The Big History of Humanity:

A Theory of Philosophy of History, Macrosociology and Cultural Evolution

Rochelle Forrester

This is a renamed edition of my book *How Change Happens: A theory of Philosophy of History, Social Change and Cultural Evolution* as the words “How Change Happens” and “Social Change” have been found to make readers believe the book is about political activism and social activism which it is not. The use of the terms “Big History” and “Macrosociology” should fix this.

Published 18 February 2022

© Rochelle Forrester 2022
The moral right of the author has been asserted.
Anyone may reproduce any part of this paper of up to five thousand words
without the permission of the author so long as a full acknowledgement
of the source of the reproduced material is made.

Best Publications Limited
14 Longcroft Terrace
Newlands
Wellington
New Zealand

ISBN 978-1-877479-74-8
for *How Change Happens: A theory of Philosophy of History, Social Change and Cultural Evolution*

Table of Contents

Preface

Since writing this book I have continued to think about the matters it raises and in a rather general conclusion I would state what can happen in history is set for us by the structure of the universe, i.e. the laws of physics, chemistry and biology, the properties of the particles, elements, compounds and mixtures making up the material of the universe and the genetics of the living matter, including humans, in the universe. The course of history i.e. the sequence of events is set for us by the order of discovery of the structure of the universe which is an order from the easiest to the more difficult or from that which is closest to us to that which is furthest from us. These two questions, what can happen in history and the order in which it happens are two quite distinct matters that should be kept separate when studying social and cultural history, social change and cultural evolution.

The other significant feature of this book is that it suggests a new way of writing social and cultural history and about social change and cultural evolution. It suggests it is possible to state not just <u>what</u> happened but also <u>why</u> it happened and why it happened at a particular point in time. An analysis can be done to show what the laws of the natural sciences and the properties of the material constituting the universe allow to happen in history and the order in which the discovery of those laws and properties provides a sequence in social and cultural history and in social change and cultural evolution. This allows the study of history to be put on a much more scientific basis than has been possible in the past. This is because both what can happen in social and cultural history and the order of events are necessary and certain and become capable of rational explanation. They are not random or dependent upon human whim or decision making. This method of writing history could provide a new way of writing thesis, articles and books in history, sociology and anthropology.

The scheme of the book is that the book is divided into two parts and an appendix. Part I contains an outline of the theory and various examples used to illustrate the theory. Part II consists of a series of case studies covering some of the most important discoveries in human history and three more detailed studies designed to illustrate the ideas proposed in Part I of the book in greater detail. It makes particular use of counterfactuals to illustrate how, if the structure of the universe e.g. laws of nature and properties of the materials in the universe, were different then how human history would have been different. It aims to show the theory explained in Part I of the book in a different way and from a different point of view. The appendix looks at a number of theories of history, social change and cultural evolution and discusses various deficiencies within them. The theories discussed are those that I consider to be reasonably similar to the theory proposed in Part I of this book.

Two words I have deliberately avoided in this book are teleological and socio-cultural. The idea I am proposing may appear to be teleological but only in a limited sense. It is teleological in the sense that we are heading to a definite end result, but we do not know what that end result is and that does not necessarily mean that human history is pre-determined in the sense of having a designer. If people want to draw that conclusion that is fine, but the idea that we are headed to a definite but unknown end result, does not necessarily mean there is a designer. That would involve a leap of faith not inherent in the theory itself. The words socio-cultural have been avoided as I consider the social and cultural to be two different concepts that I prefer to keep apart, however I do not necessarily believe this has any practical effect on the theory proposed.

Part I

The Big History of Humanity:

A THEORY OF

PHILOSOPHY OF HISTORY,

MACROSOCIOLOGY

AND CULTURAL EVOLUTION

Summary

It is proposed that the ultimate cause of much historical, social and cultural change is the gradual accumulation of human knowledge of the environment. Human beings use the materials in their environment to meet their needs and increased human knowledge of the environment enables human needs to be met in a more efficient manner. Human needs direct human research into particular areas and this provides a direction for historical, social and cultural development. The human environment has a particular structure and human beings have a particular place in it so that human knowledge of the environment is acquired in a particular order. The simplest knowledge, or the knowledge closest to us, is acquired first and more complex knowledge, or knowledge further from us is acquired later. The order of discovery determines the course of human social and cultural history as knowledge of new and more efficient means of meeting human needs results in new technology, which results in the development of new social and ideological systems. This means human history, or a major part of human history, had to follow a particular course, a course that is determined by the structure of the human environment. An examination of the structure of the human environment will reveal the particular order in which our discoveries had to be made. Given that a certain level of knowledge will result in a particular type of society, it is possible to ascertain the types of societies that were inevitable in human history. While it is not possible to make predictions about the future course of human history, it is possible to explain and understand why human history has followed a particular path and why it had to follow that particular path.

This book is about the long-term changes that have occurred in human society. It is a macro history, or a substantive theory of history and a macrosociology and a theory of social change and cultural evolution that proposes a linear progression in human knowledge and technology as the underlying cause of much social, cultural and historical change. It explains the cause of the progression and the consequences of the progression. It shows how and why humans in many environments have changed from being hunter-gatherers to being citizens of modern industrial states. It deals with the facts of scientific and technological discoveries and not with unsubstantiated or unsubstantiable speculations. It is not about events such as wars and the rise and fall of empires or dynasties, which are political events; rather it is about the intellectual and material conditions of humankind. It deals with the social and cultural history of humankind and not with political and diplomatic history.

The aim of the book is best illustrated by two quotes from Jared Diamond and A Terry Rambo. Jared Diamond states:

"The challenge now is to develop human history as a science, on par with acknowledged historical sciences such as astronomy, geology and evolutionary biology." [1]:

A Terry Rambo states concerning cultural evolution:

"there is almost no integration of research on sequence and that on process. In the absence of such integration, the study of cultural evolution remains in a state not unlike that of paleontology before Darwin. Temporal sequences were clearly evident in the fossil record but, without a plausible naturalistic mechanism to explain change, the Biblical flood was as believable an explanation as any other. The occurrence of cultural evolution is as much a fact as biological evolution, in that sequences of cultural forms can be shown beyond a shadow of reasonable doubt to exist in the archeological record. Not even Franz Boas doubted that hunting and gathering cultures preceded agricultural societies or that stone tools were invented before iron tools. In the absence of any convincing model of the causal processes that produce such temporal sequences, however, cultural evolutionism does not offer social scientists a coherent theory around which to organize further research."[2]

The aim of this book is to put human social and cultural history on a scientific basis as suggested by Jared Diamond. It also shows how social and cultural change happens and how this explains the sequence of events in social and cultural history as sought by A Terry Rambo. The book shows that to a large extent social and cultural history follows a predetermined and necessary path that can be analyzed and rationally understood and explained. Most books on history just give a narrative describing <u>how</u> one thing followed another. This book describes <u>why</u> one thing followed another. This involves going into areas where historians do not usually go, for example into the areas of science such as the chemical structure of rocks and the melting and smelting points of metals and ores. It is this which enables us to state why the stone age was followed by the bronze age which was followed by the iron age.

The causes of social, cultural and historical change proposed involve a mixture of ideological, social and material factors. Ideological factors are involved in that new ideas will often be the driving force for change. Ideas such as that of systematic experimentation and the application of quantitative methods form the basis of modern science and are the ultimate causes of much of the scientific and technological change that has occurred since the seventeenth century. Social factors are involved in that change requires openness to new ideas and technology and the absence of institutions, which may try to suppress new ideas and technology. Material factors are involved in that the particular technology available to a given society will have a powerful effect on the way in which its people live. However, behind the ideology, social system and technology of any particular society is the level of knowledge of that society. A change in the level of knowledge of a society may change the state of a society's ideology, technology and social systems.

The words "human environment" and "nature" are used more or less interchangeably in this book. It is however emphasized that human environment does not mean the natural environment such as climate, rivers, mountains and landscape but the structure of nature such as shown by the laws of physics, chemistry and biology and the properties of the materials in the natural world. The aim of this book is to show the effect that the laws of physics, chemistry and biology and the properties of the particles, elements, compounds and mixtures that make up the world we live in, have on human history. It will show that the laws of the natural sciences and the properties of the substances in our world ensure that the course of human social and cultural history proceeded along a limited range of particular paths and that it could not follow any other paths.

The theory proposed is based upon five concepts. These concepts concern human needs and desires; the level of knowledge of the environment, available in particular societies; the order in which discoveries concerning the environment take place; the properties of matter that constitute the environment and the structure of the universe in which we live. These five concepts are explained in detail later in this book but a brief explanation is appropriate here. It will be suggested that all societies have certain needs or desires and that they meet these needs by utilizing the resources in their environments. The ability to utilize those resources changes as their knowledge of their environment changes. In particular they develop knowledge of the properties of the resources in their environment and how the resources in their environment can be used to meet human needs and desires. Human knowledge of the resources is dynamic; it changes over time. Greater knowledge of the properties of the resources in the environment allows new ways in which human needs can be met by exploiting the resources in the environment. Our knowledge of our environment grows in a particular order; certain knowledge will inevitably be discovered before other knowledge. The order of our discoveries about nature determines the order of technological change and scientific discoveries in human society. The order of our discoveries of both the properties and structure of nature depend upon the relationship between nature and us. We discover these things in an order from that which is closest to us, to that which is further away, or perhaps in an order from the simplest to the more complex. It is the structure of the universe and our place in it, which determines the order in which our knowledge of nature will grow and this determines what technological and scientific options are available to meet our needs and desires.

The theory proposed is a multilateral theory of human development. It recognizes that different cultures and societies live in different environments and so will develop in different ways. Societies in arctic, mountainous, coastal and desert environments will develop different cultures. Societies in areas of mineral deposits may develop differently from those without such mineral deposits. Societies in areas where large domesticable animals are present may develop differently from those without large domesticable animals. A society's religious beliefs may be quite arbitrarily chosen by the society and be quite uninfluenced by the particular environment within which the society lives.

This book deals only with that part of human history, which changes due to changes in human knowledge. I have called that part of history, human social and cultural history, which is perhaps an imprecise description. When I refer to human social and cultural history, I mean that part of human history that changes due to changes in human knowledge of the human environment. This necessarily leaves out significant parts of human history, but it enables us to put what I call social and cultural history on a more rational and scientific basis.

Human history obviously does change in a major way due to changes in human knowledge. The domestication of plants and animals, the invention of writing, the discovery of mathematics, the development of metallurgy, the scientific revolution, the invention of the steam engine and other technologies during the industrial revolution, ideas such as evolution, the heliocentric universe and cultural relativity, motor vehicles, aircraft, television, telephones and computers are all derived from increasing knowledge of the human environment. These ideas and technologies were all based upon the acquisition of new knowledge, whether scientific or empirical, and those ideas and technologies have caused enormous changes in human history.

This is not to say that all changes in human history are caused by changes in human knowledge. There are other causes of change in history notably decisions made by people in power that can cause great historical events. However this book will only deal with changes in human society derived from changes in human knowledge of the human environment. Such changes tend to be cumulative and usually lead to linear changes in human history. Such change is usually not reversible as once knowledge is part of human culture it seldom seems to be lost. When it is lost it is usually due to changes in the human environment and such changes are rare because the behavior of materials in the natural world is usually consistent over time. There are a few examples of cumulative change, derived from increasing human knowledge, being reversed such as soil exhaustion or climate change which could lead to an abandonment of agriculture or mineral exhaustion leading to an end of metallurgy. Soil exhaustion or climate change may be the explanations for the abandonment of the Mayan and Zimbabwe civilizations. Knowledge may also be lost when societies become culturally isolated, especially when the isolation is accompanied by low population as happened to the Australian aborigines living in Tasmania. While such occasions may have occurred in human history they are comparatively rare and are vastly disproportionate to the tendency for knowledge to accumulate in human societies.

Change derived from increasing human knowledge, in other words, cumulative change can be contrasted with reversible change which can be defined as changes caused by human will and decision making. Such changes are easily reversible, they can swing as easily one way as another, like a pendulum, as they are subject to human whim and decision making. Since the discovery of agriculture there has been a great area of civilization running from China and SouthEast Asia, through India and the Middle East, North Africa and Europe which has been based on agriculture and which had metallurgy and writing. During the thousands of years these societies have practiced agriculture they have not shown any indication of changing back to hunting and gathering or losing metallurgy and writing. Yet during those thousands of years there has been a constant rising and falling of empires, dynasties and change in religious beliefs. In the Middle East the Babylonian Empire was replaced by the Assyrian which was replaced by the neo-Babylonian which fell to the Persians, who succumbed to Alexander the Great, whose empire divided into Hellenic states which were eventually conquered by the Romans. While all these empires came and went the practices of agriculture, writing and metallurgy and many other technologies and the social structures of the empires consisting of a small landowning elite, a large rural peasantry and small urban populations, remained. Rulers changed, depending on their military and diplomatic abilities and luck, but the technologies and social structure of the societies continued on. The discovery of agriculture, metallurgy and writing are cumulative changes that are not easily reversed, whereas the rise and fall of empires, dynasties and religions is a matter subject to human decision making and can go one way or another depending upon human decisions and abilities. On the other hand cumulative changes tend to be based on matters such as efficiency or what is the best solution to a problem and those matters are given and are not subject to human decision making or whims. We can of course choose to adopt the least efficient answer to a problem, but we do not normally do so. Cyclical theories of history will usually be based on and seek to explain reversible change in human history such as the rise and fall of empires. This book however deals only with cumulative change and does not attempt to explain reversible change.

A summary of the ideas proposed in this book are:

1. Human beings meet their needs by using the resources in their environment.
2. Human beings have a limited knowledge of their environment.

3. Human beings have the ability to learn and remember so their knowledge of their environment increases over time.
4. As human knowledge of the environment increases, new ways of meeting human needs become available.
5. If the new ways of meeting human needs are better than the old ways of meeting human needs they will be adopted and the old ways discarded.
6. The adoption of new ways of meeting human needs constitutes social and cultural change in itself, but also leads to further social and cultural change.
7. The order of discovery of new means of meeting human needs follows a particular path from that which is more easily discovered to that which is more difficult to discover. Many discoveries require prior discoveries before the discovery can take place. This means there is a necessary order in the discoveries that constitute and cause social and cultural change.
8. The particular order in the discoveries means social and cultural change occurs in a particular order so that the sequence of social and cultural change is inevitable and is rationally understandable.

All of the above statements appear to be obviously correct. If they are then the study of social and cultural history can be considered to be a science in the same way as biological evolution is. Social and cultural change derived from increasing human knowledge is not random and so can be scientifically understood. We cannot predict the future of social and cultural change as we do not know what future discoveries we will make. This is analogous to biological evolution where changes in living species are unpredictable as we do not know what changes will occur in the environment of those species. However biological evolution does make changes in living species rationally understandable, just as an analysis of the order of discovery of the human environment makes social and cultural change rationally understandable.

Needs

The starting point in this development is the human being itself. Human beings have the ability to learn and they have this ability above and beyond that of any other living species. This capacity is used to meet various human needs or desires. A consideration of human needs is necessary for two reasons. First, human needs direct human interests and research into particular directions or areas. This direction in combination with the opportunities our environment allows us for meeting our needs sets the course of human historical development. Secondly, human needs are a requirement for the adoption of new inventions or ideas. They will not be adopted unless a need for them exists.

Human needs can be described in various ways. One such description is that of Abraham Maslow with his hierarchy of human needs. Maslow's needs ascended from basic physiological needs (food, warmth, shelter) to safety needs (to be secure, safe, out of danger), to belongingness and love needs (to be accepted, to belong), to esteem needs (achievement, competence, respect from others), cognitive needs (to know, understand, explain) aesthetic needs (beauty, symmetry, elegance) to self-actualization (to develop and explore oneself to the full). Maslow's hierarchy of needs are somewhat controversial. Nevertheless while individual exceptions can always be shown to Maslow's hierarchy and the exact order of the needs at the top level may be arguable, there would seem to be considerable truth in his theory. Just about all human beings in all cultures appear to desire food, warmth, shelter and safety and security. A sense of belongingness to groups and for the respect of others would also appear to be common to all societies. Equally all societies appear to have cognitive needs (all societies have creation stories) and aesthetic needs (art).

We are not, however, restricted to Maslow's description of human needs. An alternative set of needs could be the basic human needs such as for light, warmth, oxygen, food, moisture, sleep, and physical safety and such needs as for love and affection, the respect of others, self-respect, power (either as a means of satisfying other needs or as an end in itself), material possessions and wealth (either as means or end), the satisfaction of intellectual curiosity, peace of mind, aesthetic satisfaction, new experience or variability of experience and for creative opportunities. The list is not necessarily exhaustive and the needs are not necessarily found in every society or individual[3]. Nevertheless such needs are found in nearly all societies and they provide a useful explanation for human exploitation of the environment.

A further set of needs, arising from the human inclination to live in societies, are for systems of communication, production, distribution, defense, member replacement and social control. These needs are often called the functional requisites of societies and are universal needs existing in all human societies.

The needs expressed above are mainly universal needs present in all, or almost all, human cultures. However there are many needs that relate only to particular cultures. These needs, however, are usually derived from the universal needs. An example of this would be the need of mine owners in Britain in the seventeenth and eighteenth centuries to pump water out of mines. This may have been a need for a particular country at a particular time but this need related to a need for the goods that would be produced by the use of the coal and other minerals. Those goods would have met a universal need that would have been common to all cultures such as the production of food, shelter or warmth. Coal obviously can be used for warmth but it may also be used for the smelting of metals that may be used for the making of agricultural implements or the production of hammers and nails for the erection of buildings that would provide shelter from the elements. The fact that derived or relative needs can usually be related back to universal needs, suggests that the direction the universal needs provide to human knowledge and research will exist in all societies.

Human needs direct human attention in particular directions. Hunter-gatherers are well known as having a very considerable knowledge of the plants and animals in their environment. They know which plants are safe to eat, where they are likely to be found and when they are best to eat. They know the behavior of the animals in their environment, where they are most likely to be found and how to trap and kill them. They would know where water is to be found in arid environments. Yet they would know little about the soils they walk on, the geology of the earth and have only a minimal knowledge of the seasons. Hunter–gatherers developed their knowledge of the plants, animals and water sources in their environments because they had a need for that knowledge.

An agrarian people would tend to lose the knowledge that hunter-gatherers have of wild animals and plants. However they would develop a considerable knowledge of what domesticated plants grow best in what soils, and if they have domesticated animals, how to care for and breed domestic animals. They would also have a considerable knowledge of the seasons and what is the best time to plant crops. The development of a calendar and the beginnings of a science of astronomy would be needed by an agrarian society to assist decisions as to when crops should be planted. An agrarian society will produce a surplus and need to record the amount and the whereabouts of the surplus. This will result in a need for writing or some other record keeping system. The need to calculate the amount of the surplus, tax owed and areas of land lead to the development of mathematics. The need to protect the surplus and to maintain law and order lead to the development of governments, bureaucracy and armies. The need for trade led to the development of improved sea and land transport, such as sailing ships and wheeled transport. Agrarian peoples developed their knowledge of agriculture and pastoralism, of calendars, astronomy, writing, mathematics and invented governments, bureaucracy, armies, sailing ships and wheeled transport because they had a need for such knowledge and inventions.

Industrial societies have their own set of needs. The agrarian farmer's knowledge of agriculture and pastoralism would be replaced by a more scientific knowledge of agriculture involving analysis of soils and deliberate selective breeding of animals. Scientific and engineering knowledge would replace the empirical building and engineering knowledge of agrarian societies. Better observations of nature with improved instruments and techniques allowed accurate and rational (whether true or not) explanations of nature to replace the mythical and religious explanations of agrarian societies. Industrial societies develop their knowledge of science and engineering, as they are the means used in industrial societies to meet human needs.

This shows how human needs, whether they be universal needs, or needs that exist in only one or some societies, focus human attention into certain areas, which involve the meeting of human needs. We see little attempt to meet the needs of other species, we are profoundly human-centric. We do not attempt to feed or tend other animals unless we have an interest in the survival of those animals. We do not tend to engage in conduct that does not meet our needs. Conduct such as standing on our heads, sleeping 20 hours a day, praying to gods we do not believe exist, (as opposed to those we do believe exist), eating food with no taste or nutritional value, betting on non-existent races, do not meet any human needs and so are not normally engaged in by human beings. There is probably an infinite range of behavior that does not meet human needs and is consequently not engaged in by humans.

The question of human needs was raised by George Bassalla[4] when he repeats a question raised by V Gordon Childe "Did a reindeer hunter in 30,000 BC or an Ancient Egyptian in 3,000 BC or an ancient Briton in 30 BC really need or want to travel a couple of hundred miles at 60mph?" Childe and Bassalla considered the answer was no and Bassalla considered "the speed of land travel appropriate to one time and culture are not necessarily appropriate to another." Childe and Bassalla are wrong. Reindeer hunters, ancient Egyptians and Britons would have found such a vehicle enormously useful and if it were available they would certainly have used it. A reindeer hunter would have found his hunting much more successful if he was hunting from such a vehicle as he could easily outrun his prey and the vehicle would be extremely useful for carrying the dead reindeer back to his camp. Ancient Egyptians and Britons would use such a vehicle for the transport of agricultural produce or goods, for hunting, for communication purposes and for military purposes. Any society that has draught animals and the cart would find the vehicle referred to by Childe and Bassalla as simply an improved version of the animal and cart. Such a vehicle would have a valuable role in helping to meet the ultimate need of the provision of food.

That technology “appropriate” to one culture can meet the needs of another culture can be seen by the modern "real life" examples of modern hunter gatherers hunting with rifles and shotguns, the desire of groups such as Maoris in New Zealand in the nineteenth century to obtain goods such as metal axes and muskets and modern reindeer herding (the animals are now domesticated) involving the use of snowmobiles. The way in which the Native Americans in North America took advantage of horses as soon as they became available shows how hunter-gatherer societies were able to make use of much enhanced speed and mobility. Such a vehicle would simply be an example of technological diffusion, which often takes place. The use by third world countries of western technology, such as telephones and computers, is a further example of this. The question is not whether the technology is "appropriate" but whether it is useful and a vehicle traveling at 60mph over hundreds of miles would be useful in all cultures other than those that have better vehicles. The vehicle referred to by Childe and Bassalla would not of itself be a universal need, even though it would be a need in all cultures, but would assist in the meeting of universal needs such as assisting in the provision of food by hunting or the trading of goods, which could meet some universal need. The point is that many human needs are the same in all cultures. A major difference between cultures lies in the extent to which they are able to meet those needs.

It is not however the case that just because a need exists, that it will be met. It is also necessary that a means by which the need can be met be known. If a new idea or invention is to be adopted then usually three conditions must be met. The first is that the knowledge as to how to create the idea or make the invention must be present; the second is that the idea or invention must meet a need; and the third is that the idea or invention must be the best way available to meet the need. The particular idea or invention must be the most economic or the most efficient way of meeting the need.

The desire that needs be met in the most efficient manner possible shows consistently throughout history. Efficiency gains can take the form of increased output, or better quality output, or the same output for fewer inputs. If one examines particular areas of economic activity such as energy production, transport, communications or the production of goods and services, it is possible to see the adoption of improvements, which continually increase the efficiency of humankind's technology. In relation to ideas, the simplest explanation consistent with the known facts, is the most efficient and is the explanation usually adopted.

The importance of a need existing before an invention or idea is adopted is shown by those inventions and ideas that could have taken place at earlier times due to their being relatively simple developments, but did not take place until later times. Such ideas or inventions could have been made without great difficulty, due to all necessary prior inventions having already been made, and yet those ideas or inventions were not immediately made. The reason for their discovery, when they were discovered, was that the need for the inventions before discovery was insufficient to justify the risk and expense of abandoning the existing practices and adopting the new invention or idea. In this situation the main determinant for when the discovery will be made is most likely to be when the need for the invention reaches a critical state, so that it becomes worthwhile to change the existing practices to adopt the new idea or invention.

There are a number of examples in history of inventions or ideas not being developed until a need arose. Prior to the development of double entry bookkeeping in Renaissance Italy, existing bookkeeping methods were adequate to record business activity. A considerable increase in trade meant that the existing bookkeeping methods were no longer adequate to cope with the increased business activity. The more sophisticated method of double-entry bookkeeping was then adopted to deal with the increasing level of business activity.

A similar situation existed with the technological improvements carried out in the textile industry in Britain in the early industrial revolution. Technological innovations such as Kay's flying shuttle, Hargreves spinning jenny, Arkwright's water frame and Crompton's mule were largely made by connecting together parts of previous inventions that had been around for centuries. They were relatively easy inventions and could be made by inventors with no special qualifications or training[5]. This suggests the timing of the inventions has more to do with market demand or a newly developed need that had not previously existed. It may be that increased demand, caused by increasing population and lower agricultural prices due to the agricultural revolution of eighteenth century Britain, required greater production than the cottage industry textile production of pre-industrial Britain could provide. Improved transport from canals and better roads may have allowed textiles to be sold over a larger area, thus allowing a larger scale of production.

The theory that it was population pressure that led to the development of agriculture is a needs based theory. This theory assumes that the knowledge required for agriculture was known to hunter-gatherers before the development of agriculture around 10,000 years ago. Before that time hunting and gathering was preferred to agriculture as it was a better life style and agriculture was only adopted when the population pressure forced humankind to adopt agriculture which was a more productive food acquiring system than hunting and gathering.

Human scientific and technological change requires the presence of both the knowledge as to how to make the change and the need for the change. If either of these factors is absent then the change will not take place. However throughout the course of human history it can be observed that the factor most commonly lacking is knowledge. This is because humans began with a full set of needs but with only a limited amount of knowledge, as knowledge, apart from that immediately available to our senses, is something that accumulates over time. In comparison, we are born with a full set of needs, the universal needs found in all cultures and only relative needs have developed over time. This means that it is knowledge that is usually the missing factor in our attempts to find better and better means of meeting our needs. It is the discovery of knowledge, which is the ultimate cause of human technological and scientific change, and such change is at the root of all fundamental historical change, social change and cultural evolution.

Knowledge

Many human societies have changed from hunting and gathering to farming and/or pastoralism and then to being industrial societies. What was necessary for this to happen? Obviously a knowledge of agricultural and pastoral practices and of the technology required for industrial society. Without this, the change from hunter gathering to farming and pastoralism and then to industrial society could not have taken place. The knowledge came from the capacity of humans to learn and from the human desire to meet certain needs in a better and more efficient manner.

