies could be purchased both locally and over longer distances, a far-reaching economic interdependence resulted. Eventually, this led to redistribution systems for luxuries and basic commodities, systems that were organized and controlled by Minoan rulers from their palaces. As time went on, the self-sufficiency of communities was replaced by mutual dependence. Interest in long-distance trade brought about some cultural homogeneity from trade and gift exchange, and perhaps even led to piracy. Thus, intensified trade and interaction, and the flowering of specialist crafts, in a complex process of positive feedback, led to much more complex societies based on palaces, which were the economic hubs of a new Minoan civilization.

Renfrew's model made some assumptions that are now discounted. For example, he argued that the introduction of domesticated vines and olives allowed a substantial expansion of land under cultivation and helped to power the emergence of complex society. Many archaeologists and paleobotanists now question this view, pointing out that the available evidence for cultivated vines and olives suggests that they were present only in the later Bronze Age. Trade, nevertheless, was probably one of many variables that led to the emergence of palace economies in Minoan Crete.

American archaeologist William Rathje developed a hypothesis that considered an explosion in long-distance exchange a fundamental cause of Mayan civilization in Mesoamerica. He suggested that the lowland Mayan environment was deficient in many

vital resources, among them obsidian, salt, stone for grinding maize, and many luxury materials. All these could be obtained from the nearby highlands, from the Valley of Mexico, and from other regions, if the necessary trading networks came into being. Such connections, and the trading expeditions to maintain them, could not be organized by individual villages. The Maya lived in a relatively uniform environment, where every community suffered from the same resource deficiencies. Thus, argued Rathje, long-distance trade networks were organized through local ceremonial centers and their leaders. In time, this organization became a state, and knowledge of its functioning was exportable, as were pottery, tropical bird feathers, specialized stone materials, and other local commodities.

Rathje's hypothesis probably explains part of the complex process of Mayan state formation, but it suffers from the objection that suitable alternative raw materials can be found in the lowlands. It could be, too, that warfare became a competitive response to population growth and to the increasing scarcity of prime agricultural land, and that it played an important role in the emergence of the Mayan states.

Now that we know much more about ancient exchange and commerce, we know that, because no one aspect of trade was an overriding cause of cultural change or evolution in commercial practices, trade can never be looked on as a unifying factor or as a primary agent of ancient civilization. Many ever-changing variables affected ancient trade, among them the demand for goods. There were also the logistics of transportation, the extent of the trading network, and the social and political environment. Intricate market networks channeled supplies along well-defined routes. Authorities at both ends might regulate the profits fed back to the source, providing the incentive for further transactions. There may or may not have been a market organization. Extensive long-distance trade was a consequence rather than a cause of complex societies.

40 – 1 Ancient Athens

One of the most important changes in Greece during the period from 800 B.C. to 500 B.C. was the rise of the polis, or city-state, and each polis developed a system of government that was appropriate to its circumstances. The problems that were faced and solved in Athens were the sharing of political power between the established aristocracy and the emerging other classes, and the adjustment of aristocratic ways of life to the ways of life of the new polis. It was the harmonious blending of all of these elements that was to produce the classical culture of Athens.

Entering the polis age, Athens had the traditional institutions of other Greek protodemocratic states: an assembly of adult males, an aristocratic council, and annually elected officials. Within this traditional framework the Athenians, between 600 B.C. and 450 B. C., evolved what Greeks regarded as a fully fledged democratic constitution, though the right to vote was given to fewer groups of people than is seen in modern times.

The first steps toward change were taken by Solon in 594 B. C., when he broke the aristocracy's stranglehold on elected offices by establishing wealth rather than birth as the basis of office holding, abolishing the economic obligations of ordinary Athenians to the aristocracy, and allowing the assembly (of which all citizens were equal members) to overrule the decisions of local courts in certain cases. The strength of the Athenian aristocracy was further weakened during the rest of the century by the rise of a type of government known as a tyranny, which is a form of interim rule by a popular strongman (not rule by a ruthless dictator as the modern use of the term suggests to us). The Peisistratids, as the succession of tyrants were called (after the founder of the dynasty, Peisistratos), strengthened Athenian central administration at the expense of the aristocracy by appointing judges throughout the region, producing Athens' first national coinage, and adding and embellishing festivals that tended to focus attention on Athens rather than on local villages of the surrounding region. By the end of the century, the time was ripe for more change: the tyrants were driven out, and in 508 B C a new reformer, Cleisthenes, gave final form to the developments reducing aristocratic control already under way.

Cleisthenes' principal contribution to the creation of democracy at Athens was to complete the long

process of weakening family and clan structures, especially among the aristocrats, and to set in their place locality-based corporations called demes, which became the point of entry for all civic and most religious life in Athens. Out of the demes were created 10 artificial tribes of roughly equal population. From the demes, by either election or selection, came 500 members of a new council, 6,000 jurors for the courts, 10 generals, and hundreds of commissioners. The assembly was sovereign in all matters but in practice delegated its power to subordinate bodies such as the council, which prepared the agenda for the meetings of the assembly, and the courts, which took care of most judicial matters. Various committees acted as an executive branch, implementing policies of the assembly and supervising, for instance, the food and water supplies and public buildings. This wide-scale participation by the citizenry in the government distinguished the democratic form of the Athenian polis from other, less liberal forms.

The effect of Cleisthenes' reforms was to establish the superiority of the Athenian community as a whole over local institutions without destroying them. National politics rather than local or deme politics became the focal point. At the same time, entry into national politics began at the deme level and gave local loyalty a new focus: Athens itself. Over the next two centuries the implications of Cleisthenes' reforms were fully exploited.

During the fifth century B. C. the council of 500 was extremely influential in shaping policy. In the next century, however, it was the mature assembly that took on decision-making responsibility. By any measure other than that of the aristocrats, who had been upstaged by the supposedly inferior "people," the Athenian democracy was a stunning success. Never before, or since, have so many people been involved in the serious business of self-governance. It was precisely this opportunity to participate in public life that provided a stimulus for the brilliant unfolding of classical Greek culture.

42 – 4 Europe in the Twelfth Century

Europe in the eleventh century underwent enormous social, technological, and economic changes, but this did not create a new Europe—it created two new ones. The north was developed as a rigidly hierarchical society in which status was determined, or was at least indicated, by the extent to which one owned, controlled, or labored on land; whereas the Mediterranean south developed a more fluid, and therefore more chaotic, world in which industry and commerce predominated and social status both reflected and resulted from the role that one played in the public life of the community. In other words, individual identity and social community in the north were established on a personal basis, whereas in the south they were established on a civic basis. By the start of the twelfth century, northern and southern Europe were very different places indeed, and the Europeans themselves noticed it and commented on it.

Political dominance belonged to the north. Germany, France, and England had large Populations and large armies that made them, in the political and military senses, the masters of Western Europe. Organized by the practices known collectively as feudalism, these kingdoms emerged as powerful states with sophisticated machineries of government. Their kings and queens were the leading figures of the age; their castles and cathedrals stood majestically on the landscape as symbols of their might; their armies both energized and defined the age. Moreover, feudal society showed a remarkable ability to adapt to new needs by encouraging the parallel development of domestic urban life and commercial networks; in some regions of the north, in fact, feudal society may even have developed in response to the start of the trends toward bigger cities. But southern Europe took the lead in economic and cultural life. Though the leading Mediterranean states were small in size, they were considerably wealthier than their northern counterparts. The Italian city of Palermo in the twelfth century, for example, alone generated four times the commercial tax revenue of the entire kingdom of England. Southern communities also possessed urbane, multilingual cultures that made them the intellectual and artistic leaders of the age.

Levels of general literacy in the south far surpassed those of the north, and the people of the south put that learning to use on a large scale. Science, mathematics, poetry, law, historical

writing, religious speculation, translation, and classical studies all began to flourish; throughout most of the twelfth century, most of the continent's best brains flocked to southern Europe.

So too did a lot of the north's soldiers. One of the central themes of the political history of the twelfth century was the continual effort by the northern kingdoms to extend their control southward in the hope of tapping into the Mediterranean bonanza. The German emperors starting with Otto I (936-973), for example, struggled ceaselessly to establish their control over the cities of northern Italy, since those cities generated more revenue than all of rural Germany combined. The kings of France used every means at their disposal to push the lower border of their kingdom to the Mediterranean shoreline. And the Normans who conquered and ruled England established outposts of Norman power in Sicily and the adjacent lands of southern Italy; the English kings also hoped or claimed at various times to be, either through money or marriage diplomacy, the rulers of several Mediterranean states. But as the northern world pressed southward, so too did some of the cultural norms and social mechanisms of the south expand northward. Over the course of the twelfth century, the feudal kingdoms witnessed a proliferation of cities modeled in large degree on those of the south. Contact with the merchants and financiers of the Mediterranean led to the development of northern industry and international trade (which helped to pay for many of the castles and cathedrals mentioned earlier). And education spread as well, culminating in the foundation of what is arguably medieval Europe's greatest invention: the university. The relationship of north and south was symbiotic, in other words, and the contrast between them was more one of differences in degree than of polar opposition.

43 – 1 The Empire of Alexander the Great

In 334 B.C. Alexander the Great took his Greek armies to the east and in only a few years completed his creation of an empire out of much of southwest Asia in the new empire, barriers to trade and the movement of peoples were removed; markets were put in touch with one another. In the next generation thousands of Greek traders and artisans would enter this wider world to seek their fortunes. Alexander's actions had several important consequences for the region occupied by the empire.

The first of these was the expansion of Greek civilization throughout the Middle East. Greek became the great international language. Towns and cities were established not only as garrisons (military posts) but as centers for the diffusion of Greek language, literature, and thought, particularly through libraries, as at Antioch (in modern Turkey) and the most famous of all, at Alexandria in Egypt, which would be the finest in the world for the next thousand years.

Second, this internationalism spelled the end of the classical Greek city-state—the unit of government in ancient Greece—and everything it stood for. Most city-states had been quite small in terms of citizenry, and this was considered to be a good thing. The focus of life was the agora, the open marketplace where assemblies could be held and where issues of the day, as well as more fundamental topics such as the purpose of government or the relationship between law and freedom, could be discussed and decisions made by individuals in person. The philosopher Plato (428-348 B.C.) felt that the ideal city-state should have about 5,000 citizens, because to the Greeks it was important that everyone in the community should know each other. In decision making, the whole body of citizens together would have the necessary knowledge in order generally to reach the right decision, even though the individual might not be particularly qualified to decide. The philosopher Aristotle (384-322 B.C.), who lived at a time when the city-state system was declining, believed that a political entity of 100,000 simply would not be able to govern itself.

This implied that the city-state was based on the idea that citizens were not specialists but had multiple interests and talents—each a so-called jack-of-all-trades who could engage in many areas of life and politics. It implied a respect for the

wholeness of life and a consequent dislike of specialization. It implied economic and military self-sufficiency. But with the development of trade and commerce in Alexander's empire came the growth of cities; it was no longer possible to be a jack-of-all-trades. One now had to specialize, and with specialization came professionalism. There were getting to be too many persons to know, an easily observable community of interests was being replaced by a multiplicity of interests. The city-state was simply too "small-time."

Third, Greek philosophy was opened up to the philosophy and religion of the East. At the peak of the Greek city-state, religion played an important part. Its gods—such as Zeus, father of the gods, and his wife Hera—were thought of very much as being like human beings but with superhuman abilities. Their worship was linked to the rituals connected with one's progress through life—birth, marriage, and death—and with invoking protection against danger, making prophecies, and promoting healing, rather than to any code of behavior. Nor was there much of a theory of afterlife.

Even before Alexander's time, a life spent in the service of their city-state no longer seemed ideal to Greeks. The Athenian philosopher Socrates (470-399 B.C.) was the first person in Greece to propose a morality based on individual conscience rather than the demands of the state, and for this he was accused of not believing in the city's gods and so corrupting the youth, and he was condemned to death. Greek philosophy—or even a focus on conscience—might complement religion but was no substitute for it, and this made Greeks receptive to the religious systems of the Middle East, even if they never adopted them completely. The combination of the religious instinct of Asia with the philosophic spirit of Greece spread across the world in the era after Alexander's death, blending the culture of the Middle East with the culture of Greece.

46 – 2 The Commercial Revolution in Medieval Europe

Beginning in the 1160s, the opening of new silver mines in northern Europe led to the minting and circulation of vast quantities of silver coins. The widespread use of cash greatly increased the volume of international trade. Business procedures changed radically. The individual traveling merchant who alone handled virtually all aspects of exchange evolved into an operation involving three separate types of merchants: the sedentary merchant who ran the "home office," financing and organizing the firm's entire export-import trade; the carriers who transported goods by land and sea; and the company agents resident in cities abroad who, on the advice of the home office, looked after sales and procurements.

Commercial correspondence, unnecessary when one businessperson oversaw everything and made direct bargains with buyers and sellers, multiplied. Regular courier service among commercial cities began. Commercial accounting became more complex when firms had to deal with shareholders, manufacturers, customers, branch offices, employees, and competing firms. Tolls on roads became high enough to finance what has been called a road revolution, involving new surfaces and bridges, new passes through the Alps, and new inns and hospices for travelers. The growth of mutual trust among merchants facilitated the growth of sales on credit and led to new developments in finance, such as the bill of exchange, a device that made the long, slow, and very dangerous shipment of coins unnecessary.

The ventures of the German Hanseatic League illustrate these advancements. The Hanseatic League was a mercantile association of European towns dating from 1159. The league grew by the end of the fourteenth century to include about 200 cities from Holland to Poland. Across regular, well-defined trade routes along the Baltic and North seas, the ships of league cities carried furs, wax, copper, fish, grain, timber, and wine. These goods were exchanged for finished products, mainly cloth and salt, from western cities. At cities such as Bruges and London, Hanseatic merchants secured special trading concessions, exempting them from all tolls and allowing them to trade at local fairs. Hanseatic merchants established foreign trading centers, the most famous of which was the London Steelyard, a walled community with warehouses,

offices, a church, and residential quarters for company representatives. By the late thirteenth century, Hanseatic merchants had developed an important business technique, the business register. Merchants publicly recorded their debts and contracts and received a league guarantee for them. This device proved a decisive factor in the later development of credit and commerce in northern Europe.

These developments added up to what one modern scholar has called "a commercial revolution." In the long run, the commercial revolution of the High Middle Ages (A D 1000-1300) brought about radical change in European society. One remarkable aspect of this change was that the commercial classes constituted a small part of the total population—never more than 10 percent. They exercised an influence far in excess of their numbers. The commercial revolution created a great deal of new wealth, which meant a higher standard of living. The existence of wealth did not escape the attention of kings and other rulers. Wealth could be taxed, and through taxation, kings could create strong and centralized states. In the years to come, alliances with the middle classes were to enable kings to weaken aristocratic interests and build the states that came to be called modern.

The commercial revolution also provided the opportunity for thousands of agricultural workers to improve their social position. The slow but steady transformation of European society from almost completely rural and isolated to relatively more urban constituted the greatest effect of the commercial revolution that began in the eleventh century. Even so, merchants and business people did not run medieval communities, except in central and northern Italy and in the county of Flanders. Most towns remained small. The nobility and churchmen determined the predominant social attitudes, values, and patterns of thought and behavior. The commercial changes of the eleventh through fourteenth centuries did however, lay the economic foundation for the development of urban life and culture.

48 – 4 The Rise of Moscow

The rise of Moscow during medieval times was a fundamental development in Russian history. Moscow began with very little and for a long time could not be compared to such flourishing principalities as Novgorod or Galicia. Even in its own area, the northeast, it was junior to old centers like Rostov and Suzdal. In accounting for Moscow's rise, historians have emphasized several factors or rather groups of factors.

First, attention may be given to the doctrine of geographic causation. It stresses the decisive importance of the location on Moscow for the later expansion of the Muscovite state (the medieval state centered in Moscow) and includes several lines of argument. Moscow lay as a crossing of three roads. The most important was the way from the historically crucial city of Kiev and the declining south to the growing northeast. In fact Moscow has been described as the first stopping and setting point in the northeast. But it also profited from moments in other directions, including the reverse. Thus it seems immigrants came to Moscow after the Mongol devastation of the lands further to the northeast. Moscow was also situated on a bend of the Moscow River that flows from the northwest to the southeast into the Oka, the largest western tributary of the Volga River. To speak more broadly of water communications which span and unite European Russia, Moscow has the rare fortune of being located near the headwaters of four major rivers: the Oka, the Volga, the Don, and the Dnieper. This offered marvelous opportunities for expansion across the flowing plain, especially as there were no mountains or other natural obstacles to hem in the young principality.

In another sense too, Moscow benefited from a central position. It stood in the midst of lands inhabited by the Russian people which, so the argument runs, provided a proper setting for a natural growth in all directions. In fact some specialists have tried to estimate precisely how close to the geographic center of the Russian people Moscow was situated, noting also such circumstances as proximity to the land dividing the two main dialects of the Great Russian language. Central location within Russia, to make an additional point, cushioned Moscow from outside invaders. Thus, for example, it was the city of Novgorod, not Moscow, that continuously had to meet enemies from the northwest, while in the southeast Ryazan absorbed the first blows from the

direction. All in all, the considerable significance of the location on Moscow cannot be denied although this geographic factor has generally been assigned less relative weight by recent scholars.