However the human capacity to learn has existed ever since *homo sapien-sapiens* have been on this planet and the needs have always been there even though previous societies have been less able to meet the needs than industrial societies. The difference is that the knowledge of how to meet the needs in a better and more efficient manner has not always existed. It has gradually accumulated over time. It is the increasing knowledge that is present in the change from hunter-gathering to farming and pastoralism and then to industrial societies that is absent from the preceding society. The knowledge required for industrial societies was not available in agrarian and pastoralist societies and the knowledge of how to domesticate plants and animals was not known to prehistoric hunter-gatherers. Yet many of the needs of hunter-gatherers are the same as for modern humans. Only the knowledge of how to meet those needs is different between the various types of societies and this can be used to explain many of the differences between those different types of societies.

However the knowledge differences between those societies are not limited to knowledge of how to grow crops and herd animals and of various industrial processes. Agrarian societies usually have a knowledge of writing, metallurgy, transport (e.g. sailed and wheeled), and mathematics and in many other areas that does not exist amongst hunter-gatherers. Equally industrial societies have a knowledge of scientific matters that does not normally exist in agrarian societies, except by diffusion, and in the one agrarian society modern science existed in, it was an agrarian society on the verge of turning into an industrial society. Indeed, it was the growth in the knowledge of science in Europe from the time of Galileo to the beginnings of the industrial revolution in late eighteenth century Britain that was the necessary precursor to the industrial revolution.

The changes from hunter-gathering to agrarian/pastoralist to industrial societies were caused by changes in the methods used by humans to produce the goods and services that meet human needs. These were changes in the technology used by humans but behind the changes in technology were changes in knowledge. It was the changes in knowledge that caused changes in technology, which caused the historical development from hunter-gathering to agrarian/pastoralist and then to industrial societies. The idea that increasing human knowledge is a major cause of social, cultural and historical change can be traced back to Comte and J. S. Mill.

Changes in human knowledge resulted from the basic nature of human beings. The human ability to learn, to understand, to remember and human curiosity plus a desire to meet human needs resulted in humans gradually learning more and more about their environment. This ever increasing knowledge of human-kinds environment was the ultimate reason for the changing nature of human society, of human historical, social and cultural development and the replacement of hunter-gathering by agrarian/pastoral societies and in many cases of those societies by industrial societies.

While human beings have certain needs, those needs can only be met to the extent allowed by the knowledge available in the particular society. Originally human beings were hunter-gatherers, the same as our close relatives the great apes and all other animals. In common with some other animals, humans have made tools to assist in their hunting and gathering. However such human beings were limited in their tool making capacity by their knowledge being restricted to the use of stone, bone and wood. Such wooden tools as may have been used in Paleolithic times have long since decayed. However the stone and bone tools do survive and provide a record of increasing sophistication and efficiency. However, not only did tools get more and more efficient as humans learnt to make better and better tools, but the range of tools available to humans also expanded considerably as human knowledge of the properties of the materials in the environment increased.

There are two types of human knowledge that can be used to meet human needs. The first, which has been around as long as *homo sapien-sapiens*, is that of empirical experience, where humans have observed the results of certain behavior or processes. When certain behavior has produced a certain result in many cases in the past people have learnt that it will usually do so in the future. Stone tool manufacturers learnt that certain stones, especially flint, when chipped a certain way would produce a sharp edge, without any knowledge of the chemical structure of the material they were dealing with. Equally, early metal workers found they could shape metals and produce alloys, such as bronze, with no knowledge of why the metals behaved as they did.

The other way in which knowledge can be used to meet human needs is by logical reasoning from scientific laws or knowledge. This is a recent phenomenon existing only since science itself has existed. Modern inventions such as the internal combustion engine, television, radio, nuclear power and bombs arose partially or wholly from reasoning from scientific knowledge. This use of human knowledge would now be the primary means of technological development in industrial societies, but empirical observation still retains a role in modern technology and perhaps an important role.

Our knowledge of the environment does not include unsubstantiated speculations. Good guesses as to how our world is such as the atomic theory of the Greek philosophers Leucippus and Democritus, the heliocentric astronomy of Aristarchus of Samos and the suggestion by Giordano Bruno that the sun was a star did not constitute knowledge. At the time these ideas were made, the evidence was against them, and they were not accepted at that time. Only ideas that are accepted constitute part of the knowledge of any given society.

Properties and structure of the environment

Human knowledge is of course knowledge of the human environment. It can scarcely be of anything else. The objects in our environment, including ourselves, have certain properties which determine whether those objects are able to meet human needs or may be processed in such a way that they will meet human needs. The nature of human biology determines where we live and what our needs are. We cannot fly or breathe under water, so we live on the surface of the earth. We have a need for freshwater and as water is a heavy item, relative to human strength, we have spent most of our history living close to supplies of freshwater. We have a need for food and as this need is not as easily met as other human needs, such as for oxygen, humans have spent a great deal of time and effort in searching for or growing food. It is only since the industrial revolution, in some societies, that the production of food has become a lesser part of human activity.

However it is not just human biology that determines how we live. The biological nature of the plants and animals in our environment determines which we live on and which we do not. Some plants are poisonous to us and some animals are too fast for us to catch. However the wide range of food humans can consume has allowed humans to spread over the entire planet. Some plants and animals may be relatively easy to domesticate, others cannot be domesticated at all. It is the property of some plants that they are capable of domestication that enabled the development of agriculture. Plants ideal for human consumption may be sown, fertilized, watered, protected from competing plants by weeding and will grow and provide the food necessary to feed human populations. Some animals may be domesticated and may serve as draught animals as well as their meat, hides and milk being utilized to meet human needs. If plants and animals were incapable of domestication or, if domesticated, they were not able to meet human needs, then they would not have been domesticated and human history would be quite different.

A further determinant of how humans live is the properties of non-living matter that makes up the human environment. It is because wood and flint can be easily manipulated and altered, by chipping in the case of flint and breaking or cutting in the case of wood and because they are hard and can be made sharp that they have been important materials for tools and weapons. Materials such as bone and ivory have similar properties and have also been used for such purposes. It is the properties of some metals, such as hardness, malleability and that they can be mixed together to produce alloys, such as bronze and steel, that allowed them to supersede wood, flint, bone and ivory as the principal material for tools and weapons. If these materials did not have the appropriate properties they would never have been used to make tools or weapons.

It is the property of clay that it occurs naturally as a sticky but plastic lump and as a lumpy liquid. The structure of clay is that the particles of clay are flat and plate-like and the addition of water enables them to slide over each other without breaking apart. This enables clay to be formed into almost any shape, making it ideal for the creation of pottery.

It is the properties of sand, soda and lime, when melted together that they will form an opaque or transparent substance, as desired, which we call glass. It is the properties of glass, that it can be transparent or colored and can be molded into different shapes, which makes it useful to meet human needs as windows, ornaments and vessels of various kinds.

It is also the particular properties of hides, wool, fur and cotton and other products that enable them to be fashioned into clothes capable of keeping people warm. If these products did not have those properties they would not have been used for the purpose for which they were used. If there were no products with the properties required for clothing then the area of human habitation of the planet would have been severely restricted to the warmer and temperate areas of the planet.

Certain products in the natural world are also used for the construction of buildings, most particularly, wood, stone, mud and bricks. It is because these materials are the most suitable materials available to create buildings and structures that they were used for those purposes. They have the right properties for use as building materials. If these materials had not existed or it was not possible to make them, then either other less suitable materials would have been used with less satisfactory buildings being created or if there were no suitable materials, then no "permanent" buildings would have been built.

The objects in our environment will be in a state of being a gas, a liquid or a solid. Gases have the property of being able to expand and fill any available space. Gas molecules are only loosely connected. They assert pressure on the wall of anything they are held in. If the container of the gas is reduced in size, the pressure of the gas on the container's walls will increase. If the size of the container is increased the gas pressure on the container walls will reduce. If the gas is heated, the gas pressure will also increase and the gas will expand if it can. If it cools the gas pressure will fall. Hot expanding gas has been used to drive steam engines, to fire bullets, cannon balls and accelerate rockets.

It is the property of gases that when heated, their pressure increases. This is what caused the piston to rise in the early steam engines. It is also the property of gases that when their volume increases their pressure reduces so that as the piston rises the pressure of gas beneath the piston would fall. When the atmospheric pressure on the top of the piston is greater than the pressure beneath the piston, the piston will fall causing the gases beneath the piston to compress. This will cause the pressure underneath the piston to increase, which will cause the piston to rise again and so on. It is this property of gases that they expand when heated and that their pressure falls when their volume rises and the pressure rises when their volume falls that made the early steam engines possible.

Liquids have no fixed shape but do have a fixed volume. Liquid molecules slide over each other so as to fill any available space but they do not move as freely as gas molecules. Solids have a fixed space and are more strongly bound together than liquids. Different solids tend to have different properties depending on their composition and structure. Solids such as metals, bones, computer chips and gemstones are crystals and have a regular array of atoms tightly packed together. Plastics are formed from long chains of molecules linked by carbon atoms while glass has a largely random structure.

Whether matter is solid, liquid or a gas affects their properties, but each mixture, compound and element in nature has its own individual properties. Metals tend to have certain properties in common. They conduct heat well; they have high electrical conductivity; they have high reflectivity and a shiny metallic luster; they are malleable and ductile; other than mercury they are solid at room temperatures and they emit electrons when exposed to high energy and heat. Nonmetals tend to be poor conductors of heat and electricity; they may be gas, liquids or solids at room temperature; when solid they tend to be brittle and fracture under stress. Different metals of course have different properties. Iron has a melting point of 1535°C, copper's melting point is 1083°C, aluminum's is 660°C and lead's is 327°C. The density in g cm^{-3} of aluminum is 2.71, iron is 7.86, copper is 8.97 and lead is 11.4. It is the low density or weight of aluminum that is the reason it is used in aircraft and space vehicles. It is the third most abundant element on the earth's surface so it is relatively inexpensive, and it is used for beer and soft drink cans and household utensils. Iron is also fairly common and its alloy steel, which is much stronger and harder than iron, is used in buildings, bridges, cars, machinery and in many other areas. Copper was one of the first metals to be used by humans, as, with gold and silver, it exists on earth in its pure state so no smelting is required to release it from its ore. Furthermore when smelting was developed the low melting temperature of copper meant it was the first extensively used metal. Copper has a very high electrical conductivity and is soft and ductile so it can be drawn into thin wires and is widely used for electrical wiring. Lead has a low melting point and so is easily extracted from its ore. Due to this it has been used for a long time. It was used by the Romans for lead pipes for the supply of water. These days lead is used for making batteries and in type metal and solder.

Our environment has a particular structure as is revealed by the laws and facts of physics, chemistry and biology. Curved space time, gravity, the laws of motion, the structure of atoms, electro-magnetism, the chemical bonds between atoms, our biological and non-biological needs and our physical and mental capabilities all go to make up the structure of our environment.

Order of discovery

Human knowledge of the properties and structure of nature is acquired in a particular order. Certain things will necessarily be discovered before other things. Fire had to be discovered before metallurgy, as it is a necessary part of the metallurgical process. Copper was inevitably the first metal to be extensively used by human beings as it has a relatively low melting point. This meant it could be more easily released from its ores and shaped and reshaped than other metals. However the working of copper requires a furnace and molds so that inevitably it could only be done by a sedentary people. It is obviously not practicable for hunter-gatherers to carry round furnaces and molds. This meant that metallurgy could only develop after the domestication of plants and animals. The occasional example of sedentary hunter-gatherers, such as those on the north-west coast of America, do not seem to have developed metallurgy. Copper is a soft metal which limits its uses; a much stronger metal, bronze, can be made by mixing copper with another metal such as tin. Inevitably bronze was discovered after copper, as the use of copper is a necessary part of the manufacture of bronze. Bronze could not be made without the earlier discovery of how to produce copper and tin. The next metal to come into common use was iron. Iron has a melting point of 1535°C, about 500°C higher than copper. This means a bellows is required to produce the necessary heat for the smelting and working of iron. Inevitably the metals that cannot be worked without bellows only came into common use after the invention of the bellows. They would also only come into common use at a later time than the use of such metals as copper and bronze, which did not require the use of bellows. Iron came into use after bronze, as the process of creating an alloy is a relatively simpler process than the creation of heat of 1535°C which is required to work iron. Iron was followed by steel, an alloy of iron and carbon. Obviously steel could not be made until after it had been discovered how to work iron, as iron is a necessary part of the production of steel.

The process of one thing necessarily being followed by another, either because the earlier thing is a necessary ingredient in the later thing, or because the earlier thing requires a simpler technology, such as fire with a lesser heat, can be seen throughout the history of science and technology. Inevitably, the steam engine had to be invented before it could be given rotary motion, and it had to be given rotary motion before it could drive the new machinery being developed in the industrial revolution and steam locomotives and ships. The sedentary lifestyles produced by the agricultural revolution were a necessary part of a great host of scientific and technological discoveries. Permanent buildings, metallurgy and writing are just three of the more important developments that would not have happened without the prior development of sedentism. The domestication of animals was a necessary precondition to developments such as wheeled transport and plough agriculture. The discovery that the earth and other planets orbit the sun could not be made, or at least confirmed, without the prior invention of the telescope. Without the telescope there would have been insufficient information about the movement of extra-terrestrial bodies to support the helio-centric theory. The development of more complex mathematics such as calculus and differential equations was necessarily dependent upon the earlier development of number systems and simple operations such as addition, multiplication, subtraction and division. The discovery of electricity had to take place before electrical heating and lighting and computers. The splitting of the atom by Rutherford had to take place before the development of nuclear power and nuclear bombs. These are just a few of the more obvious examples of the way in which certain discoveries or inventions could not have been made without prior discoveries or inventions being made.

There are lines of development through which the increases in human knowledge inevitably move. Many discoveries could not be made, without a succession of prior discoveries having been made. The line of development would be the simplest way in which any given discovery could be made. It may be that there are more difficult ways in which a discovery could be made, but in fact discoveries are most likely to be made in the simplest way possible, along the simplest line of development.

A line of development does not mean the continual improvement of a particular invention or idea such as the improvements in the steam engine during the Industrial Revolution or the change from the Ptolemaic theory of the universe to Newton's theory and then to general relativity. Rather it involves a series of discoveries that have to be made before an idea or invention is adopted by a society. It will for example include ideas and inventions that are not directly a part of the invention or idea that is being developed. The line of development of the steam engine for example included the invention of the air pump and the subsequent discovery of some of the properties of gases. These discoveries were necessary before a steam engine could be developed. The line of development of humankind's view of the universe included such inventions as the telescope and the prior discoveries of how to make glass and that glass could be shaped in such a way as to magnify objects seen through the glass. Further discoveries that were part of the development of the human view of the universe were mathematical ideas such as calculus, an important part of Newtonian physics, and non-Euclidean geometry, which provided support for general relativity.

Lines of development grow much as the branches of a tree. Inventions and ideas will often be developed due to prior developments in a wide range of areas, totally unrelated to the invention or idea that is subsequently developed.

The following table[6] shows the approximate dates for the development of various new technologies in six different areas. Some of the dates are controversial and are a simplification of complex events about which little detail is known. Dates for animal domestication, concern food producing animals, rather than dogs, which were domesticated before food producing animals.

	Fertile Crescent	China	Andes	Amazonia	Meso-am erica	Eastern U.S.
Plant domestication	8500 bc	by 7500 bc	By 3000 bc	3000 bc	By 3000 bc	2500 bc
Animal domestication	8000 bc	by 7500 bc	3500 bc	?	500 bc	_
Pottery	7000 bc	by 7500 bc	3100-1800 bc	6000 bc	1500 bc	2500 bc
Villages	9000 bc	by 7500 bc	3100-1800 bc	6000 bc	1500 bc	500 bc
Chiefdoms	5500 bc	4000 bc	By 1500 bc	ad 1	1500 bc	200 bc
Widespread use of copper bronze tools	4000 bc	2000 bc	ad 1000	_	_	_
States	3700 bc	2000 bc	ad 1	_	300 bc	_
Writing	3200 bc	By 1300 bc	_	_	600 bc	_
Widespread iron tools	900 bc	500 bc	_	_	_	_

The table shows a more or less consistent pattern, with plant and animal domestication, villages and pottery occurring around the same time, with chiefdoms and non-iron metal tools occurring later and states, iron tools and writing being developed still later. Insofar as the order varies such as in Amazonia where pottery and villages occurred substantially before agriculture it could be due to local conditions such as unusually abundant wild plants and animals, which allows the existence of sedentary hunter-gatherer communities.

The question needs to be asked, why are certain discoveries made before other discoveries and certain discoveries could not happen without prior discoveries being made? The answer is that the universe has a particular structure and particular properties. The structure of the universe and its properties becomes known to us in a particular order. This order could be described as either from the simpler to the more complex or perhaps from that which is closest to us to that which is further from us. We learn about the world in a particular order and that order is due to the relationship between ourselves and the world. Our usual way of observing our world is with our naked senses and this gives us certain information about the world. We learn additional information by means of practical empirical, trial and error experiments, such as when we learnt that if flint were chipped in a particular way, it would produce a useful tool. We gain increased knowledge about the world either through changing the method of observation, such as using telescopes or microscopes or by making empirical experiments that show the relationship of one thing to another.

Ease and difficulty of discovery

The order of discovery of human knowledge of nature is determined by how easy it is to make that discovery. What determines whether a discovery is easy to make or more difficult? If there is direct sensory experience of something then the discovery of that thing is fairly easy. There are many examples of this. We have direct sensory experience of air, for example with the wind, leading to it being included within the four elements of ancient Greek philosophy. Yet there is no direct sensory experience of oxygen and nitrogen in the air as these gases are colorless, odorless and tasteless and make no sound or cause any feeling distinguishable from the air as a whole. It was not until the late 18th century that oxygen and nitrogen were discoverable as a result of a series of experiments carried out by scientists such as Lavioiser, Priestly, Scheele and Cavendish.

A further factor in whether a discovery is easy or not depends upon whether other prior discoveries need to be made before the discovery is made. In metallurgy, native metals, which do not have to be separated from an ore, were used earlier in history than metals from ores as it was not necessary to discover a prior smelting process to get the metals from their ores. A further example from metallurgy is that copper metallurgy developed before iron metallurgy as copper could be smelted using an ordinary kiln while iron smelting required higher temperatures than copper smelting, so a kiln with a bellows was necessary for iron smelting. Iron smelting required the prior discovery of a kiln with bellows before it could be developed.

Many modern inventions and discoveries required a considerable number of prior inventions before they could be made. A modern personal computer would have required discoveries such as electricity and how to control electricity, how to control electrons on a computer monitor, discoveries in metallurgy and in the production of plastic materials and developments in mathematics and computer programming before its invention. The prior inventions of writing, printing and keyboards were also required while in mathematics the invention of a number system, simple operations such as adding, subtracting, multiplication and division and the invention of binary number systems were required for the invention of the modern computer. The list of discoveries required prior to the invention of the personal computer is enormous and ultimately goes back to the discovery of fire.

A discovery will be difficult to make and so will take place at a later point in history, if cause and effect are not closely linked. Many hunter gatherer and tribal societies are unaware of the relationship between sex and pregnancy. This is not surprising as there is a time gap of some months between when sex takes place and the first clear signs of pregnancy occur. There are also many occasions when sex takes place and there is no pregnancy. This makes the discovery that sex causes pregnancy quite difficult. However if for example pregnancy was clearly present within a few hours or days of having sex then the discovery that sex causes pregnancy would have been made much earlier in history. A similar situation exists with the discovery that plants grow from seeds. Seeds do not look at all like plants and they will usually be in the ground for some time before the plants emerge. In addition there are many occasions when seeds are in the ground and they do not turn into plants. If however seeds fell to the ground and within hours or even days plants began to grow then the discovery that seeds grow into plants would have occurred much earlier in human history.

A discovery may also be easier if that which is to be discovered exists in abundance than if it is available in more limited quantities. The noble gases such as argon, helium and neon were discovered a lot later than nitrogen and oxygen as they make up only 1% of the atmosphere, while nitrogen and oxygen make up the other 99% of the atmosphere.

Yet another factor that could make a discovery easier or harder is whether the thing being discovered has properties that are easily detectable by scientific instruments. Charged particles such as electrons and protons were discovered before the neutron as scientific instruments were able to detect the charge on charged particles. A further problem was that the neutron, unlike electrons and protons, cannot survive outside the atomic nucleus as when it is outside the nucleus it decays into a proton, electron and an antineutrino. These properties make the detection of the neutron rather more difficult than the detection of protons and electrons.

The above five factors are some of the more obvious things which affect the ease or difficulty by which discoveries concerning nature are made. The list is not exhaustive and further study may reveal other factors which affect the ease or difficulty of discovery.

Multiples

A lot can be learnt about the order of discovery of things in our environment by a study of the phenomena of "multiples". Multiples concern the multiple and independent discovery of the same scientific idea or invention. Considerable work was done on multiples by William Ogburn and Dorothy Thomas who established a list of 148 independently duplicated scientific and technological discoveries. They suggested these discoveries became virtually inevitable as knowledge accumulated within any given society and the needs of that society caused attention to be directed towards problems associated with meeting those needs.[7]

The history of science and technology provides many examples of multiples. Some of the better known examples are:

1. Agriculture and the domestication of animals were invented independently in the old world and the new world. It may be that there were a number of independent inventions of agriculture and the domestication of animals in both the new and old worlds. It has been suggested that agriculture was an almost simultaneous yet completely independent development in South West Asia, China, Southeast Asia, Mesoamerica, South America and the Eastern United States.[8].
2. Calculus may have been invented independently by both Newton and Leibnitz leading to conflicting claims as to who was first. However it may have been the case that Leibnitz had seen Newton's work before it was published.
3. The theory of evolution was invented separately by both Darwin and Wallace. Both had read Malthus's *Essay on Population* and had been studying flora and fauna in Darwin's case in the Galapagos Islands and in Wallace's case in Burma.
4. The periodic table was proposed by Mendeleev in 1869 and a year later a similar idea, developed independently, was put forward by Lothar Meyer.
5. The discovery of oxygen was made by Carl Scheele in 1771, but his work was not published until 1777. Joseph Priestly independently discovered the gas in 1774 and informed Antoine Lavoisier and both Priestley and Lavoisier continued to work on the gas until Lavoisier concluded the gas was a separate component of air.
6. The discovery of Neptune was made by Adams and Leverrier in 1846.
7. Genetics was discovered by Mendel in the 19th century and then independently by Hugo Marie de Vries, Erich von TSchermak and Carl Correns in 1900.
8. Non- Euclidean geometry was independently invented by Carl Gauss, who did not publish his work and the Russian Niolai Lobachevsky in 1829 and by a Hungarian Janos Bolyai.
9. The wave theory of light was developed independently by Thomas Young in England and Augustin Fresnel in France.
10. Visual pigments were independently discovered by German physiologists Franz Boll and Wilhelm Kuhe.

There are many more examples of multiples; Robert Merton came up with 264.[9]

Merton considered that the pattern of independent multiple discoveries in science is the dominant pattern of scientific discovery and that discoveries made only once in science, known as singletons, are the more unusual case. More particularly he considered that all scientific discoveries were, in principle, multiples. Merton gives ten reasons for that belief.

The first is that many discoveries considered to be singletons turn out to be rediscoveries of previous unpublished work. He gives the example of the physicist and chemist Cavendish and the mathematician Gauss both of whom were reluctant to publish their work and their discoveries were made later by others with the discoveries being considered to be singletons. When Cavendish and Gauss's work was later discovered and published it was realized that the cases were multiples rather than singletons. Merton's second reason for believing all scientific discoveries are potential multiples is that there are many examples of scientists discontinuing inquires when they become aware that someone else has published the same work. Merton’s third reason is that even when scientists are beaten to publication by others they still report their own work. His fourth reason involved cases of unnecessary duplication of scientific work. When such duplication is discovered, one set of work is stopped, so the work is eventually considered to be a singleton. Merton's fifth reason concerns scientists often believing their work is original until being informed that another had already written on the subject. His sixth reason is where scientists, he gives the example of Lord Kelvin, give lectures only to be informed by the audience that his work had already been discovered and published by others. Merton's seventh reason is where a scientist with a clearly developed program of investigation gives up the investigation due to interference by others. All these cases involve situations which are singletons, but would have been multiples but for the scientists discovering others had done the same work.

Merton's last three reasons for suggesting all singletons are potential multiples, concern the behavior of the scientists themselves. Merton considers that this behavior shows that the scientists themselves believe that all scientific discoveries are potential multiples. His eighth reason is the race scientists engage in to get published. Their assumption is that they must publish quickly or someone else will publish and get the credit for the discovery. The ninth reason is that scientists are known to advise each other to publish quickly or someone else will publish earlier and gain credit for the discovery. Merton's last reason is the practices used by scientific institutions to protect scientists' priority for discoveries. Practices such as the depositing of sealed and dated manuscripts, containing an outline of an idea, with scientific societies and academies show that scientists believe that their discoveries will usually be under threat of being discovered by others. Merton considers that all singletons are singletons only because one discoverer published his or her work before others were able to complete their work. If publication were delayed long enough someone else would eventually make the same discovery. Scientists' own behavior confirms they also believe this to be the case.

The consequences of the occurrence of multiples in the history of science is expressed by Merton as:

“Such occurrences suggest that discoveries become virtually inevitable when prerequisite kinds of knowledge and tools accumulate in man’s cultural store and when the attention of an appreciable number of investigators become focused on a problem by emerging social needs, by developments internal to the science, or by both."[10].

Multiples suggest that discoveries are inevitable because if one scientist does not make the discovery, another one will. This was also the view of Ogburn and Thomas and has become the standard interpretation of multiples. This suggests there is an inevitable element in the progress of science and technology, so long as it is not interfered with by external forces such as governments and religious authorities.

Multiples also suggest that discoveries are not only inevitable, but that they must take place in a particular order. Thousands of years of human history may go by without something being discovered, and then several scientists or inventors make the same discovery at the same time. This suggests that certain prior developments were necessary before a discovery could be made. This is what Merton was referring to in the above quote when he mentioned "prerequisite kinds of knowledge and tools [must] accumulate in man's cultural store" before a discovery could take place. Only when that knowledge and those tools have been discovered is it possible for certain later discoveries to be made.

The existence of multiples is exactly what would be expected if there were a specific order of discovery for science and technology. A particular scientific fact or technological achievement may remain uncovered for thousands of years and then be discovered separately by two or more individuals suggests it could not have been discovered until certain other scientific facts or technological achievements had been discovered. It also suggests that when those other facts and achievements have been uncovered then the discovery of further scientific facts and technological achievements will be almost inevitable. This however is conditional upon the state of society being conducive to scientific and technological discovery. In particular there should be no institutions, such as church or state, interfering with the process or communication of the discovery.