The economic argument is linked in part to the geographic. The Moscow River served as an important trade artery, and as the Muscovite principality expanded around its waterways, it profited by and in turn helped to promote increasing economic intercourse. One school of thought has treated the expansion of Moscow largely in terms of the growth of a common market. Another economic approach emphasizes the success of the Muscovite princes in developing agriculture in their domains and supporting colonization. These princes clearly outdistance their rivals in obtaining peasants to settle on their lands. As a further advantage, they managed to maintain in their realm a relative peace and security highly beneficial to economic life. The last view introduces another key factor in explaining the Muscovite rise: the role of the rulers of Moscow. Moscow has generally been considered fortunate in its princes. Sheer luck constituted an important part of the picture. For several generations, the princes of Moscow had the advantage of male succession without interruption or conflict. In particular, for a long time the sons of the princes of Moscow were lucky not to have uncles competing for the Muscovite seat. When the classic power struggle between royal uncles and nephews finally erupted under Basil II (reigned 1425-1462), direct succession from father to son possessed sufficient standing and support in the principality of Moscow to overcome the challenge. The principality has also been considered fortunate because its early rulers, descending from the youngest son of Alexander Nevskii (1220?-1263) and thus representing a junior princely branch, found it expedient to devote themselves to their small holdings instead of neglecting them for more ambitious undertakings elsewhere.

48 – 1 Chinese Population Growth

Increases in population have usually been accompanied (indeed facilitated) by an increase in trade. In the Western experience, commerce provided the conditions that allowed industrialization to get started, which in turn led to growth in science, technology, industry, transport, communications, social change, and the like that we group under the broad term of “development.” However, the massive increase in population that in Europe was at first attributed to industrialization starting in the eighteenth century occurred also and at the same period in China, even though there was no comparable industrialization.

It is estimated that the Chinese population by 1600 was close to 150 million. The transition between the Ming and Qing dynasties (the seventeenth century) may have seen a decline, but from 1741 to 1851 the annual figures rose steadily and spectacularly, perhaps beginning with 143 million and ending with 432 million. If we accept these totals, we are confronted with a situation in which the Chinese population doubled in the 50 years from 1790 to 1840. If, with greater caution, we assume lower totals in the early eighteenth century and only 400 million in 1850, we still face a startling fact: something like a doubling of the vast Chinese population in the century before Western contact, foreign trade, and industrialization could have had much effect.

To explain this sudden increase we cannot point to factors constant in Chinese society but must find conditions or a combination of factors that were newly effective in this period. Among these is the almost complete internal peace maintained under Manchu rule during the eighteenth century. There was also an increase in foreign trade through Guangzhou (southern China) and some improvement of transportation within the empire. Control of disease, like the checking of smallpox by variolation may have been important. But of most critical importance was the food supply.

Confronted with a multitude of unreliable figures, economists have compared the population records with the aggregate data for cultivated land area and grain production in the six centuries since 1368. Assuming that China's population in 1400 was about 80 million, the economist Dwight Perkins concludes that its growth to 700 million or more in the 1960s was made possible by a steady increase in the grain supply, which evidently grew five or six

times between 1400 and 1800 and rose another 50 percent between 1800 and 1965. This increase of food supply was due perhaps half to the increase of cultivated area, particularly by migration and settlement in the central and western provinces, and half to greater productivity—the farmers' success in raising more crops per unit of land.

This technological advance took many forms: one was the continual introduction from the south of earlier-ripening varieties of rice, which made possible double-cropping (the production of two harvests per year from one field). New crops such as corn (maize) and sweet potatoes as well as peanuts and tobacco were introduced from the Americas. Corn, for instance, can be grown on the dry soil and marginal hill land of North China, where it is used for food, fuel, and fodder and provides something like one-seventh of the food energy available in the area. The sweet potato, growing in sandy soil and providing more food energy per unit of land than other crops, became the main food of the poor in much of the South China rice area.

Productivity in agriculture was also improved by capital investments, first of all in irrigation. From 1400 to 1900 the total of irrigated land seems to have increased almost three times. There was also a gain in farm tools, draft animals, and fertilizer, to say nothing of the population growth itself, which increased half again as fast as cultivated land area and so increased the ratio of human hands available per unit of land. Thus the rising population was fed by a more intensive agriculture, applying more labor and fertilizer to the land.

50 – 2 The Achievement of Brazilian Independence

In contrast to the political anarchy, economic dislocation, and military destruction in Spanish America, Brazil's drive toward independence from Portugal proceeded as a relatively bloodless transition between 1808 and 1822. The idea of Brazilian independence first arose in the late eighteenth century as a Brazilian reaction to the Portuguese policy of tightening political and economic control over the colony in the interests of the mother country. The first significant conspiracy against Portuguese rule was organized from 1788-1799 in the province of Minas Gerais, where rigid governmental control over the production and prices of gold and diamonds, as well as heavy taxes, caused much discontent. But this conspiracy never went beyond the stage of discussion and was easily discovered and crushed. Other conspiracies in the late eighteenth century as well as a brief revolt in 1817 reflected the influence of republican ideas over sections of the elite and even the lower strata of urban society. All proved abortive or were soon crushed. Were it not for an accident of European history, the independence of Brazil might have been long delayed.

The French invasion of Portugal in 1807 followed by the flight of the Portuguese court (sovereign and government officers) to Rio de Janeiro brought large benefits to Brazil. Indeed, the transfer of the court in effect signified achievement of Brazilian independence. The Portuguese prince and future King Joao VI opened Brazil's ports to the trade of friendly nations, permitted the rise of local industries, and founded the Bank of Brazil. In 1815 he elevated Brazil to the legal status of a kingdom coequal with Portugal. In one sense, however, Brazil's new status signified the substitution of one dependence for another. Freed from Portuguese control, Brazil came under the economic dominance of England, which obtained major tariff concessions and other privileges by the Strangford Treaty of 1810 between Portugal and Great Britain. The treaty provided for the importation of British manufactures into Brazil and the export of Brazilian agricultural produce to Great Britain. One result was an influx of cheap machine-made goods that swamped the handicrafts industry of the country.

Brazilian elites took satisfaction in Brazil's new role and the growth of educational, cultural, and economic opportunities for their class. But the feeling was mixed with resentment toward the

thousands of Portuguese courtiers (officials) and hangers-on who came with the court and who competed with Brazilians for jobs and favors. Thus, the change in the status of Brazil sharpened the conflict between Portuguese elites born in Brazil and elites born in Portugal and loyal to the Portuguese crown.

The event that precipitated the break with the mother country was the revolution of 1820 in Portugal. The Portuguese revolutionaries framed a liberal constitution for the kingdom, but they were conservative or reactionary in relation to Brazil. They demanded the immediate return of King Joao to Lisbon, an end to the system of dual monarchy that he had devised, and the restoration of the Portuguese commercial monopoly. Timid and vacillating, King Joao did not know which way to turn. Under the pressure of his courtiers, who hungered to return to Portugal and their lost estates, he finally approved the new constitution and sailed for Portugal. He left behind him, however, his son and heir, Pedro, and in a private letter advised him that in the event the Brazilians should demand independence, he should assume leadership of the movement and set the crown of Brazil on his head.

Soon it became clear that the Portuguese parliament intended to set the clock back by abrogating all the liberties and concessions won by Brazil since 1808. One of its decrees insisted on the immediate return of Pedro from Brazil. The pace of events moved more rapidly in 1822. On January 9, urged on by Brazilian advisers who perceived a golden opportunity to make an orderly transition to independence without the intervention of the masses, Pedro refused an order from the parliament to return to Portugal, saying famously, "I remain." On September 7, regarded by all Brazilians as Independence Day, he issued the even more celebrated proclamation, "Independence or death!" In December 1822, having overcome slight resistance by Portuguese troops, Dom Pedro was formally proclaimed constitutional Emperor of Brazil.

51 – 3 Population Growth in Nineteenth-Century Europe

Because of industrialization, but also because of a vast increase in agricultural output without which industrialization would have been impossible, Western Europeans by the latter half of the nineteenth century enjoyed higher standards of living and longer, healthier lives than most of the world's peoples. In Europe as a whole, the population rose from 188 million in 1800 to 400 million in 1900. By 1900, virtually every area of Europe had contributed to the tremendous surge of population, but each major region was at a different stage of demographic change.

Improvements in the food supply continued trends that had started in the late seventeenth century. New lands were put under cultivation, while the use of crops of American origin, particularly the potato, continued to expand. Setbacks did occur. Regional agricultural failures were the most common cause of economic recessions until 1850, and they could lead to localized famine as well. A major potato blight (disease) in 1846-1847 led to the deaths of at least one million persons in Ireland and the emigration of another million, and Ireland never recovered the population levels the potato had sustained to that point. Bad grain harvests at the same time led to increased hardship throughout much of Europe.

After 1850, however, the expansion of foods more regularly kept pace with population growth, though the poorer classes remained malnourished. Two developments were crucial. First, the application of science and new technology to agriculture increased. Led by German universities, increasing research was devoted to improving seeds, developing chemical fertilizers, and advancing livestock. After 1861, with the development of land-grant universities in the United States that had huge agricultural programs, American crop-production research added to this mix. Mechanization included the use of horse-drawn harvesters and seed drills, many developed initially in the United States. It also included mechanical cream separators and other food-processing devices that improved supply.

The second development involved industrially based transportation. With trains and steam shipping, it became possible to move foods to needy regions within Western Europe quickly. Famine (as opposed to malnutrition) became a

thing of the past. Many Western European countries, headed by Britain, began also to import increasing amounts of food, not only from Eastern Europe, a traditional source, but also from the Americas, Australia, and New Zealand. Steam shipping, which improved speed and capacity, as well as new procedures for canning and refrigerating foods (particularly after 1870), was fundamental to these developments.

Europe's population growth included one additional innovation by the nineteenth century: it combined with rapid urbanization. More and more Western Europeans moved from countryside to city, and big cities grew most rapidly of all. By 1850, over half of all the people in England lived in cities, a first in human history. In one sense, this pattern seems inevitable: growing numbers of people pressed available resources on the land, even when farmwork was combined with a bit of manufacturing, so people crowded into cities seeking work or other resources. Traditionally, however, death rates in cities surpassed those in the countryside by a large margin; cities had maintained population only through steady immigration. Thus rapid urbanization should have reduced overall population growth, but by the middle of the nineteenth century this was no longer the case. Urban death rates remained high, particularly in the lower-class slums, but they began to decline rapidly.

The greater reliability of food supplies was a factor in the decline of urban death rates. Even more important were the gains in urban sanitation, as well as measures such as inspection of housing. Reformers, including enlightened doctors, began to study the causes of high death rates and to urge remediation. Even before the discovery of germs, beliefs that disease spread by "miasmas" (noxious forms of bad air) prompted attention to sewers and open garbage; Edwin Chadwick led an exemplary urban crusade for underground sewers in England in the 1830s. Gradually, public health provisions began to cut into customary urban mortality rates. By 1900, in some parts of Western Europe life expectancy in the cities began to surpass that of the rural areas. Industrial societies had figured out ways to combine large and growing cities with population growth, a development that would soon spread to other parts of the world.

50 – 1 American Railroads

In the United States, railroads spearheaded the second phase of the transportation revolution by overtaking the previous importance of canals. The mid-1800s saw a great expansion of American railroads. The major cities east of the Mississippi River were linked by a spider web of railroad tracks. Chicago's growth illustrates the impact of these rail links. In 1849 Chicago was a village of a few hundred people with virtually no rail service. By 1860 it had become a city of 100,000, served by eleven railroads. Farmers to the north and west of Chicago no longer had to ship their grain, livestock, and dairy products down the Mississippi River to New Orleans; they could now ship their products directly east. Chicago supplanted New Orleans as the interior of America's main commercial hub.

The east-west rail lines stimulated the settlement and agricultural development of the Midwest. By 1860, Illinois, Indiana, and Wisconsin had replaced Ohio, Pennsylvania, and New York as the leading wheat-growing states. Enabling farmers to speed their products to the East, railroads increased the value of farmland and promoted additional settlement. In turn, population growth in agricultural areas triggered industrial development in cities such as Chicago, Davenport (Iowa), and Minneapolis, for the new settlers needed lumber for fences and houses and mills to grind wheat into flour.

Railroads also propelled the growth of small towns along their routes. The Illinois Central Railroad, which had more track than any other railroad in 1855, made money not only from its traffic but also from real estate speculation. Purchasing land for stations along its path, the Illinois Central then laid out towns around the stations. The selection of Manteno, Illinois, as a stop of the Illinois Central, for example, transformed the site from a crossroads without a single house in 1854 into a bustling town of nearly a thousand in 1860, replete with hotels, lumberyards, grain elevators, and gristmills. By the Civil War (1861-1865), few thought of the railroad-linked Midwest as a frontier region or viewed its inhabitants as pioneers.

As the nation's first big business, the railroads transformed the conduct of business. During the early 1830s, railroads, like canals, depended on financial aid from state governments. With the onset of economic depression in the late 1830s, however, state governments scrapped overly

ambitious railroad projects. Convinced that railroads burdened them with high taxes and blasted hopes, voters turned against state aid, and in the early 1840s, several states amended their constitutions to bar state funding for railroads and canals. The federal government took up some of the slack, but federal aid did not provide a major stimulus to railroads before 1860. Rather, part of the burden of finance passed to city and county governments in agricultural areas that wanted to attract railroads. Such municipal governments, for example, often gave railroads rights-of-way, grants of land for stations, and public funds.

The dramatic expansion of the railroad network in the 1850s, however, strained the financing capacity of local governments and required a turn toward private investment, which had never been absent from the picture. Well aware of the economic benefits of railroads, individuals living near them had long purchased railroad stock issued by governments and had directly bought stock in railroads, often paying by contributing their labor to building the railroads. But the large railroads of the 1850s needed more capital than such small investors could generate. Gradually, the center of railroad financing shifted to New York City, and in fact, it was the railroad boom of the 1850s that helped make Wall Street in New York City the nation's greatest capital market. The stocks of all the leading railroads were traded on the floor of the New York Stock Exchange during the 1850s. In addition, the growth of railroads turned New York City into the center of modern investment firms. The investment firms evaluated the stock of railroads in the smaller American cities and then found purchasers for these stocks in New York City, Philadelphia, Paris, London, Amsterdam, and Hamburg. Controlling the flow of funds to railroads, the investment bankers began to exert influence over the railroads' internal affairs by supervising administrative reorganizations in times of trouble.

52 – 3 Early Food Production in Sub-Saharan Africa

At the end of the Pleistocene (around 10,000 B.C.), the technologies of food production may have already been employed on the fringes of the rain forests of western and central Africa, where the common use of such root plants as the African yam led people to recognize the advantages of growing their own food. The yam can easily be resprouted if the top is replanted. This primitive form of "vegiculture" (cultivation of root and tree crops) may have been the economic tradition onto which the cultivation of summer rainfall cereal crops was grafted as it came into use south of the grassland areas on the Sahara's southern borders.

As the Sahara dried up after 5000 B.C., pastoral peoples (cattle herders) moved southward along major watercourses into the savanna belt of West Africa and the Sudan. By 3000 B.C., just as ancient Egyptian civilization was coming into being along the Nile, they had settled in the heart of the East African highlands far to the south. The East African highlands are ideal cattle country and the home today of such famous cattle-herding peoples as the Masai. The highlands were inhabited by hunter-gatherers living around mountains near the plains until about 3300 B.C., when the first cattle herders appeared. These cattle people may have moved between fixed settlements during the wet and dry seasons, living off hunting in the dry months and their own livestock and agriculture during the rains.

As was the case elsewhere, cattle were demanding animals in Africa. They required water at least every 24 hours and large tracts of grazing grass if herds of any size were to be maintained. The secret was the careful selection of grazing land, especially in environments where seasonal rainfall led to marked differences in graze quality throughout the year. Even modest cattle herds required plenty of land and considerable mobility. To acquire such land often required moving herds considerable distances, even from summer to winter pastures. At the same time, the cattle owners had to graze their stock in tsetse-fly-free areas. The only protection against human and animal sleeping sickness, a disease carried by the tsetse fly, was to avoid settling or farming such areas—a constraint severely limiting the movements of cattle-owning farmers in eastern and central Africa. As a result, small cattle herds spread south rapidly in areas where they could be grazed. Long before cereal agriculture took hold far south

of the Sahara, some hunter-gatherer groups in the savanna woodlands of eastern and southern Africa may have acquired cattle, and perhaps other domesticated animals, by gift exchange or through raids on herding neighbors.

Contrary to popular belief: there is no such phenomenon as "pure" pastoralists, a society that subsists on its herds alone. The Saharan herders who moved southward to escape drought were almost certainly also cultivating sorghum, millet; and other tropical rainfall crops. By 1500 B.C., cereal agriculture was widespread throughout the savanna belt south of the Sahara. Small farming communities dotted the grasslands and forest margins of eastern West Africa, all of them depending on what is called shifting agriculture. This form of agriculture involved clearing woodland, burning the felled brush over the cleared plot, mixing the ash into the soil, and then cultivating the prepared fields. After a few years, the soil was exhausted, so the farmer moved on, exploiting new woodland and leaving the abandoned fields to lie fallow. Shifting agriculture, often called slash-and-burn, was highly adaptive for savanna farmers without plows, for it allowed cereal farming with the minimal expenditure of energy.

The process of clearance and burning may have seemed haphazard to the uninformed eye, but it was not. Except in favored areas, such as regularly inundated floodplains: tropical Africa's soils were of only moderate to low fertility. The art of farming was careful soil selection, that is, knowing which soils were light and easily cultivable, could be readily turned with small hoes, and would maintain their fertility over several years' planting, for cereal crops rapidly remove nitrogen and other nutrients from the soil. Once it had taken hold: slash-and-burn agriculture expanded its frontiers rapidly as village after village took up new lands, moving forward so rapidly that one expert has estimated it took a mere two centuries to cover 2,000 kilometers from eastern to southern Africa.

52 – 4 England's Economy in sixteenth century

In the last half of the sixteenth century England emerged as a commercial and manufacturing power in Europe due to a combination of demographic, agricultural and industrial factors. The population of England and Wales grew rapidly from about 2.5 million in the 1520s to more than 3.5 million in 1580, reaching about 4.5 million in 1610. Reduced mortality rates and increased fertility, the latter probably generated by expanding work opportunities in manufacturing and farming (leading to earlier marriage and more children), explained this rapid rise in population. While epidemics and plague occasionally took their toll, the people in England still suffered less than did those in continental Europe. Furthermore, the country had been pulled out of the war that occurred in France and central Europe during the same period.

England provides the prominent example of the expansion of agricultural production well before the general European agricultural revolution of the eighteenth and nineteenth centuries. A larger population stimulated the increased woollen through crop civilization. English agriculture became more efficient and market-oriented than almost anywhere else on the continent. Between 1450 and 1640 the yield of grain per acre increased by at least thirty percent. In sharp contrast with farming in Spain, English land owners brought more dense marshes and woodlands into cultivation.

The great land estates of the English society largely remained intact and many wealthy land owners aggressively increased the size of their holdings, a precondition for increased productivity. Marriages between the children of landowners also increased the size of land estates. Primogeniture (the full inheritance of land by the eldest son) helped prevent land from being subdivided. Younger sons of independent land owners left the family and went to find other respective locations. Larger farms were conducive more to commercialized farming at the time when an expanding population pushed up demand and prices. Farmland owners turned part of their land into pasture land for sheep in order to adapt to developing woollen trade.

Some of the great land owners as well as Yeomen (farmers whose holdings and security of land tenure guaranteed their prosperity and status),

organized their holdings in the interest efficiency. Many farmers selected crops for sales in growing London market. In their quest for greater profits, many land owners put their squeeze on their tenants. Between 1580 and 1620 land lords raised rents and altered conditions of land tenure in their favor, preferring shorter phases and forcing tenants to pay an entry fee before agreeing to rent them land. Landlords evicted those who could not afford annual, more onerous terms. But they also pushed tenants toward more productive farming methods, including crop rotation.

England's exceptional economic development also drew the country's natural resources, including iron, timber, and coal, extracted in far greater quantity than elsewhere in the continent. New industrial development expanded the production of iron and pewter in and around the city of Birmingham. But above all textile manufacturing transformed English economy. Woolens, which accounted for eighty percent of the exports, worsteds (sturdy yarn spun from combed wool fibers), and other cloth found eager buyers in England as well as in the continent. Moreover, late in the sixteenth century as English merchants began making forays across the Atlantic these textiles were also sold in the Americas. Cloth manufacturers undercut production by urban craftspeople by "putting out" work to the villages and farms of the countryside. In such domestic industry poor rural women could spin and make cading (combing fibers in preparation for spin) in their homes.

The English textile trade was closely tied to Antwerp, in the Spanish Netherlands, where workers dyed English cloth. The entrepreneur Sir Thomas Gresham became England's representative there. He so enhanced the reputation of English business in that region that English merchants could operate on credit---the most prominent achievement for sixteenth century. He also advised the government to explore the economic possibilities of Americas, which led to the first concerted efforts at colonization, undertaken with commercial profits in mind.

51 – 1 Memphis: United Egypt's First Capital

The city of Memphis, located on the Nile near the modern city of Cairo, was founded around 3100 B.C. as the first capital of a recently united Egypt. The choice of Memphis by Egypt's first kings reflects the site's strategic importance. First, and most obvious, the apex of the Nile River delta was a politically opportune location for the state's administrative center, standing between the united lands of Upper and Lower Egypt and offering ready access to both parts of the country. The older predynastic (pre-3100 B.C.) centers of power, This and Hierakonpolis, were too remote from the vast expanse of the delta, which had been incorporated into the unified state. Only a city within easy reach of both the Nile valley to the south and the more spread out, difficult terrain to the north could provide the necessary political control that the rulers of early dynastic Egypt (roughly 3000-2600 B.C.) required.

The region of Memphis must have also served as an important node for transport and communications, even before the unification of Egypt. The region probably acted as a conduit for much, if not all, of the river-based trade between northern and southern Egypt. Moreover, commodities (such as wine, precious oils, and metals) imported from the Near East by the royal courts of predynastic Upper Egypt would have been channeled through the Memphis region on their way south. In short, therefore, the site of Memphis offered the rulers of the Early Dynastic Period an ideal location for controlling internal trade within their realm, an essential requirement for a state-directed economy that depended on the movement of goods.

Equally important for the national administration was the ability to control communications within Egypt. The Nile provided the easiest and quickest artery of communication and the national capital was, again, ideally located in this respect. Recent geological surveys of the Memphis region have revealed much about its topography in ancient times. It appears that the location of Memphis may have been even more advantageous for controlling trade, transport, and communications than was previously appreciated. Surveys and drill cores have shown that the level of the Nile floodplain has steadily risen over the last five millennia. When the floodplain was much lower, as it would have been in predynastic and early dynastic times, the outwash fans (fan-shaped deposits of sediments)

of various wadis (stream-beds or channels that carry water only during rainy periods) would have been much more prominent features on the east bank. The fan associated with the Wadi Hof extended a significant way into the Nile floodplain, forming a constriction in the vicinity of Memphis. The valley may have narrowed at this point to a mere three kilometers, making it the ideal place for controlling river traffic.

Furthermore, the Memphis region seems to have been favorably located for the control not only of river-based trade but also of desert trade routes. The two outwash fans in the area gave access to the extensive wadi systems of the eastern desert. In predynastic times, the Wadi Digla may have served as a trade route between the Memphis region and the Near East, to judge from the unusual concentration of foreign artifacts found in the predynastic settlement of Maadi. Access to, and control of, trade routes between Egypt and the Near East seems to have been a preoccupation of Egypt's rulers during the period of state formation. The desire to monopolize foreign trade may have been one of the primary factors behind the political unification of Egypt. The foundation of the national capital at the junction of an important trade route with the Nile valley is not likely to have been accidental. Moreover, the Wadis Hof and Digla provided the Memphis region with accessible desert pasturage. As was the case with the cities of Hierakonpolis and Elkab, the combination within the same area of both desert pasturage and alluvial arable land (land suitable for growing crops) was a particularly attractive one for early settlement; this combination no doubt contributed to the prosperity of the Memphis region from early predynastic times.

54 – 4 Documenting the Incas

The Incas ruled a vast empire in western South America when the Spaniards encountered them in the sixteenth century. Although the Incas had no writing system of their own, historical information about Incas is available to researchers because early Spaniards wrote documents about them. However, there are drawbacks to use the written record. First, the Spanish writers were describing activities and institutions that were very different from their own, but they often described Inca culture in terms of their own society. As an example, consider the list of kings given by the Incas. As presented in the historical chronology, Spanish sources indicate there were thirteen kings who ruled sequentially. The names were given to them by Inca informants.

However, one school of thought in Inca studies suggests that the names were not actual people, but, rather, titles filled by different individuals. Thus, the number of actual kings may have been fewer, and several titles may have been filled at the same time. The early Spanish writers, being unfamiliar with such a system of titles, simply translated it into something they were familiar with (a succession of kings). Given that the Inca empire expanded only during the time of the last four kings, or as a result of the actions of the individuals in those four positions, this question is not deemed significant for an understanding of the Incas. But the example shows that biases and inaccuracies may have been introduced inadvertently from the very beginning of the written Spanish reports about the Incas. Moreover, early writers often copied information from each other - so misinformation was likely to be passed on and accepted as true by later scholars.

Second, both Spanish writers and Incan informants sometimes had motives for being deliberately deceitful. For example, in an effort to gain status in the Spaniards' eyes, Incas might say that they formerly had been more important in the Inca empire than they actually were. Spanish officials as well were occasionally untruthful when it served their purposes. For example, Spaniards might deliberately underreport the productivity of a region under their authority so they could sell the additional products and keep the money, rather than hand it over to the Spanish Crown.

Third, it should be noted that the Spaniards' main sources of information were the Incas themselves,

often members of the Inca ruling class. Therefore, what was recorded was the Incas' point of view about their own history and empire. Some modern authorities question whether the history of Incas happened as they said it did. Although some of their history is certainly more myth than truth, many, if not most, scholars agree that the history of the last four Inca kings is probably accurate. The same is true of other things told to the Spanish writers: the more recently an event is said to have occurred, the more likely it is to have actually happened.

A fourth problem relates to the nature of the Inca conquests of the other people in the Americas before the Spanish arrived and how accurate the accounts of those conquests are - whether related by the Spaniards or by the Incas on whom they relied. It was certainly in the Inca's interest to describe themselves as invincible and just. However, lacking accounts by conquered people about their interactions with the Incas, it is unknown how much of the information of the Inca conquest as related by the ruling class is factual.

Finally, there is a certain vagueness in the historical record regarding places and names. Many Spanish writers listed places they had visited within the empire, including both provinces and towns. However, other writers traveling along the same routes sometimes recounted different lists of places. In addition, it is difficult to identify the exact locations of towns and other geographic points of reference because of the widespread movements of people over the past five centuries.

For all these reasons, the historical record must be carefully evaluated to determine whether it is accurate and to verify the locations of past events. One approach is to cross-check information from a number of authors. Another approach is to conduct archaeological research. Regardless of the problems, historical documents review some important information about the Incas.

54 – 1 The Commercialization of Lumber

In nineteenth-century America, practically everything that was built involved wood. Pine was especially attractive for building purposes. It is durable and strong, yet soft enough to be easily worked with even the simplest of hand tools. It also floats nicely on water, which allowed it to be transported to distant markets across the nation. The central and northern reaches of the Great Lakes states—Michigan, Wisconsin, and Minnesota—all contained extensive pine forests as well as many large rivers for floating logs into the Great Lakes, from where they were transported nationwide.

By 1860, the settlement of the American West along with timber shortages in the East converged with ever-widening impact on the pine forests of the Great Lakes states. Over the next 30 years, lumbering became a full-fledged enterprise in Michigan, Wisconsin, and Minnesota. Newly formed lumbering corporations bought up huge tracts of pineland and set about systematically cutting the trees. Both the colonists and the later industrialists saw timber as a commodity, but the latter group adopted a far more thorough and calculating approach to removing trees. In this sense, what happened between 1860 and 1890 represented a significant break with the past. No longer were farmers in search of extra income, the main source for shingles, firewood, and other wood products. By the 1870s, farmers and city dwellers alike purchased forest products from large manufacturing companies located in the Great Lakes states rather than chopping wood themselves or buying it locally.

The commercialization of lumbering was in part the product of technological change. The early, thick saw blades tended to waste a large quantity of wood, with perhaps as much as a third of the log left behind on the floor as sawdust or scrap. In the 1870s, however, the British-invented band saw, with its thinner blade, became standard issue in the Great Lakes states' lumber factories. Meanwhile, the rise of steam-powered mills streamlined production by allowing for the more efficient, centralized, and continuous cutting of lumber. Steam helped to automate a variety of tasks, from cutting to the carrying away of waste. Mills also employed steam to heat log ponds, preventing them from freezing and making possible year-round lumber production.

For industrial lumbering to succeed, a way had to be found to neutralize the effects of the seasons on production. Traditionally, cutting took place in the winter, when snow and ice made it easier to drag logs on sleds or sleighs to the banks of streams. Once the streams and lakes thawed, workers rafted the logs to mills, where they were cut into lumber in the summer. If nature did not cooperate—if the winter proved dry and warm, if the spring thaw was delayed—production would suffer. To counter the effects of climate on lumber production, loggers experimented with a variety of techniques for transporting trees out of the woods. In the 1870s, loggers in the Great Lakes states began sprinkling water on sleigh roads, giving them an artificial ice coating to facilitate travel. The ice reduced the friction and allowed workers to move larger and heavier loads.

But all the sprinkling in the world would not save a logger from the threat of a warm winter. Without snow the sleigh roads turned to mud. In the 1870s, a set of snowless winters left lumber companies to ponder ways of liberating themselves from the seasons. Railroads were one possibility. At first, the remoteness of the pine forests discouraged common carriers from laying track. But increasing lumber prices in the late 1870s combined with periodic warm, dry winters compelled loggers to turn to iron rails. By 1887, 89 logging railroads crisscrossed Michigan, transforming logging from a winter activity into a year-round one.

Once the logs arrived at a river, the trip downstream to a mill could be a long and tortuous one. Logjams (buildups of logs that prevent logs from moving downstream) were common—at times stretching for 10 miles—and became even more frequent as pressure on the northern Midwest pinelands increased in the 1860s. To help keep the logs moving efficiently, barriers called booms (essentially a chain of floating logs) were constructed to control the direction of the timber. By the 1870s, lumber companies existed in all the major logging areas of the northern Midwest.

56 – 2 Representative Government in Colonial North America

Before 1750, colonists in North America had little occasion to think of themselves as a distinct people. There was no American government, no single political organization in which all the colonies joined to manage their common concerns. There was not even a wish for such an organization except among a few eccentric individuals America, to the people who lived in it, was still a geographic region, not a frame of mind.

Asked about nationality, the typical American colonist of 1750 would have said English or British. In spite of substantial numbers of Dutch, Germans, and Scotch-Irish, English people and English institutions prevailed in every colony, and most colonists spoke of England as home even though they had never been there. Yet no American institutions were quite like their counterparts in England: the heritage of English ideas that went with these institutions was so rich and varied that colonists were able to select and develop those that best suited their situation and forget others that meanwhile were growing prominent in the mother country. This variety sometimes led to regional differences: in some ways New Englanders were set off from Virginians even more than from people in England. But some ideas, institutions, and attitudes became common in all the colonies and remained uncommon in England. Although colonial Englishmen were not yet aware that they shared these Americanisms with one another or that English people in England did not share them, many of the characteristic ideas and attitudes that later distinguished United States nationalism were already present by the mid-eighteenth century.

English people brought with them to the New World the political ideas that still give English and American government a close resemblance. But American colonists very early developed conceptions of representative government that differed from those in England. Representative government in England originated in the Middle Ages, when the king called for men to advise him. They were chosen by their neighbors and informed the king of his subjects' wishes. Eventually, their advice became so compelling that the king could not reject it, and the representatives of the people, organized as a legislature known as the House of Commons, became the most powerful branch of the English government.

At first, the House of Commons consisted of representatives from each county, or shire, and from selected boroughs. Over the centuries many of these boroughs became ghost towns with only a handful of inhabitants, and great towns sprang up where none had existed before. Yet the old boroughs continued to send members to the House of Commons, and the new towns sent none. Moreover, only a fraction of the English population could participate even in county elections. In order to vote, a man had to own property that would, if rented, yield him at least 40 shillings yearly. Few could meet the test. A number of English people thought the situation absurd and said so. But nothing was done to improve it; in fact, a theory was devised to justify it. A member of the House of Commons, it was said, represented not the people who chose him but the whole country, and he was not responsible for any particular constituency. Not all Englishmen could vote for representatives, but all were virtually represented by every member of the Commons.

The assemblies of American colonial representatives were more democratic. Although every colony had property qualifications for voting, probably the great majority of adult white males owned enough land to meet them. Moreover, the system for apportioning representation was more balanced. New England colonies gave every town the right to send delegates to the legislature. Outside New England, the unit of representation was usually the county. The political organization of new counties and the extension of representation seldom kept pace with the rapid advance of settlement westward, but nowhere was representation so uneven or irrational as in England.

American colonists knew nothing of virtual representation. A colonial representative was supposed to be an agent of the people who chose him. He was supposed to look after their interests first and those of the colony second. In New England, where town meetings could be called at any time, people often gathered to tell their delegate how to vote on a particular issue.

58 – 3 The Rise of Classic Maya Civilization

Prior to the beginning of the Late Preclassic period in 300 B.c, Maya ceremonial centers of relatively substantial size had already appeared in Central America. The ceremonial center was a distinctive feature of Maya culture, acting as a focus for the community. Generally speaking, these centers were not what we would call cities. Although they did consist of a number of large and varied buildings, they did not have a substantial resident population. Some scholars have even labeled these Maya centers “vacant towns.” Their permanent population consisted mainly of rulers, priests, and their attendants plus a limited number of artisans. The elite lived in big houses or in palaces in and around the center. The bulk of the peasant population lived in much more modest wood and thatch homes in the areas surrounding the centers. At certain times of the year, on the occasion of major religious festivals such as the one at the time of the planting of their crops, scholars hypothesize, by analogy to historical and modern practices, that the peasants would flock to the centers to observe and participate in the ceremonies. At other times of the year, some of the peasants would be called into the centers to help in the construction of new temples and palaces dedicated to the glory of the gods and to the comforts of their earthly representatives, the priestly rulers. The peasants also would provide the food to help sustain the elite in the centers.

What the peasants received from the elite was certainly not as tangible as the services they provided. In return for food and labor, the peasants were offered a psychologically and spiritually secure and ordered world, as well as access to some trade goods. Apparently, this was enough. Agriculture in the tropical Maya lowlands was at best a chancy business even slight shifts in the onset of the rainy season or the dry season could mean disaster for that year's harvest. The religion of the ancient Maya helped the peasants cope with their precarious lives. If the gods were properly propitiated, the crops would be good-as would life in general.

One archaeologist, William Haviland, argues that it was the centralizing effects of Maya religion that led to the rise of Classic Maya civilization. He believes that the religious centers acted as magnets to peoples living in the surrounding areas. To support the growing populations around the centers, Haviland argues, the agricultural systems

became intensified. This led to the evolution of a complex state. Haviland believes that as early as 200 B.c., the “vacant” ceremonial centers at Tikal had begun transformation toward urban centers. Moreover, by this time or even earlier, other centers with huge, labor-intensive buildings had begun to emerge. Sites such as Nakbe, El Mirador, and Lamanai may have approached urban dimensions in the last few centuries of the first millennium B.C.