Guttman Scale Analysis

The idea that societies acquire social and cultural traits in a particular order is also shown by Guttman scale analysis. Guttman scale analysis is a method of assembling data that can show the order in which social and cultural traits were acquired. When repeated over a number of societies it can suggest there is an order of acquisition of traits that commonly occurs and occurs far too often to be regarded as a statistical coincidence. A detailed explanation of how Guttman Scale Analysis works is provided in Appendix 3. Guttman Scale Analysis has been carried out in a number of studies to examine the order in which societies acquire social and cultural traits by Robert Carneiro and most recently by Peter Peregrine and Carol and Melvin Ember.

The Peregrine-Ember study[11] looked initially at eight social and cultural traits being inter-societal trade, subsistence economy based on food production, social stratification or slavery, full-time government specialists, full-time craft specialists, political states of 10,000 in population, towns exceeding 1,000 in population and writing. Using scale analysis Peregrine and the Embers concluded that the scale analysis suggested there were general sequences in cultural evolution and a comparison of how these traits developed in eight societies being the Yellow River Valley, Nile River Valley, West Africa, Mesopotamia, Indus River Valley, Highland Peru, Lowland Peru and Highland Mesopotamia confirmed the conclusion of general sequences in cultural evolution that applied to a wide variety of societies.

Only eight traits were used for the study, so, to avoid the possibility of chance affecting the results, a further study was made using fifteen traits. Those traits in order in which they scaled and in which societies developed them were ceramic production, presence of domesticates, sedentarism, inegalitarianism, population density of more than one person per square mile, reliance on food production, villages of more than 100 people, metal production, presence of social classes, towns of more than 400 persons, states of 3 or more levels of hierarchy, population density of more than 25 people per square mile, wheeled transport, writing and money. The sequence in which these traits were developed was compared in the same eight societies used in the first study and the results showed very similar scaling between those societies indicating a universal pattern in cultural evolution. The scaling was not perfect, for example in five of the eight sequences ceramics was not present before domesticates, but the overall results show a pattern that was not random and could not have arisen through chance.

Robert Carneiro made a more detailed series of studies of a larger number of traits and societies, using scale analysis, than was used in the Peregrine-Ember studies. These studies consistently showed societies developing traits in a particular order that could not be explained by chance. One study[12] involved fifty traits ranging from special religious practitioners, the most common trait, to the least common which was temples extracting tithes. The societies studied ranged from the aboriginal Tasmanians, the society with fewest traits, to New Kingdom Egypt. The scale analysis showed that societies with the most traits had the same traits as the other societies, plus additional traits and the most common traits existed in nearly all societies. This by itself does not show sequence but Carneiro then compared[13] thirty four of the traits whose order of development could be identified in Anglo-Saxon England with the order of development suggested by scale analysis. The comparison resulted in 84.9% of the traits in Anglo-Saxon England arose in the order suggested by scale analysis, while 15.1% did not. Such a result, given that historical information may not be perfect and that scaling may be affected by the description and diffusion of traits, suggests the order in which traits were acquired was not random. Carneiro made further studies involving a comparison of two traits such as agriculture and cities over many societies and found, not surprisingly, that agriculture preceded cities in every case. A similar comparison involving taxation and sumptuary laws found that while neither trait was very common in the societies studied, indicating they were developed later in time, sumptuary laws always followed taxation, indicating a definite order of development. Caneiro also studied cases of differential evolution where evolution within one area of culture develops independently of other areas of culture. He considered that as traits from a particular part of culture, such as political organization are more closely related to each other than they are related to other traits there will be a greater degree of scaling between traits concerning political organization than with other cultural traits. When Carneiro scaled the traits concerning political organization he found they showed a higher degree of scaling than was obtained by scaling all cultural traits together.[14]

The results of the Peregrine-Ember studies and Carneiro's studies indicate that societies develop cultural traits in a particular order. This is shown over a wide variety of traits and over a wide variety of societies. The results of the Guttman scaling show the accumulation of cultural traits is not random as random accumulation of traits would produce quite different results in scale analysis.

The rate of historical change

A study of history reveals that the rate of change varies from one period to another. Before the domestication of plants and animals there were many tens of thousands of years when the rate of change, in the way humans lived, was very slow. Improvements in the technology employed by human beings were made, but only very slowly. After the domestication of plants and animals there was a period of rapid change as sedentism allowed the development of many new technologies and the beginnings of science and mathematics. This was followed by a period of slow change, sped up somewhat by the achievements of the classical Greeks. The golden age of classical Greece was followed by a period of slow intellectual and technological change. A period of more rapid change began with the development of modern science in late Renaissance Europe and was accelerated by the industrial revolution beginning in the late eighteenth century. This period of rapid change has continued to the present day. The picture is one of both science and technology growing unevenly, with periods of rapid change giving way to periods of slow change or even stagnation.

In technology a distinction is sometimes made between macro and micro inventions. Macro-inventions involve radical new ideas, without clear precedent and emerge more or less *ab ninito*. Micro-inventions are small incremental steps that improve, adapt and streamline existing techniques, reduce costs, improve form and function, increase durability and reduce energy and raw material requirements.[15] In practice macro and micro inventions are on a continuum and there are many inventions that are somewhere in the middle between macro and micro inventions.

The development of macro-inventions was difficult and they are comparatively rare. They require a considerable leap in human imagination; they involve a major new discovery of how nature can be utilized to meet human needs. Micro-inventions are relatively easier to develop and more or less inevitably follow the development of macro-inventions.

It is this situation that explains the uneven growth in technology. Where a major macro-invention has been made it will often stimulate or allow the development of many other inventions producing periods of rapid technological change. When the inventions stimulated or allowed by the macro-invention have run their course and all have been made, then this will lead to a period of slow or no technological change. Major macro-inventions such as the domestication of plants and animals allowed sedentism and this allowed the development of metallurgy, permanent buildings and writing. Writing and other forms of record keeping allowed the development of government and bureaucracy. The steam engine had a similar effect, allowing the driving of the machinery invented in the industrial revolution and new transport systems such as the steam ships and railways.

However other periods such as those of classical Greece and Rome were periods of little technological development. It was certainly not the nature of Greek and Roman society that caused their poor record for producing new technology. Both societies were wealthy, had considerable trade that produced large amounts of capital; they had relatively large numbers of educated, literate people, they had reasonably secure property rights and substantial legal systems and religions that were generally tolerant and open to new ideas. Their failure to produce substantial technological developments was, not because of slavery as is sometimes suggested, but because the macro-inventions, their society used had been improved as much as possible by micro-inventions and they were unable to produce more macro-inventions as that would have involved a leap that was beyond their societies. They produced no macro-inventions and little in the way of micro-inventions so their societies were comparatively limited in producing new technology.

It is sometimes suggested that the classical world failed to reach some fairly obvious solutions to technical problems that they faced.[16] However what is an obvious solution in hindsight is not necessarily obvious to those without the benefit of hindsight. If a generally intelligent, literate people such as the Greeks and Romans were unable to come up with answers to problems, then it seems likely that the solution to the problems were difficult rather than easy. There may have been problems such as poor workmanship or materials that would have made solutions, which are obvious to us, impossible in classical times. Alternatively technological solutions available in classical times may not have been used for economic reasons, as there were cheaper solutions to the problems than the use of the particular technology.

The same situation that applies to technology and macro-inventions applies to more intellectual developments. Science has its own macro-discoveries, perhaps the most important being the development of the modern methodology of science. The development of the practice of systematic experimentation and the application of quantitative approaches to science were macro-inventions that have led to a dramatic growth in scientific progress since the seventeenth century. In mathematics, the Greek discovery of abstract theoretical mathematics was a macro-discovery that led to considerable progress in geometry. Similarly, the discovery of the zero and Hindu-Arabic numerals was a macro-discovery that resulted in considerable improvements in mathematics since the Renaissance. The scientific revolutions described by Thomas Kuhn in *The Structure of Scientific Revolutions* could also be considered to be macro-discoveries. Newton's revolution in physics and Lavoisier's in chemistry produced radical changes within those sciences and led to periods of what Kuhn called normal science. Normal science involves problem solving within the context of a particular view of science called a paradigm and is broadly similar to the idea of micro-discoveries.

The concepts of macro and micro-discoveries in both science and technology explain the varying rate of historical change. Periods of macro-discoveries are periods of rapid change, periods of micro-discoveries are periods of steady change and periods when the micro-discoveries derived from particular macro-discoveries have run their course are periods of stagnation. Macro-discoveries occur when there has been a great leap in human knowledge, which is able to be built on and expanded by the acquisition of more easily acquired knowledge

A map of the facts of the universe

A map shows the location in space of different places, such as countries, cities, streets and other geographic entities. If a person knows where they are located on the map they are then able to work out where they are in relation to other places and through what places they would have to pass to arrive at any other place. It should be equally possible to produce a "map" showing where the facts of the human environment are in relationship to human beings and to all the other facts of the human environment. This is a direct consequence of the human environment having a particular structure and that human knowledge of the environment grows in a particular order with certain discoveries inevitably being made before certain other discoveries. Such a map will not show the location of facts in space, rather it will show their location in relation to each other and to humankind.

The basis of such a map is that some facts (say facts B) will not be obtainable without the prior discovery of other facts (say facts A). This means that facts B will lie beyond or are further away from us than facts A. Obviously the discovery of planets such as Neptune, Uranus and Pluto would not have been made without the prior discovery of some means of observing them, such as the telescope. This is because they cannot be seen by unaided sensory observation. Equally, metallurgy, pottery and glass making could not have been discovered without the prior discovery of fire, as fire is a necessary ingredient in metallurgy, pottery and glass making. The discovery of Neptune, Uranus and Pluto lie beyond the discovery of the telescope or some other means of extending human sense perception and the discovery of metallurgy, pottery and glass making lies beyond the discovery of fire.

A further way of locating facts on such a map is where certain facts are relatively easily acquired such as how to make fire and certain other facts such as how to do calculus, are less easily acquired. This is because the discovery of calculus is more complex than the discovery of fire. Calculus requires a number of prior discoveries to be made before it could be discovered.

The knowledge of fire is not a precondition to the discovery of calculus, but calculus was always going to be discovered after the discovery of fire and so could be located on a map as being much further from human beings than the discovery of fire. Calculus would be located on a different line of development from fire, being on a line of development requiring the invention of a number system and the ability to do simple mathematics such as addition, subtraction, multiplication and division.

Certain facts are obvious to the naked senses. The four elements of classical Greece, air, fire, water and earth are obvious to the naked senses and are widespread in nature and so were the first explanation of the constituents of matter. Indian science had the same four elements of classical Greece. The Chinese had five elements being water, fire, earth, metal and wood. The difference between the Chinese elements and the Greek and Indian elements can be put down to neither theory being correct, the correct understanding of the constituents of matter being beyond classical Greek, Indian and Chinese science. Naked sense observations of matter were always going to produce theories like the Greeks, Indians and Chinese held but as there was no way they could produce a conclusive answer to the constituents of matter, the theories could always be a little different.

A further Greek explanation of the nature of matter was the mathematical theories of Pythagoras and Plato. Such theories could not be developed until a society had reached a certain level of mathematical knowledge, so they will lie further away from human kind than the facts immediately available to the naked senses. The classical Chinese never had such geometric theories of matter as their geometry was never as sophisticated as that of the Greeks.

The traditional Greek view of fire, air, water and earth as the basic elements of matter continued to be at least partially accepted in Europe until the revolution in chemistry that occurred in the late eighteenth century. The decomposition of air and water brought about by the use of new scientific instruments and techniques lead to the modern concept of elements as matter that could not be broken down into constituent parts. Lavoisier’s list of 33 elements, despite some mistakes, was the first modern list of elements. The list of elements was subsequently corrected and added to when new elements were discovered. Dalton’s atomic theory suggested different elements were made up of different atoms and this explained the different properties of the elements. The eighteenth and nineteenth century concepts of elements and atoms could not have been developed without the prior decomposition of air and water which showed they were not elements but were made up of other substances. The discovery of the elements was necessary before the atomic theory, which explained the different elements as being made up of different atoms.

Atoms remained the basic constituents of nature until 1897 when J J Thompson discovered the electron. The nucleus of the atom was then discovered by Ernest Rutherford, which made a negatively charged electron and the positively charged nucleus the basic constituents of matter. The neutron was added in 1932 with its discovery by James Chadwick, so the basic constituents of matter were the proton, neutron and electron. In the 1960’s protons and neutrons were discovered to be made up of quarks, so the smallest constituent parts of matter could be considered to be electrons and quarks. There is considerable current debate as to whether quarks and electrons are made up of tiny vibrating strings called superstrings.

There was an order of discovery running from the elements of ancient Greece, India and China to the mathematical theories of the Greeks, to the elements as discovered in the late eighteenth century, to Dalton's atoms, to the nucleus of the atom and electrons, to protons, neutrons and electrons, to quarks and electrons and possibly to strings. The particular order in which these discoveries were made was inevitable. This enables us to say that in some sense that those things we can see with unaided sense perception are closer to us and that successively the mathematical ideas for the constitution of matter by Pythagoras and Plato, the idea of the elements, atoms, the nucleus and electrons, protons, neutrons and electrons and quarks and then strings are located further from us.

A similar situation applies in astronomy. The unaided sensory view is that the earth is not moving and the sun orbits the earth. When more sophisticated observations were made of the heavens the Greeks created the Ptolemaic system with a stationary Earth being the center of the universe and being orbited by the sun and the planets in circular orbits with epicycles being used to further describe the planet's movements.

The classical Chinese cosmology also considered the earth to be the motionless center of the universe with various theories of the sun and the planets orbiting the earth. The Chinese theory however differed from the Greek by not having the Greek geometric schemes of planetary motion. Indian cosmology also involved a stationary earth orbited by sun and planets and seems to have been as geometric as the Greek cosmology.

The Ptolemaic system survived in Europe, until Copernicus published his helio-centric theory and Kepler showed the Earth and other planets orbited the sun in elliptical orbits. Kepler had the benefit of improved observations of planetary movements from Tycho Bathe and his theory could be confirmed with observations made using the newly invented telescope. The work of Copernicus and Kepler was ultimately completed by Newton with his laws of gravity and motion with the help of new mathematical tools such as calculus.

Observations of planetary motions continued to improve and it was observed that Mercury did not move in accordance with the Newtonian system. Eventually the Newtonian system was replaced by Einstein's law of general relativity, which had the planets orbiting the sun in circular orbits in curved space-time. Improved mathematical tools such as non-Euclidean geometry helped the establishment of general relativity.

The order of discovery from a motionless Earth orbited by the Sun, to the Ptolemaic and classical Chinese and Indian systems, to the Newtonian system to Einstein's system was fixed. Each system gave way to its successor due to improved observations and/or mathematical tools. Each successive system can be considered to be further away from humankind than its predecessor so that the closest to humankind is the sun orbiting the earth, followed by the

Ptolemaic and classical Chinese and Indian systems, then the Newtonian system with Einstein's system being the furthest away.

It should be possible to create a "map" that shows where every fact of the universe lies in relation to human beings and in relation to every other fact. Such maps would show the various lines of development through which human knowledge of the universe grew and had to grow. They would show the order in which human knowledge of the universe developed which has a great effect on the type of society available to human beings

Effect of scientific and technological change on society

The development of science and technology obviously has a substantial effect on human society. However it does not affect all elements of human society equally. Leslie White in *The Science of Culture*[17] proposes a three way sub-division of culture into the technological, the sociological and the ideological. The technological consists of the material, mechanical, physical and chemical instruments and the techniques for their use by which human beings live in their environment. It includes the tools of production, the means of subsistence, the materials of shelter and the instruments of hunting and war. The sociological system consists of the interpersonal relationships expressed in individual and collective patterns of behavior. This includes the social, kinship, economic, ethical, political, military, religious, occupational and recreational systems of a culture. The ideological system consists of the ideas, beliefs and knowledge of a culture. This includes the mythologies, theologies, literature, philosophy, science and common sense knowledge of a culture.

These three aspects make up the culture of a society. They are interrelated, each affects the others and is affected by the others. However the effect they have on each other is not equal. The technological plays a primary role, as human beings must first obtain food and protection from the elements and enemies. The technological represents the lower needs of Maslow's hierarchy of needs. These are the most basic of human needs, the ones that must be satisfied before all other needs.

The sociological system is secondary and subsidiary to the technological system. It is a function of the technological system, the technology is the independent variable, the sociological is the dependent variable. The sociological is determined by the technological system. If the technology changes so will the sociological system.

The ideological system is also powerfully conditioned by the technological system. There is a type of ideological system appropriate to each type of technological system. However it is not just the technological system that affects the ideological system, it is also affected by the sociological system.

White sums up his system as follows:

“We may view a cultural system as a series of three horizontal strata: the technological layer on the bottom, the philosophical on the top, the sociological stratum in between. These positions reflect their respective roles in the cultural process. The technological system is basic and primary. Social systems are functions of technologies; and philosophies express technological forces and reflect social systems. The technological factor is therefore the determinant of the cultural system as a whole. It determines the form of social systems, and technology and society together determine the content and orientation of philosophy. This is not to say, of course, that social systems do not condition the operation of technologies, or that social systems and technological systems are not affected by philosophies. They do and are. But to condition is one thing; to determine, quite another.”[18].

White's system is hardly new and has certain obvious similarities to Marx's ideas concerning the infrastructure and superstructure of societies. It is also very similar to what Marvin Harris calls a universal pattern within cultures consisting of an infrastructure (White's technological system), a structure (White's sociological system), and a superstructure (White's ideological system).[19] It is possible to quibble about the exact extent to which the various elements in White's system affect each other, but it seems quite clear that technological systems have a major determining effect on sociological and ideological systems.

Given that the technology available to a society will determine its sociological and ideological states, then societies with similar technologies will tend to have similar sociological and ideological states. This situation is referred to by J H Plumb in *Encounter* for June 1971 when he said:

“... the present world ... is witnessing the close of an epoch that began roughly ten thousand years ago: the end indeed of societies based primarily on agriculture and craftsmanship, in which towns were rarely more than centers for the organizing and servicing of these activities, or of religion or government. ... Of course within these millennia societies differed greatly in complexity, in extent of power, in achievement and in sophistication. And yet there are basic similarities, whether we compare Tang China with the France of Louis XIV or the Peru of the Incas. In general the same social institutions are common to them all - family, organized religion, warrior castes, an elite of bureaucrats, more often than not a semi-divine authority. And between the literate societies there are many resemblances in ideology; a symbolization of natural forces that clearly links the rituals of the agrarian year or reflects the family structure; often this is combined with a religion of personal salvation and hope. Often the presence is sanctified by an interpretation of the past. Men of authority possess the genealogies which confirm their power; and the history of their societies is theirs.”

The situation that Plumb describes, that a given technology will tend to produce a particular social and ideological system, applies for all technological systems, so that it is possible to produce a table such as that below which shows various technological systems and the social and ideological systems that tend to accompany the particular technological system. The four technological systems described are hunter-gathering, pastoralism, agriculture and industrial technology.

	Hunter-gathering	Pastoralism	Agriculture	Industrial technology
Settlement pattern	nomadic	nomadic	sedentary-mainly rural	sedentary-mainly urban
Number of people	dozens	thousands to hundreds of thousands	millions	millions
Basis of relationship	kin	kin based clans	class and residence	class and residence
Ethnicity and language	one	one	one or more	one or more
Government	egalitarian	chiefs	centralized and hereditary	centralized and democratic
Bureaucracy	none	none	many levels	many levels
Monopoly of force	no	yes	yes	yes
Conflict resolution	informal	centralized	law/judges	law/judges
Religion	yes	yes	yes	secular & scientific
Food production	no	Yes-principle area of economy	Yes-principle area of economy	Yes-minor area of economy
Food surplus	no	yes-but occasional shortages	Yes-but occasional shortages	yes-generally no shortages

Division of Labor	limited to gender and age	wider than for hunter gatherers but still limited	yes-extensive with large farming population	highest level of division of labor
Control of land	band-but limited due to nomadism	horde-but limited due to nomadism	ruler & aristocracy	spread amongst population, but unequally
Society stratified	no	to some extent by kin	to a great extent by kin	yes by class
Slavery	no	no	yes	no
Luxury goods for the elite	no	only to a minor extent	yes	yes
Public Architecture	no	no	yes	yes
Literacy	no	no	yes-but usually not widespread	widespread

The above table shows the social consequences of the technological states referred to in the top line. It is designed to give a general overview of the sorts of social and ideological situations that will exist for any given technological state. There are the occasional exceptions to what is mentioned in the table, for example the agrarian Incas did not have writing, although they did have other forms of record keeping. It is also true that there are many different types of hunter-gathering, pastoralist, agrarian and industrial societies and not all can be included in a relatively simple table. The societies practicing shifting agriculture are obviously not included in the agriculture column of the table. The agricultural column is concerned with the civilizations that existed in China, South East Asia, India, the Middle East, North Africa, Europe, Meso-America and South America. There are also a variety of industrial societies, for example not all industrial societies are democratic. It is however considered that democracy will be the typical form of government of industrial societies, much like absolute monarchy was the typical form of government for agrarian societies. However with agrarian societies we have many thousands of years of experience with these societies, so we know a lot about them. We have only a brief experience of industrial societies so the type of social and ideological states that may exist in such societies may be arguable.

There is only a limited range of cultural forms available, so it is hardly unlikely that the same types will evolve independently. Yet of the forms available it is quite clear that certain institutions are much more clearly associated with certain technological states than others. There is no reason in principle why hunter-gatherer bands should not be ruled by hereditary monarchs, but they never are. Equally there is no reason why agrarian states, especially if they are not excessively large, could not be democratic, but they hardly ever are. Rather, hunter-gatherers usually have informal egalitarian leadership and agrarian states are usually ruled by a hereditary monarch. There is a disproportionate co-relation between certain technological states and the type of social institutions and ideological beliefs that accompany those technological states.

The important point is that changes in human knowledge cause changes in technology and through the effect that technology has on the sociological and ideological systems of a society, the change in human knowledge will affect all elements in that society. Changes in human knowledge may also directly affect the sociological and ideological elements in human society. Ideas such as biological evolution and cultural relativity have affected human society, without producing any technological innovations. Human history in all its elements will be affected by the increase in knowledge that gradually accumulates in human culture.

It is necessary to describe not only why societies and cultures have certain similarities, but also why they differ one from the other. Many of the differences are a direct result of differences in the physical environments occupied by the various societies or cultures. Societies or cultures located in the arctic or in temperate zones or in the tropics will all be different from each other, as will inland groups and coastal groups and groups in areas of good rainfall will differ from those in deserts. Such differences will be much greater among agrarian and hunter-gatherer societies located in different environments, than between industrial societies located in different environments. This is because hunter-gatherer and agrarian societies are much more dependent on the immediate physical environment than industrial societies.

The range and type of crops and animals capable of domestication will also cause variations among agrarian societies. The lack of any large domesticable animal in the New World ensured that it never developed the plough agriculture and wheeled transport that existed through large parts of the Old World.

The level of knowledge and technology available to different cultures will also be a reason for the variations that exist between cultures. This is very much the theme of this book. Such variations tend to be used (for example in this book) to categorize various cultures rather than to explain the differences between the same types of cultures. Nevertheless within any category of culture (such as hunter-gatherer, pastoralist, agrarian, or industrial) there will always be variations in the knowledge and technology available to different societies and this will explain some of the variations that exist within a particular category of culture.

There are also certain aspects of a culture where the knowledge and technology does not influence or determine that aspect of the culture or does not particularly favor one outcome over another. This is where chance, the activities of great men and women and differences in traditions may determine aspects of a culture. An aspect of culture such as religion in agrarian societies can vary from the tolerant pluralism of pre-Christian Rome and the Chinese Empire to the severe attitude to dissent displayed in ancient Israel, Christian Europe and some Islamic societies. The power of priestly classes could vary from very strong in medieval Europe and pre-British India to relatively weak in pre-Christian Rome and in China. Whether a society drives on the left or right hand side of the road is a matter which is not likely to be influenced by the rest of the society's culture. The choice is purely arbitrary.

The important point is that the variations can all be explained in ways that do not contradict the model of social, cultural and historical change being driven by increases in human knowledge and that this model can apply to all human cultures and societies. Environmental differences, prior cultural traditions and chance will explain cultural variations, and similar knowledge of the human environment will explain the similarities between cultures.

A Law of Social, Cultural and Historical Change

The course of human social and cultural history is determined by our increasing knowledge of our environment. As our knowledge grows we are able to make better and better use of our environment to meet our needs. Our increasing knowledge results in technology that produces changes in the overall state of society. Our increasing knowledge of the environment may also produce changes in the state of society without involving technological change. The discovery of the idea of biological evolution, resulted in changes in society without involving technological change.

The growth in our knowledge of our environment takes place in a particular order. Some things in our environment will inevitably be discovered before other things. This is because the prior discovery is necessary before the later discovery can be made. It also may be because the prior discovery is a simpler matter involving a limited number of previous discoveries, while a later discovery may require a much larger number of more difficult discoveries to precede it. Our knowledge of our environment grows from the simple to the more complex or from that which is closer to us to that which is further from us. The phenomena of multiples shows both that discoveries are inevitable, so long as social conditions allow them, and that they have to take place in a particular order.

Given that the human discovery of the environment is inevitable and takes place in a particular order, and that the discoveries produce new technologies to better meet human needs and that new technologies produce new social and ideological systems, then it can be said the course of human social and cultural history was inevitable. That history and social change and cultural evolution could take only one course is because humans can only discover that which is there. We cannot discover things that do not exist; all knowledge must be about that which exists. This means that the nature of reality, or what we can find out about, is the determining factor in human history. The consequence is that human history could only take one broad course. That course is from hunter-gathering to agriculture or pastoralism to industrial society. Some societies, of course, did not take this course due to local environmental conditions, such as arctic conditions, deserts or lack of domesticable plants and animals. But all societies that could, did undertake the progression, and no society was able to miss out a step, except where there were cases of diffusion of knowledge and technology.

This means that it is possible to produce a law of historical, social and cultural development that will apply to all human societies and cultures. Such a law might be as follows:

If you have a being that is:

(a) intelligent
The requirement that the being be intelligent refers to its ability to learn and also its ability to remember what it has learnt. This is the capacity to retain knowledge, so that knowledge will accumulate in the culture of that being.
(b) rational
(c) exists in a situation where new ideas may be freely communicated to others
(d) desires to meet its needs
(e) that meets its needs by using the resources in its environment
(f) all aspects of its environment are not immediately obvious to that being
(g) there exists in the environment the means by which the being is better able to meet its needs

then the greater its knowledge of the environment the better it is able to meet its needs. Such a being will attempt to learn about its environment in order to meet its needs in an improved manner. As the being learns more about its environment, new means of meeting its needs become available to it and the adoption of improved means of meeting its needs, results in changes throughout the beings society and is the ultimate cause of social and cultural change.