Other scholars attribute the growth of the ceremonial centers to trade. William Rathje argues that the basic cause for the rise of Maya civilization was the necessity for the Maya, who lived in resource-poor lowlands, to trade with adjacent highlanders for materials such as obsidian (a rock used to make ornaments and cutting edges), salt, and hard stone for grinding implements. Rathje maintains that lowland sites such as Tikal, which were quite distant from the highland resource areas, were made bigger and architecturally magnificent to attract highland merchants and their trade. In order to undertake all the building, the elite had to attract more artisans and bring more laborers into the community to do the work. These population increases led in turn to even more building, population growth, greater population density, greater social differentiation, and occupational specialization. Critics have argued, however, that there were local substitutes for the external “necessities” and that foreign trade was present well before the rise of complex architecture.

Yet other factors beyond ideology and trade must have been important in the emerging Classic Maya civilization. Competition among the growing number of centers may also have played a key role in the growth of social, economic, and political complexity, as the organization of the centers grew to meet the pressures of other centers for new agricultural land and control of rising populations.

60 – 4 A Model of Urban Expansion

In the early twentieth century, the science of sociology found supporters in the United States and Canada partly because the cities there were growing so rapidly. It often appeared that North American cities would be unable to absorb all the new comers arriving in such large numbers. Presociological thinkers like Frederick Law Olmsted, the founder of the movement to build parks and recreation areas in cities, and Jacob Riis, an advocate of slum reform, urged the nation's leaders to invest in improving the urban environment, building parks and beaches, and making better housing available to all. These reform efforts were greatly aided by sociologists who conducted empirical research on the social conditions in cities. In the early twentieth century, many sociologists lived in cities like Chicago that were characterized by rapid population growth and serious social problems. It seemed logical to use empirical research to construct theories about how cities grow and change in response to major social forces as well as more controlled urban planning.

The founders of the Chicago school of sociology, Robert Park and Ernest Burgess, attempted to develop a dynamic model of the city, one that would account not only for the expansion of cities in terms of population and territory but also for the patterns of settlement and land use within cities. They identified several factors that influence the physical form of cities. As Park stated, among them are "transportation and communication, tramways and telephones, newspapers and advertising, steel construction and elevators—all things, in fact, which tend to bring about at once a greater mobility and a greater concentration of the urban populations."

Park and Burgess based their model of urban growth on the concept of "natural areas"—that is, areas such as occupational suburbs or residential enclaves in which the population is relatively homogeneous and land is used in similar ways without deliberate planning. Park and Burgess saw urban expansion as occurring through a series of "invasions" of successive zones or areas surrounding the center of the city. For example, people from rural areas and other societies "invaded" areas where housing was inexpensive. Those areas ended to be close to the places where they worked. In turn, people who could afford better housing and the cost of commuting "invaded" areas farther from the business district.

Park and Burgess's model has come to be known as the "concentric-zone model" (represented by the figure). Because the model was originally based on studies of Chicago, its center is labeled "Loop," the term commonly applied to that city's central commercial zone. Surrounding the central zone is a "zone in transition," an area that is being invaded by business and light manufacturing. The third zone is inhabited by workers who do not want to live in the factory or business district but at the same time need to live reasonably close to where they work. The fourth or residential zone consists of upscale apartment buildings and single-family homes. And the outermost ring, outside the city limits, is the suburban or commuters' zone; its residents live within a 30- to 60-minute ride of the central business district.

Studies by Park, Burgess, and other Chicago-school sociologists showed how new groups of immigrants tended to be concentrated in separate areas within inner-city zones, where sometimes experienced tension with other ethnic groups that had arrived earlier. Over time, however, each group was able to adjust to life in the city and to find a place for itself in the urban economy. Eventually many of the immigrants moved to unsegregated areas in outer zones; the areas they left behind were promptly occupied by new waves of immigrants.

The Park and Burgess model of growth in zones and natural areas of the city can still be used to describe patterns of growth in cities that were built around a central business district and that continue to attract large numbers of immigrants. But this model is biased toward the commercial and industrial cities of North America, which have tended to form around business centers rather than around palaces or cathedrals, as is often the case in some other parts of the world. Moreover, it fails to account for other patterns of urbanization, such as the rapid urbanization that occurs along commercial transportation corridors and the rise of nearby satellite cities.

63 – 1 The Sumerians and Regional Interdependence

With the emergence of the Sumerian civilization in about 3100 B.C., a new era in human experience began—one in which the economic, political, and social mechanisms created by humans began to affect the lives of cities, towns, and villages located hundreds and perhaps thousands of miles apart. In a real sense, a rapidly evolving world system linked hundreds of Southwest Asian societies all the way from modern-day eastern Iran to the eastern Mediterranean and the Nile Valley. This nascent world system developed as a result of insatiable demands for nonlocal raw materials in different ecological regions where societies were developing along very similar evolutionary tracks toward greater complexity. In each area, social developments and technological innovations were triggered not only by basic economic needs but also by the competitive instincts of newly urbanized elites, who used lavish display and exotic luxuries to reaffirm their social prestige and authority. Sumerian civilization is a mirror of this developing regional interdependence.

Sumerian civilization came into being as a result of a combination of environmental and social factors. The Sumerians lived in a treeless, lowland environment with fertile soils but no metal, little timber, and no semiprecious stones. They obtained these commodities by trading with areas where such items were in abundance. Sumerian rulers controlled not only large grain surpluses that could be moved in river craft but also a flourishing industry in textiles and other luxuries. The trade moved up and down the great rivers, especially the placid Euphrates. Ancient overland trade routes linked the Tigris and Euphrates rivers with the distant cities and ports in the Levant (eastern Mediterranean area). Even as early as Sumerian times, caravans of pack animals joined Anatolia to the Euphrates, the Levant to Mesopotamia, and Mesopotamia to isolated towns on the distant Iranian highlands to the east.

Bronze technology produced tougher-edged, more durable artifacts that could be used for more arduous day-to-day tasks. One resulting innovation was the metal- and wood-tipped plow, an implement dragged by oxen that was capable of digging a far deeper furrow than the simple hoes and digging sticks of earlier times. The plow was developed as irrigation agriculture assumed greater importance in Sumer, and the combined

innovations increased agricultural yields dramatically. These yields not only supported larger urban and rural populations but also provided a means for the rulers of city-states both in Sumer and farther afield to exercise more control over food surpluses and over the wealth obtained by long-distance exchange.

An intricate and ever-changing system of political alliances and individual obligations of friendship linked community with community and city-state with city-state. In time, financial and logistical checks and balances were maintained by an administrative system based in the temples to bring order to what had begun as in formal bartering. Specialized merchants began to handle such commodities as copper and lapis lazuli, a semiprecious stone. There was wholesaling and contracting, loans were floated, and individual profit was a prime motivation. Increasingly, every city-state, and even entire civilizations, came to depend on the world system, not so much for political stability but for survival.

As the volume of long-distance trade increased dramatically, so competition over resources intensified. Each state raised an army to defend its water rights, trade routes, and city walls. The onerous tasks of defense and military organization passed to despotic kings supposedly appointed by the gods. Such Sumerian city-states as Erech, Kish, and Ur had periods of political strength and prosperity when they dominated their neighbors. Then, just as swiftly, the tide of their fortunes would change and they would sink into obscurity.

Inevitably, the ambitions of some proud Sumerian leaders led them to entertain bolder visions than merely the control of a few city-states in the lowlands. They were well aware that the control of lucrative sources of raw materials and trade routes was the secret of vast political power. In about 2400 B.C., a monarch named Lugalzagesi boasted of overseeing the entire area from the Persian Gulf to the Mediterranean. This boast was probably false. It is likely that Sumerian cities dominated the overland routes that linked Mesopotamia, Anatolia, and the Levant, but their influence was never permanent, their control probably illusory.

63 – 3 The Roman Empire

After the formation of the Roman Republic in 509 B.C., the Romans expanded the borders of their realm through near-continuous warfare. At its greatest extent, in the early second century A.D., the Roman Empire reached from the Euphrates River in southwest Asia to Scotland in the West. The vast territory ringed the Mediterranean Sea—*mare nostrum*, or “our sea,” the Romans called it. As the Romans absorbed the peoples they conquered, they imposed on them a legal, administrative, and cultural structure that endured for some five centuries—in the eastern Mediterranean until the fifteenth century AD—and left a lasting mark on the civilizations that emerged in Europe.

Conquering and maintaining a vast empire required not only inspired leadership and tactics but also careful planning, massive logistical support, and great administrative skill. Some of Rome's most enduring contributions to Western civilization—its system of law, its governmental and administrative structures, and its sophisticated civil engineering and architecture—reflect these qualities.

To facilitate the development and administration of the empire, as well as to make city life comfortable and attractive to its citizens, the Roman government undertook building programs of unprecedented scale and complexity, mandating the construction of central administrative and legal centers (forums and basilicas), recreational facilities (racetracks and stadiums), theaters, public baths, roads, bridges, aqueducts (bridge-like structures for carrying fresh water from the mountains), middle-class housing, and even new towns. To accomplish these tasks without sacrificing beauty, efficiency, and human well-being, Roman builders and architects developed rational plans using easily worked but durable materials and highly sophisticated engineering methods. The architect Vitruvius described these accomplishments in his *Ten Books of Architecture*.

To move their armies about efficiently, speed communications between Rome and the farthest reaches of the empire, and promote commerce, the Romans built a vast and sophisticated network of roads and bridges. Many modern European highways still follow the lines laid down by Roman engineers, and foundations dating from the Roman era underlie the streets of many cities. Roman bridges are still in use, and remnants of Roman aqueducts need only repairs and connecting links to enable them to function again. The Pont du Gard, near Nîmes in southern France, for example, is a powerful reminder of Rome's rapid spread and enduring impact. Entirely functional, the Roman aqueduct conveys the balance,

proportion, and rhythmic harmony of a great work of art and fits naturally into the landscape, a reflection of the Romans' attitude toward the land.

Despite their sophistication as city-dwellers, Romans liked to portray themselves as simple country folk who had never lost their love of nature. The middle classes enjoyed their town-home gardens, wealthy city dwellers maintained rural estates, and Roman emperors had country villas that were both functioning farms and places of recreation. Wealthy Romans even brought nature indoors by commissioning artists to paint landscapes on the interior walls of their homes.

The Romans particularly admired Greek art. Historians have even suggested that although Rome conquered the Greek world, Greek culture conquered Rome. The Romans used Greek designs and Greek styles of columns in their architecture, imported Greek art, and employed Greek artists. In 146 B.C., for example, they stripped the Greek city of Corinth of its art treasures and shipped them back to Rome. Ironically, this love of Greek art was not accompanied by admiration for its artists. In Rome, as in Greece, professional artists were generally considered little more than skilled laborers.

Although the Romans had gods of their own, they also adopted many Greek gods and myths and assimilated Greek religious beliefs and practices into a form of state religion. To these gods they added their own deified emperors, in part to maintain the allegiance of the culturally diverse populations of the empire. Worship of ancient gods mingled with homage to past rulers, and oaths of allegiance to the living ruler made the official religion a political duty—increasingly ritualized, perfunctory, and distant from the everyday life of the average person. As a result, many Romans adopted the more personal religious beliefs of the people they had conquered. These unauthorized religions flourished, despite occasional government efforts to suppress them.

64 – 3 The Development of Complex Societies in Ancient Mexico

Between 9,000 and 4,000 years ago, most of Mexico was inhabited mainly by hunter-foragers who lived in small bands that moved with the seasons to exploit cactus fruits, deer herds, nuts, and the hundreds of other plant and animal species in their range, depending on the season. Since these bands were small in size and never stayed in one place for a sufficient period of time to have much long-term effect on the plant and animal populations on which they subsisted, the hunter-foragers' overall impact on their environment was low. A few groups along the margins of the lake in the Valley of Mexico may have been sedentary villagers, as were some groups along the coasts, and their role in the domestication of plants and animals and the eventual spread of agriculture is unclear. Some researchers suggest, however, that by about 4,000 years ago, maize cob size had become large enough that people over large areas of the Mexican highlands could subsist mainly on maize.

The recent re-dating of some of the supposedly earliest domesticated maize in Mesoamerica (Ancient Mexico) to about 3500 B.C. raises the possibility that initial agriculture evolved out of intensified foraging by groups of people who were relatively sedentary, perhaps living all or most of the year in one or a few places, and that they were perhaps even in the process of developing social differences (social hierarchy) that increased the intensity of their foraging. However, maize was domesticated, and by whom, maize appears to have reached sufficient productivity to permit the village-farming way of life soon after about 2000 B.C., and agricultural communities appeared at about this time in many different areas.

From the hot, wet coastal lowlands to the arid Tehuacan Valley, the earliest villages were quite similar in size and contents. Almost all houses were built using the wattle-and-daub method-sticks, branches, and cane were woven in-and-out between vertical wall poles, then covered with a mud plaster, which was dried by the hot sun. Houses, which were seldom larger than four by six meters, featured thatched roofs and tamped clay floors on which fine sand was scattered.

Most of the earliest farming communities were tiny hamlets-villages of ten to twelve houses that were home to about fifty to sixty people-but some

communities were larger. Most houses that have been excavated have yielded the same remains, mainly grinding stones, storage pits, pieces of large ceramic storage jars, bones of cottontail rabbits, carbonized maize fragments, and broken pieces of ceramic charcoal braziers. In addition, ovens, middens, and graves are very common. While the proportion of plant and animal foods varied somewhat, all villages probably grew maize, beans, squash, peppers, and some other crops, and hunted deer and rabbits. Each village, or each extended family, may have had a specialist who did pressure flaking of stone (to make tools), leather-working, or a similar craft, and individual villages may have concentrated on specialties like salt production, feather-weaving, shell-working, and grinding stone manufacture.

As in Mesopotamia, China, and elsewhere, the background to the origins of complex society in Mesoamerica was a great scatter of relatively simple agricultural villages in which the mechanics of producing a reliable, expandable food supply had been mastered. An early radical break with the simple village farming tradition of Mesoamerica occurred in the sweltering lowlands of the South Gulf Coast of Mexico. Here, beginning at about 1000 B.C., people built massive clay pyramids and platforms, lived in small town groups of hundreds or even thousands, intensively farmed a variety of ecological zones, and produced what is one of the world's most valued examples of stone sculpture.

These people are known to us as the Olmec, a name derived from an ancient American word for rubber-doubtless a reference to the rubber trees that grow in this area-but a name they themselves probably did not use. Some scholars have considered the possibility that the Olmec culture was the mother culture of all later complex societies in Mesoamerica and that the Olmec were directly responsible for transforming their neighbors by military, political religious, or economic means into complex societies. Other scholars, however, have argued convincingly that the Olmec represent only one of several largely independent cases of the evolution of social complexity in Mesoamerica.

64 – 4 Nineteen-century Politics in the United States

The development of the modern presidency in the United States began with Andrew Jackson who swept to power in 1829 at the head of the Democratic Party and served until 1837. During his administration, he immeasurably enlarged the power of the presidency. "The President is the direct representative of the American people," he lectured the Senate when it opposed him. "He was elected by the people, and is responsible to them." With this declaration, Jackson redefined the character of the presidential office and its relationship to the people.

During Jackson's second term, his opponents had gradually come together to form the Whig party. Whigs and Democrats held different attitudes toward the changes brought about by the market, banks, and commerce. The Democrats tended to view society as a continuing conflict between "the people"—farmers, planters, and workers—and a set of greedy aristocrats. This "paper money aristocracy" of bankers and investors manipulated the banking system for their own profit, Democrats claimed, and sapped the nation's virtue by encouraging speculation and the desire for sudden, unearned wealth. The Democrats wanted the rewards of the market without sacrificing the features of a simple agrarian republic. They wanted the wealth that the market offered without the competitive, changing society; the complex dealing; the dominance of urban centers; and the loss of independence that came with it.

Whigs, on the other hand, were more comfortable with the market. For them, commerce and economic development were agents of civilization. Nor did the Whigs envision any conflict in society between farmers and workers on the one hand and businesspeople and bankers on the other. Economic growth would benefit everyone by raising national income and expanding opportunity. The government's responsibility was to provide a well-regulated economy that guaranteed opportunity for citizens of ability.

Whigs and Democrats differed not only in their attitudes toward the market but also about how active the central government should be in people's lives. Despite Andrew Jackson's inclination to be a strong President, Democrats as a rule believed in limited government. Government's role in the economy was to promote competition by destroying

monopolies' and special privileges. In keeping with this philosophy of limited government, Democrats also rejected the idea that moral beliefs were the proper sphere of government action. Religion and politics, they believed, should be kept clearly separate, and they generally opposed humanitarian legislation.

The Whigs, in contrast, viewed government power positively. They believed that it should be used to protect individual rights and public liberty, and that it had a special role where individual effort was ineffective. By regulating the economy and competition, the government could ensure equal opportunity. Indeed, for Whigs the concept of government promoting the general welfare went beyond the economy. In particular, Whigs in the northern sections of the United States also believed that government power should be used to foster the moral welfare of the country. They were much more likely to favor social-reform legislation and aid to education.