Such a law would apply to all intelligent, rational beings, meeting their needs from their environment, but lacking a complete knowledge of their environment. The reason cultures have evolved in similar fashion is because their environments are similar; to the extent they have evolved differently, it is because their environments are different.

Human beings cannot know now, what they will only discover in the future. This means to some extent the above law cannot be used to make predictions about the future of human society. It should really be seen as an explanatory law, explaining why and how human societies have developed throughout history.

There is, however, an exception to this. While future scientific discoveries cannot be known, human needs are known. Of course, just because needs exist, does not guarantee they will be met. But it is predictable humans will attempt to meet their needs and if they are met it is possible to predict in theory at least, what their effects on society or culture may be. This is the basis of the more sound attempts to predict the future that have been made. Predictions of faster computers with larger memories and hard disks could safely be made in the 1990's. Developments in nanotechnology would also fit into this category.

A further exception to the rule that the above law cannot be used to predict the future is where you have a God like view of a being and its environment. Such a view will provide full information on a being, and its environment, including that part of its environment that is unknown to the being. We obviously do not have such a God like view of our own selves and our own environment. However there is no reason in principle why human beings could not have such a God like view of other societies, or other beings could not have a God like view of human society. It is theoretically possible that human astronauts visiting a distant planet could find a society whose future we could predict if we knew enough about the beings in the society and the environment they live in. Equally, sufficiently knowledgeable space travelers, visiting earth, could predict our future.

There would be nothing to stop human astronauts or non-human space travelers from revealing their knowledge to our intelligent, rational, but not all knowing beings. This would simply be a case of the diffusion of knowledge and is similar to the diffusion of knowledge that has occurred on earth, when western science and technology has been spread around the world. This has had the predicted result of the development of urbanized, industrial and democratic societies, most clearly shown in some East Asian countries. It can be predicted that in time other non-western societies will eventually become democratic, industrial and urbanized.

A final point is that if our intelligent, rational being was to know everything that can be known about its environment, or everything about its environment that enables it to meet its needs, so that its needs are met in the best way that they can ever be met, then in some sense at least that will amount to the end of history. Cultures will cease to evolve, societies will no longer change in any fundamental way. Rulers may come and go, laws can continue to be changed and rechanged, wars may possibly be fought, but the fundamental nature and structure of society will cease to change.

Popper’s criticisms of historicism

One critic of theories of history and social and cultural change, such as is offered in this book, is Karl Popper. He calls the making of such philosophies "historicism" and in *The Poverty of Historicism* he claims that it is not possible to produce "laws of historical development". This is because the evolution of human society is a unique historical process. Laws make assertions about all processes of a particular kind. If I heat a particular metal it will always melt at the same temperature, and this can be predicted by a law. Popper says we are hardly likely to be able to formulate such laws with just one instance available of what the law deals with. Nor is it possible to test the law with future examples and we cannot foresee the future of human society so as to be sure the future will not proceed in a manner inconsistent with the law.

Popper suggests there are two arguments that can be made against his argument. It could be argued that the process of social, cultural and historical change is not unique, but that such change is cyclical in that civilizations, cultures and states rise and fall in a repetitive process. The second argument is that even if human social, cultural and historical development is unique we can discern a trend and formulate a hypothesis that states the trend and can be tested against future experience.[20] Popper rejects the argument that the historical process is cyclical. He considered the claimed instances of repetition to be quite dissimilar and are based on a selective use of facts.[21]

Popper deals with the second argument by claiming that while trends exist, trends do not amount to laws. Trends may exist at a certain place or time, but they are singular statements not laws from which predictions may be made. We cannot base predictions on trends, as trends (Popper gives the example of population growth) which might have existed for thousands of years may rapidly change. Popper considers the confusion of trends for laws, together with the intuitive observation of trends, such as technical progress, is the basis of historicism. Trends, Popper notes, are dependent on initial conditions and this point is usually overlooked by historicists. If the initial conditions change then the trends may alter or disappear. If the historicists were to find and explicitly state the conditions the trend is dependent on there would be no problem, but they seldom, if ever, do this. Such conditions are easily overlooked and may be quite numerous and difficult to state completely.[22]

A statement of the form "whenever there are conditions of the kind c, there will be a trend of the kind t" would be acceptable to Popper. It would however be necessary to test such a law by trying to produce conditions under which it does not hold. We may for example try to show conditions of the kind c are insufficient and that even when they exist, the trend t will not always occur.[23]

Popper provides an example of an historicist theory to illustrate his concerns. It is a theory of scientific and industrial process advocated by Comte and Mill in which they claim such progress is reducible to laws of human nature. Comte believed there was a tendency in human beings that impels them to perfect their nature to an ever-increasing degree. Mill considered the impelling force to be the desire for ever-increasing comforts. This enables us to deduce the phases of history without observation or data. In principle, though probably not in practice, the whole course of history would be deductible.

Popper claims that even if Comte and Mills premises and deductions are correct, it will not mean that progress will necessarily result. Difficult natural environments or certain elements of human nature such as forgetfulness or indolence could destroy the prospect of progress. Popper also claims that progress depends on conditions such as freedom of thought and expression. An epidemic of mysticism might also hinder or eliminate the progress Comte and Mill would expect. Popper considers that it is the human element in historicist theories that ensures their unreliability. Humans cannot be relied on to act rationally and any attempt to reduce historical theory to theories of human nature or psychology may have quite unpredictable results.

In the preface to the second edition of *The Poverty of Historicism* Popper suggests that while he shows historicism is unreliable he does not actually refute it. In the preface, however he does produce a refutation of historicism, which he considers shows for strictly logical reasons it is impossible to predict the future course of history. The refutation is as follows:

"(1) The course of human history is strongly influenced by the growth of human knowledge.
(2) We cannot predict, by rational or scientific methods, the future growth of our scientific knowledge.
(3) We cannot, therefore, predict the future course of human history.
(4) This means we must reject the possibility of a theoretical history; that is to say, of an historical social science that would correspond to theoretical physics. There can be no scientific theory of historical development serving as a basis for historical prediction.
(5) The fundamental aim of historicist methods is therefore misconceived; and historicism collapses."

Popper states the above argument only refutes the possibility of making predictions of historical developments to the extent to which they are influenced by the growth of our knowledge. Other predictions such as those claiming certain things will happen under certain conditions are still possible. The most important part of the argument is (2) which Popper justifies by stating, "if there is such a thing as growing human knowledge, then we cannot anticipate today what we shall know only tomorrow." He claims "no scientific predictor... can possibly predict by scientific methods its own future results." No society can predict its own future knowledge.

The first of Popper's criticisms of historicism is that human history, social change and cultural evolution is a unique process and we have just a single instance of this process before us. This is not necessarily the case if we can see the same situations occurring time after time in history in a variety of societies. If we could see many examples of scientific and technological progress occurring throughout history such as the development of agriculture, of the steam engine, Newtonian physics, automobiles, aircraft, the theory of evolution, the structure of DNA and genetics we would have many examples of scientific and technological progress. If all these and other examples reveal the same conditions under which the progress was achieved, then we may be able to formulate a law of scientific and technological progress. Human history, social change and cultural evolution does not have to be seen as a single event; it can be divided up into parts that can be studied separately.

Popper is quite right to distinguish between trends and laws. In order to constitute a law the conditions under which it operates must be stated, otherwise we will only be dealing with a trend. The law stated in the previous part of this book does state the conditions listed (a) to (g) for the operation of the law.

Popper, when considering Comte and Mills' theory of scientific and technological progress, gives a number of examples where their theories may fail. Comte and Mills theories on the philosophy of history are in many ways similar to those suggested in this book, so Popper's objections to their theories could apply to the theory suggested in this book.

Popper cites difficult natural environments as possibly destroying the prospects of progress. However people live in a variety of environments, some difficult, some easier and if progress fails to occur in difficult environments such as deserts or arctic conditions, it may well happen in other environments and by a process of diffusion reach the more difficult environments. Some might also suggest that difficult environments may stimulate progress. More particularly, condition (g) requires that there exists in the environment the means by which the being may meet its needs in a more efficient manner. Popper's objection seems to suggest that the environment may not allow new knowledge to be obtained, but this is already dealt with in the proposed law.

Popper also claims forgetfulness, indolence, mysticism, human irrationality and a lack of freedom of expression and thought may destroy the possibility of progress. The law stated earlier in this paper requires that the beings to which it applies be rational, intelligent, must desire to meet their needs and be able to freely communicate to others. These are the conditions for the law to operate and if they exist then Popper's objections cannot apply to the theory. The objections he states are already ruled out by the law.

However while the law may be true, given its conditions of intelligent, rational, free beings, desiring to meet their needs, the question arises, do such beings exist? In particular do human beings meet the conditions stated in the law, so the law could be used to explain human scientific and technological progress? The answer is that humans come in a great variety, some are intelligent and rational, and others are less intelligent or rational. But in all societies there will be some people who are intelligent so progress will be possible. There have of course been some societies where freedom of thought and expression has not existed. However there have been other societies where it does exist and in those societies scientific and technological progress will often take place. In short the conditions stated in the law will sometimes exist and when they do there is a possibility of scientific and industrial progress taking place. In particular where progress has occurred it has been under the conditions stated in the law. The law also states the situation where progress will not take place, most particularly the requirement for ideas being able to be freely communicable to others. If this requirement is not there, progress will be impossible. The law not only states under what conditions progress can be achieved, but also under what conditions it will not be achieved.

Popper's final criticism of historicism is the refutation contained in the preface to the second edition of *The Poverty of Historicism*. This refutation is certainly correct as regards predictions of future events with one significant exception. While we cannot know scientific knowledge that will only be discovered in the future, we do know what the ultimate human needs are. These needs transcend cultures and are the same for all times even though quite different methods may be used to meet the needs. This provides an element of predictability as to the future in that we can predict that humans will use the best means available, as is culturally acceptable to them, to meet those ultimate needs.

A further point concerning Popper's refutation of historicism is that it is limited to denying the viability of scientific theories that predict the future. Prediction is not the only role of scientific theories. Scientific theories may also serve the purpose of explaining events and facts. A theory such as evolution by natural selection does not predict the future development of species. This is because species evolve by adapting to changes in the natural environment and the future changes in the natural environment are unknown and unpredictable. This means the theory of evolution is limited to explaining, rather than predicting, the evolution of species.

The same situation applies to the theory expressed in this book. The theory is not able to predict the future, as the knowledge that we will obtain in the future is unknown to us in the present. This situation is similar to that existing in the theory of evolution where future changes in the natural environment are unknowable. However, like evolutionary theory is able to explain how organisms evolve, the theory proposed is able to explain how historical, social and cultural change takes place, without being able to make predictions about the future. However, just as evolutionary theory is unable to make predictions as to the future, but is certainly a scientific theory, the social, cultural and historical theory proposed would also be a scientific theory even though it is unable to make predictions as to the future.

The Challenge

An interesting factor in the proposed theory is that it enables human historical development to be studied scientifically and objectively. The natural sciences are the same for all societies on earth, and they have an element of certainty or truth about them, which is not available in the social sciences. By tying human historical development to our discovery of the facts of the natural sciences, we are tying human historical development to the most well established facts we have available. This will provide a solid objective basis to any theory of human historical, social and cultural development based on the order of the discovery of the facts of the human environment.

Given the scientific and objective nature of the physical sciences it should be possible to construct a theoretical map of the facts of the human environment, such as they are known to us. This map should show which facts are closer to us and which facts are further away from us. Such a map may require some means of measuring how far particular items of knowledge are from us. This would involve an analysis of the complexity of the facts and in particular what prior facts would need to be known before the particular fact could become known to us. If this were done, you would have a theoretical order of discovery of the facts of the universe. This theoretical order could then be compared to the actual order of discovery of the facts of the universe in order to test the theory expressed in this paper.

In order to create a theoretical order of discovery and to compare it with the actual order of discovery it will be necessary to write a new type of history. Histories of science and technology are usually written in narrative form, copying the form of political and diplomatic history. If they were written so as to involve an analysis of the ingredients that went into the making of a particular scientific or technological discovery, then it would be possible to analyze the order of discovery, that lead to the discovery of any particular invention or scientific discovery. This would provide a new insight into scientific and technological discoveries; an insight that is not apparent from a simple narrative describing how one invention or discovery followed another invention or discovery. This would enable a complete description of the order of discovery made by humankind and would show all the intellectual and technological states humankind passed through in order to reach any particular intellectual or technological state. Such empirical studies could then be compared with a theoretical map of the facts of the universe to see if the order of discovery, suggested by such a map, has in fact been followed in the actual course of human history. Local environmental factors, chance and the activities of great men and women, insofar as they may be applicable, would need to be taken into account, but once this was done, it should be possible to compare a theoretical order of discovery, with the actual historical order of discovery. If they match up, or discrepancies may be explained by local conditions, chance or the activities of great individuals, then the theory proposed may be correct. Such a procedure would amount to a test of the proposed theory, making it potentially falsifiable.

Local conditions, chance and the activities of great individuals may be seen as a bit of a cop out, in that they can be used to excuse any failure of the actual order of discovery, to fit in with the theoretical order of discovery. Yet such local environmental conditions, chance and great individual achievements plainly do exist. The lack of large domesticable animals in Meso-America is the probable explanation for the lack of plough agriculture and wheeled transport in Meso-America. It may also explain the fragility of Meso-American civilization, the Mayan civilization disappearing with the people abandoning their cities, science and mathematics and returning to shifting agriculture. Many discoveries, for example, penicillin, were discovered by chance while the activities of a Newton or Einstein will certainly have advanced their discoveries. If they had not existed their discoveries would have to have been made by others at a later time.

Some puzzles are less easily explained. The Mayan discovery of the zero in mathematics in a civilization lacking the plough agriculture and sophisticated metallurgy of the old world and which was so fragile it was soon to disappear seems odd. The zero was first clearly used in the old world in India around 600 CE with some evidence of earlier use in India and South East Asia. The zero only reached Europe through diffusion through the Islamic world. Why did the, in many ways less advanced, Mayan culture develop the zero, while the old world cultures developed the zero at a much later stage of their cultural development? The probable answer to this is individual brilliance by one or more Mayan priests.

It is of course necessary to confirm that part of the theory that holds that increasing human knowledge will affect human technology, social institutions and beliefs. The very fact that certain levels of human knowledge, technology, social institutions and beliefs tend to co-exist is some evidence that human knowledge affects technology and both of these will affect social institutions and beliefs. However it is possible to do detailed analysis that will show that certain technology, social institutions and beliefs could not, or were unlikely to, exist without certain levels of human knowledge and that certain social institutions and beliefs could only exist or were likely to exist with certain prior technologies.

Many examples of knowledge affecting technology and of one technology affecting other technologies can be offered, but probably the most obvious are those technologies that arose with the knowledge and practice of agriculture. Agriculture allowed sedentism and sedentism allowed the development of permanent buildings, pottery, writing and metallurgy. Permanent buildings meet an obvious human need for people to live in and for goods to be stored in. They were not, however, practical for nomadic hunter-gatherers who had to move away when the food supplies in the area were exhausted. So permanent buildings were practical only for sedentary societies and were never made by hunter-gatherers. Pottery was only developed by sedentary peoples, as it was not practical for hunter-gatherers to carry pottery around with them. Pottery is easily broken and its weight would discourage hunter-gatherers from carrying it around with them. More particularly pottery is often used to store food, a practice agrarian people engage in, but hunter gatherers do not. Writing, or a substitute like the Incas quipu, was a virtual necessity for agrarian populations of a particular size, to enable record keeping of the storage of produce and the payment of taxes and dues. Hunter-gatherers did not have this need so they did not develop writing. Further, the carrying around of writing implements and written records would have been a significant burden for hunter-gatherers. Metallurgy was never developed by hunter-gatherers as the furnaces and bellows required could not have been carried around by hunter-gatherers.

There are a number of social institutions that may be shown to be possible for agrarian societies, but could not exist in hunter-gatherer societies. Slavery was not a good option for hunter-gatherers as slaves could easily run off and would have the same hunting and gathering skills required for survival as the rest of a hunting and gathering community. In agrarian societies running away is not so easy as wherever one runs, unless one is near a border area, there is continuous territory under the control of the slave owning class. Runaway slaves may also lack the skills to survive while on the run in an agrarian society. Slavery was a more practical solution for agrarian societies and it also provided a way to control labor in societies that lacked or had only a limited monetary system. Industrial societies almost totally lack slavery, even though slavery would be quite practical in industrial societies. The reason for the absence of slavery in industrial societies would possibly be due to beliefs about the equality and dignity of humankind and a monetary system that allows labor to be efficiently controlled and directed.

Feminism arose in the twentieth century, in industrial societies, rather than during 5,000 years of agrarian civilization when women were in a substantially unequal position in relation to men. The reason feminism arose in industrial societies may be because of the invention of effective birth control and because of the much increased life expectancy in industrial societies. In agrarian societies a woman with little control over her fertility would be restricted to child bearing and rearing and other domestic activities and by the time menopause arrived, her life expectancy would be virtually at its end, leaving little opportunity for work outside the home. Since the development of modern birth control women have been able to control their fertility allowing work outside the home and the greater life expectancy of people in industrial societies has meant that women have a good twenty or thirty years of post-menopausal working life outside the home available to them. Modern educational systems requiring children's attendance at school and that heavy physical work was an important part of work in agrarian societies, but is only a minor part of work performed in industrial societies would also have enabled greater female participation in work outside the home.

Outlined above are various ways in which human knowledge and technology have determined or influenced human social and cultural history. It is possible to trace the development of social movements such as feminism through inventions such as birth control to scientific developments in biology and chemistry that allowed the invention of modern birth control. These scientific developments were dependent on earlier scientific discoveries, but ultimately the whole process is determined by the structure of the universe and its relationship to human beings. The increase in life expectancy since the industrial revolution has many causes such as increased and better food supplies, medical knowledge and sanitary practices in modern cities. All these ultimately were dependent upon increasing human knowledge, both scientific and non-scientific, of the human environment. The same can be argued for the reduction of hard physical work since the industrial revolution and the development of modern educational systems. The result is that a social movement such as feminism is ultimately dependent upon increases in human knowledge and the increases in knowledge take place in a particular order as is determined by the structure of the universe and its relationship to human beings. The origins of feminism lie in the biological and chemical facts of the universe that enabled the production of modern birth control and in the facts of nature that enabled the creation of the technologies that constitute the industrial revolution.

The same could be said about slavery which only became widespread with the development of agriculture. Slavery worked best after the discovery of agriculture which required certain prior discoveries, such that plants grow from seeds and that plants if tendered properly will produce a normally reliable food supply. These discoveries are not obvious to ordinary observation and would have required considerable observation, thought and experimentation. They represent a great leap in human knowledge of how the environment can be utilized to meet human needs.

It should be possible to trace the development of all elements of human technological, sociological and ideological systems through changes in human knowledge of the environment. At any given stage the state of a society's technological, sociological and ideological systems will depend upon that society's level of knowledge of the human environment. The state of human knowledge of the environment will depend upon the ease with which humans can discover the facts of the human environment. This is dependent upon the structure of the environment and its relationship to humankind. This means in principle all human technological, sociological and ideological systems are ultimately based on the conditions of the human environment and on our knowledge of the human environment. We should be able to trace a causal chain from the structure of our environment to our knowledge of the environment to human technological, sociological and ideological systems. The effects should be able to be traced backwards from the human technological, sociological and ideological systems through to the state of human knowledge of the environment to the structure of the human environment and our relationship to it.

Part II

CASE STUDIES

Introduction

These case studies are written to show in some detail the ideas outlined in Part 1 of this book. It is intended to show there is a pattern in at least a part of human history that can be rationally understood and which shows that to some extent human history followed an inevitable and necessary path. It aims to show this by investigating what I call human social and cultural history that is the history of human knowledge, beliefs, technology and social systems. The investigation involves looking at the course human social and cultural history actually took and then showing how it had to take that course. It will also show how the course of human social and cultural history proceeded in accordance with the theory outlined in Part 1 of this book. It will particularly look at the way in which there is a fixed or likely order of discovery for the various areas of human technology.

There are three ways in which the order of discovery may be fixed or likely. The first is where the order of discovery is absolutely necessary. It was not possible to develop bronze tools without first learning how to melt copper as melted copper is a necessary ingredient in making bronze. In this situation the step of producing bronze is completely dependent on the prior step of being able to melt copper. The later step could not take place without the earlier step. This rule is dependent upon the natural sciences for example the laws of physics, chemistry and biology and cannot be broken.

The second way in which the order of discovery is affected is where the second step would not normally take place without the earlier step as it would be irrational or uneconomic to take the second step without the prior first step having taken place. Nomadic hunter-gatherers could for example have built large stone buildings but it would be irrational for them to do so, if they are continually moving around. The same would also apply to the use of pottery which is heavy and easily broken, so it would not be rational for nomadic hunter-gatherers to use pottery. This is like a law of economics, it can be broken, but one would not normally expect it to be broken as breaking the law would be irrational.

The third rule for the order of discovery is that improvements in technology usually move from the simple to the complex. This is like a law of psychology, it is the manner in which humans normally learn things. This move from the simple to the complex is very likely to happen but occasionally one can get odd results if for example a step is missed out when something simple is missed and something more complicated is discovered.

A further method by which human social and cultural history can be understood is by an examination of the physical and chemical structure of those items that have a significant role in human social and cultural history. The effect of the structure of the items on the course of human social and cultural history can be assessed by looking at counterfactuals which would show how human social and cultural history would be different if the physical or chemical structure of the items that affect human social and cultural history were different.

The case studies contained in this part of the book generally take a particular form. They first describe how a particular discovery took place and what the social and cultural consequences of the discovery were. Then there is an analysis of why the discovery took place at all and then an analysis of why it took place when it did. This is a study of why the discovery was made at a particular point in human history and not at another time in human history.

Natural environment

The genus hominids have been on this planet for approximately 5 million years. Anatomically modern humans, *homo sapien-sapiens* have been on the planet for around 200,000 years. The environment our hominid and *homo sapien-sapien* ancestors lived in was as provided for them by their senses. Sight is by far the most important of the senses for human beings. Human sight however has thresholds which limit what can be seen. Very small objects such as micro-organisms and far away objects such as some planets and many stars are not observable by human beings. Equally fast moving objects such as bullets cannot be observed by human beings. Our ancestors had no knowledge of cells, molecules, atoms, electrons, protons and quarks, as they had no way of observing or detecting these objects. The other human senses are also subject to similar limitations and thresholds. The human perception of the environment determines what knowledge of the environment humans possess. A different being with a different sensory apparatus will have a different perception of the environment and a different knowledge of the environment. The human ability to understand the environment would also have been limited by human intellectual power. This limitation would have been greater for our hominid ancestors but also applies to modern *homo sapien-sapiens*.

Human attention when observing the environment would inevitably have been directed at those objects that are vital for human survival such as plants and animals. Plants can be easily studied to determine whether they are suitable for food or for other purposes such as the curing of sickness. As a result early hominids would have some knowledge of what we call botany. They would also have some knowledge of the animal world, although that knowledge would not be as extensive as that of the plant world. This is because animals are mobile and often avoid humans making their study difficult and often leading to erroneous beliefs about animal behaviour. Our hominid ancestors would have also had the opportunity to make some observations of the geological world leading to the making of stone tools. Observations of the heavens lead to the beginnings of astronomy. The beginnings of anatomy and physiology were made from observations of the human body. An understanding of the seasons and the weather would have also been available to our hominid ancestors.

The overall quantity of knowledge within a modern industrial society is vastly greater than the quantity of knowledge held by our hunter-gatherer ancestors. In a hunter-gatherer band most of the adult members of the band would know or carry most of the knowledge of the band. The entire knowledge of the band would be held by a band which might consist of 30 or 40 individuals. With no way of recording knowledge, knowledge would be restricted to that which could be held in human memories. In agrarian societies no individual would hold anywhere near all the knowledge of the society. The peasant farmer would know about his crops and when to plant them and how to tend them but he would have no or only a very limited knowledge of metallurgy, law, theology, war, leather working or dozens of other specialist activities. Any individual would know only a small fraction of the total knowledge held by the society. In an industrial society a single individual will know only a tiny fraction of the knowledge held by the society. If she is an accountant she may know how to create a set of accounts and how to work out how much tax a person owes. She will probably know nothing or next to nothing about how her computer, television, telephone, car and microwave oven works. She will know nothing, or next to nothing about how galaxies form, the behaviour of amoeba when exposed to light, how to make nuclear explosions and how to build a laser. If her television or car breaks down she will get an expert to fix it. The knowledge of all these things is contained within industrial societies yet an individual within the society will only possess a tiny fraction of the knowledge of the society. Clearly there has been a vast increase in knowledge throughout human history. This increase is the cause of the increasing specialization and division of labour that has occurred throughout human history.

Studies of modern hunter-gatherers reveal a substantial set of beliefs involving creation theories and a supernatural world. It is not known whether our hominid ancestors such as *Australopithecus* and *homo habilis* had such beliefs. Practices such as burial of the dead, especially with objects buried with the deceased, usually considered as evidence of belief in an afterlife, seem to have been practiced only by modern humans and *homo-sapien-neanderthals*. Given such comprehensive ideas require sophisticated thought and such thought would require language which only anatomically modern humans and possibly to some extent *homo-sapien-neanderthals* possess, our early hominid ancestors would be unlikely to have such sophisticated beliefs.

Stone Tools

One thing our hominid ancestors certainly had was the ability to make tools. Many animals such as birds and chimpanzees engage in tool-making so it is hardly surprising that our hominid ancestors made and used tools. Many of the earliest tools would have been made of wood and bone and other materials which decay and leave no trace for archaeological inspection.

The earliest tools which we have available for archaeological study are stone tools from about 2.6 million years ago made by *homo habilis*. These tools are known as the Oldowan stone tool industry. The Oldowan tools were made by chipping flakes off an unmodified core with another stone that acted as a hammer. Both the flakes and the core provided useful tools, the flakes being used mainly as cutters for cutting up or scraping dead animal carcasses or for stripping plants. The cores may have been used for food processing that involved bashing or pounding. The tools were so simple that it was often difficult to distinguish them from naturally created objects. The tools were usually made from quartz, quartzite, flint or chert, all of which are crystalline rocks.