In some ways the social makeup of the two parties was similar. To be competitive in winning votes, Whigs and Democrats both had to have significant support among farmers, the largest group in society, and workers. Neither party could win an election by appealing exclusively to the rich or the poor. The Whigs, however, enjoyed disproportionate strength among the business and commercial classes. Whigs appealed to planters who needed credit to finance their cotton and rice trade in the world market, to farmers who were eager to sell their surpluses, and to workers who wished to improve themselves. Democrats attracted farmers isolated from the market or uncomfortable with it, workers alienated from the emerging industrial system, and rising entrepreneurs who wanted to break monopolies and open the economy to newcomers like themselves. The Whigs were strongest in the towns, cities, and those rural areas that were fully integrated into the market economy, whereas Democrats dominated areas of semisubsistence farming that were more isolated and languishing economically.

67 – 2 Human Activity and the Archaeological Record

Archaeologists distinguish the "cultural" formation processes of the archaeological record(those involving some kind of human activity) from the "natural" formation processes (those involving nonhuman activity, such as river action disturbing archaeological material). There are two kinds of cultural formation processes: those that reflect the original human behavior and activity before a find or site became buried and those(such as looting)that came after burial. Of course, most major archaeological sites are formed as the result of a complex sequence of use, burial, and reuse so that a simple twofold division of cultural formation processes may not be so simple to apply in practice, but the distinction is still useful.

Original human behavior is often reflected archaeologically in at least four major activities. In the case of a tool, for example, on of the raw material, (2)manufacture, (3)use, and(4)disposal or discard when the tool is worn out or broken.The tool may of course be reworked and recycled, thus repeating stages 2 and 3.)Similarly, a food crop such as wheat will be acquired(harvested), manufactured(processed), used(eaten), and discarded(digested and the waste products excreted), although here one might add a common intermediate stage of storage before use. From the archaeologist's point of view, the critical factor is that remains can enter the archaeological record at any one of these stages. A tool may be lost or thrown out for being of inferior quality during manufacture, or a crop may be accidentally burnt and thus preserved during processing. In order to reconstruct the original activity accurately it is therefore crucial to try to understand which of the stages one is looking at. It may be quite easy to identify the first stage for stone tools, for instance, because stone quarries can often be recognized by deep holes in the ground with piles of associated waste flakes that survive well. But it is much more difficult to know beyond reasonable doubt where a sample of charred plant remains comes from, whether from an area where harvested wheat was taken for threshing or from the area where the grain was eaten. This may also make it difficult to reconstruct the true plant diet, since certain activities may favor the preservation of certain species of plant.

Deliberate burial of valuables or the dead is another human behavior that has left its mark on

the archaeological record. In times of conflict or war, people often deposit prized possessions in the ground, intending to reclaim them at a later date; but sometimes, for one reason or another, they fail to do so. These hoards are a prime source of evidence for certain periods, such as the European Bronze Age, for which hoards of metal goods are common, or later Roman Britain, which has yielded buried treasures of silver and other precious metals. The archaeologist, however, may not find it easy to distinguish between hoards originally intended to be reclaimed and valuables buried-perhaps as offerings to supernatural powers-with no reclamation intended.

In addition to burial hoards, a major source of archaeological evidence comes from burial of the dead, whether in simple graves, elaborate burial mounds, or giant pyramids, usually with grave goods such as ceramic vessels or weapons, and sometimes with painted tomb-chamber walls as in ancient Mexico or Egypt. The Egyptians indeed went so far as to mummify their dead-to preserve them, they hoped, for eternity-as did the Incas of Peru whose kings were kept in the Temple of the Sun at Cuzco and brought outside for special ceremonies.

Human destruction of the archaeological record might be caused by burials of the kind just described being dug into earlier deposits. But people in the past deliberately or accidentally obliterated traces of their predecessors in innumerable other ways Rulers often destroyed monuments or erased inscriptions belonging to previous chiefs or monarchs. On the other hand, some human destruction meant to obliterate has inadvertently preserved material for the archaeologist to find. Burning, for example, may not always destroy. Clay daubing and adobe usually decay, but if a structure has been burned, the mud is baked to the consistency of a brick.

68 – 1 Salt and the Rise of Venice

The city of Venice, on Italy's coastline, achieved commercial dominance of southern Europe during the Middle Ages largely because of its extensive trade in the valuable commodity of salt. At first, Venice produced its own salt at its Chioggia saltworks. For a time, its principal competitor in the region was the town of Cervia, with Venice having the advantage because Chioggia was more productive. But Chioggia produced a fine-grained salt, so when Venetians wanted coarser salt, they had to import it. Then, in the thirteenth century, after a series of floods and storms destroyed about a third of the salt-producing ponds in Chioggia the Venetians were forced to import even more salt.

That was when the Venetians made an important discovery. More money could be made buying and selling salt than producing it. Beginning in 1281, the government paid merchants a subsidy on salt landed in Venice from other areas. As a result of this assistance, shipping salt to Venice became so profitable that the salt merchants could afford to ship other goods at prices that undersold their competitors. Growing fat on the salt subsidy Venice merchants could afford to send ships to the eastern Mediterranean, where they picked up valuable cargoes of Indian spices and sold them in western Europe at low prices that their non-Venetian competitors could not afford to offer. That meant that Venetians were paying extremely high prices for salt, but they did not mind expensive salt if they could dominate the spice trade and be leaders in the grain trade. When grain harvests failed in Italy, Venice would use its salt income to subsidize grain imports from other parts of the Mediterranean and thereby corner the Italian grain market.

Unlike the Chinese salt monopoly, the Venetian government never owned salt but simply took a profit from regulating its trade. Enriched by its share of sales on high-priced salt, the salt administration could offer loans to finance other trade. Between the fourteenth and sixteenth centuries, a period when Venice was a leading port for grains and spices, 30 to 50 percent of the tonnage of imports to Venice was in salt. All salt had to go through government agencies. The salt administration issued licenses that told merchants not only how much salt they could export but also to where and at what price. The salt administration also maintained Venice's palatial public buildings and the complex hydraulic system that prevented the metropolis from washing away. Many of

Venice's grand statues and ornamental buildings were financed by the salt administration.

Venice carefully built its reputation as a reliable supplier, and so contracts with the merchant state were desirable. Venice was able to dictate terms for these contracts. In 1250, when Venice agreed to supply Mantua and Ferrara with salt, the contract stipulated that these cities would not buy salt from anyone else. This became the model for Venetian salt contracts. As Venice became the salt supplier to more and more countries, it needed more and more salt producers from which to buy. Merchants financed by the salt administration went farther into the Mediterranean, buying salt from many distant sources. Wherever they went, they tried to dominate the supply, control the saltworks, and even acquire them if they could.

Venice manipulated markets by controlling production. In the late thirteenth century, wishing to raise the world market price, Venice had all saltworks on the Greek island of Crete destroyed, and it banned the local production of salt. The Venetians then brought in all the salt needed for local consumption, built stores to sell the imported salt, and paid damages to the owners of the saltworks. The policy was designed to control prices and at the same time keep the locals happy. Aiding its ability to ruthlessly manipulate commerce and control territory, Venice maintained the ships of the merchant fleet as a naval reserve and called them into combat when needed. The Venetian fleet patrolled the Adriatic Sea, stopped ships, inspected cargo, and demanded licensing documents to make sure all commercial traffic was conforming with its regulations.

70 – 1 The Development of Chinese Dynasties

The centralized dynasties that ruled ancient China were a product of the terrain and climate of the Asian continent. China is isolated from the rest of the Eastern hemisphere by formidable natural barriers of mountains on the west and southwest as well as the Gobi Desert on the north. To the east lies the Pacific Ocean. Although China's separation was not total – trade goods, people and ideas moved back and forth between China, India, and Central Asia – in many respects its development was distinctive.

In its formative years, Chinese civilization had been influenced by the contributing factors of the Yellow River region – the land of the loess. It is generally understood that the river running through a vast area, where the fine and porous soil accounts for the large silt content in its current, has a constant tendency to clog its own course and cause dikes to break and enormous inundations to occur. The problem cannot be dealt with locally. This predicament has a historical origin that can be traced at least to the Spring and Autumn period (722 – 481 B.C.), when in 652 B.C., Duke Huan of Qi convened his league of states. A mutual pledge by all participants is recorded in various sources as “not to execute improper dikes and not to hinder the water flow.” The stone inscriptions of his own achievements erected by the first emperor of the Qin dynasty (221 – 207 B.C.) are reproduced in The Records of History. One of the meritorious deeds that he credited to himself was the neutralization of the barriers that obstructed water flows. These and other references are evidence that China's political unification, achieved in the B.C. era had come under the pressing demand for a coordinated effort in dealing with the flood problem which over and above technical considerations has remained in force for the past 2,000 years.

A second factor compelling China to proceed to an early national unification and to follow it with political centralization was the effect of monsoon rain on agriculture. The summer monsoon in China comes from the direction of the Philippine Sea. It by itself does not give rain. The moisture in the air current depends upon winds flowing from west to east and northeast, lifting it to an altitude to be cooled, only then does rain come down from condensation. This climatic arrangement subjects agricultural drops of the nation to the synchronization of two sets of variables. When the

two kinds of currents miss each other, drought sets in. When they repeatedly converge over a specific area, flood and inundation are the inevitable result. It is not uncommon for lack of rain and too much rain simultaneously to victimize two or more parts of China. Only an enormous empire in control of vast resources can deal with the situation. The requirement imposed by the power of nature was felt too, during the B.C. era. In one ancient text we read many stories about wars between the principalities during times of natural disaster. The scholar Yao Shanyu, working from ancient summaries, calls attention to the fact that in the 2,117 years under study, 1,621 floods and 1,392 droughts caused damage serious enough to be reported by the imperial courts.

Further ruling out the possibility of a decentralized China was the potential threat of nomads. The traditional line of defense that China erected in the north along the steppe – a vast grass-covered plain – customarily referred to as the Great Wall was by no means fixed. The territories north and west of the Great Wall, having an annual rainfall of less than fifteen inches and therefore inadequate for cultivation, remained a grazing ground for the nomads. In times of bad weather and in periods of China's dynasty, these nomads had a tendency to execute large-scale invasion. This was a problem whose cause was so deeply rooted in geography that the Chinese were not able to solve it merely through military offensives. Long-term experience taught them to put the regional and local government under a strong leader, allowing national defense to dictate a degree of homogeneity and uniformity in order to survive.

HUMAN BIOLOGY

18 – 2 The mystery of yawning

According to conventional theory, yawning takes place when people are bored or sleepy and serves the function of increasing alertness by reversing, through deeper breathing, the drop in blood oxygen levels that are caused by the shallow breathing that accompanies lack of sleep or boredom. Unfortunately, the few scientific investigations of yawning have failed to find any connection between how often someone yawns and how much sleep they have had or how tired they are. About the closest any research that has come to supporting the tiredness theory is to confirm that adults yawn more often on weekdays than at weekends, and that school children yawn more frequently in their first year at primary school than they do in kindergarten.

Another flaw of the tiredness theory is that yawning does not raise alertness or physiological activity, as the theory would predict. When researchers measured the heart rate, muscle tension and skin conductance of people before, during and after yawning, they did detect some changes in skin conductance following yawning, indicating a slight increase in physiological activity. However, similar changes occurred when the subjects were asked simply to open their mouths or to breathe deeply. Yawning did nothing special to their state of physiological activity. Experiments have also cast serious doubt on the belief that yawning is triggered by a drop in blood oxygen or a rise in blood carbon dioxide. Volunteers were told to think about yawning while they breathed either normal air, pure oxygen, or an air mixture with an above-normal level of carbon dioxide. If the theory was correct, breathing air with extra carbon dioxide should have triggered yawning, while breathing pure oxygen should have suppressed yawning. In fact, neither condition made any difference to the frequency of yawning, which remained constant at about 24 yawns per hour. Another experiment demonstrated that physical exercise, which was sufficiently vigorous to double the rate of breathing, had no effect on the frequency of yawning. Again the implication is that yawning has little or nothing to do with oxygen.

A completely different theory holds that yawning assists in the physical development of the lungs early in life, but has no remaining biological function in adults. It has been suggested that yawning and hiccupping might serve to clear out

the fetuses airways. The lungs of a fetus secrete a liquid that mixes with its mother's amniotic fluid. Babies with congenital blockages that prevent this fluid from escaping from their lungs are sometimes born with deformed lungs. It might be that yawning helps to clear out the lungs by periodically lowering the pressure in them. According to this theory, yawning in adults is just a developmental fossil with no biological function. But, while accepting that not everything in life can be explained by Darwinian evolution, there are sound reasons for being skeptical of theories like this one, which avoid the issue of what yawning does for adults. Yawning is distracting, consumes energy and takes time. It is almost certainly doing something significant in adults as well as in fetuses. What could it be?

The empirical evidence, such as it is, suggests an altogether different function for yawning— namely, that yawning prepares us for a change in activity level. Support for this theory came from a study of yawning behavior in everyday life. Volunteers wore wrist-mounted devices that automatically recorded their physical activity for up to two weeks: the volunteers also recorded their yawns by pressing a button on the device each time they yawned. The data showed that yawning tended to occur about 15 minutes before a period of increased behavioral activity. Yawning bore no relationship to sleep patterns, however. This accords with anecdotal evidence that people often yawn in situations where they are neither tired nor bored, but are preparing for impending mental and physical activity. Such yawning is often referred to as “incongruous” because it seems out of place, at least on the tiredness view: soldiers yawning before combat, musicians yawning before performing, and athletes yawning before competing. Their yawning seems to have nothing to do with sleepiness or boredom—quite the reverse—but it does precede a change in activity level.

24 – 2 Breathing during Sleep

Of all the physiological differences in human sleep compared with wakefulness that have been discovered in the last decade, changes in respiratory control are most dramatic. Not only are there differences in the level of the functioning of respiratory systems, there are even changes in how they function. Movements of the rib cage for breathing are reduced during sleep, making the contractions of the diaphragm more important. Yet because of the physics of lying down, the stomach applies weight against the diaphragm and makes it more difficult for the diaphragm to do its job. However, there are many other changes that affect respiration when asleep.

During wakefulness, breathing is controlled by two interacting systems. The first is an automatic, metabolic system whose control is centered in the brain stem. It subconsciously adjusts breathing rate and depth in order to regulate the levels of carbon dioxide (CO₂) and oxygen (O₂), and the acid-base ratio in the blood. The second system is the voluntary, behavioral system. Its control center is based in the forebrain, and it regulates breathing for use in speech, singing, sighing, and so on. It is capable of ignoring or overriding the automatic, metabolic system and produces an irregular pattern of breathing.

During NREM (the phase of sleep in which there is no rapid eye movement) breathing becomes deeper and more regular, but there is also a decrease in the breathing rate, resulting in less air being exchanged overall. This occurs because during NREM sleep the automatic, metabolic system has exclusive control over breathing and the body uses less oxygen and produces less carbon dioxide. Also, during sleep the automatic metabolic system is less responsive to carbon dioxide levels and oxygen levels in the blood. Two things result from these changes in breathing control that occur during sleep. First, there may be a brief cessation or reduction of breathing when falling asleep as the sleeper waxes and wanes between sleep and wakefulness and their differing control mechanisms. Second, once sleep is fully obtained, there is an increase of carbon dioxide and a decrease of oxygen in the blood that persists during NREM.

But that is not all that changes. During all phases of sleep, several changes in the air passages have been observed. It takes twice as much effort to

breathe during sleep because of greater resistance to airflow in the airways and changes in the efficiency of the muscles used for breathing. Some of the muscles that help keep the upper airway open when breathing tend to become more relaxed during sleep, especially during REM (the phase of sleep in which there is rapid eye movement). Without this muscular action, inhaling is like sucking air out of a balloon — the narrow passages tend to collapse. Also there is a regular cycle of change in resistance between the two sides of the nose. If something blocks the "good" side, such as congestion from allergies or a cold, then resistance increases dramatically. Coupled with these factors is the loss of the complex interactions among the muscles that can change the route of airflow from nose to mouth.

Other respiratory regulating mechanisms apparently cease functioning during sleep. For example, during wakefulness there is an immediate, automatic, adaptive increase in breathing effort when inhaling is made more difficult (such as breathing through a restrictive face mask). This reflexive adjustment is totally absent during NREM sleep. Only after several inadequate breaths under such conditions, resulting in the considerable elevation of carbon dioxide and reduction of oxygen in the blood, is breathing effort adjusted. Finally, the coughing reflex in reaction to irritants in the airway produces not a cough during sleep but a cessation of breathing. If the irritation is severe enough, a sleeping person will arouse, clear the airway, then resume breathing and likely return to sleep.

Additional breathing changes occur during REM sleep that are even more dramatic than the changes that occur during NREM. The amount of air exchanged is even lower in REM than NREM because, although breathing is more rapid in REM, it is also more irregular, with brief episodes of shallow breathing or absence of breathing. In addition, breathing during REM depends much more on the action of the diaphragm and much less on rib cage action.

HUMAN BEING / ANTHROPOLOGY / PSYCHOLOGY / MEMORY / ADULTHOOD AND IMMATURITY

06 – 3 Infantile Amnesia

What do you remember about your life before you were three? Few people can remember anything that happened to them in their early years. Adults' memories of the next few years also tend to be scanty. Most people remember only a few events—usually ones that were meaningful and distinctive, such as being hospitalized or a sibling's birth.

How might this inability to recall early experiences be explained? The sheer passage of time does not account for it; adults have excellent recognition of pictures of people who attended high school with them 35 years earlier. Another seemingly plausible explanation—that infants do not form enduring memories at this point in development—also is incorrect. Children two and a half to three years old remember experiences that occurred in their first year, and eleven month olds remember some events a year later. Nor does the hypothesis that infantile amnesia reflects repression—or holding back—of sexually charged episodes explain the phenomenon. While such repression may occur, people cannot remember ordinary events from the infant and toddler periods either.

Three other explanations seem more promising. One involves physiological changes relevant to memory. Maturation of the frontal lobes of the brain continues throughout early childhood, and this part of the brain may be critical for remembering particular episodes in ways that can be retrieved later. Demonstrations of infants' and toddlers' long-term memory have involved their repeating motor activities that they had seen or done earlier, such as reaching in the dark for objects, putting a bottle in a doll's mouth, or pulling apart two pieces of a toy. The brain's level of physiological maturation may support these types of memories, but not ones requiring explicit verbal descriptions.

A second explanation involves the influence of the social world on children's language use. Hearing and telling stories about events may help children store information in ways that will endure into later childhood and adulthood. Through hearing stories with a clear beginning, middle, and ending children may learn to extract the gist of events in ways that they will be able to describe many years later. Consistent with this view, parents and children

increasingly engage in discussions of past events when children are about three years old. However, hearing such stories is not sufficient for younger children to form enduring memories. Telling such stories to two year olds does not seem to produce long-lasting verbalizable memories.

A third likely explanation for infantile amnesia involves incompatibilities between the ways in which infants encode information and the ways in which older children and adults retrieve it. Whether people can remember an event depends critically on the fit between the way in which they earlier encoded the information and the way in which they later attempt to retrieve it. The better able the person is to reconstruct the perspective from which the material was encoded, the more likely that recall will be successful.

This view is supported by a variety of factors that can create mismatches between very young children's encoding and older children's and adults' retrieval efforts. The world looks very different to a person whose head is only two or three feet above the ground than to one whose head is five or six feet above it. Older children and adults often try to retrieve the names of things they saw, but infants would not have encoded the information verbally. General knowledge of categories of events such as a birthday party or a visit to the doctor's office helps older individuals encode their experiences, but again, infants and toddlers are unlikely to encode many experiences within such knowledge structures.

These three explanations of infantile amnesia are not mutually exclusive; indeed, they support each other. Physiological immaturity may be part of why infants and toddlers do not form extremely enduring memories, even when they hear stories that promote such remembering in preschoolers. Hearing the stories may lead preschoolers to encode aspects of events that allow them to form memories they can access as adults. Conversely, improved encoding of what they hear may help them better understand and remember stories and thus make the stories more useful for remembering future events. Thus, all three explanations—physiological maturation, hearing and producing stories about past events, and improved encoding of key aspects of events—seem likely to be involved in overcoming infantile amnesia.

12 – 1 Which Hand Did They Use?

We all know that many more people today are right-handed than left-handed. Can one trace this same pattern far back in prehistory? Much of the evidence about right-hand versus left-hand dominance comes from stencils and prints found in rock shelters in Australia and elsewhere, and in many Ice Age caves in France, Spain, and Tasmania. When a left hand has been stenciled, this implies that the artist was right-handed, and vice versa. Even though the paint was often sprayed on by mouth, one can assume that the dominant hand assisted in the operation. One also has to make the assumption that hands were stenciled palm downward—a left hand stenciled palm upward might of course look as if it were a right hand. Of 158 stencils in the French cave of Gargas, 136 have been identified as left, and only 22 as right; right-handedness was therefore heavily predominant.

Cave art furnishes other types of evidence of this phenomenon. Most engravings, for example, are best lit from the left, as befits the work of right-handed artists, who generally prefer to have the light source on the left so that the shadow of their hand does not fall on the tip of the engraving tool or brush. In the few cases where an Ice Age figure is depicted holding something, it is mostly, though not always, in the right hand.

Clues to right-handedness can also be found by other methods. Right-handers tend to have longer, stronger, and more muscular bones on the right side, and Marcellin Boule as long ago as 1911 noted the La Chapelle-aux-Saints Neanderthal skeleton had a right upper arm bone that was noticeably stronger than the left. Similar observations have been made on other Neanderthal skeletons such as La Ferrassie I and Neanderthal itself.

Fractures and other cut marks are another source of evidence. Right-handed soldiers tend to be wounded on the left. The skeleton of a 40- or 50-year-old Nabatean warrior, buried 2,000 years ago in the Negev Desert, Israel, had multiple healed fractures to the skull, the left arm, and the ribs.

Tools themselves can be revealing. Long-handed Neolithic spoons of yew wood preserved in Alpine villages dating to 3000 B.C. have survived; the signs of rubbing on their left side indicate that their users were right-handed. The late Ice Age rope found in the French cave of Lascaux consists of

fibers spiraling to the right, and was therefore tressed by a right-hander.

Occasionally one can determine whether stone tools were used in the right hand or the left, and it is even possible to assess how far back this feature can be traced. In stone tool making experiments, Nick Toth, a right-hander, held the core (the stone that would become the tool) in his left hand and the hammer stone in his right. As the tool was made, the core was rotated clockwise, and the flakes, removed in sequence, had a little crescent of cortex (the core's outer surface) on the side. Toth's knapping produced 56 percent flakes with the cortex on the right, and 44 percent left-oriented flakes. A left-handed toolmaker would produce the opposite pattern. Toth has applied these criteria to the similarly made pebble tools from a number of early sites (before 1.5 million years) at Koobi Fora, Kenya, probably made by *Homo habilis*. At seven sites he found that 57 percent of the flakes were right-oriented, and 43 percent left, a pattern almost identical to that produced today.

About 90 percent of modern humans are right-handed: we are the only mammal with a preferential use of one hand. The part of the brain responsible for fine control and movement is located in the left cerebral hemisphere, and the findings above suggest that the human brain was already asymmetrical in its structure and function not long after 2 million years ago. Among Neanderthals of 70,000 – 35,000 years ago, Marcellin Boule noted that the La Chapelle-aux-Saints individual had a left hemisphere slightly bigger than the right, and the same was found for brains of specimens from Neanderthal, Gibraltar, and La Quina.

13 – 1 Types of Social Groups

Life places us in a complex web of relationships with other people. Our humanness arises out of these relationships in the course of social interaction. Moreover, our humanness must be sustained through social interaction—and fairly constantly so. When an association continues long enough for two people to become linked together by a relatively stable set of expectations, it is called a relationship.

People are bound within relationships by two types of bonds: expressive ties and instrumental ties. Expressive ties are social links formed when we emotionally invest ourselves in and commit ourselves to other people. Through association with people who are meaningful to us, we achieve a sense of security, love, acceptance, companionship, and personal worth. Instrumental ties are social links formed when we cooperate with other people to achieve some goal. Occasionally, this may mean working with instead of against competitors. More often, we simply cooperate with others to reach some end without endowing the relationship with any larger significance.

Sociologists have built on the distinction between expressive and instrumental ties to distinguish between two types of groups: primary and secondary. A primary group involves two or more people who enjoy a direct, intimate, cohesive relationship with one another. Expressive ties predominate in primary groups; we view the people as ends in themselves and valuable in their own right. A secondary group entails two or more people who are involved in an impersonal relationship and have come together for a specific, practical purpose. Instrumental ties predominate in secondary groups; we perceive people as means to ends rather than as ends in their own right. Sometimes primary group relationships evolve out of secondary group relationships. This happens in many work settings. People on the job often develop close relationships with coworkers as they come to share gripes, jokes, gossip, and satisfactions.

A number of conditions enhance the likelihood that primary groups will arise. First, group size is important. We find it difficult to get to know people personally when they are milling about and dispersed in large groups. In small groups we have a better chance to initiate contact and establish rapport with them. Second, face-to-face contact

allows us to size up others. Seeing and talking with one another in close physical proximity makes possible a subtle exchange of ideas and feelings. And third, the probability that we will develop primary group bonds increases as we have frequent and continuous contact. Our ties with people often deepen as we interact with them across time and gradually evolve interlocking habits and interests.

Primary groups are fundamental to us and to society. First, primary groups are critical to the socialization process. Within them, infants and children are introduced to the ways of their society. Such groups are the breeding grounds in which we acquire the norms and values that equip us for social life. Sociologists view primary groups as bridges between individuals and the larger society because they transmit, mediate, and interpret a society's cultural patterns and provide the sense of oneness so critical for social solidarity.

Second, primary groups are fundamental because they provide the settings in which we meet most of our personal needs. Within them, we experience companionship, love, security, and an overall sense of well-being. Not surprisingly, sociologists find that the strength of a group's primary ties has implications for the group's functioning. For example, the stronger the primary group ties of a sports team playing together, the better their record is.

Third, primary groups are fundamental because they serve as powerful instruments for social control. Their members command and dispense many of the rewards that are so vital to us and that make our lives seem worthwhile. Should the use of rewards fail, members can frequently win by rejecting or threatening to ostracize those who deviate from the primary group's norms. For instance, some social groups employ shunning (a person can remain in the community, but others are forbidden to interact with the person) as a device to bring into line individuals whose behavior goes beyond that allowed by the particular group. Even more important, primary groups define social reality for us by structuring our experiences. By providing us with definitions of situations, they elicit from our behavior that conforms to group-devised meanings. Primary groups, then, serve both as carriers of social norms and as enforcers of them.

13 – 3 Methods of Studying Infant Perception

In the study of perceptual abilities of infants, a number of techniques are used to determine infants' responses to various stimuli. Because they cannot verbalize or fill out questionnaires, indirect techniques of naturalistic observation are used as the primary means of determining what infants can see, hear, feel, and so forth. Each of these methods compares an infant's state prior to the introduction of a stimulus with its state during or immediately following the stimulus. The difference between the two measures provides the researcher with an indication of the level and duration of the response to the stimulus. For example, if a uniformly moving pattern of some sort is passed across the visual field of a neonate (newborn), repetitive following movements of the eye occur. The occurrence of these eye movements provides evidence that the moving pattern is perceived at some level by the newborn. Similarly, changes in the infant's general level of motor activity—turning the head, blinking the eyes, crying, and so forth—have been used by researchers as visual indicators of the infant's perceptual abilities.

Such techniques, however, have limitations. First, the observation may be unreliable in that two or more observers may not agree that the particular response occurred, or to what degree it occurred. Second, responses are difficult to quantify. Often the rapid and diffuse movements of the infant make it difficult to get an accurate record of the number of responses. The third, and most potent, limitation is that it is not possible to be certain that the infant's response was due to the stimulus presented or to a change from no stimulus to a stimulus. The infant may be responding to aspects of the stimulus different than those identified by the investigator. Therefore, when observational assessment is used as a technique for studying infant perceptual abilities, care must be taken not to overgeneralize from the data or to rely on one or two studies as conclusive evidence of a particular perceptual ability of the infant.

Observational assessment techniques have become much more sophisticated, reducing the limitations just presented. Film analysis of the infant's responses, heart and respiration rate monitors, and nonnutritive sucking devices are used as effective tools in understanding infant perception. Film analysis permits researchers to carefully study the infant's responses over and over and in slow motion. Precise measurements can be

made of the length and frequency of the infant's attention between two stimuli. Heart and respiration monitors provide the investigator with the number of heartbeats or breaths taken when a new stimulus is presented. Numerical increases are used as quantifiable indicators of heightened interest in the new stimulus. Increases in nonnutritive sucking were first used as an assessment measure by researchers in 1969. They devised an apparatus that connected a baby's pacifier to a counting device. As stimuli were presented, changes in the infant's sucking behavior were recorded. Increases in the number of sucks were used as an indicator of the infant's attention to or preference for a given visual display.

Two additional techniques of studying infant perception have come into vogue. The first is the habituation-dishabituation technique, in which a single stimulus is presented repeatedly to the infant until there is a measurable decline (habituation) in whatever attending behavior is being observed. At that point a new stimulus is presented, and any recovery (dishabituation) in responsiveness is recorded. If the infant fails to dishabituate and continues to show habituation with the new stimulus, it is assumed that the baby is unable to perceive the new stimulus as different. The habituation-dishabituation paradigm has been used most extensively with studies of auditory and olfactory perception in infants. The second technique relies on evoked potentials, which are electrical brain responses that may be related to a particular stimulus because of where they originate. Changes in the electrical pattern of the brain indicate that the stimulus is getting through to the infant's central nervous system and eliciting some form of response.

Each of the preceding techniques provides the researcher with evidence that the infant can detect or discriminate between stimuli. With these sophisticated observational assessment and electro-physiological measures, we know that the neonate of only a few days is far more perceptive than previously suspected. However, these measures are only "indirect" indicators of the infant's perceptual abilities.

14 – 1 Children and Advertising

Young children are trusting of commercial advertisements in the media, and advertisers have sometimes been accused of taking advantage of this trusting outlook. The Independent Television Commission, regulator of television advertising in the United Kingdom, has criticized advertisers for “misleadingness”—creating a wrong impression either intentionally or unintentionally—in an effort to control advertisers’ use of techniques that make it difficult for children to judge the true size, action, performance, or construction of a toy.

General concern about misleading tactics that advertisers employ is centered on the use of exaggeration. Consumer protection groups and parents believe that children are largely ill-equipped to recognize such techniques and that often exaggeration is used at the expense of product information. Claims such as “the best” or “better than” can be subjective and misleading; even adults may be unsure as to their meaning. They represent the advertiser’s opinions about the qualities of their products or brand and, as a consequence, are difficult to verify. Advertisers sometimes offset or counterbalance an exaggerated claim with a disclaimer—a qualification or condition on the claim. For example, the claim that breakfast cereal has a health benefit may be accompanied by the disclaimer “when part of a nutritionally balanced breakfast”. However, research has shown that children often have difficulty understanding disclaimers: children may interpret the phrase “when part of a nutritionally balanced breakfast” to mean that the cereal is required as a necessary part of a balanced breakfast. The author George Comstock suggested that less than a quarter of children between the ages of six and eight years old understood standard disclaimers used in many toy advertisements and that disclaimers are more readily comprehended when presented in both audio and visual formats. Nevertheless, disclaimers are mainly presented in audio format only.

Fantasy is one of the more common techniques in advertising that could possibly mislead a young audience. Child-oriented advertisements are more likely to include magic and fantasy than advertisements aimed at adults. In a content analysis of Canadian television, the author Stephen Kline observed that nearly all commercials for character toys featured fantasy play. Children have strong imaginations and the use of fantasy brings

their ideas to life, but children may not be adept enough to realize that what they are viewing is unreal. Fantasy situations and settings are frequently used to attract children’s attention, particularly in food advertising. Advertisements for breakfast cereals have, for many years, been found to be especially fond of fantasy techniques, with almost nine out of ten including such content. Generally, there is uncertainty as to whether very young children can distinguish between fantasy and reality in advertising. Certainly, rational appeals in advertising aimed at children are limited, as most advertisements use emotional and indirect appeals to psychological states or associations.

The use of celebrities such as singers and movie stars is common in advertising. The intention is for the positively perceived attributes of the celebrity to be transferred to the advertised product and for the two to become automatically linked in the audience’s mind. In children’s advertising, the “celebrities” are often animated figures from popular cartoons. In the recent past, the role of celebrities in advertising to children has often been conflated with the concept of host selling. Host selling involves blending advertisements with regular programming in a way that makes it difficult to distinguish one from the other. Host selling occurs, for example, when a children’s show about a cartoon lion contains an ad in which the same lion promotes a breakfast cereal. The psychologist Dale Kunkel showed that the practice of host selling reduced children’s ability to distinguish between advertising and program material. It was also found that older children responded more positively to products in host selling advertisements.

Regarding the appearance of celebrities in advertisements that do not involve host selling, the evidence is mixed. Researcher Charles Atkin found that children believe that the characters used to advertise breakfast cereals are knowledgeable about cereals, and children accept such characters as credible sources of nutritional information. This finding was even more marked for heavy viewers of television. In addition, children feel validated in their choice of a product when a celebrity endorses that product. A study of children in Hong Kong, however, found that the presence of celebrities in advertisements could negatively affect the children’s perceptions of a product if the children did not like the celebrity in question.

21 – 3 Autobiographical Memory

Think back to your childhood and try to identify your earliest memory. How old were you? Most people are not able to recount memories for experiences prior to the age of three years, a phenomenon called infantile amnesia. The question of why infantile amnesia occurs has intrigued psychologists for decades, especially in light of ample evidence that infants and young children can display impressive memory capabilities. Many find that understanding the general nature of autobiographical memory, that is, memory for events that have occurred in one's own life, can provide some important clues to this mystery. Between ages three and four, children begin to give fairly lengthy and cohesive descriptions of events in their past. What factors are responsible for this developmental turning point?

Perhaps the explanation goes back to some ideas raised by influential Swiss psychologist Jean Piaget — namely, that children under age two years represent events in a qualitatively different form than older children do. According to this line of thought, the verbal abilities that blossom in the two year old allow events to be coded in a form radically different from the action-based codes of the infant. Verbal abilities of one year olds are, in fact, related to their memories for events one year later. When researchers had one year olds imitate an action sequence one year after they first saw it, there was correlation between the children's verbal skills at the time they first saw the event and their success on the later memory task. However, even children with low verbal skills showed evidence of remembering the event; thus, memories may be facilitated by but are not dependent on those verbal skills.

Another suggestion is that before children can talk about past events in their lives, they need to have a reasonable understanding of the self as a psychological entity. The development of an understanding of the self becomes evident between the first and second years of life and shows rapid elaboration in subsequent years. The realization that the physical self has continuity in time, according to this hypothesis, lays the foundation for the emergence of autobiographical memory.

A third possibility is that children will not be able to tell their own "life story" until they understand

something about the general form stories take, that is, the structure of narratives. Knowledge about narratives arises from social interactions, particularly the storytelling that children experience from parents and the attempts parents make to talk with children about past events in their lives. When parents talk with children about "what we did today" or "last week" or "last year", they guide the children's formation of a framework for talking about the past. They also provide children with reminders about the memory and relay the message that memories are valued as part of the cultural experience. It is interesting to note that some studies show Caucasian American children have earlier childhood memories than Korean children do. Furthermore, other studies show that Caucasian American mother-child pairs talk about past events three times more often than do Korean mother-child pairs. Thus, the types of social experiences children have do factor into the development of autobiographical memories.