The Acheulean stone tool industry began about 1.5 million years ago and is largely associated with *Homo erectus* (except in east Asia) and *Homo ergaster*. The Acheulean tools are more complex than the Oldowan tools in that the core was prepared before flaking took place and tools were produced that had bifacial cutting edges. Bifacial tools are flaked on both sides so that they are sharper than Oldowan tools. A further improvement was the use of bone or wood hammers that provided better control over the flaking process so as to produce sharper cutting edges. Stone hammering was used to give tools an initial shape but finishing work was done with wood or bone hammers.

Acheulean tools included hand axes, cleavers, picks, choppers and flakes. Tools were used for cutting up large animals, or with *Homo erectus' use* of fire for cutting branches of trees to provide fuel for fires. They may also have been used for digging up the edible roots of plants and for woodworking. The tools were mainly made of flint, quartzite, chert and obsidian. Acheulean tools almost certainly included spears and clubs but evidence for this is rare. There is some evidence for wooden spears from Clacton in England and Schoningen in Germany between 600,000 and 300,000 years ago.

One puzzle is that Acheulean tools were not found in East Asia. Among the explanations suggested for that is that the quality of raw materials was not good enough, in that fine grain rocks were rare. A further explanation was that different materials such as bamboo allowed alternative tools to be produced in place of stone tools. Alternatively, hominids of East Asia had different needs from those elsewhere so Acheulean tools were not required.

The effects of the improved tools used by *Homo erectus* would have been to allow some population increase due to the greater ability of *Homo erectus* to hunt and to protect his or herself from wild animals. Meat derived from hunting large animals was a much greater part of the diet of *Homo erectus* than it was for earlier hominids. The improved hunting ability would have come both from the use of better tools and from the use of fire by *Homo erectus*. A further effect of the use of improved tools and the use of fire was that *Homo erectus* was the first hominid to live not just in Africa but also in Europe and Asia.

The Mousterian stone tool industry began about 200,000 years ago and lasted till about 40,000 years ago. It is particularly associated with *Homo sapiens neanderthalensis* but the tools were also used by *Homo sapiens sapiens*. The Mousterian stone tool working techniques involved the careful preparation of a stone core before a flake was struck from the core. This could involve shaping the core into a round surface by trimming the edges of the core and then further trimming to shape the flake that is to be struck off. Only then would the flake be struck off. An alternative system was to shape the core into a prism and then to strike off triangular shaped flakes. Flakes would then be worked with additional trimming to sharpen their edges to produce a better cutting edge. Flakes were produced for many specialized purposes. Hand axes and tools for cutting up meat similar to earlier times were used but were better made and more efficient. New tools such as points for spear heads were made which were attached to a wooden shaft being the first evidence of composite tools being used by hominids.

The Upper Paleolithic tool industry ran from roughly 40,000 years ago to 12,000 years ago. The Upper Paleolithic period comprised a series of tool making periods known as the Aurignacian (40,000 to 28,000 year ago), the Gravettian (28,000 to 22,000 years ago), the Solutrean (22,000 to 19,000 years ago) and the Magdalenian (18,000 to 12,000 years ago). The Aurignacian was associated with both *Homo sapiens neanderthalensis* and *Homo sapiens sapiens* (more particularly Cro-Magnon man). The other three periods were exclusively those of *Homo sapiens sapiens* due to the extinction of *Homo sapiens neanderthalensis*.

The rate of improvement in the quality and variety of tools was much faster in the Upper Paleolithic than in the earlier periods. These improvements included better techniques for the working of raw materials. Before this time technology largely involved the use of only four techniques, those of percussion, whittling, scraping and cutting all of which required only a limited range of hand motions. In the Upper Paleolithic new techniques were added including pressure flaking, drilling, twisting, grinding and others, which involved different motor abilities than those previously used. Secondly, in the earlier period the main raw materials used were stone, wood and skin. Later on bone, ivory and antler and less importantly shell and clay were added to the original materials. Thirdly, the number of components in composite tools expanded considerably in the Upper Paleolithic, increasing the complexity of the tools used. Fourthly, the number of stages involved in manufacturing artifacts significantly increased in the Upper Paleolithic. Before the Upper Paleolithic manufacturing involved only a short series of single stage operations, while later there were often several stages of manufacture to produce the final product. The number of processes and techniques had increased as had the degree of conceptualization required to manufacture the product.[24]

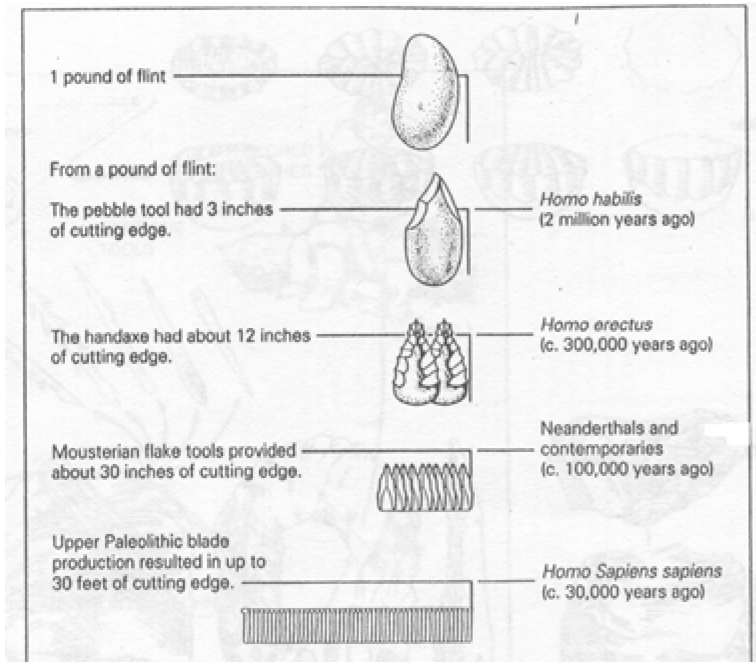
In the Upper Paleolithic there were substantial improvements in the artifacts available to people. Hunting equipment improved by the use of narrow bone or ivory points for spears which had greater penetrating power than earlier flint tipped spears. Spear-throwers and the bow and arrow were also introduced, allowing prey to be killed from a greater distance. Cooking was made more effective through the use of cobble-lined hearths which allowed heat to be retained longer and at a more even temperature. Improvements in clothing seem to have been made between the Middle and Upper Paleolithic, providing humans with much better protection against the elements. Eyed needles seem to have been invented around this time. Housing became more sophisticated in the Upper Paleolithic with many structures being made of mammoth bones suggesting that some sort of sophisticated transport device such as sledges were used to move the bones. Art, which played little role in earlier periods, became much more extensive in the Upper Paleolithic. Cave paintings appeared in Europe, Australia and North and South Africa. Many artifacts such as bone needles, ivory beads, spear throwers and bows had engravings or carvings performed on them. Artistic objects such as Venus figurines were traded over considerable distances suggesting the Upper Paleolithic had much improved trade and communications than the Middle Paleolithic[25]. Technology developed by hunter-gatherers in the Middle East to utilize wild cereals, such as stone sickles and underground storage pits, were useful to early cereal farmers in the Middle East. The substantial improvements in the tools, clothing, art and general culture of humankind between the lower and upper Paleolithic could only have taken place with a gradually increasing knowledge of how to make better and better use of the materials in the environment.

The improvements in stone tools involves a progression from the simple to the complex. Earlier Stone Age technologies were both simpler and less efficient than later technologies. As time went by, or as human mental facilities developed, the technology became more efficient and complex. In *People of the Earth: An Introduction to World Prehistory* Brian Fagan says:

"There is a basic continuum in stone working skills that begins in the Lower Paleolithic and continues through the Middle and the Upper Paleolithic and even later in prehistory. Even the more efficient technological changes associated with the spread of *Homo sapiens sapiens* after 40,000 years ago have a strong basis in much earlier, simpler technologies."

He also said:

"The growing efficiency of stone age technology is shown by the ability of ancient stoneworkers at producing ever larger numbers of cutting edges from a pound of flint or other fine grained rock. The Neanderthals were far more efficient stone artisans than their predecessors. By the same token, *Homo sapiens sapiens* used a blade technology which produced up to 30 feet (9.1 metres) per pound of flint."[26] (See diagram below).



The trend from the simpler less efficient stone tools to more efficient complex tools was inevitable. Our hominid ancestors were always going to find the simplest way to make stone tools before ways to make more complex tools were learnt. This is because it is always easier to learn something simple, than something that is more complicated. The Oldowan tools were so simple they were sometimes difficult to distinguish from naturally created objects and would produce only 3 inches of cutting edges from a pound of flint. The Acheulean tools were often bifacial and could produce 12 inches of cutting edge from a pound of flint. Mousterian tools have a still greater degree of complexity involving considerable preparation of the core before a flake was struck and substantial finishing work being done on the tools. Increased complexity can also be seen in the development of composite tools. The Upper Paleolithic tools reveal even more complexity with new manufacturing techniques and still more composite tools. The order of improvement in Paleolithic stone tools was inevitable as our ancestors were always going to learn stone tool manufacture in the order from the simple to the complex. It is easier to learn how to knock a flake from a stone, than to knock it in particular ways to produce a flake of a particular size and shape. It was also inevitable that people would learn how to knock a flake from a stone, before they could learn that preparatory work on the stone could produce a more desirable flake. One also had to learn to knock a flake from a stone before you could realize that finishing work on the flake could make it a more desirable flake. The order of discovery of how to make better and better stone tools was inevitable and the social and cultural consequences of better tools such as higher population was equally inevitable. The development of better tools was probably dependent upon the increasing brain capacity of our hominid ancestors. Only when new species of hominids evolved were improvements able to be made in tool manufacture and efficiency, until the arrival of *homo-sapiens* when the improvements began to happen much faster.

Stone tools developed before metal tools as the stone and rocks were plentiful and widespread and the process of hitting one stone with another is a simple and relatively easily developed process. On the other hand native metal (pure metal not in an ore) is very rare and the techniques for working it are more difficult involving heating and hammering. Obtaining metal from an ore usually involves kilns and a complex process of obtaining sufficient heat to separate the metal from its ore. Even further heat was required to melt the metals for making alloys or for casting the metals. Compared to the difficulties of metallurgy, the production of stone tools was a relatively straightforward process. Metal tools eventually took over from stone tools as metal tools, or at least bronze, iron and steel tools were superior to stone tools. Cooper was somewhat soft and was not an ideal material for tools, so there is a stone age, bronze age and iron age but not really a copper age.

Paleolithic tools that have survived for modern archaeological inspection are mainly made of stone. The tools were largely made of flint, quartz, quartzite, basalt, chert and obsidian. These materials were particularly suitable for manufacturing tools because their chemical structure is cryptocrystalline, which means they are made up of minute crystals. When cryptocrystalline stones are hit by another stone they break in a manner known as a conchoidal fracture. The conchoidal fracture results in sharp edged blades because the cryptocrystalline stones have no preferential fracture planes so blades of any size and shape can be made. These desirable qualities resulted in flint, chert and obsidian being favored rocks for Paleolithic tools. Where these stones were not available similar stones such as quartz, which also breaks in a conchoidal fracture and which is a very common mineral, were used.

If the properties of cryptocrystalline stones were different then they might not have been an important material for our hunter-gatherer ancestors. If cryptocrystalline stones could not be chipped to produce a sharp edge then there may have been no stone age based upon the use of stone tools. Either human beings would have had to do without stone tipped tools or a less efficient substitute such as bone would have had to be used. A less efficient substitute would inevitably have certain social effects like a reduced ability to kill wild animals leading to less population growth due to reduced results from hunting and a greater mortality from wild animal attacks. This shows that the cryptocrystalline structure of the rocks in the human environment, which break in a conchoidal fracture, have had a major effect on human social and cultural history.

Fire

The earliest known use of fire was by *homo erectus* about 500,000 years ago. It appears first in colder climates in northern China and Europe and only much later in the warmer climates of Africa. Human kind almost certainly learnt the use of fire due to the observation of natural fires caused by lightning strikes, volcanic eruptions or some other natural cause. Fire could be obtained from natural fires and kept alive by adding additional fuel and then could be put to use by our ancestors. The ability to actually make fire was only learnt much later after fire had been used for a considerable time.

The principal difficulty with the making of fire is the problem of ignition. This is why human kind learnt to use fire long before it learnt to make fire. Around 12,000 BCE humans were able to make fire by rubbing certain stones such as iron pyrites against flint which caused sparks, which could set alight dry leaves or grass. Around 8,000 BCE fire could be made by rapidly turning a stick in a hole in another piece of wood. Later a bow was used to spin the stick, making the whole process somewhat easier. Such Stone Age methods of producing fire are difficult and unreliable and it was not until the 19th century after the discovery of phosphorus, a highly inflammable substance, that fire could be easily produced.

Once humans had learnt to control fire it soon developed a wide range of uses. Fire was used to keep humans warm, especially at night and in colder climates. It was used to provide light allowing humans to work after dark and to explore the depths of caves. Fire was used to keep predators away as other animals are afraid of fire. Fire was also used in hunting to drive prey over cliffs or into swamps where they could be more easily killed. It was also used to destroy old vegetation to produce re-growth that provides good grazing for the animals humans hunt. Fire was also used for cooking and the sharpening of spears. After the development of agriculture, fire was to play an essential role in the development of pottery, metallurgy and glass.

The effect of the uses of fire was certainly to allow some increase in human population due to a greater food supply from better hunting and cooking and a reduced mortality from wild animal attacks. Hominids were able to occupy territories with colder climates such as Europe and Northern China. Human activities could continue at night and access to dark caves became possible. It is likely the use of fire turned human beings into the leading predator on the planet.

Fire is the result of a chemical reaction between oxygen and an organic (e.g. carbon based) compound. Oxygen is contained in the earth's atmosphere and organic material which includes all plant life is widespread so fire can be used nearly anywhere on the planet. The chemical reaction which causes fire produces both heat and light which are valuable products for human beings.

It is to be noted that fire was first used where there was the greatest need for it, in cold and dark Europe and northern China. The use of fire initially involved the use of natural fire, a much simpler process than the difficult task of actually working out how to make fire, a development which occurred much later than the first use of fire.

Fire has had a major effect on human history. If it were possible to easily make fire, then human history would have been different as the making of fire would have occurred much earlier in human history. Equally, human history would be different if it was impossible to make fire and we always had to rely on natural fire. However, the most significant change in human history would be if fire could not happen at all. If oxygen simply did not react with organic matter to produce fire, or the reaction only took place at a very high temperature so that fire could not be made or even occur naturally then human history would be radically different. The development of pottery, metallurgy and glass would have occurred much later in history or possibly not at all. If fire had different properties, for example if it burnt at much higher temperatures, say 2,000ºC, then the entire history of metallurgy would have been much different. Special kilns and ovens needed to produce high temperatures for metallurgy would not have been required. Human beings would have been able to smelt and melt iron at a much earlier stage in history so there would have been no bronze age and hunter-gatherers could have used iron tools and weapons.

Agriculture and Pastoralism

The domestication of plants and animals is one of the most important events in human history. It is also one of the most controversial with much debate as to why humans began to practice agriculture and whether agriculture was a good thing. For the great majority of their existence humans were hunter-gatherers and then beginning about 10,000 years ago some humans in South West Asia began farming.

The change from hunting and gathering to farming did not take place overnight. Almost certainly there was a transitional stage between hunting and gathering, and farming. This transitional stage is often called proto-agriculture. Proto-agriculture occurs when hunter-gatherers engage in practices which assist the growth of wild plants. This may involve burning off unwanted foliage to encourage regrowth, weeding, irrigation and the re-planting of plants such as wild yams after removing most of the edible part of the plant. The development of proto-agriculture reflects hunter-gatherers' increasing knowledge of how to make plants grow which was eventually to lead to the knowledge required for full scale agriculture. The knowledge would have been acquired over hundreds and possibly thousands of years. It is sometimes considered that the knowledge of how to grow crops was always possessed by hunter-gatherers. This assertion is answered in my paper *The Discovery of Agriculture* contained in Appendix 2 of this book.

When agriculture began it would have involved a considerable time when both agriculture and hunting and gathering would have been practiced together. Eventually, however, in South West Asia agriculture became the primary form of subsistence. One reason for this may have been that the technologies used in agriculture would have gradually improved during the early periods of agriculture when both agriculture and hunting and gathering were practiced together. New flint-bladed sickles for harvesting grains, grinding slabs to remove husks, underground storage pits and the practice of roasting grains to prevent them from sprouting when stored would all have improved the practice of agriculture. A further factor is that the crops themselves evolved to become more suitable for agriculture.

The plants evolved because humans selected the wild plants most useful to themselves for planting and this caused the preservation of certain mutations within those plants. The mutations involved concerned the size and taste of the edible parts of the plants, a high fruit to seed ratio within fruits and oily fruits or seeds. These selections were made more or less consciously by early farmers. However other selections which affected the plant's methods of seed dispersal, germination and reproduction were made quite unconsciously.

When early farmers selected wild plants they chose plants with large edible parts. Plants with large edible parts, for example fruit and berries, produce seeds which when planted are likely to produce further plants with large edible parts. Crops such as peas, corn and many fruits are much larger than the wild plants they evolved from.

Human selection also affected the taste of seeds. Many wild seeds taste bad to prevent them being eaten by animals. However, the occasional mutant plant will produce pleasant tasting seeds which can be planted and produce further pleasant tasting seeds. Almonds, lima beans, watermelons, cabbages and potatoes all had wild ancestors with an unpleasant taste or were poisonous. But when nice and safe mutants were cultivated by early farmers, valuable crops were produced.

Fruits with much flesh and small or no seeds were also selected by early farmers. This led to such qualities becoming standard among domesticated plants. Oily fruits and seeds were selected by early farmers so cultivated plants such as olives became much more oily than their wild ancestors. Plants used for producing textiles like flax and hemp were selected for long stems as the fiber used to produce the textiles came from the plant's stems.

Domesticated plants differ from wild plants in other ways. Many wild plants such as wheat and barley have mechanisms for seed dispersal which involves their stalks automatically shattering to spread the seeds on the ground. However there are mutant varieties of wheat and barley whose stalks do not shatter and these tend to be gathered by humans as it is easier to collect the mutants than the normal plants. When the mutants were planted they tended to produce more mutants so that non-shattering stalks became the norm for domesticated wheat and barley. A similar situation exists with wild peas that have exploding pods to disperse the seeds. Mutant peas in which the pods did not explode were most easily collected by early farmers and when the seeds were planted they produced crops in which the pods did not explode.

Annual plants in the wild often have germination inhibitors to stop the seeds all germinating at the same time making them vulnerable to a single frost or drought. When people first planted the seeds, mutants without germination inhibitors would sprout first and be harvested by humans so as to select the mutants without the germination inhibitors.

Most wild plants reproduce by fertilization from other plants. When this happens, mutations desirable for domesticated plants would be lost. Early farmers' selection of crops favored self-fertilizing mutants, as those plants retained the qualities desirable for farming, so that domesticated crops became self-fertilizing. This meant desirable attributes in domesticated species were not lost by fertilization from wild plants without those attributes.

There are roughly 200,000 wild plant species, but of these only a few thousand can be eaten by humans. Only a few hundred plant species have actually been domesticated and there are about a dozen plant species that make up over 80% of the crops people eat. These twelve species are the cereals wheat, barley, rice, corn and sorghum, the pulse soybean, the roots, manioc, potato and sweet potato and sugar cane and sugar beet and the banana.

The first crops to be domesticated were wheat, barley and peas in South West Asia about 10,000 years ago. The reason why these crops were the first to be domesticated was because the qualities of their wild ancestors made them the easiest crops to domesticate. The wild ancestors of these crops were edible and gave good yields and were easily planted or sown. They were quick growing and could be harvested a few months after planting and they were easily stored. Relatively minor genetic change was required before they were domesticated. They were usually self-fertilizing so desirable qualities were not lost by reproduction with other plants. The main genetic changes were the development of non-shattering stalks and consistent rapid germination.

More difficult crops were domesticated somewhat later. Some fruit trees such as grapes, figs, dates and olives were domesticated about 4,000 BCE. These crops do not provide food until more than three years after planting and may take as long as ten years to reach maximum production. Compared with other domesticated trees, these crops are easy to plant as they can be grown from seeds or cuttings.

Trees such as pears, plums, cherries and apples were quite difficult to domesticate as they could only reliably be grown by grafting. Grafting was a difficult technique to develop as it could only be discovered by deliberate experimentation. The wild ancestors of these trees had the additional problem of not being self-pollinating so farmers had to plant other trees nearby or find self-pollinating mutants.

A number of plants became domesticated after first evolving as weeds in cultivated fields. These crops known as “secondary crops” only became domesticated in West Asia and Europe in the second and first millennia BCE. Such crops included oats, turnips and probably lettuce, leeks and beets.

Strawberries and raspberries were not domesticated until medieval times. This was because wild strawberries and raspberries have very small berries that are of only limited value to people. It was not possible to select plants with larger berries to produce domesticated strawberries and raspberries with larger berries as birds which eat the small wild berries would defecate wild berry seeds everywhere and would interfere with human selection of plants with larger berries. Strawberries and raspberries were only domesticated when greenhouse or protective nets were used to protect selected plants from birds, allowing plants with extra large berries to be produced.

The first crops to be domesticated were those most easy to domesticate. The wild ancestors of wheat, barley and peas had the right qualities concerning taste, yields, ease of planting, rapid growth, ease of storage and minimal genetic changes needed for domestication, so they were the earliest plants to be domesticated. It was the qualities those particular crops had which made them easier to domesticate than other crops so they became the first crops to be domesticated. The next crops to be domesticated were fruit trees, which while they could be grown easily enough, did not provide food for some years after planting. Secondary crops which evolved from weeds in cultivated fields had to be domesticated after the earlier crops had been domesticated. Trees that could only be grown by using the difficult technique of grafting inevitably were domesticated after trees that could be grown from seeds or cuttings. Finally strawberries and raspberries were domesticated last due to undersize berries and the difficulty in selecting and growing plants with larger berries due to wild strawberries and raspberries interbreeding with the selected plants.

The plants with the best qualities for domestication were domesticated first and those that were more difficult to domesticate or those that were less desirable, for example with a long period of growth required before food was produced, were domesticated later. The order of domestication was rational and was the order in which domestication was always going to take place.

A further important point concerning the domestication of plants is that plants, being living organisms, will evolve to fit in with the new environment the plants were put into. Larger edible parts, non-shattering stalks, consistent rapid germination and self-fertilization were attributes domestic plants developed in response to the new environment created for the plants by humans. This ability to evolve into more useful plants than their wild ancestors made the development and spread of agriculture much easier. There must be considerable doubt as to whether agriculture would have lasted if plants were not capable of evolving and humans could only grow the wild ancestors of domesticated plants. Certainly agriculture would never have become so widespread as it did if we could only grow wild plants. Again one sees a particular quality of plants, their ability to adapt to new environments, having a major effect on human history. If plants could not evolve there may have been little or no agriculture, sedentism may have been impossible apart from in a few environments that are endowed with an unusual abundance of food and human history would have been radically different.

Agriculture only became possible because, of the few thousand plants that people can eat, a few hundred of them were capable of domestication. The other edible plants could not be domesticated due to characteristics of the plants that made them unsuitable for domestication. Some plants are just so slow at growing they are uneconomic to grow. In other plants undesirable qualities, such as bitter tasting or small fruits or nuts, or shattering stalks and delayed germination, are controlled by a single gene and can be bred out of the plant by human selection. Where undesirable qualities are controlled by a number of genes it is far more difficult or even impossible to get rid of those qualities by selective breeding. The manner of seed dispersal used by certain plants can also make domestication difficult or impossible. Where seed dispersal is by animals such as squirrels it is very difficult for humans to select and isolate trees with desirable qualities. This is because the squirrels are constantly spreading seeds everywhere, including those with undesirable qualities, so it is not possible to prevent pollination of trees with good qualities by trees with undesirable qualities. The same problem existed for strawberries and raspberries whose seeds are spread by thrushes. Only when nets and glasshouses were used to isolate mutant strawberries and raspberries with desirable qualities were those plants able to be domesticated.

It is quite apparent that whether a particular plant can be domesticated depends on the particular characteristics of that plant. The length of time a plant takes to grow, its method of seed dispersal and whether it can evolve qualities humans desire for example if only a single gene controls the particular quality. All these qualities are ultimately controlled by the genetic make-up of the plant, so that whether a plant can be domesticated or not ultimately depends on the genetic make-up of the plant.

If the genetic make-up of all plants prohibited domestication, then agriculture would never have occurred. If agriculture had not been possible humans would have remained hunter-gatherers and sedentism would have been impossible except possibly for a few areas of very abundant food supplies. This almost certainly would have meant cities, civilization, writing and the industrial society many of us live in would never have existed. Obviously some plants, due to their genetic make-up, were suitable for domestication so agriculture was possible. But only about ten percent of the plants edible by humans were capable of domestication.

Different crops and different combinations of crops could have different effects on societies. Certain crops such as wheat require a system of field rotation due to nitrogen exhaustion in the soil. In Roman times a two field rotation system was used with half the land being left fallow each year to allow the nitrogen to be replenished by natural processes. By the 8th century CE crops such as winter wheat, rye and legumes, such as peas and beans, began to be used in a three field rotation system. The three field system allowed two thirds of the land to be used each year with legumes being planted in one of the three fields as they restored nitrogen to the land. The increase in land use and the better nutrition provided by a wider range of crops allowed an increase in population in Europe that ended only with the famine and disease (the black death) of the fourteenth century.

Rice on the other hand is normally grown in paddy fields, where the water is muddy and the mud restores the fertility of the soil so that it is not necessary to leave any land to lie fallow. Rice is often grown with sophisticated irrigation systems which require considerable organization to create and maintain. Wittfogel suggested that the need for building and maintaining the irrigation systems inevitably led to substantial state control and firm social discipline. It is also possible, from about the thirteenth century onwards, to grow two or sometimes three harvests of rice per year. Given that no land is left to lie fallow and that several harvests could be produced per year, rice growing areas such as Southern China, tended to have a high population density compared to Europe.

The principal crop that supported the Aztec and Inca civilizations in the New World was maize. Maize grows quickly and produces extremely high yields and it is sometimes possible to have two harvests per year. It also requires little work to produce, leaving people free to engage in other activities. In *Capitalism and Material Life 1400-1800* Fernand Braudel suggests:

"Maize on the irrigated terraces of the Andes or on the lakesides of the Mexican plateaux brought about theocratic totalitarian systems and all the leisure at the disposal of the countryside was used for immense Egyptian style public works. …. Without maize the giant Mayan or Aztec pyramids, the cyclopean wall of Cuzco or the wonders of Machu Picchu would have been impossible. They were achieved because maize virtually produces itself."[27]

It seems likely that the type of crops, available to a particular society, will have a significant impact on the type of society that uses the crop. The type of crop available is determined by what nature provides us in terms of wild plants whose characteristics are determined by the genetic make-up of those wild plants. This means the genetic make-up of the wild plants on this planet has had a major effect on the type of human societies that have existed and on human history.