A final suggestion is that children must begin to develop a "theory of mind" — an awareness of the concept of mental states (feelings, desires, beliefs, and thoughts), their own and those of others — before they can talk about their own past memories. Once children become capable of answering such questions as "What does it mean to remember?" and "What does it mean to know something?" improvements in memory seem to occur.

It may be that the developments just described are intertwined with and influence one another. Talking with parents about the past may enhance the development of the self-concept, for example, as well as help the child understand what it means to "remember". No doubt the ability to talk about one's past represents memory of a different level of complexity than simple recognition or recall.

30 – 1 Role of Play in Development

Play is easier to define with examples than with concepts. In any case, in animals it consists of leaping, running, climbing, throwing, wrestling, and other movements, either along, with objects, or with other animals. Depending on the species, play may be primarily for social interaction, exercise, or exploration. One of the problems in providing a clear definition of play is that it involves the same behaviors that take place in other circumstances — dominance, predation, competition, and real fighting. Thus, whether play occurs or not depends on the intention of the animals, and the intentions are not always clear from behaviors alone.

Play appears to be a developmental characteristic of animals with fairly sophisticated nervous systems, mainly birds and mammals. Play has been studied most extensively in primates and canids (dogs). Exactly why animal's play is still a matter debated in the research literature, and the reasons may not be the same for every species that plays. Determining the functions of play is difficult because the functions may be long-term, with beneficial effects not showing up until the animal's adulthood.

Play is not without considerable costs to the individual animal. Play is usually very active, involving movement in space and, at times, noisemaking. Therefore, it results in the loss of fuel or energy that might better be used for growth or for building up fat stores in a young animal. Another potential cost of this activity is greater exposure to predators since play is attention-getting behavior. Great activities also increase the risk of injury in slipping or falling.

The benefits of play must outweigh costs, or play would not have evolved, according to Darwin's theory. Some of the potential benefits relate directly to the healthy development of the brain and nervous system. In one research study, two groups of young rats were raised under different conditions. One group developed in an "enriched" environment, which allowed the rats to interact with other rats, play with toys, and receive maze training. The other group lived in an "impoverished" environment in individual cages in a dimly lit room with little stimulation. At the end of the experiments, the results showed that the actual weight of the brains of the impoverished rats was less than that of those raised in the enriched environment (though they were fed the same diets). Other

studies have shown that greater stimulation not only affects the size of the brain but also increases the number of connections between the nerve cells. Thus, active play may provide necessary stimulation to the growth of synaptic connections in the brain, especially the cerebellum, which is responsible for motor functioning and movements.

Play also stimulates the development of the muscle tissues themselves and may provide the opportunities to practice those movements needed for survival. Prey species, like young deer or goats, for example, typically play by performing sudden flight movements and turns, whereas predator species, such as cats, practice stalking, pouncing, and biting.

Play allows a young animal to explore its environment and practice skills in comparative safety since the surrounding adults generally do not expect the young to deal with threats or predators. Play can also provide practice in social behaviors needed for courtship and mating. Learning appropriate social behaviors is especially important for species that live in groups, like young monkeys that needed to learn to control selfishness and aggression and to understand the give-and-take involved in social groups. They need to learn how to be dominant and submissive because each monkey might have to play either role in the future. Most of these things are learned in the long developmental periods that primates have, during which they engage in countless play experiences with their peers.

There is a danger, of course, that play may be misinterpreted or not recognized as play by others, potentially leading to aggression. This is especially true when play consists of practicing normal aggressive or predator behaviors. Thus, many species have evolved clear signals to delineate playfulness. Dogs, for example, will wag their tails, get down their front legs, and stick their behinds in the air to indicate "what follows is just for play".

31 – 2 Early Childhood Education

Preschools - educational programs for children under the age of five - differ significantly from one country to another according to the views that different societies hold regarding the purpose of early childhood education. For instance, in a cross-country comparison of preschools in China, Japan, and the United States, researchers found that parents in the three countries view the purpose of preschools very differently. Whereas parents in China tend to see preschools primarily as a way of giving children a good start academically, Japanese parents view them primarily as a way of giving children the opportunity to be members of a group. In the United States, in comparison, parents regard the primary purpose of preschools as making children more independent and self-reliant, although obtaining a good academic start and having group experience are also important.

While many programs designed for preschoolers focus primarily on social and emotional factors, some are geared mainly toward promoting cognitive gains and preparing preschoolers for the formal instruction they will experience when they start kindergarten. In the United States, the best-known program designed to promote future academic success is Head Start. Established in the 1960s when the United States declared the War on Poverty, the program has served over 13 million children and their families. The program, which stresses parental involvement, was designed to serve the "whole child", including children's physical health, self-confidence, social responsibility, and social and emotional development.

Whether Head Start is seen as successful or not depends on the lens through which one is looking. If, for instance, the program is expected to provide long-term increases in IQ (intelligence quotient) scores, it is a disappointment. Although graduates of Head Start programs tend to show immediate IQ gains, these increases do not last. On the other hand, it is clear that Head Start is meeting its goal of getting preschoolers ready for school. Preschoolers who participate in Head Start are better prepared for future schooling than those who do not. Furthermore, graduates of Head Start programs have better future school grade. Finally, some research suggests that ultimately Head Start graduates show higher academic performance at the end of high school, although the gains are modest.

In addition, results from other types of preschool readiness programs indicate that those who participate and graduate are less likely to repeat grades, and they are more likely to complete school than readiness program, for every dollar spent on the program, taxpayers saved seven dollars by the time the graduates reached the age of 27.

The most recent comprehensive evaluation of early intervention programs suggests that, taken as a group, preschool programs can provide significant benefits, and that government funds invested early in life may ultimately lead to a reduction in future costs. For instance, compared with children who did not participate in early intervention programs, participants in various programs showed gains in emotional or cognitive development, better educational outcomes, increased economic self-sufficiency, reduced levels of criminal activity, and improved health-related behaviors. Of course, not every program produced all these benefits, and not every child benefited to the same extent. Furthermore, some researchers argue that less-expensive programs are just as good as relatively expensive ones, such as Head Start. Still, the results of the evaluation were promising, suggesting that the potential benefits of early intervention can be substantial.

Not everyone agrees that programs that seek to enhance academic skills during the preschool years are a good thing. In fact, according to developmental psychologist David Elkind, United States society tends to push children so rapidly that they begin to feel stress and pressure at a young age. Elkind argues that academic success is largely dependent upon factors out of parents' control, such as inherited abilities and a child's rate of maturation. Consequently, children of a particular age cannot be expected to master educational material without taking into account their current level of cognitive development. In short, children require development appropriate educational practice, which is education that is based on both typical development and the unique characteristics of a given child.

35 – 2 The Development of Social Complexity

For most of human history, we have foraged (hunted, fished, and collected wild plants) for food. Small nomadic groups could easily supply the necessities for their families. No one needed more, and providing for more than one's needs made little sense. The organization of such societies could be rather simple, revolving around age and gender categories. Such societies likely were largely egalitarian, beyond distinctions based on age and gender, virtually all people had equivalent rights, status, and access to resources.

Archaeologist Donald Henry suggests that the combination of a rich habitat and sedentism (permanent, year-round settlement) led to a dramatic increase in human population. In his view, nomadic, simple foragers have relatively low levels of fertility. Their high-protein, low-carbohydrate diets result in low body-fat levels, which are commonly associated with low fertility in women. High levels of physical activity and long periods of nursing, which are common among modern simple foragers, probably also contributed to low levels of female fertility if they were likewise common among ancient foragers.

In Henry's view, the adoption of a more settled existence in areas with abundant food resources would have contributed to higher fertility levels among the sedentary foragers. A diet higher in wild cereals produces proportionally more body fat, leading to higher fertility among women. Cereals, which are easy to digest, would have supplemented and then replaced mother's milk as the primary food for older infants. Since women are less fertile when they are breast-feeding, substituting cereals for mother's milk would have resulted in closer spacing of births and the potential for a greater number of live births for each woman. A more sedentary existence may also have lowered infant mortality and perhaps increased longevity among the aged. These more vulnerable members of society could safely stay in a fixed village rather than be forced regularly to move great distances as part of a nomadic existence, with its greater risk of accidents and trauma.

All of these factors may have resulted in a trend of increasing size among some local human populations in the Holocene (since 9600 B C E). Given sufficient time, even in very rich habitats, human population size can reach carrying capacity, the maximum population an area can sustain within

the context of a given subsistence system. And human population growth is like a runaway tram once it picks up speed, it is difficult to control. So even after reaching an area's carrying capacity, Holocene human populations probably continued to grow in food-rich regions, overshooting the ability of the territory to feed the population, again within the context of the same subsistence strategy. In some areas, small changes in climate or minor changes in plant characteristics may have further destabilized local economies.

One possible response to surpassing the carrying capacity of a region is for a group to exploit adjoining land. However, good land may itself be limited—for example, within the confines of a river valley where neighbors are in the same position, having filled up the whole of the desirable habitat available in their home territories, expansion is also problematic. Impinging on the neighbors' territory can lead to conflict, especially when they too are up against the capacity of the land to provide enough food.

Another option is to stay in the same area but to shift and intensify the food quest there. The impulse to produce more food to feed a growing population was satisfied in some areas by the development of more-complex subsistence strategies involving intensive labor and requiring more cooperation and greater coordination among the increasing numbers of people. This development resulted in a change in the social and economic equations that defined those societies. Hierarchies that did not exist in earlier foraging groups but that were helpful in structuring cooperative labor and in organizing more-complex technologies probably became established, even before domestication and agriculture, as pre-Neolithic societies (before the tenth millennium B C E) reacted to the population increase.

62 – 3 Is Perceptual Development an Innate or Socially Acquired Process?

Most developmental scientists now agree that both nature (genetics) and nurture (environment) are essential for the normal development of perception. However, there is still much dispute about the extent to which either nature or nurture is the more important factor. Points of view on this issue are more than just philosophical musings; they affect the kinds of experiments that are undertaken. I argue here that classifying particular aspects of perceptual development as either innate or learned presents us with an overly passive view in which either genes or environment imposes structure on the developing brain. In contrast, I suggest that perceptual development is better characterized as an activity-dependent process involving complex and subtle interactions at many levels.

To begin to illustrate my point, let's consider some recent neurobiological work on the prenatal (before-birth) development of the brain in rodents. Neurons are specialized cells that transmit impulses or messages to other neurons, glands, and muscles. The neurons studied in these experiments are those involved in binocular vision. Experiments show that the prenatal tuning (training) of these neurons arises through their response to internally generated waves of electrical activity. In other words, the response properties of these visual neurons are shaped by a "virtual environment" generated by cells elsewhere in the brain and eye. Although the term "innate" can be stretched to cover this example of development, we could equally well describe this process as the cells "learning" from the input provided. Further, after birth the same neurons continue to be tuned in the same way except that now their input also reflects the structure of the world outside. When we examine development in detail, it becomes harder to argue, as some theorists do, that "innate knowledge" is fundamentally different from learning.

Another example of the role of activity-dependent processes in perceptual development comes from the ability to detect and recognize faces. Because regions of the human brain are specialized for processing faces, some researchers have argued that this ability is innate. However, experiments with infants reveal a more complex story. The tendency for newborns to look more toward faces turns out to be based on a very primitive, reflex like system that is triggered by a stimulus as simple as

three high-contrast blobs in the approximate locations of the eyes and mouth. This simple bias is sufficient to ensure that newborns look much more at faces than at other objects and patterns over the first weeks of life. One consequence of this is that developing circuits on the visual recognition pathway of the brain get more input related to faces and thus are shaped by experience with this special type of visual stimulus. We can now study this process by using new brain-imaging methods. Such studies have shown that the brains of young infants show less-localized and less-specialized processing of faces than do the brains of adults. It is not until they are one year old that infants show the same patterns of brain specialization for processing faces as do adults, by which time they have had as much as a thousand hours of exposure to human faces.

Another example comes from the study of infants' eye movements to visual targets. Although newborns are capable of some primitive reflexive eye movements, only much later in the first year can they make most of the kinds of complex and accurate eye movements seen in adults. One view is that the very limited ability present in newborns is just sufficient to allow them to practice and develop new brain circuits for the more complex integration of visual and motor information necessary for adult eye movements. Once again, it appears that infants actively contribute to their own subsequent development.

These considerations should make us skeptical about the many claims that are made for innate perceptual abilities based on experiments with babies of four months and older. In fact, when the same experiments were done with younger infants, quite different results have often been obtained, suggesting dramatic changes in perceptual abilities over the first few weeks and months after birth.

Infants are not passively shaped by either their genes or their environment. Rather, perceptual development is an activity-dependent process in which, during postnatal life, the infant plays an active role in generating the experience it needs for subsequent development.

64 – 2 Auditory Perception in Infancy

Because they have had some practice in hearing before birth, it is not surprising that infants have reasonably good auditory perception after they are born. In fact, for certain very high and very low frequencies, infants actually are more sensitive to sound than adults—a sensitivity that seems to increase during the first two years of life. On the other hand, infants are initially less sensitive than adults to middle-range frequencies. Eventually, however, their capabilities within the middle range improve.

It is not fully clear what leads to the improvement during infancy in sensitivity to sounds, although it may be related to the maturation of the nervous system. More puzzling is why, after infancy, children's ability to hear very high and low frequencies gradually declines. One explanation may be that exposure to high levels of noise may diminish capacities at the extreme ranges.

In addition to the ability to detect sound, infants need several other abilities in order to hear effectively. For instance, sound localization permits infants to pin point the direction from which a sound is emanating. Compared to adults, infants have a slight handicap in this task, because effective sound localization requires the use of the slight difference in the times at which a sound reaches our two ears. Because infants' heads are smaller than those of adults; the difference in timing of the arrival of sound at the two ears is less than it is in adults. However, despite the potential limitation brought about by their smaller heads, infants' sound localization abilities are actually fairly good even at birth, and they reach adult levels of success by the age of one year. Interestingly, their improvement is not steady: although we don't know why, the accuracy of sound localization actually declines between birth and two months of age, but then begins to increase.

Even more important to their ultimate success in the world, young infants are capable of making the fine discriminations that their future understanding of language will require. For instance, in one classic study, a group of one to four-month-old infants sucked on nipples that activated a recording of a person saying “ba” every time they sucked. At first, their interest in the sound made them suck vigorously. Soon, though, they became acclimated to the sound and sucked with less energy. On the other hand, when the experimenters changed the

sound to “pa”, the infants immediately showed new interest and sucked with greater vigor once again. The clear conclusion: infants as young as one month old could make the distinction between the two similar sounds.

Even more intriguing is the fact that young infants are able to discriminate certain characteristics that differentiate one language from another. Some evidence suggests that even two-day-olds show preferences for the language spoken by those around them over other languages. In the first few months, the ability to discriminate between languages develops rapidly. By the age of five months, they can distinguish the difference between English and Spanish passages, even when the two are similar in meter, number of syllables, and speed of recitation.

Given their ability to discriminate a difference in speech as slight as the difference between two consonants, it is not surprising that infants can distinguish different people on the basis of voice. In fact, from an early age they show clear preferences for some voices over others. For instance, in one experiment newborns were allowed to suck a nipple that turned on a recording of a human voice reading a story. The infants sucked significantly longer when the voice was that of their mother than when the voice was that of a stranger.

How do such preferences arise? One hypothesis is that prenatal exposure to the mother's voice is the key. As support for this conjecture, researchers point to the fact that newborns do not show a preference for their fathers' voices over other male voices. Furthermore, newborns prefer listening to melodies sung by their mothers before they were born to melodies that were not sung before birth. It seems, then, that the prenatal exposure to their mothers' voices—although muffled by the liquid environment of the womb—helps shape infants' listening preferences.

68 – 3 Research into Aging and Extending Life Span

The mounting evidence for age genes that influence the aging process is by no means conclusive, but it is quite impressive, coming from a variety of independent research from aging in worms and fruit flies to antioxidants and gene repair mechanisms, and human mutations. Still, the connections are circumstantial.

Christopher Wills, professor of biology at the University of California in San Diego, thinks that by 2025 science will likely isolate the mammalian age genes in mice. We share roughly 75 percent of our genes with mice and have much the same body chemistry; this is a strong reason to believe that an age gene found in mice could also be at work in humans. If such genes are located, the next step would be to find out if these age genes have their counterparts in humans. Wills believes that if they are found in humans, they may extend the human life span perhaps to 150 years.

But by 2020, when personalized DNA sequencing becomes widespread, a second tactic may prove fruitful as well. By analyzing populations of healthy individuals in their nineties and beyond, scientists will find it possible to use computers to compare their genetic backgrounds and cross-check for similarities in key genes that are suspected of influencing aging. A combination of studies on the DNA of long-lived animals and on the personalized DNA sequences of elderly individuals may considerably narrow down the search for the age gene.

As yet, none of these methods can prove that we can increase the human life span. Indeed, the only theory with a proven track record of extending the life span of animals is the caloric restriction theory, which states that animals which consume calories just above starvation levels live significantly longer than the average. Although this theory flies in the face of common sense (a well-fed animal is well nourished and healthy, and should have greater resistance to disease and aging), it has held up under repeated testing among a wide range of animals. Scientists have consistently increased the life span of rats and mice in the laboratory by 50 to 100 percent. It is the only laboratory-tested theory of age extension for animals that has held up under decades of careful scrutiny. Why?