Humankind has domesticated fourteen species of large herbivorous and omnivorous mammals. The five most important of these are cattle, sheep, horses, goats and pigs. All of these animals are now distributed worldwide. Nine other large herbivorous mammals, being two species of camel, reindeer, donkey, lama, water buffalo, bali cattle, mithan and yak have also been domesticated but are confined to particular areas. Ten of these fourteen animals became domesticated between 8,000 and 2,500 BCE, the remaining four animals having no clear date of domestication. Given that no large herbivores have been domesticated since 2,500 BCE even with modern scientific methods, it seems that all the animals that can be domesticated, have been domesticated.

In *Guns, Germs and Steel* Jared Diamond refers to there being 148 large wild herbivorous animals that could be considered for domestication. However, only fourteen of these were domesticated.[28] This is because in order for animals to be domesticated they must meet certain criteria. These criteria relate to the animal's diet, breeding habits, rate of growth, social structure, inclination to panic and viciousness.

Domesticated animals must eat and if they consume an excessive quantity of food or have fussy eating habits they may be uneconomic to keep. Ideally they should consume low quantities of easily produced food such as grass or some other easily grown food. A domesticated animal will always eat far more food than it is able to produce for its human owners. It may take ten tons of food to produce one ton of herbivore. If the food is grass then that is not necessarily a problem but if the animal is a koala that only eats eucalyptus leaves or a panda who eats bamboo shoots it might not be worthwhile to keep the animal. The most important domestic animals eat easy to produce food, such as grass, and are not fussy about what they eat.

The reason no carnivore is domesticated as food for people is because the economics is even worse than for herbivores. If you want to produce a one ton carnivore you may have to feed it ten one ton herbivores. However the herbivores themselves would have to be fed with ten tons of feed each, so the total cost of feed for a one ton carnivore would be one hundred tons of food. This means domesticated animals were always going to be herbivores or omnivores.

An animal's breeding habits may also affect its suitability for domestication. Many animals will simply not breed in captivity, often because they require elaborate courtship rituals which are not possible in captivity. If an animal cannot breed in captivity, then it cannot be domesticated.

Some animals are not worth domesticating due to the length of time it takes them to reach maturity. If you have to feed your animal for ten or fifteen years before it is fully grown it may not be worth domesticating the animal.

The vast majority of domesticated animals have wild ancestors with three particular social behavior traits. These are that the wild ancestors have a dominance hierarchy, they live in herds and the herds do not occupy exclusive territories. Animals with a dominance hierarchy are easier to domesticate as they treat their human owner as the dominant animal and are easily able to be led around by the human. Animals used to living in herds are easier to domesticate as they are comfortable with being penned in small spaces with many other animals. Animals that do not occupy exclusive territories are easier to domesticate as they can be easily mixed in with animals from other herds without fighting.

Some animals have an inclination to panic when they feel threatened. If put in fenced enclosures they might charge the fence in an attempt to escape and either injure or kill themselves. Alternatively they might die of shock if kept penned up in an enclosed area.

The final problem with domesticating animals is that many are so vicious that it is dangerous for humans to be around them. Animals like large bears, the African buffalo and zebras cannot be domesticated as they are just too dangerous to have living among humans.

Unless large herbivores meet all of the criteria of eating the right foods, being able to be bred in captivity, having a good rate of growth, the right social habits, are comfortable with being enclosed and are not too dangerous to people then they cannot be domesticated. If any of these factors are not present, then the animal cannot be domesticated. This is why only fourteen out of the one hundred and forty eight large herbivores have been domesticated.

The qualities animals need in order to be domesticated are dependent upon the genetic make-up of those animals. If those fourteen large domesticated animals did not have the right genetic make-up then they would not have been able to be domesticated. If there were no large domestic animals all those societies we describe as pastoralist would not have existed. There would have been no Mongol Empire, no Mongol conquests of China and most of Asia. The Roman Empire may not have fallen, it certainly would not have been invaded by Goths and Huns mounted on horses. Many of the infectious diseases that cause illness and death in humans appear to have originated from domestic animals. Smallpox, tuberculosis and measles appear to have come from cattle, while the flu has come from pigs and ducks. If these animals had not been domesticated humans may well have never caught those diseases. If there had been no large domesticated animals there would have been little or no plough agriculture as there would have been no horses or oxen to pull the plough. With no horses, oxen or other large domestic animals to pull carts or wagons or to serve as pack animals, the transport of goods and people over land would have been much more difficult. Everything would have had to be carried by people.

A good idea of what societies without large domestic animals would be like can be seen by looking at the Aztec and Inca civilizations in the New World. The Aztecs had no large domestic animals at all and the Incas only had the lama which is nowhere near as large or as strong or as co-operative as horses or oxen. The Aztecs had the idea of the wheel which they used on children's toys but they had no carts or wagons due to having no animals to pull them. Both the Aztecs and the Incas practiced hoe agriculture, breaking the ground with a digging stick before planting the seeds. With no large animals to pull ploughs, plough agriculture was not practicable. If large domestic animals had existed in the New World, they almost certainly would have been used as can be seen from the way in which the American plains Indians took the horse when it became available. One of the major factors in the Spanish conquest of the Aztec and Inca empires was the lack of large domesticated animals in the New World. The Spanish had cavalry, while the Aztecs and Incas had none and the Spanish came with smallpox to which the people of the New World had no immunity. The disease ran through the Aztec and Inca populations killing millions and greatly assisting the Spanish conquest of those empires. Clearly the presence or absence of large domesticable animals can have a great effect on a society. The presence or absence of such animals is determined by the genetic make-up of the large, wild herbivorous animals nature has evolved. This means the genetic make-up of such animals can determine the types of human societies that have existed and human history.

Pottery

One of the consequences of human kind becoming sedentary was the development of pottery. Hunter-gatherers do not use pottery as it is fragile and easily broken when being moved. One of pottery's major uses is food storage which is of no interest to hunter-gatherers who do not usually store food. With the beginnings of widespread sedentism resulting from the discovery of agriculture, the use of pottery became a practical possibility.

Pottery is made principally from clay which is widespread throughout the surface of the earth. The first human use of fired clay seems to have been in the production of Venus figurines during the Upper Paleolithic between 30,000 and 15,000 years ago. This suggests humans probably had enough knowledge to produce pottery long before it was extensively used. The first known use of pottery appears to have been by the Jomon in the area of modern day Japan. The Jomon seem to have been semi-sedentary hunter-gatherers and to have used pottery as early as 10,500 BCE. Pottery seems to have been independently invented in North Africa around 8,000 BCE and in South America around 5,000 BCE.

It is the particular properties of clay that allow it to be used to manufacture pottery. Clay is plastic in that it will retain any reasonable shape it is molded into. If water contained in the clay is allowed to evaporate, for example by drying it in the sun, the plasticity is lost but the shape of the clay is retained. However if water is later added to the clay the plasticity returns. This meant sun dried pottery could only be used for storing dry products and in climates where humidity and rainfall were low. Egypt and Mesopotamia had such climates and sun-dried pottery was used in those places.

The need to produce pottery capable of holding water led to fire dried pottery. When fire with a heat of over 450°C is applied to clay, the clay will lose its plasticity even if it comes in contact with water. Its shape will be permanently retained and such fire-dried pottery can be used to store water and to boil water, for example in cooking. Temperatures of just over 450°C however only produce earthenware which are porous in that water can slowly percolate through the walls of the vessel. This can have the beneficial effects of cooling the contents of the vessel or alternatively if this effect is not desired then the vessel can be glazed or if heated to a much higher temperature stoneware which is not porous may be produced. A glaze is a glass-like substance which if applied to a vessel stops water percolating through the walls of a vessel. Stoneware can be produced by mixing the clay with a fusible stone which under high temperatures vitrifies and produces a non-porous vessel.

The earliest method of manufacturing pottery vessels was to place the clay around the inside of a basket or sack and then placing the vessel in the sun or fire. Molds were also used in the production of pottery from an early date. The potter’s wheel was a later manufacturing technique which was well established by the first millennium BCE.

The earliest decoration done on pottery involved scratching the clay before the pottery was placed in the fire. Pigments were later used on Neolithic pots and colored glazes were used in China and the Persians introduced the use of luster in the 9th century.

Metallurgy

In, *Before Civilization: The Radiocarbon Revolution and Prehistoric Europe,* Colin Renfrew described the development of copper and bronze metallurgy in the Near East.

"This development in general tends to follow a series of steps. These have been well documented in the Near East by Theodore Wertime, and can be listed as follows:

1. *Simple use of native copper.* Native copper - almost pure copper as found in nature - occurs fairly widely in many regions where there are copper ores. In most areas copper may first have been valued simply as another attractive mineral or stone - just as meteoric iron was used in the Near East to make cylinder seals, along with a whole range of attractive stones long before its metallic properties were exploited.
2. *Cold hammering of native copper.* It would soon be realized that this new mineral did not fracture on hammering so easily as other stones. Shaping by hammering was an obvious way of working.
3. *Annealing of native copper.* Repeated cold hammering makes the copper brittle so that the object fractures. By heating it in an open fire, and hammering while hot, this brittleness can be avoided. Cold hammering can then be used to finish the object, and to give harder cutting edges if desired.
4. *Smelting of copper from its ores.* This represents a notable advance. The ores themselves are often brightly colored, like azurite (blue) and malachite (green). The oxide and carbonate ores are more easily reduced than the sulphide ones, and a temperature of about 700°C is needed, which can be attained without the construction of a complicated oven. Only fairly small and irregular pieces can be obtained in this way, however, unless the copper is allowed to run off at a higher temperature.
5. *Casting the copper in an open mould.* Casting requires heating to the melting point of copper, 1,083°C, and allows the production of good thick blanks in roughly the required shape. These can then be further worked by annealing and cold hammering.
6. *Casting-in, and the use of the two-piece mould.* More complicated shapes can be obtained by these methods. Shaft-holes, for instance, can be produced during casting by inserting a charcoal core in the mold. A two-piece mold allows a more elaborate shape than in a one-piece mold where the upper surface of the casting is always flat.
7. *Alloying with arsenic or tin.* Arsenic bronze and tin bronze are much stronger than pure copper, so that the objects are less likely to snap in use. Alloying can also improve the hardness, and also the process of casting, avoiding the formation of blow-holes made by gases dissolved in the melt as they come out of solution on cooling.

8. *Lost wax casting.* A wax model is made in the shape of the desired bronze casting, and coated with clay which forms the mold. The wax melts as the molten bronze is poured in to replace it in the mold. The mold itself is broken and removed when the bronze cools. In this way, castings of much more elaborate shapes can be produced.

Further developments are possible of course, and the story does not stop there. It is important to note that each step is to some extent dependent on the preceding one, and indeed the sequence can really be regarded as one of increasing competence in pyro technology, in the handling of materials at high temperatures. Increasingly specialized skills are involved at each stage, and efficient casting of bronze usually requires some sort of oven where the flow of air can be controlled.

In the Near East, stages 1 and 2, and probably 3 as well, were reached very early over a wide area. Many of the earliest Neolithic settlements known, including Ali Kosh in Iran, and Catal Huyuk and Cayonu in Turkey (the last perhaps not even a farming community), have yielded finds of native copper. Stage 4 is reportedly documented at Catal Huyuk around 6000 b.c. in radiocarbon years. Stages 5 and 6 come later - the earliest reported instance, not yet documented by metallurgical analysis, is a mace head from Can Hasan in Turkey dated around 5000 b.c. in radiocarbon years.

Alloying with tin, stage 7 in this sequence, was a much later development and is seen around 3000 BC. (c. 2400 b.c. in radiocarbon years) in the Near East, the Aegean and the Balkans. It is about this time also that lost wax casting, stage 8, is first seen.

A similar sequence of development can be demonstrated also in the Balkans. Stage 1 is documented by the find of beads at the cemetery of Cernica in Romania, described as of 'copper mineral', which in this case implies ore rather than pure native copper, worked in the same manner as beads of stone or shell. Cernica is a contemporary of the earlier Vinca culture, and must be dated back almost to 5000 B.C. in calendar years. A little after this time, but still before 4700 B.C., awls and small objects of native copper are found in the Vinca culture and its Balkan contemporaries.

The earliest scientifically documented indication of stage 3, hot working, comes from a site in the western U.S.S.R., dated before 4000 B.C. It is a copper fish-hook which had been heated to 300°C and worked to shape. Tools made from smelted copper, which can be recognized by their greater content of minor impurities, occur at about the same time.

The most striking advances, illustrating stages 5 and 6, are seen in the Gumelnitsa culture. There, certainly before 4000 B.C. in calendar years, impressive axes were cast, with the shaft-hole already in position. Examination by J. A. Charles shows that these were indeed cast in open molds, with the shaft-hole cast-in rather than being drilled out subsequently. Several have been found stratified at sites in Bulgaria including Chotnitsa, and one was included in a hoard of flat axes or chisels at a Vinca culture tell in Jugoslavia.

From this form developed the axe-adze, with its working edge at each end (Plate 7). Some of these are magnificent objects, and their manufacture may have begun before 4000 B.C., and must have continued for a long period after this time. But alloying was apparently not practiced in the Balkans until the bronze age, from around 2500 B.C. in calendar years, at much the same time as it began in the Aegean and the Near East.

This gradual and logical development, which took at least a thousand years, from the first tentative use of copper and copper ore to the accomplished casting of the shaft-hole tools, clearly reflects considerable advances in pyro technology. But it is important to realize that, in the copper age at least, it was pottery rather than metallurgy which led the way in pyro technological innovation.

Already the very first Neolithic farmers in Europe had ovens for parching grain and baking bread; examples of these were excavated at the very early Neolithic site of Nea Nikomedeia in north Greece. And from the very beginning the Balkan farmers were accomplished potters. In the earlier Vinca culture temperatures as high as 700°C or 800°C may have been reached for the firing of pottery. It is particularly significant that the attractive graphite-decorated pottery of the Gumelnitsa culture required even more exacting firing conditions. Graphite will burn off, if it is fired in oxidizing conditions where the supply of air is not limited, at a temperature above 700°C. It is clear that the pottery was in fact fired at around this temperature in conditions where the flow of air was carefully regulated. Whether or not this involved the use of some more elaborate potter's kiln is not yet clear, but it certainly does indicate an increasing mastery in the control of materials at high temperatures.

All this had come about in Bulgaria and south Romania, where graphite decorated pottery was being produced, already before 4500 B.C. And the development of ceramic technology seems a logical one, for which no outside influence need be invoked. The exciting thing is that these conditions were not so far from those needed for the smelting and casting of copper - a temperature of 1100°C and the control of air to provide a reducing atmosphere. Seen purely in technological terms, the development of copper metallurgy in the Balkans was already heralded by the skills of the potter.

Technically, then, it is entirely possible that metallurgy developed independently in the Balkans. The natural resources were available, and so was the pyro technological skill. But this alone does not demonstrate that metallurgy was something worked out locally, without essential ideas from the earliest metal workers of the Near East."[29]

The Renfrew quote only covers part of the history of metallurgy. Arsenic bronze was developed before tin bronze probably because copper and arsenic are sometimes found in the same ores while tin is a relatively rare metal in Europe, North Africa and South Asia, although it is found in present day Iran. However tin bronze soon became the preferred form of bronze as arsenic bronze had a tendency to slowly poison those working with the metals.

The next major metallurgical development, after the use of bronze, was the use of iron. The earliest use of iron was iron obtained from meteorites. Iron trinkets were found in Ancient Egyptian tombs dating from 4,000-3,000 BCE. However, such iron was rare and had little effect on human societies.

The first major impact of iron on human civilization was when the Hittites began smelting iron around 1500 BCE. Iron is smelted from its ores at around 1200°C and melts at 1528°C. The Hittites used furnaces lined with clay to produce the temperatures required to smelt the iron ore. The ore and wood was placed in the furnace to burn and the wood became charcoal. Carbon in the charcoal combined with oxygen in the iron ore to produce an oxide of carbon and the iron metal in the form of a spongy mass. The spongy mass contained earthy slag which could mostly be removed by considerable hammering to produce wrought iron. Additional hammering when heated would allow the metal to be shaped into tools and weapons. The furnaces available to ancient metallurgists did not produce enough heat to melt the iron to produce cast iron.

The situation was different in China where better furnaces and iron ore with a high phosphorus content which produced iron which melted at relatively low temperatures allowed the production of cast iron from around the 3rd century BCE. In Europe an improved furnace was invented about 700 CE in Catalonia. A bellows was used to force air through a nozzle called a tuyere into the charcoal to produce higher temperatures. The temperatures however were not sufficient to melt the iron and allow the production of cast iron.

It was not until the 14th century that iron smelting furnaces capable of melting iron were built in Europe. These furnaces were known as blast furnaces and were substantially larger than earlier furnaces. The blast furnaces had water powered bellows which produced much higher furnace temperatures as the bellows produced a continuous and strong flow of air through the tuyeres into the furnace. The higher temperatures allowed the iron to absorb a small quantity of carbon, which lowered the melting point of the iron to a temperature which the blast furnace could obtain. The melted iron, known as pig iron, could be poured into molds or could be remelted and cast into any shape. The carbon in the pig iron could be removed to produce wrought iron which was more malleable than pig iron.

Substantial improvements were made to blast furnaces between 1500 and 1700. Reverberatory furnaces, with no chimneys and using underground pipes to bring in air, achieved higher temperatures with domed shaped roofs lined with clay reflecting the heat back into the furnace. Continuous smelting processes, which involved ore and fuel being continuously fed into the furnace to provide a continuous supply of iron greatly increased efficiency and production.

The use of coke, purified bituminous coal, in blast furnaces began around 1709 and greatly increased after 1760 when a method was found to get rid of silicon from iron produced from blast furnaces using coke. The silicon made it costly to convert pig iron into wrought iron. In the late 18th century coke replaced charcoal in most British blast furnaces. Blast furnaces produce pig iron but for many products the more malleable wrought iron was more suitable. The conversion of pig iron into wrought iron involved eliminating the carbon from the pig iron. An improved method of getting the carbon out of the pig iron was invented by Henry Cort in 1784. Cort's puddling process melted the pig iron in a reverberatory furnace which burnt the carbon and other impurities out of the iron and produced a mixture of iron and slag. The slag was removed by hammering to produce the wrought iron.

A further improvement to blast furnaces allowing still higher temperatures and reduced fuel use was invented by James Neilsen in 1829. Neilsen's invention involved using the furnaces' own gases to preheat air before it entered the furnace. The air entered the furnace through a red-hot tube heated by the furnace's own gases and the hot air allowed the furnace to reach temperatures not previously obtainable. The pre-heating of the air blast was further improved by Edward Cowper in 1860 when he invented the hot-blast stove. Waste gases from the furnace were fed into a brick-lined stove and heated the stove. Air entering the furnace is passed through the stove so it is heated before it reaches the furnace.

Wrought iron was the principal material of the Industrial Revolution. Steel was a better material but was too expensive for widespread use during the Industrial Revolution. Steel is chemically mid-way between wrought iron which contains almost no carbon and pig iron which contains about 4% carbon. Steel usually contains between 0.2% carbon and 1.5% carbon. It was not until the second half of the 19th century that a process for creating cheap steel was invented. The Bessemer process was patented in 1856 and used a vessel called a converter into which molten pig iron was poured. Air was blown through holes in the base of the converter. The oxygen from the air combines with some of the iron to produce iron oxide which reacts with the carbon in the pig iron to produce carbon monoxide which releases some of the carbon from the pig iron. The remaining carbon is removed when the oxygen in the air is combined with silicon and manganese which form a slag. The resulting metal was brittle so manganese was added to remove the brittleness and then carbon was added to bring the steel up to the desired carbon content. The same process was independently invented in America by William Kelly.

An alternative method of making steel, known as the open-hearth process was invented in 1864 by William and Frederick Siemens and then improved by Pierre and Emile Martin. The open–hearth process involved pre-heating the air going into the furnace in two chambers that operated alternatively. The chambers, known as regenerators, contained a fire brick checker work and were alternatively heated by the furnace gases so the air passing into the furnace through the regenerators was heated, resulting in higher furnace temperatures. As with the Bessemer process iron oxide was used to remove carbon and other impurities and manganese was added to remove brittleness and if necessary carbon was added to obtain the desired carbon levels.

The invention of electrical generators led to the use of electricity for heating furnaces. The first electric arc furnace began operation in 1902 and, while more expensive than the Bessemer and the open-hearth processes, was able to produce better quality steel due to it having fewer impurities than steel which had been in contact with fuel. Electric furnaces were able to produce greater heat and the temperatures could be more easily controlled than with ordinary furnaces. The use of electric furnaces was to result in the large-scale production of metals such as tungsten, chromium and manganese which when added to steel gave it useful properties such as improved hardness and resistance to wear. The electric furnace also allowed the mass production of aluminum. Aluminum is widespread on the Earth but it was difficult and expensive to extract from its ore, bauxite, before the invention of the electric furnace. The electric furnace produces aluminum by a process of high temperature electrolysis which produces molten aluminum in large quantities, although the process used substantial quantities of electricity.

It had been long recognized that the use of oxygen, rather than air, in steel making would produce higher temperatures, faster production and reduce fuel costs. The high cost of producing oxygen stopped its use in steel making, until the price fell substantially and in 1948 the L-D process for using oxygen in steel making was developed. The L-D process involves blowing a jet of nearly pure oxygen at supersonic speed onto the surface of molten iron. The oxygen quickly burns out the carbon and other impurities resulting in faster production and reduced fuel costs.

The social and cultural consequences of the discovery of metallurgy were initially quite minor. Copper was initially used mainly for ornaments and jewelry as it was too soft a material to replace the stone tools and weapons used in Neolithic times. It was only when bronze was invented that metal tools and weapons replaced stone tools and weapons to create a Bronze Age. Bronze however was a reasonably expensive metal and when iron smelting was discovered by the Hittites the new metal soon replaced bronze as the principal material for tools and weapons. Iron ores are reasonably widespread and iron is a harder material than bronze making it better for both tools and weapons.

Iron was used for a wide variety of purposes such as nails and tools, cooking pots and kitchen utensils, axes for clearing land and for the tips of ploughs. The use of iron tools and weapons gave humankind greater control of their environment leading to increased population and larger settlements. Iron became the principal material for the Industrial Revolution being used in steam engines, industrial machinery, in railways for rails and locomotives, for bridges, buildings and in iron ships.

The Bessemer and open-hearth steel making processes lead to a great reduction in the price and increase in production of steel. Cheap steel replaced iron in a great variety of applications. Steel was used in railways and for ships and in bridge building. Motor vehicles became one of the biggest users of steel in the 20th century and different types of steel began to be developed for different purposes. Cutting tools were made from steel containing chromium and tungsten as that steel remains hard even at high temperatures. Excavating machinery was made from wear resistant manganese steels and transformers, generators and motors were made from silicon steel due to its magnetic quality. Stainless steel containing chromium and nickel was widely used in kitchens and in industrial plants vulnerable to corrosion as it does not rust. Steel coated in zinc or tin also resists rust and is used for cans containing food and for equipment used around the home.

Metallurgy has had a great effect on human societies, certainly since the Bronze Age and increasingly since the Iron Age and particularly with the modern Steel Age where a vast range of products and structures contain metals. If metals did not exist at all then we would be restricted to stone, bone and wood tools. This would have had an enormous effect on human history. It is doubtful whether the Industrial Revolution and the industrial world that emerged from it would have been possible without metals. It is hard to conceive of wooden or stone steam engines or internal combustion engines. Wooden engines would catch fire while it is doubtful that stone could be worked in a way that could create pistons and cylinders. Without metals it is doubtful that there would be usable electricity, as the transfer of electricity over significant distances would be difficult or impossible.

Even if there were metals, the properties of those metals would have had a major effect on human history. If the smelting and melting points of metals were different then human history would have been different. This can be seen by the use of counterfactuals. If for example there was a metal with all the properties of iron except that it could be smelted at say 800°C and melted at 900°C then the course of human history would be different. Given iron's superior qualities to copper and bronze, iron would be used in preference to those two metals for most purposes, so there would have been no copper and bronze ages. Or alternatively if such a metal could be smelted at 400°C and melted at 500°C then such operations could take place on open fires without furnaces or other special equipment. In this case hunter-gatherers could or would have developed iron and steel weapons and tools so that there would have been no stone age. However, as the smelting point of iron was around 1200°C and its melting point was 1528°C, inevitably the human use of iron was limited until temperatures of 1200°C were possible and the iron age followed the earlier stone, copper and bronze ages.

The quote from Colin Renfrew illustrates a number of points. The first is that copper and bronze metallurgy in the Near East developed through a series of steps each to some extent dependent on the preceding step. The development of metallurgy took place in a particular order and the order of development was a necessary and inevitable order. The order involved a move from simpler metallurgy to more complex metallurgy involving increasing specialization and skills as the metallurgy developed. The reasons for this is that simpler forms occur to humans before more complex forms and the complex forms are often refinements or improvements of the simpler forms. In this sense the simpler forms will always come before the more complex forms.

The progress of metallurgy started with the use of native copper and iron from meteorites as the metals were obtainable without smelting the metals from ores. It was soon discovered that copper could be shaped by hammering a fairly easy discovery simply involving hitting the copper with a hard object. Annealing was soon discovered as it involved heating the copper in a fire and then hammering it, a relatively easy discovery as fire had been known to humans for hundreds of thousands of years.

A more complex discovery was how to extract copper from its ores. This requires temperatures of around 700°C so that some form of furnace or oven is required. As this involves an extra and reasonably complex element (the building of furnaces) it makes sense that metallurgy involving smelted copper took place sometime after the use of native copper. The casting of copper in open molds requires a temperature of 1083°C which requires more complex furnaces and bellows to get the required temperature. This inevitably means that it occurred after the development of smelting and the use of native copper.

The use of casting-in and two-piece moulds inevitably followed the use of simpler casting with an open mould. More complex casting techniques could only be developed after simpler techniques had been mastered and had become well understood. The creation of bronze, an alloy of either copper and tin or copper and arsenic requires the ability to heat the metals to their melting points. This meant bronze could only be created after it was discovered how to produce heat of 1083°C, the melting point of copper which had the highest melting point of the three metals. Tins melting point is 232°C and arsenics is 818°C. To produce heat of 1083°C required sophisticated furnaces and bellows and then to acquire the knowledge that the alloy was stronger and harder than copper would have ensured that the development of bronze took place later than copper smelting and the more sophisticated copper casting techniques were developed.

The last step mentioned by Renfrew was lost wax casting. This is a quite sophisticated form of casting far less obvious than casting in or the use of two piece moulds so that lost wax casting was developed later than the other two techniques.

The development of iron metallurgy proceeded in a similar manner to that of copper. The first use of iron involved the use of meteorite iron which is also the simplest use of iron as no smelting, involving the use of complex kilns with bellows, was needed. When furnaces were built that could achieve temperatures capable of smelting iron, the iron age began and iron replaced bronze as the principal material for tools and weapons. Temperatures capable of melting iron were eventually produced when furnaces were improved, the most important development being the introduction of the blast furnace. This required the prior invention of the water wheel. The water wheel was invented in Roman times and was steadily improved with cams and cranks to convert its circular motion into reciprocating motion so it could be used for a wide variety of purposes including powering bellows. Once the water wheel was used to drive bellows, the new blast furnaces were able to reach temperatures that could melt iron and produce cast iron. Further improvements were made to blast furnaces such as the use of reverberatory furnaces and the preheating of air before it entered the furnace, which led to still higher temperatures being obtained. Advances in the study of chemistry lead to methods for the mass production of steel such as the Bessemer process and the open-hearth process. Even higher temperatures were produced by electric furnaces and the use of oxygen rather than air for steel making and for the production of other metals.

The progress of metallurgy was partly based on the ability to produce higher and higher temperatures to smelt and melt metals. The use of open fires to allow the hammering of heated metals, then of furnaces and of furnaces with bellows, then of furnaces with bellows driven by water wheels, then of reverberatory furnaces, then of the pre-heating of air before it enters the furnace, then of electric furnaces and of furnaces using oxygen rather than air lead to ever higher temperatures, which allowed a wider range of metals to be smelted and melted. These developments took place in a logical order in that the simplest ways of smelting and melting ores and metals were invented before the more complex ways. The gradual increase in temperatures available for metallurgy allowed metals to be smelted and melted in a particular order which was set by the particular properties of the ores and metals concerned. The particular properties were the smelting and melting points of those ores and metals so the order of development from the bronze age to the iron age was inevitable in human history. The Steel age inevitably occurred later than the Iron age as to produce cheap pig iron or wrought iron is a much easier process than to produce cheap steel with its requirements for relatively exact amounts of carbon to be mixed with the iron to produce steel.

Metallurgical processes that required prior inventions or discoveries were made after the prior inventions or discoveries. The discovery of how to melt iron (in Europe) was made only after the invention of the blast furnace, which was dependent upon the prior discovery of the water wheel and how to convert circular motion into reciprocal motion. The invention of the electric furnace was made only after the discovery of how to make, control and use electricity. The widespread use of aluminum occurred only after the invention of the electric furnace. The use of oxygen in metallurgy occurred only after the discovery of oxygen as a separate element and after it became possible to cheaply produce oxygen for industrial use. The whole development of metallurgy followed a logical process which was inevitable given the properties of the metals and ores available for human use.

Writing and record keeping

Writing was first invented by the Sumerians in ancient Mesopotamia before 3,000 BCE. It was also independently invented in Meso-America before 600 BCE and probably independently invented in China before 1,300 BCE. It may have been independently invented in Egypt around 3,000 BCE although given the geographical proximity between Egypt and Mesopotamia the Egyptians may have learnt writing from the Sumerians.

There are three basic types of writing systems. The written signs used by the writing system could represent either a whole word, a syllable or an individual sound. Where the written sign represents a word, the system is known as logographic as it uses logograms which are written signs that represent a word. The earliest writing systems such as the Sumerian cuneiform, Egyptian hieroglyphics and Mayan glyphs are predominantly logographic, as are modern Chinese and Japanese writing systems. Where the written sign represents a syllable, the writing system is known as syllabic. Syllabic writing systems were more common in the ancient world than they are today. The Linear A and B writing systems of Minoan Crete and Mycenaean

Greece are syllabic. The most common writing systems today are alphabetical. These involve the written sign (a letter) representing a single sound (known as a phoneme). The earliest known alphabetical systems were developed by speakers of Semitic languages around 1700 BCE in the area of modern day Israel and Palestine. All written languages predominantly use one or other of the above systems. They may however partly use the other systems. No written language is purely alphabetic, syllabic or logographic but may use elements from any or all systems.

In order for there to be writing three criteria must be met:

-the writing must consist of artificial graphical marks on a durable surface
-the purpose of the marks must be to communicate something
-the purpose must be achieved due to a conventional relationship between marks and language

Such fully developed writing only emerged after development from simpler systems. Talley sticks with notches on them to represent a number of sheep or to record a debt have been used in the past. Knotted strings have been used as a form of record keeping, particularly in the area around the Pacific rim. They reached their greatest development with the Inca quipus where they were used to record payment of tribute and to record commercial transactions. A specially trained group of quipu makers and readers managed the whole system. The use of pictures for the purpose of communication was used by native Americans and by the Ashanti and Ewe people in Africa. Pictures can show qualities and characteristics which cannot be shown by tally sticks and knot records. They do not, however, amount to writing as they do not bear a conventional relationship to language. Even so, the Gelb dictum (from its originator Ignace Gelb), that “At the basis of all writing stands the picture” has been widely accepted.

An alternative idea was that a system by which tokens, which represented objects like sheep, were placed in containers and the containers were marked on the outside indicating the number and type of tokens within the container gave rise to writing in Mesopotamia. The marks on the outside of the container were a direct symbolic representation of the tokens inside the container and an indirect symbolic representation of the object the token represented. The marks on the outside of the containers were graphically identical to some of the earliest pictograms used in Sumerian cuneiform. However, cuneiform has approximately 1,500 signs and the marks on the outside of the containers can only explain the origins of a few of those signs.

The first written language was the Sumerian cuneiform. Writing mainly consisted of records of numbers of sheep, goats and cattle and quantities of grain. Eventually clay tablets were used as a writing surface and were marked with a reed stylus to produce the writing. Thousands of such clay tablets have been found in the Sumerian city of Uruk. The earliest Sumerian writing consists of pictures of the objects mentioned such as sheep or cattle. Eventually the pictures became more abstract and were to consist of straight lines that looked like wedges.

The earliest cuneiform was an accounting system consisting of pictograms representing commodities such as sheep and a number. The clay tablets found might for example simply state “ten sheep”. Such writing obviously has its limitations and would not be regarded as a complete writing system. A complete writing system only developed with the process of phonetization. This occurs when the symbol ceases to represent an object and begins to represent a spoken sound, which in early cuneiform would be a word. This process was assisted when the symbols which initially looked very like the object they represented gradually became more abstract and less clearly related to an object. However while the symbol became more closely connected to words, it was words dealing with objects, such as sheep, bird or pot. It was still not possible to write more abstract ideas such as father, running, speech or foreigner.

The solution to this problem was known as the rebus principle. Words with the same or similar pronunciation to an abstract word could be used to represent the abstract word. The sign for eye could be used to represent the word “I”. The sign for deer could represent the word “dear”. Which word is referred to by the picture is decided by an additional sign. Pictographs which originally represented a word began to represent the sound of the word. The rebus principle is used to represent abstract words in all word writing systems in Sumer, Egypt, China and in the Aztec and Mayan writing in central America.

The Rebus principle led to cuneiform becoming a form of logo-syllabic writing consisting of both logograms and syllabic writing. The effect of the change from logographic to logo-syllabic writing was substantial. Logographic writing cannot produce normal prose and is restricted to nouns, numbers, names and adjectives.. The vast majority of early Sumerian writing consisted of bureaucratic records of products received or products distributed. Only when syllabic writing was introduced into cuneiform did it become possible to write prose such as myths and royal propaganda.

The next major development in writing in the old world was the development of the alphabet. The alphabet was developed out of Egyptian hieroglyphs which contained 24 signs for 24 Egyptian consonants. About 1700 BCE Semites who knew Egyptian hieroglyphs began making certain changes in their writing system. They put the letters in a particular sequence and gave them simple names to assist learning and ease of memory. They also dropped the logograms and other signs used in hieroglyphs and just kept the Egyptian consonants and restricted the signs to those for individual consonants. Finally they introduced vowels into their alphabet. Alphabets were soon to spread over most of the world as they provided both flexibility and simplicity for a writing system.

Writing also developed independently in Meso-America about 650 BCE. The earliest Meso-American writing seems to be logographic writing produced by the Olmecs. Later the Maya developed a writing system that was based partly on logographics and was partly syllabic. There are broad similarities between the Maya writing and the logo-syllabic writing systems used in the old world. Mayan syllabic signs are pictographs of objects whose pronunciation begins with that syllable, which is the same system used in early Semitic alphabets. The rebus principle is used for logograms for abstract words as it is used in Sumerian cuneiform. Mayan syllabic signs were commonly signs for syllables of a single constant and one vowel as in the Linear B writing system of Mycenaean Greece. Similar problems which emerged while developing a writing system were solved in similar ways in both the old world and the new world.

Writing developed both in the old and the new worlds as a movement from the simple to the complex in the form of increasing abstraction. The simplest way to make a record of something is to draw a picture of it, so that the earliest writing was logographic. However the limitations of logographic writing were to result in the development of syllabic writing to allow for the writing of prose and for a reduction in the number of signs used. The most difficult system to invent due to its high level of abstraction was the alphabet system which was why it was the last system invented in the old world and why it had not been invented in the new world by the time of the Spanish conquest. Coulmas in *The Writing Systems of the World* states:

"The general tendency of development is roughly from pictogram to alphabet via word writing first and then syllabic writing."[30].

Later he states:

"Syllables are clearly more abstract and more difficult to conceive of than words, and accordingly syllabic writing appears historically later than word writing."[31].

Still later he states:

"The alphabet is the logical conclusion of a development of ever increasing abstraction. As its units are minute and highly abstract it is in principle, universally applicable. … It is a generally accepted view that the alphabet is the teleological goal of the history of writing."[32].

The move from logographic, to syllabic, to alphabetical writing was a move from the easiest form of writing to invent to the most difficult form of writing to invent. The increasing difficulty is caused by the increasing level of abstraction with the move from logographic, to syllabic and then to alphabetic writing. However, as the level of abstraction increases, the ease of use of the writing system increases. This is because the number of signs used falls, with logographic systems typically employing thousands of signs, syllabic systems anywhere from 50 to many hundreds and an alphabetic system like the Roman alphabet, 26 signs. This is because the number of words in a language always exceeds the number of syllables and the number of syllables will normally exceed the number of phoneme, upon which alphabetic writing is based.

If writing, or a similar record keeping system like the Inca quipu, had not been invented, then it is doubtful whether states as a form of government could have existed. States such as the

Aztec, Inca or the Mesopotamian and Egyptian civilizations require a bureaucracy and record keeping system to account for the payment of taxation (in goods rather than money) and the distribution of produce to government bureaucrats, soldiers and supporters. If writing did not exist it would be difficult or impossible to run or control any political entity larger than a chiefdom. Somewhere as societies got bigger and bigger, writing or a similar record keeping system was needed to control the administration of the government of that society.

Glass

The earliest known glass making was in ancient Egypt around 3,000 BCE although Mesopotamia followed quickly thereafter. The basic requirements for making glass are silica (sand), an alkaline substance such as soda and lime (calcium carbonate). The earliest and simplest technique for glass manufacture was baked glazing. Baked glazing involved melting the silica, soda and lime and then coloring the product which was then poured into molds to produce small statues or jewelry. It was later discovered that a bottle or vase could be made by placing the molten glass around a core that could be removed when the glass cooled.

In the first century CE glass blowing was developed, probably in the area of present-day Syria. Glass blowing involved the molten glass being placed at the end of a metal pipe and a person blowing through the pipe which blew the glass up to the desired size and then the glass was shaped and decorated. The development of glass blowing opened up new technical and artistic possibilities for glass makers. The quality of the glass improved as its texture became more refined and it became more transparent and colorless. New molding techniques developed with the glass being blown into the molds. Glass began to be used to make plates, pitchers and vases, objects which had previously been made mainly of metal or clay. The Roman author Pliny referred to the use of manganese oxide to rid glass of impurities so as to ensure that it was transparent.

Chemically glass is an amorphous noncrystalline solid. This means the atoms are arranged in a random rather than a regular pattern. This accounts for the optical qualities of glass such as transparency and causes the glass not to have a definite melting point. When heated the silicon in glass causes it to go into a state where it is a soft solid or viscous liquid. This state exists over a wide temperature range and explains why glass is able to be molded and blown into a wide variety of shapes.

If it had not been possible to invent glass, for example if no combination of materials could produce a solid transparent substance, then the effect on society would have been considerable. Eye glasses to correct defective vision would have been impossible and the microscope and the telescope may not have been invented. This would have meant our discovery of the world of microorganisms may not have happened or may have been delayed until the 20th century and the development of other transparent products such as plastics. Progress in medical research and the discovery of disease would have been delayed or never happened. It would also have meant that our discovery of the universe from new planets to new galaxies would have been delayed until other transparent materials could have been developed. The earth-centered astronomies of the pre–telescope age would have continued to be believed at least until some substitute for glass had been developed and could be used to create telescopes.

Astronomy

Virtually every culture has theories concerning the movement of the earth, of the sun and of the planets. Before the beginnings of modern science nearly all such theories were geocentric in that they put a stationary earth at the center of the universe with the sun, moon and planets moving around the earth. Given that the earth people stand on appears to be fixed and unmoving, and that the sun moves across the sky, it seems obvious that the earth was not moving and that the sun and other heavenly bodies revolve around the earth.

The first people who left us with a systematic and detailed description of the behavior of the heavenly bodies were the ancient Greeks. The system was called the Ptolemaic system after the Greek astronomer Ptolemy who lived around 200 CE. The Ptolemaic system had the earth at the center of the universe with the moon, sun, and planets orbiting the earth in circular orbits. Only five planets, not counting the earth, were known to the ancient Greeks and they were believed to orbit the earth in a special motion known as epicycles. Epicycles were circles which planets were considered to move in when orbiting the earth, in circular orbits called deferents. The Ptolemaic system required epicycles to explain the irregular movement of the planets which from observations made from earth, which was moving around the sun, the other planets appeared to periodically change direction. The ancient Greeks considered circles to be a perfect shape and that as everything in the heavens was perfect the heavenly bodies must move in circular orbits.

The Ptolemaic system with an unmoving earth at the center of the universe seemed superior to the idea of the earth and other planets orbiting the sun as:

-common sense-suggested everything moved around the earth while the idea of the earth and planets orbiting the sun is contrary to common sense
-awareness of motion-we are not aware of any motion so the earth appears not to be moving
-falling objects-objects thrown in the air drop straight down-if the earth was moving they would fall to the ground in a different place
-stellar parallax-if the earth was moving, at different times the stars would appear to be in different positions relative to each other-such changes were not observable to the ancient Greeks

(due to insufficient instruments and the great distance of the stars from earth) so it was assumed the earth was not moving
-planetary orbits-with deferents and epicycles the Ptolemaic system was more accurate than simple circular orbits around the sun

The evidence available to the ancient Greeks was very much in favor of the Ptolemaic system with the earth at the center of the universe. Not surprisingly the suggestion that the earth orbited the sun and spun on its axis was not accepted as the evidence was very much against that system.

The Ptolemaic system was challenged in the 16th century by the Polish astronomer and mathematician Copernicus. Copernicus proposed that the sun was at the center of the universe and the earth and planets orbited the sun in circular orbits. The apparent movement of the sun and the stars and some of the planets was caused by the movement of the earth, both spinning on its axis and orbiting the sun, rather than the movement of the sun stars and planets. Copernicus still had to use epicycles to describe the planet's orbits. Copernicus's system did not really get adopted as it did not fit the observation any better than the already well-established Ptolemaic system. Kepler, using observations made by the Danish astronomer Tycho Brathe, came to the conclusion that the earth and planets orbited the sun in elliptical orbits with the sun at one foci of the ellipse. This enabled Kepler to get rid of the epicycles so as to create a simpler system. Kepler's system worked, it fitted well with Tycho Brathe's observations and also with observations that began to be made using the telescope which was invented about 1600.

It was the invention of the telescope that allowed Kepler's system to be confirmed. All astronomers were able to study the sun, planets and stars with the telescope and observations were made which matched Kepler's theory. Galileo was the first to make extensive observations of the cosmos with a telescope. He found the moon was heavily cratered, whereas previously it had been believed it was a smooth, perfect sphere. He also discovered the moons of Jupiter showing not everything in the cosmos orbited the earth. Galileo also made observations of the phases of Venus which were consistent with the Copernican theory and were inconsistent with the Ptolemaic theory. The new system which was to replace the Ptolemaic system was completed by Sir Isaac Newton and is normally known as the Newtonian system. Newton's invention of the law of gravity and his three laws of motion described why the planets and the earth moved as they did. In particular the first law of motion that a body in motion will continue in motion unless a force acts upon it describes why the earth keeps turning on its axis and why it and the planets keep orbiting the sun. Gravity explains what holds the earth and planets in orbit around the sun and stops them heading off into outer space while the earth and planets motion stops gravity causing the earth and planets to be sucked into the sun. Gravity also explained the old problem of why objects not attached to the earth, such as clouds and objects thrown into the air moved with the earth. They were held in place by the earth's gravity. Newton's system was dependent upon the prior invention of calculus, independently invented by both Newton and

Leibnitz, which enabled accurate calculations to be made of moving objects such as planets, when the starting position was known.

The Newtonian system was confirmed when Uranus was found not to be behaving in accordance with the Newtonian system which lead to the discovery of Neptune. The telescope had led to the discovery of Uranus, in 1781 by William Herscel, the first planet discovered since ancient times. Observations of Uranus made it clear the planet was not behaving in the way predicted by Newton's laws. Calculations were made independently by both Urbain Leverrier and John Adams that suggested the gravitational effect of another planet could explain the behavior of Uranus. This led to the discovery of Neptune in 1846, exactly where the calculations predicted. When the gravitational force of Neptune was taken into account the whole solar system corresponded exactly to Newton's theory.

In the 19th century the development of spectroscopy which enables the identification of elements from the light they reflect revealed that the sun had the same chemical makeup as the stars and that the sun was just another star. It only appears to be different to people due to the sun being so much closer to us than the other stars.

Newton's system had been well established for several hundred years when it was discovered, with new and more accurate measurements that the planet Mercury did not move in accordance with Newton's calculations. In 1905 Einstein published his special theory of relativity which described the universe but did not involve gravity. It dealt mainly with light and how the universe would look to observers traveling close to the speed of light. The special theory of relativity suggested the space occupied by an object would shrink in line with the movement of the object relative to the observer. It also predicted an increase in the mass of a moving object as its speed approaches the speed of light and time dilation, which involves clocks that are moving running slower than stationary clocks. It also states that no object moving less than the speed of light could accelerate past the speed of light. The special theory states that all motion is relative to the observer and that an observer not in a state of acceleration (i.e. in a constant velocity) can consider themselves at rest and are able to measure motion relative to themselves. The speed of light however was the same for all observers. The special theory of relativity has been confirmed in many experiments for example involving particles accelerated to close to the speed of light in particle accelerators. Special relativity was developed further when Herman Minkowski suggested special relativity meant space and time should be combined and understood as a four dimensional geometry of space-time.

The general theory of relativity was published in 1916 and was a development from special relativity in that it dealt with accelerations, while special relativity applied only to objects moving at constant speeds and in straight lines. General relativity states there is no difference between acceleration and gravity. Einstein provided a set of equations that shows that gravity and acceleration were equal to each other. The mathematics were based on non-Euclidean geometry which dealt with the geometry of curved surfaces and had only been developed in the 19th century. The general theory concerns the interaction of four dimensional space-time and matter.

It states that matter produces a curvature in space-time and that all objects including electro-magnetic radiation move in accordance with the curvature of space-time. The curvature of space-time is caused by the presence of massive objects such as the sun. The planets traveling around the sun move in accordance with the curvature of space-time caused by the sun. Due to this general relativity considers the planets move in circular orbits in four dimensional curved space-time rather than in elliptical orbits in three dimensional space as was assumed by the Newtonian system. Gravity according to general relativity is the curvature of space-time, a quite different concept from Newton's idea of gravity. Newton's concept of gravity was that everything with mass exerts gravity and the gravitational force depends on the amount of mass an object has and its distance from another object.

The calculations involved with general relativity exactly explained the orbit of Mercury which was something that the Newtonian system could not explain. General relativity has been confirmed by some other experiments and observations and is now the generally accepted explanation of the universe. One problem with general relativity was that it assumed the universe must be either contracting or expanding at a time when Einstein and others believed the universe was stable. To solve this problem Einstein invented a mathematical device called the cosmological constant. However in the 1920's Edwin Hubble discovered the universe was expanding, the galaxies were moving further and further apart and Einstein was able to drop the cosmological constant. Eventually the expanding universe led to the big bang theory that all the matter in the universe once occupied a single point in space which then exploded scattering matter over the entire universe.

The changes in the view of the universe from the earth centered Ptolemaic system, to the sun centered Newtonian system to General Relativity were inevitable. Every pre-scientific society believes the earth is stationary and the sun moves around it. It is the obvious and common sense view of the universe to people who could only view the universe with the naked eye. The Ancient Greeks, who had a highly sophisticated knowledge of geometry developed with the Ptolemaic system a comprehensive and heavily geometric explanation of the universe that was to survive nearly 1,500 years. It was only to be displaced when a new means of observation, the telescope, became available to produce new information about the universe that was inconsistent with the Ptolemaic system. The sun centered system had been previously suggested by the Ancient Greek astronomer Aristarchus but had not been accepted as the evidence before the invention of the telescope supported the earth centered view of the universe. When Copernicus proposed a sun centered universe and Kepler, using Tycho Bathe's observations, improved it with the elliptical orbits, the sun centered system was still not accepted. This is because before the invention of the telescope, there was no way of confirming the correctness of the sun centered system. Kepler's system was based on accurate observations, but they were still only one mans observations and until the invention of the telescope there was no way of confirming their accuracy. Without the invention of the telescope Copernicus and Kepler's system would almost certainly have been rejected in the same way as Aristarchus' system had been rejected. It was only when the telescope was used to see:

-the phases of Venus which were consistent with the sun centered system and inconsistent with the earth centered system
-the moons of Jupiter which were inconsistent with the earth centered system as everything was meant to circle the earth
-imperfections on the sun and moon, as the Ptolemaic system considered everything in the heavens to be perfect

and Newton’s law of motion and gravity which explained:

-why objects thrown in the air dropped straight down
-why the planets continued in orbit around the sun

that it became rational to believe in the sun centered system. The invention of the telescope was dependent upon prior discoveries of the refraction of light as it moves through different mediums, particularly glass and how to manufacture glass of sufficient transparency. The invention of calculus also played a vital role in the establishment of the Newtonian system. The telescope was to lead to the discovery of Uranus and a combination of telescope and calculus was to lead to the discovery of Neptune.

Observations inconsistent with the Ptolemaic system caused the abandonment of that system and observations consistent with the Newtonian system led to the adoption of that system. The same process can be seen with the adoption of general relativity. The orbit of Mercury was inconsistent with the Newtonian system but was perfectly explained by general relativity. General relativity was also confirmed when it was discovered that during a total solar eclipse light passing close to the sun was deflected by the sun's gravitational field to an amount exactly as predicted by general relativity. The initial results of observations of a total solar eclipse in 1919 were somewhat ambiguous but were confirmed by subsequent eclipses. New observations were made either as the results of new instruments such as the telescope or improved instruments revealed Mercury’s orbit failing to conform to the predictions of the Newtonian system. New mathematical systems such as calculus and non-Euclidean geometry led to a greater understanding of the universe. As new knowledge became available it would either support the existing explanation of the universe or conflict with that explanation. When new knowledge conflicted with the existing explanation of the universe it would lead to the development of new theories to explain the universe. It was the ever increasing knowledge of the universe that led to the abandonment of the Ptolemaic system and its replacement by the Newtonian system and the abandonment of the Newtonian system and its replacement by General Relativity. Each of the three systems was an historical inevitability as the earth centered

Ptolemaic system was the common sense and obvious system for societies dependent upon naked eye observation. The Newtonian system with the sun at the center of the universe was the inevitable replacement for the Ptolemaic system given the accuracy and power of 17th, 18th, and 19th century telescopes. It was not until the 19th century that it was realized that Mercury did not orbit the sun in accordance with Newtonian theory. This was because the variation from the expected Newtonian orbit was extremely small and it was not until the 19th century that there were telescopes capable of discovering the variation. If the variation had been significantly larger then the Newtonian system may well have been replaced by general relativity somewhat earlier and with someone other than Albert Einstein being the inventor of the new system. If telescopes, when invented, were much more powerful and more able to show Mercury and other planets did not orbit the sun as predicted by Newton then the Newtonian system may not have existed at all. Possibly a system such as general relativity may have directly replaced the Ptolemaic system or maybe there would have been a period of confusion at least until the invention of non-Euclidean geometry. It is the structure of nature that determines the power of telescopes when invented and their rate of improvement is affected by the structure of nature and the social institutions and culture of a society.

Microscopes and Telescopes

The transparency of glass combined with the way in which light changes direction as it passes through one transparent medium to another, a phenomenon known as refraction, allows glass to be used for the purposes of magnification. When two (or with poorer results, one) lenses are used, an object seen through those lenses is magnified. This is because the shape of the lenses causes light going through the lenses to converge at a particular focus or focal point in accordance with the laws of refraction. This focal point is different from the normal human focal point and allows the object to be magnified without blurring. The apparent size of an object increases as it is brought closer to the eye but if it is brought too close, blurring occurs. The blurring occurs because the lens in our eye cannot bend (or refract) light from an object enough to bring it into proper focus on the retina if it is too close. The lenses magnify by starting the refraction or bending process before the light enters the eye. This enables objects closer than the usual human focal point to be examined without losing focus. This was to result in the invention of glasses to correct bad vision and in the invention of the telescope and the microscope.

The particular shape of the lenses used in microscopes can be worked out by using the law of refraction (known as Snell's law after Willebrord Snellius (1580-1626)) and trigonometry which was being developed by the mathematician Rheticus (1514-76). An index of refraction establishes the angle at which light bends when going from one medium to another. When light passes from air through glass the refractive index is approximately 1.52. Armed with this knowledge it is possible to manufacture both microscopes and telescopes. Alternatively the earliest microscopes and telescopes may have been developed simply by experiment and observation.