Across the animal kingdom, the life span of animals is roughly inversely correlated to the metabolism rate. The slower their normal metabolism rate, the longer their normal life span. In 1996, in a study that reduced the caloric intake of 200 monkeys by 30 percent, the monkeys were shown to have a slower metabolism rate, a longer life span, and reduced rates of cancer, heart disease, and diabetes. "We have known for 70 years that if you feed laboratory mice less food, they age slower, they live longer, and they get diseases less frequently. We find that monkeys respond in the same way as rodents and that the same biological changes may be in play here," says George Roth of the National Institute of Aging.

There is still room for scientific debate on the question "why?" Ron Hart, a scientist at the National Center for Toxicological Research, believes that the answer may lie in the high body temperature of mammals, and humans in particular. "Heat causes pieces of the DNA molecule to split off randomly, and it must be repaired," Hart says. "Under caloric restriction, though, the engine runs cooler and there's less damage. Merely reducing caloric intake by 40 percent reduced this form of spontaneous DNA damage almost 24 percent!" Furthermore, at a higher internal body temperature, oxygen is being burned at a greater rate, creating more free radicals, which also speed up the aging process. Cooling the body, on the other hand, increases the amount of antioxidants in the body. Hart found a fourfold increase in the enzyme catalase and a threefold increase in superoxide dismutase in animals on a restricted diet. "What's fascinating," Hart concludes, "is that reduced food intake is the only experimental paradigm ever found that enhances DNA repair." Hart is so convinced of the importance of this work that in 1993 he began the first systematic studies of caloric restrictions in humans.

70 – 3 Functionalism

In biology, different parts of an organism may be described by their functions or the parts they play in maintaining the life of the whole organism. Functionalism in social science similarly looks for the role (function) some aspect of culture or social life plays in maintaining a system. Two quite different schools of functionalism arose in conjunction with two British anthropologists – Bronislaw Malinowski (1884 – 1942) and A.R. Radcliffe-Brown (1881 – 1955).

Malinowski's version of functionalism assumes that all cultural traits serve the needs of individuals in a society – that the function of a cultural trait is its ability to satisfy some basic or derived need of the members of the group. The basic needs include nutrition reproduction, bodily comfort, safety, relaxation movement, and growth. Some aspects of the culture satisfy these basic needs. In doing so, they give rise to derived needs that must also be satisfied. For example, cultural traits that satisfy the basic need for food give rise to the secondary or derived need for cooperation in food collection or production. Societies will in turn develop forms of political organization and social control that guarantee the required cooperation. How did Malinowski explain such things as religion and magic? He suggested that since humans always live with a certain amount of uncertainty and anxiety, they need stability and continuity. Religion and magic are functional in that they serve those needs.

Unlike Malinowski, Radcliffe-Brown felt that the various aspects of social behavior maintain a society's social structure rather than satisfying individual needs. By social structure he meant the total network of existing social relationships in a society. Since Radcliffe-Brown's version of functionalism emphasized the social structure as the system to be maintained through the development of supporting rules, practices, and customs, the term "structural functionalism" is often used to describe his approach.

An example of Radcliffe-Brown's structural-functionalist approach is his analysis of the ways in which different societies deal with the tensions that are likely to develop among people related through marriage. To reduce potential tension between in-laws, he suggested societies do one of two things. They may develop strict rules forbidding the

persons involved ever to interact face to face (as do the Navajos, for example, in requiring a man to avoid his mother-in-law) or they may allow mutual disrespect and teasing between the in-laws. Radcliffe-Brown suggested that avoidance is likely to occur between in-laws of different generations, whereas disrespectful teasing is likely between in-laws of the same generation. Both avoidance and teasing he suggested, are ways to avoid real conflict and help maintain the social structure.

The major objection to Malinowski's functionalism is that it cannot readily account for cultural variation. Most of the needs he identified – such as the need for food – are universal, all societies must deal with them if they are to survive. Thus, while the functionalist approach may tell us why all societies engage in food getting, it cannot tell why different societies have different food-getting practices. In other words, functionalism does not explain why certain specific cultural patterns arise to fulfill a need that might be fulfilled just as easily by any of a number of alternative possibilities.

A major problem of the structural-functionalist approach is that it is difficult to determine whether a particular custom is in fact functional in the sense of contributing to the maintenance of the social system. In biology the contribution an organ makes to the health or life of an animal can be assessed by removing it. But we cannot subtract a cultural trait from a society to see if the trait really does contribute to the maintenance of the group. It is conceivable that certain customs within the society may be neutral or even detrimental to its maintenance. We cannot assume that all of a society's customs are functional merely because the society is functioning at the moment. And even if we are able to assess whether a particular custom is functional, this theoretical orientation fails to deal with the question of why a particular society chooses to meet its structural needs in a particular way. A given problem does not necessarily have only one solution. We still must explain why one of the possible solutions is chosen.

JOB / CAREER

09 – 2 Reflection in Teaching

Teachers, it is thought, benefit from the practice of reflection, the conscious act of thinking deeply about and carefully examining the interactions and events within their own classrooms. Educators T. Wildman and J. Niles (1987) describe a scheme for developing reflective practice in experienced teachers. This was justified by the view that reflective practice could help teachers to feel more intellectually involved in their role and work in teaching and enable them to cope with the paucity of scientific fact and the uncertainty of knowledge in the discipline of teaching.

Wildman and Niles were particularly interested in investigating the conditions under which reflection might flourish – a subject on which there is little guidance in the literature. They designed an experimental strategy for a group of teachers in Virginia and worked with 40 practicing teachers over several years. They were concerned that many would be “drawn to these new, refreshing” conceptions of teaching only to find that the void between the abstractions and the realities of teacher reflection is too great to bridge. Reflection on a complex task such as teaching is not easy. The teachers were taken through a program of talking about teaching events, moving on to reflecting about specific issues in a supported, and later an independent, manner.

Wildman and Niles observed that systematic reflection on teaching required a sound ability to understand classroom events in an objective manner. They describe the initial understanding in the teachers with whom they were working as being “utilitarian ... and not rich or detailed enough to drive systematic reflection.” Teachers rarely have the time or opportunities to view their own or the teaching of others in an objective manner. Further observation revealed the tendency of teachers to evaluate events rather than review the contributory factors in a considered manner by, in effect, standing outside the situation.

Helping this group of teachers to revise their thinking about classroom events became central. This process took time and patience and effective trainers. The researchers estimate that the initial training of the teachers to view events objectively took between 20 and 30 hours, with the same

number of hours again being required to practice the skills of reflection.

Wildman and Niles identify three principles that facilitate reflective practice in a teaching situation. The first is support from administrators in an education system, enabling teachers to understand the requirements of reflective practice and how it relates to teaching students. The second is the availability of sufficient time and space. The teachers in the program described how they found it difficult to put aside the immediate demands of others in order to give themselves the time they needed to develop their reflective skills. The third is the development of a collaborative environment with support from other teachers. Support and encouragement were also required to help teachers in the program cope with aspects of their professional life with which they were not comfortable. Wildman and Niles make a summary comment: “Perhaps the most important thing we learned is the idea of the teacher-as-reflective-practitioner will not happen simply because it is a good or even compelling idea.”

The work of Wildman and Niles suggests the importance of recognizing some of the difficulties of instituting reflective practice. Others have noted this, making a similar point about the teaching profession’s cultural inhibitions about reflective practice. Zeichner and Liston (1987) point out the inconsistency between the role of the teacher as a (reflective) professional decision maker and the more usual role of the teacher as a technician, putting into practice the ideas of theirs. More basic than the cultural issues is the matter of motivation. Becoming a reflective practitioner requires extra work (Jaworski, 1993) and has only vaguely defined goals with, perhaps, little initially perceivable reward and the threat of vulnerability. Few have directly questioned what might lead a teacher to want to become reflective. Apparently, the most obvious reason for teachers to work toward reflective practice is that teacher educators think it is a good thing. There appear to be many unexplored matters about the motivation to reflect – for example, the value of externally motivated reflection as opposed to that of teachers who might reflect by habit.

TPO-75-1: Clutch Size in Birds

Each year the emperor penguin lays one egg, the pigeon lays one or two eggs, the gull typically lays three eggs, the Canada goose four to six eggs, and the American merganser ten or eleven eggs. What determines clutch size in birds? The ultimate factors that determine clutch size are the requirements for long-term (evolutionary) survival. Clutch size is viewed as an adaptation under the control of natural selection.

Natural selection will favor those birds that leave the most descendants to future generations. At first thought we might hypothesize that natural selection favors a clutch size that is the physiological maximum the bird can lay. We can test this hypothesis by taking eggs from nests as they are laid. When we do this, we find that some birds, such as the common pigeon, are determinate layers; they lay a given number of eggs, no matter what. The pigeon lays two eggs; if you take away the first, it will incubate the second egg only. If you add a third egg, it will incubate all three. But many other birds are indeterminate layers; they will continue to lay eggs until the nest is “full.” If eggs are removed once they are laid, these birds will continue laying. When this subterfuge was used on a mallard female, she continued to lay one egg per day until she had laid 100 of them. Evidence from other similar experiments suggests that most birds under normal circumstances do not lay their physiological limit of eggs but that ovulation is stopped long before this limit is reached.

The British ornithologist David Lack was one of the first ecologists to recognize the importance of evolutionary thinking in understanding adaptations in life history traits. In 1947 Lack put forward the idea that clutch size in birds was determined by the number of young that parents can provide with food. If his hypothesis is correct, the total production of young ought to be highest at the normal clutch size. And if one experimentally increased clutch size by adding eggs to nests, increased clutches should suffer greater losses because the parents could not feed the extra young in the nest.

One way to think about this problem of optimum clutch size is to use a simple economic approach. Everything an organism does has some costs and some benefits. The benefits of laying more eggs are very clear—more descendants in the next generation. The costs are less clear. There is an energy cost to make each additional egg, and there is a further cost to feed each additional nestling. If the adult birds must work harder to feed their young, there is also a potential cost in adult survival—the adults may not live until the next breeding season. If adults are unable to work harder, there is a risk of reduction in offspring quality. Models of this type are useful because they help us think about the costs and benefits for a particular ecological strategy. No organism has an infinite amount of energy to spend on its activities. The reproductive rate of birds can be viewed as one sector of a bird’s energy balance, and the needs of reproduction must be maximized within the constraints of other energy requirements. The total requirements involve metabolic maintenance, growth, energy used for predator avoidance, competitive interactions, and reproduction. According to Lack’s hypothesis, if enough additional eggs are placed in a bird’s nest, reproductive energy requirements escalate and the whole brood will suffer from starvation so that, in fact, fewer young birds will fledge from nests containing large numbers of eggs.

In England, the blue tit normally lays a clutch of nine to eleven eggs. What would happen if blue tits had a brood of twelve or thirteen? A researcher artificially manipulated broods at hatching by adding or subtracting chicks and found that the survival of the young blue tits in manipulated broods was poor. Blue tits feed on insects and apparently cannot feed additional young adequately, so more of the young starve. Consequently, it would not benefit a blue tit in the evolutionary sense to lay more eggs and the results are consistent with Lack’s hypothesis. Individual birds appear to produce the clutch size that maximizes their reproductive potential.

TPO-75-2: The Palaces of Minoan Crete

The palaces built by the Minoan people of Crete (an island in the eastern Mediterranean) between 1950 B.C. and 1450 B.C. were centers of power, both sacred and secular, with wealth based on the stores of agricultural produce from the fertile hinterland that they controlled. What happened within them? Presumably they were busy places. The receipt, recording, and storage of goods brought in from the surrounding countryside must have been a major job, particularly at those times of the year when the corn was harvested, the olive ripe, or the wine newly made. Administration of the goods and commodities leaving the storerooms for consumption in the palace, for use in the workshops, or for wider distribution must have taken place throughout the year. We may imagine the storage areas as busy by day, and probably carefully guarded at night.

The craft workers creating precious works in such materials as gold, ivory, and semiprecious stones may well have been fulltime craft specialists, totally dependent on the palaces and working constantly within them, their products in demand both for immediate use and possibly of exchange quite far afield. Their raw materials were part of palatial wealth, and must have been kept under strict control. Similarly bronze workers, whose raw material would have been carefully recorded, no doubt worked close by. Pottery workshops seem to have been further afield, situated near the clay beds and fuel for the kilns (pottery ovens) rather than within palace confines, though their products could still have been largely or totally under palace control. The production of textiles was probably always an important palace-based activity, though we do not know in which rooms, precisely, spinning and weaving took place.

Some revisionist interpretations of the palace of Knossos have questioned whether people lived within it to any great extent. We must face squarely the fact, implicit in the above account, that unequivocal evidence for human residence as opposed to human use, is hard to find. Nonetheless, it seems overwhelmingly probable that all the palaces were lived in by some, if not all, of the people engaged in the many activities for which they show evidence. Many nonspecific rooms could have been lived in by various dependents—craft workers, guards, textile workers, servants, or slaves. This remains true even if we follow those interpretations that would prefer “temple” to “palace” as the buildings’ prime identification.

In fact, the choice of “temple” rather than “palace” is a change of emphasis rather than a change of interpretation. From the early twentieth-century work of archaeologist Arthur Evans onwards, the ingrained and embedded religious nature of the palaces has been accepted, as has the fact that “palace” is inadequate shorthand for the buildings’ complexities. There are, though, good reasons not to change wholesale to “temple” which in truth is similarly inadequate. The most obvious problem with such a change is the fact that, while a society based on large palaces with no temples has always been recognized as anomalous in the ancient world, large-scale temples with no palaces would be equally so. Unless there is a truly huge gap in our excavated evidence—possible, but as time goes on increasingly unlikely—these complex buildings are all we have. The most logical conclusion remains that they fulfilled two sets of functions simultaneously. Accurate terminology that did not privilege one function over another would require the coinage of a portmanteau term such as “palace-temple” or “temple-palace,” but this seems an undesirable deviation from terminology that is long established, whatever its shortcomings.

What, though, of the ruling elite, whether priests, kings, or both? Is there evidence that the palaces housed a royal family? The attempt to see royal apartments throws into high relief not only the difficulty of identifying precisely the function of a room from its form, but also the fact that we do not really know who, in Minoan society, held the reins of power. The Minoan tradition indicates a king—yet power may have been wielded by a priest, or by a priest-king, combining functions that we would divide into sacred and secular. Nonetheless, it still seems overwhelmingly likely that the ruling elite lived in the palaces, or perhaps in both the palaces and the important houses in their immediate vicinity.

TPO-75-3: Seismic Waves

Seismic waves—energy waves produced by earthquakes—permit scientists to determine the location, thickness, and properties of Earth’s internal zones. They are generated when rock masses are suddenly disturbed, such as when they break or rupture. Vibrations spread out in all directions from the source of the disturbance, traveling at different speeds through parts of Earth’s crust and interior that differ in chemical composition and physical properties. The principal categories of these waves are primary, secondary, and surface. All three types of waves are recorded on an instrument called a seismograph.

Primary waves, or P-waves, are the speediest of the three kinds of waves and therefore the first to arrive at a seismograph station after there has been an earthquake. They travel through the upper crust of Earth at speeds of 4 to 5 kilometers per second, but near the base of the crust they speed along at 6 or 7 kilometers per second. In these primary waves, pulses of energy are transmitted as a succession of compressions and expansions that parallel the direction of propagation of the wave itself. Thus, a given segment of rock set in motion during an earthquake is driven into its neighbor and bounces back. The neighbor strikes the next particle and rebounds and subsequent particles continue the motion. Vibrational energy is an accordion-like push-pull movement that can be transmitted through solids, liquids, and gases. Of course, the speed of P-wave transmission will differ in materials of different density and elastic properties.

Secondary waves, or S-waves, travel 1 to 2 kilometers per second slower than do P-waves. Unlike the movement of P-waves, rock vibration in secondary waves is at right angles to the direction of propagation of the energy. This type of wave is easily demonstrated by tying a length of rope to a hook and then shaking the free end. A series of undulations will develop in the rope and move toward the hook—that is, in the direction of propagation. Any given particle along the rope, however, will move up and down in a direction perpendicular to the direction of propagation. It is because of their more complex motion that S-waves travel more slowly than P-waves. They are the second group of oscillations to arrive at a seismograph station. Unlike P-waves, secondary waves will not pass through liquids or gases.

Both P- and S-waves are sometimes also termed body waves because they are able to penetrate deep into the interior or body of our planet. Body waves travel faster in rocks of greater elasticity, and their speeds, therefore, increase steadily as they move downward into more elastic zones of Earth’s interior and then decrease as they begin to make their ascent toward Earth’s surface. The change in velocity that occurs as body waves invade rocks of different elasticity results in a bending or refraction of the wave. The many small refractions cause the body waves to assume a curved travel path through Earth.

Not only are body waves subjected to refraction, but they may also be partially reflected off the surface of a dense rock layer in much the same way as light is reflected off a polished surface. Many factors influence the behavior of body waves. An increase in the temperature of rocks through which body waves are traveling will cause a decrease in velocity, whereas an increase in confining pressure will cause a corresponding increase in wave velocity. In a fluid where no rigidity exists, S-waves cannot propagate and P-waves are markedly slowed. Surface waves are large-motion waves that travel through the outer crust of Earth. Their pattern of movement resembles that of waves caused when a pebble is tossed into the center of a pond. They develop whenever P- or S-waves disturb the surface of Earth as they emerge from the interior.

Surface waves are the last to arrive at a seismograph station. They are usually the primary cause of the destruction that can result from earthquakes affecting densely populated areas. This destruction results because surface waves are channeled through the thin outer region of Earth, and their energy is less rapidly scattered into the large volumes of rock traversed by body waves.