The earliest lenses produced were eye glasses to correct defective vision and these were first developed in 13th century Italy. They were clearly developed from experimentation and observation without the benefit of Snell's law or trigonometry. The earliest microscopes were invented by the Dutch spectacle makers Hans and Zacharus Janssen about 1590.

The social and cultural consequences of the invention of the microscope, was the discovery of whole new worlds. An immense variety of microorganisms were discovered, the leading microscopist being Antoni van Leeuwenhoek (1632-1723). Van Leeuwenhoek discovered protozoa in water, bacteria, blood corpuscles, capillaries, striations in skeletal muscle, the structure of nerves and spermatozoa. The microscope was soon to destroy the idea of spontaneous generation which held that many animals arose from spontaneous generation from particular environments. Mosquitoes came from stagnant water, bees from the carcasses of oxen and cattle, shellfish from mud and slime and snails from the purification of fallen leaves. The work of van Leeuwenhoek and others showing the life style and sexual apparatus of such animals showed the idea of spontaneous generation was wrong, although it was not until the 19th century with the work of Louis Pasteur that the idea was finally put to an end.

The 17th and 18th century microscope had its thresholds so that while it could reveal certain things previously unknown, there was much it could not reveal. This led to some theories that would not last due to more powerful microscopes proving them wrong. One such theory was that spermatozoa was the essential instrument of reproduction which fitted in with the belief in the pre-existence of organisms. Each new organism was considered to contain all the characteristics of all its predecessors. The spermatozoon was considered to be the means of transmission of all those characteristics to the new organism. This idea however failed to understand the role of the spermatozoon in fertilizing the egg and the contribution of the egg to the characteristics of the new organism. The idea however, based on the information available to 17th and 18th century scientists, was reasonable enough for the times. It was not until the 19th century when improved microscopes showed the spermatozoon and the egg contributed equally to the characteristics of the new organism.

The telescope appears to have been invented by Hans Lipperhey, a spectacle maker in the Dutch town of Middelburg, who applied for a patent for it in 1608. Two other Dutchmen, Jacob Adriaenzoon and Sacharias Janssen also claimed to have invented the telescope, so a patent was refused. The Italian scientist Galileo heard about the Dutch invention and constructed his own telescope achieving a magnification of 20x, a better magnification than was to be achieved until 1630. Galileo's telescope had two lenses, an objective lense at one end of the telescope and an ocular lense at the other end to which the eye was applied. The objective lense was a convergent or biconvex lense while the ocular lense was a divergent or biconcave lense. The effect of light passing through the lenses was to change the focal point of the light providing for a wider visual angle in which to view the object under observation. The telescope while operating a bit differently from a microscope, like the microscope, magnifies images through manipulating the focal point of light to create a wider visual angle in accordance with the laws of refraction.

Galileo, having created his telescope, used it to look at the sky. He discovered a large number of previously unseen stars (the Milky Way), that the moon had an irregular surface, the sun was spotty and impure (sunspots), that Jupiter had four moons, there were rings around Saturn and the moon like phases of Venus. The observations were contrary to the astronomical theories of Ptolemy which had largely been accepted from classical times. Galileo's observations plus those of other scientists using even better telescopes were to result in the ending of the Ptolemaic astronomy and its eventual replacement with the Newtonian system.

Prior to the invention of the telescope six planets (the earth itself, Mercury, Venus, Mars, Jupiter and Saturn) were known to human beings and less than 5,000 stars were visible to the naked eye. The telescope led to the discovery of Uranus in the 18th century, Neptune in the 19th century and Pluto in the 20th century. The invention of photography assisted the telescope in revealing the universe as it allowed objects too dim to be seen through a telescope to be photographed on a photographic plate over a long exposure time. The long exposure time allowed the photographic plate to record the existence of very faint objects as the plate will accumulate the effect of each photon hitting the plate over a period of time.

By the start of the 20th century it had become clear that our solar system was part of the Milky Way but it was not clear whether the Milky Way was the whole universe. It was not until the 1920's when Edwin Hubble conclusively showed there were other galaxies and these galaxies were moving away from us with the furthest galaxies moving the fastest.

New forms of telescopes which detected different forms of electro-magnetic energy were developed. However, most electro-magnetic energy, other than visible light and radio waves, is blocked by the earth's atmosphere. The development of space rockets led to telescopes being placed in space, particularly the Hubble space telescope to allow detection of electro-magnetic radiation in frequencies other than those of visible light and radio waves. Telescopes using frequencies other than those of visible light have detected radio wave evidence of planets in other solar systems, x-ray evidence of black holes, radio-wave evidence of supernova explosions, and gamma ray and x–ray evidence of gamma rays originating from deepest space. Dark matter that could not be detected by any telescope operating on any electro-magnetic wavelength was detected due to its gravitational effect on matter that was visible to telescopes.

Printing

Printing in its simplest forms has existed for thousands of years, in the form of signet rings, royal seals and punches used by gold and silversmiths. The Phaistos disk, dating from around 1700 BCE and containing writing in an unknown language was discovered in Crete in 1908. The disk, made of hard clay, had writing punched into it by at least 45 stamps making it the world's earliest printed document.

The Chinese have been using seals for stamping documents since the 13th century BCE and produced the earliest known printed book in the 9th century CE. The book was in the form of a scroll and was made by letterpress printing which involved printing from raised letters. A passage of text was carved out of a block of wood and the raised areas of the wood were coated with ink and paper was pressed on the inked wood by hand. Movable type was invented by the Chinese in the 11th century CE when baked clay characters, each representing a word in Chinese script, were organized into sentences and pages. The Chinese script however has thousands of signs so that while a stock of clay characters representing common words could be kept, unusual words were made for each book. The alphabetic scripts used in Europe were far more suitable for printing with movable type as only a few dozen letters and signs had to be made for the movable type. Metal type, which had a longer life than clay type, began to be made in Korea in the early 15th century. However the same problem the Chinese had of a huge number of characters being needed as each character represented a word limited the use of printing. A new phonetic alphabet taken from the Sanskrit language of ancient India was then developed so that words could be made up from a small number of letters, allowing a small quantity of type representing the letters to be used again and again in different printing jobs.

The invention of modern printing by Johann Gutenberg did not take place until the 15th century in Europe. It involved the combination of six technological advances being the use of paper, inks, presses, movable type, metallurgy and alphabetic scripts. Paper was invented in China towards the end of the first century CE and eventually spread to Europe via the Middle East and North Africa reaching Europe in the 12th century and Germany in the 14th century. Paper was necessary for the development of modern printing as the alternatives such as parchment are difficult to handle and costly while papyrus is hard and brittle and unsuitable for printing. The inks used were oil-based and developed from inks recently used in painting. The press used by Guttenberg was probably derived from presses used in agriculture and industry for pressing products such as grapes, olives, oil seeds and herbs or papyrus. The press had a large wooden screw attached to a flat wooden “platen” which pressed the paper onto the inked type.

The movable type invented by Guttenberg consisted of stamps for each letter of the alphabet and punctuation marks and other symbols. To create a page of type Guttenberg selected the letters to make the words, placed them in a frame and clamped them together. The typeface was then inked and a sheet of paper was pressed against the typeface to produce the page.

It was likely that Guttenberg’s biggest problem was how to produce identical sized type. If type was not of the same size, problems such as some letters not printing and lines not being straight and the type not being able to be properly clamped together would arise. Type also needed to be reasonably durable to ensure printing remained cost effective and the need for consistent size and durability meant the type had to be made of metal. It was discovered that type of a consistent size could be produced if all the type were cast in the same mold. At the bottom of the mold would be a piece of metal such as copper, with an imprint of the letter to be cast. This piece of metal was known as the matrix and there would be a matrix for each letter of the alphabet. The type would be made of an alloy of lead, tin and zinc and after sufficient type was cast by pouring molten metal into the mold, the matrix could be replaced with the matrix of another letter to produce type of that letter.

The use of metal for making type and the molds within which the type was made required sophisticated metallurgy. The correct metals for making the type and molds had to be developed, the metals often being alloys for which considerable experimentation was required to work out which metals to use and the proportion of each metal.

The last essential element for the introduction of printing was an alphabetic script. This had been present for some time in Europe but its absence in China had certainly made printing by movable type impractical in China.

Gutenberg printed his earliest books around 1450 and his 42 line Latin Bible was printed around 1453. In the fifty years after Guttenberg's invention, improvements were made providing for color printing, the use of new typefaces such as Roman and Italic and woodblock printing was combined with movable type printing to produce books with pictures. Guttenberg's original printing process was quite slow with about 16 copies being produced per hour. In 1620 Willem Blaeu added a counterweight to the press which helped raise the platen and increased the speed of printing to 150 copies per hour. In 1642 the mezzotint process for printing graduated tones was patented by Jakob Le Blon. Stereotyping, which made it possible to create a copy of a page of type using a mold, was invented in 1727. This made the mass production of printing practicable as it was no longer necessary to reset the type of a page that was to be reprinted in the future or on another press. Molds were initially made of plaster of Paris and then of clay and then of *papier mache*. Lithography, a means of printing from a flat surface, which was faster than printing from a raised surface, was invented in 1798. The iron framed press invented in 1800, by increasing the force the platen applied to the paper, allowed bigger sheets and more pages to be printed at a time so that 250 pages could be produced per hour. The quality of the printed work increased due to the evenness of the impression on the pages. Steam power was applied to presses in 1810 by Frederick Konig and in 1812 he produced a press in which the form which contained the type and the paper moved below a roller pressing the paper to the type. This enabled copies to be produced at a rate of 1,100 per hour. Electrotyping was invented by four independent inventors from Great Britain, Russia and the United States and involved the making of metal copies of woodcut printing blocks. These copies could withstand the force of the steam press and could be made of complete pages and give a higher quality printed page than stereotypes. The rotary press was invented in the 1840's allowing printing of 8,000 copies per hour and the use of multiple columns. In the second half of the 19th century the printing of photographs was developed by putting a photographic image of the photograph onto a copper plate which is then etched with acid to create recesses on the copper plate which are filled with ink and then the plate and ink are applied to the paper. New methods of typesetting were invented in the late 19th century and in the 20th century when the linotype and monotype machines were invented and allowed much faster and higher quality typesetting. Photo composition and computer typesetting were introduced in the 20th century.

Printing was to provide the world with the first form of mass communication. Before the development of printing, each book had to be written out by hand. It would take months to copy a single book. A New Testament in the 14th century would take 6 months to copy and books would sometimes contain errors. European libraries before printing would often contain no more than 500 books.

The invention of printing was to result in a vast increase in the number of books available and a great reduction in their price. It has been estimated that in the fifty years after the invention of printing 15 to 20 million books consisting of at least 35,000 editions had been printed. 77% of these were in Latin, 45% were religious books and 236 cities were involved in their production. The Bible and works of classical literature by writers such as Cicero, Virgil and St Augustine were produced in large numbers. Eventually more and more books were printed in the native languages of Europe. This resulted in those languages and their spelling becoming more standardized. Scientific works by Copernicus, Kepler, Galileo, Bacon and Newton spread the scientific revolution of the 16th century throughout Europe with a speed that would have been quite impossible before the development of printing. Galileo's work got him into serious trouble with the church and printing was to play a significant role in the religious and political changes in modern Europe.

The ideas of the Reformation spread rapidly throughout Europe as works by Luther, Calvin and others were printed and rapidly became widespread in northern Europe. In an attempt to control dangerous ideas the Church produced its *Index of Prohibited Books* in 1559 which continued until 1966. In England printers could not operate without Royal approval and the Court of the Star Chamber could fine and imprison printers. Such controls did not last and in 1685 laws to control printing in England were abolished. Printing was to have major political effects with the spread of the works of the philosophes such as Voltaire, Montesquieu and Rousseau before the French Revolution. The publication of newspapers, pamphlets and leaflets during the American and French Revolutions such as *Common Sense* by Tom Paine and *What is the Third Estate?* by the Abbe Sieyes played significant roles during those revolutions.

The development of universal education in 19th century Europe meant a great demand for school books while the literacy that resulted from the education meant a much greater demand for books of all kinds. The first newspapers began in the 17th century and became more common as the cost of printing fell due to technological improvements. Advertising became a substantial part of newspaper content and was to help keep the cost of newspapers down.

The massive expansion in the production of books after the invention of printing shows the need within Europe for a cheap and efficient means of disseminating information. However, the need could not be met until the six requirements to make the invention work came together.

These requirements were paper, ink, a press, movable type, alphabetic script and sophisticated metallurgy. The metallurgy required for producing a constant size type involved a tin, zinc and lead alloy for the type, brass or bronze alloys for dies, steel for letter punches and lead for molds. Such sophisticated metallurgy only developed in the years before Guttenberg invented printing, while paper only reached Europe from China in late medieval times and the inks required for printing were also developed only in the period immediately before the invention of printing. Only the alphabetic script and presses had been available in Europe long before the invention of printing. As a result it would not have been possible to invent printing much before it was actually invented. Printing with movable type could not, for example, have been invented in Roman times. It is quite possible that printing with movable type could have been developed in China, but for the lack of an alphabetic script. An alphabetic script could have been developed in China, as it was in Korea, but for the conservative nature of the Chinese government and society under the Ming and Manchu dynasties. Given the requirements for printing with movable type, it could only have been developed when and where it did develop.

Once the macro-invention of printing had been made it was followed by a series of micro-inventions that improved the efficiency and lowered the cost of printing. These micro-inventions followed the invention of printing in a more or less logical order. Relatively simple processes like color printing, new typefaces and producing books with pictures were quickly added to the technology of printing. More complicated processes such as the mezzotint process used for producing graduated tones were developed later. Improvements, such as the steam powered press, could only be introduced after the invention of the steam engine. The printing of photographs could only develop after the invention of photography and computer typesetting could only be introduced after the invention of computers.

The Discovery of Steam Power

The earliest human knowledge of the power of steam comes from the classical world when Heron of Alexandria described various machines using steam for such purposes as opening temple doors or to blow a horn. The engines were used to amuse or astonish rather than for practical or economic purposes.[33] Claims have been made that the presence of slavery in Roman world ensured that the steam engine was not used in industry as slaves were a cheaper manual source of power. This argument can hardly be confirmed as we know little of the price of slaves and how much they cost to keep in classical times so we can hardly say that slavery made it uneconomic to develop an industrial steam engine. It may well have been that for long periods in the classical world slaves may have been expensive and a steam engine may well have been more economic than slaves especially for difficult jobs such as getting water out of mines. In any event it was impossible for the Romans to calculate the costs both for the development and manufacturing of a steam engine until they had actually produced one. The most likely reason the Romans never developed a steam engine was that the materials available were not strong enough or finely worked enough to allow an industrial steam engine and their lack of understanding of the principles of vacuums, atmospheric pressure and the properties of gas such as steam.

In the modern period from the Renaissance onwards the earliest attempts to harness steam power were toys or perhaps laboratory experiments similar to those Heron described in classical times. Such devices were produced by Giambattista della Porta (1536-1605) in Naples and Salomon de Caus (1576-1626) in England.[34] It is not known whether either of these men knew of the classical steam engines described by Heron. The first sign of any attempt to use steam power for industrial purposes were patents taken out in 1631 by a prolific patentee David Ramsay "To raise water from lowe pitts by fire", "To make any sort of mills to goe on standing waters by continual moc'on without the helpe of windes, waite of horse", "To make boates, shippes and barges to goe against the wind and tyde."[35] (sic) All these aims were eventually to be achieved by steam power, but only the first of them represented a pressing social need of the times. This was the problem that water was getting into mines and making the mining operations difficult or impossible. The extent of the problem can be seen in that of 182 patents granted between 1561-1642 one in seven was for the raising of water.[36] The details of Ramsay's patents are unknown and there is no evidence any machines were actually produced. The first attempt to actually produce a large scale machine was made by the Marquis of Worcester in the 1660's but there is some doubt as to whether it was a genuine steam engine and in any event it was not a practical success.[37] Worcester was followed by Sir Samuel Morland who described a steam engine in a book he wrote which may or may not have been the same machine mentioned in the diaries of a Roger North. Parliament seems to have been supportive of these inventors granting both Worcester and Morland a patent for their inventions, such as they were.

While this was happening in England Evangelista Torricelli of Faenza (1608-1647), Blaise Pascal (1623-1662) and Otto von Guericke (1602-1686) engaged in a series of scientific experiments which showed the effects of atmospheric pressure and that if a vacuum could be created the weight of the atmosphere could be a useful source for the transmission of power. A further discovery relevant to the development of steam power was Boyle's Law which states the volume of a given mass of gas varies inversely with its pressure when its temperature remains constant. It is the pressure from the steam which lifts the piston, in post Newcomen steam engines, and as the volume of the steam increases in the cylinder as the piston rises its pressure falls allowing atmospheric pressure to force the piston back down. As the piston falls the steam pressure in the cylinder increases, giving the steam its "spring" which then forces the piston back up. The Newcomen engine worked by a weight attached to a beam which was attached to the piston and which caused the piston to rise. The piston would then be forced down when a vacuum was created in the cylinder under the piston, which would cause the piston to fall due to atmospheric pressure above the piston being greater than the pressure below the piston. A knowledge of Boyle's Law, how to create vacuums and the effects of atmospheric pressure, were crucial to the development of the steam engine.

Denis Papin (1647-1712), a French Huguenot refugee from Louis XIV's France, while in London working for the Royal Society, produced the first working model of a steam engine operated by atmospheric pressure. He placed water in a cylinder and lit a fire under the cylinder. The steam in the cylinder caused a piston to raise to the top of the cylinder and drive the air out of the cylinder. The fire was then removed, the steam condensed and a vacuum was created within the cylinder and the piston was driven down into the vacuum causing a weight attached to the piston to rise.

The first to come up with a practical working, although rather limited, steam engine, was Thomas Savery. He was from Devon, a fellow of the Royal Society, and was granted a patent for "rising water by the impellant force of fire". Savery's engine worked by steam, alternatively entering two chambers and forcing water out. The steam is then condensed to create a vacuum which then draws more water into the chamber which is again forced out by the steam entering the chambers. Savery produced a practical steam pump capable of continuous operation but with the unfortunate defect of being unsuitable for pumping water out of mines as it could only pump water to a height of twenty feet, not enough to get water out of most mines. Furthermore the machine lacked a safety valve and was inclined to explode on occasion due to the pressure of steam on the boiler. Nevertheless, Savery's engine was the first steam engine to be sold commercially.

The first really successful steam engine was that produced by Thomas Newcomen, who like Savery was from Devon. H. W. Dickinson comments that Newcomen's engine "was little more than a combination of known parts" with one or two additional ideas of Newcomen's added to it.[38] Burstall considered Newcomen's engine:

"came about as the culmination of a series of advances that had been made during the preceding two centuries and it is most likely that if Newcomen had not built the first engine of this kind someone else would have done so very soon afterwards; indeed Denis Papin very nearly did, for he was experimenting with the condensation of steam in a cylinder, a few years earlier, but he was not a practical mechanic and he was defeated by the mechanical difficulties."[39]

Another interpretation by L. T. C. Rolt is that:

"The wonder is not that Newcomen spent anything from ten to fourteen years on his invention before he achieved success but that such a staggering advance could have been made by one man in a lifetime. ... Seldom in the history of technology has so momentous an invention been developed by one man so rapidly to so definitive a form. When, in addition we remind ourselves of the, to us, unbelievably primitive means at Newcomen's disposal in 1712, then we can scarcely fail to regard his achievement with a wonder akin to awe."[40]

Rolt does seem to go over the top in his admiration for Newcomen's achievement.

The Newcomen engine worked by using a weight attached to a beam to force a piston to rise. The piston would then fall due to a vacuum being created under the piston by the injection of water into the cylinder which caused the steam to condense, reducing pressure under the piston to a level below atmospheric pressure, which forced the piston down. The beam attached to the piston operated a pump to pump water from the mine.

In the years following 1712 Newcomen's engines began operating in mines all over England and also in Scotland, Wales and in Hungary, France, Belgium and possibly in Germany and Spain.[41] Later in the 18th Century, after a scientific study, various improvements were made to the Newcomen engines by John Smeaton which considerably increased their efficiency. An important reason for the success of Newcomen's engine over the Savery engine was that Newcomen's was an atmospheric engine that did not need to use steam pressure any higher than that of the atmosphere.[42] Savery's engine in order to lift water from mines required a steam pressure greater than that which the boilers built in his time were capable of withstanding. Increasing the steam pressure would cause Savery's engine to explode.

The next significant step in the evolution of the steam engine came when James Watt was asked to repair a model of a Newcomen engine. He studied the model and realized there was a great wastage of steam resulting from the heating of the cylinder and its cooling at each stroke. In 1765 it occurred to Watt that if a separate vessel containing a vacuum was connected to the cylinder the steam would rush into the separate vessel and could be condensed without cooling the cylinder. The separate vessel, commonly called the separate condenser, was patented in 1769 and the patent was later extended by Parliament for an extra 25 years. The partnership of Boulton and Watt was formed in 1773, trials were made and the new engine was found to have extra power and to use one quarter of the fuel of the Newcomen engine.[43] Sales were soon being made to mine owners around Great Britain and Europe.

However, for the Watt engine to become truly revolutionary it had to be capable of rotary motion which would allow it to drive all kinds of machinery. Experiments had been made to get rotary motion out of Newcomen engines with some limited success.[44] Watt was eventually able to create a rotary engine although it required many changes in mechanism; steam was required to act on both sides of the piston and new mechanisms were invented to connect the beam to a rod to turn a shaft which gave the rotary motion.[45] The first rotary engine was created in 1783 and by 1787 the design was standardized. This ensured the applications of the steam engine were greatly increased and in its use in the textile industry the rotary steam engine was to become the driving force of the industrial revolution.

A further improvement initiated by James Watt involved the use of expanding steam. In the early Watt engines, steam was admitted throughout the whole fore-stroke and energy was wasted when steam, still under pressure at the end of the stroke, left the cylinder and entered the condenser. To solve this problem Watt stopped the admission of steam into the cylinder when the piston had made only part of its stroke, the rest of the stroke being performed by the steam expanding from boiler pressure to the low pressure of the condenser. This resulted in better fuel economy for the engine.[46] The Cornish beam engines developed by Richard Trevithick after 1812 were similar to Watt engines but used steam at a much higher pressure (40lbs per square inch rather than 5) than was used in the Watt engines. This enabled a much earlier cut-off for the admission of steam into the cylinder at about one ninth of the stroke, allowing a still greater expansion of the steam.

Yet another development concerned the invention of a compound engine with two cylinders by Jonathan Hornblow in 1781 which was developed by Arthur Woolf in 1803. Steam first enters a small cylinder, where it expands from boiler high pressure to an intermediate pressure and then enters a larger cylinder, where it expands down to condenser pressure while performing work against a piston in each cylinder. This gives a better uniform motion and reduces loss of energy caused by the alternative heating and cooling of the cylinder walls.[47]

The improvements made to the steam engine increased its thermal efficiency as shown by the table below.[48]

Date	Type	Thermal efficiency (percent)
1750	Newcomen	0.5
1767	Modified by Smeaton	0.8
1774	Further modified by Smeaton	1.4
1775	Watt	2.7
1792	Watt Expansive	4.5
1816	Woolf Compound	7.5
1834	Trevithick Cornish	17.0

As the steam engine improved its uses grew from pumping water out of mines, to driving machinery in factories, to its use in transport such as railways and steam ships.

The social and cultural effects of the steam engine were immense. The steam engine was to be the driving force of the Industrial Revolution. It was to become the main power source for the factories that arose initially in England and eventually in the rest of Europe and the USA during the 19th century. It was to result in a massive transfer of labour from working in agriculture to working in manufacturing and industry. Steam engines were to power textile mills in England and were later used for digging and moving coal, for smelting and manufacturing iron and steel and in the printing industry. Steam power was also used in agriculture to power threshing machines.

The use of steam power in industry was to result in a massive expansion in the size of towns and cities, so that the majority of the population of industrial states were to live in an urban environment. Conditions in the rapidly growing cities for the workers and the poor were often crowded and filthy leading to the growth of trade unions and political ideas such as socialism and communism. Living standards of the urban poor in industrial societies were soon to improve and democracy was to become the principal political system for industrial states replacing more or less absolute monarchy which had been the principal political system of societies based on agriculture.

Steam power was also used in transport, especially in railways and in shipping. Railways spread throughout Europe and the United States in the 19th century leading to a much more mobile population and to the more efficient movement of trade goods. Steam ships were to lead to a massive expansion in international trade in the 19th century and to vast migrations of people, especially from Europe to America. Improved transport was to help make famine, often a problem in agricultural societies, a rarity in industrial societies, likely to happen only in unusual situations such as war time.

There were a range of factors that influenced the development of the steam engine. A real and significant need was a major factor in the development of steam power. The problem of water getting into mines would have existed all over Europe, but was particularly bad in Great Britain which had the largest mining industry in Europe. By 1650 British coal mines were producing five times as much coal as the rest of the world and mines were becoming deeper and extending further underground.[49] Output is estimated to have expanded from 200,000 tons in the 1550's to nearly 3 million tons in 1700.[50] In these circumstances the need to get water out of mines was largely a British problem so that it is not surprising that the British were the people to solve it. That this was a considerable problem can be seen in that one in seven of the patents granted between 1561-1642 related to the need to get water out of mines. This need was obviously not an ultimate need, but the coal was used either for the ultimate need of providing warmth or in processes that eventually related to an ultimate need.

Considerable scientific progress had been made in the years preceding the invention of the steam engine. The principles of vacuums, atmospheric pressure and the properties of gases such as steam had been discovered by Boyle, Torricelli and von Guericke immediately before the invention of the steam engine. We do not know the exact process by which Newcomen invented his engine, but it seems hardly possible that he could have invented the engine without a knowledge of the properties of vacuums, atmospheric pressure and of gases. Papin certainly knew of Boyle's Law and of the properties of gases, vacuums and atmospheric pressure. From 1675-1679 he worked as Boyle's assistant and he was an expert designer of air pumps and air pump experiments. Air Pumps were used to create vacuums and to control the air pressure within a container. Air pumps were a key element in the discovery of the properties of gases, atmospheric pressure and vacuums. The principle that the heating of gases under a piston would force the piston to rise, is not something that can be obtained by simple observation. It could only be obtained by experiment. Watt and the other improvers of the steam engine would hardly have attempted to use a gas to lift a piston unless they knew that gases expand when heated. It is this vital bit of knowledge, unknown before the propagation of Boyle's Law in the 17th century that allowed the invention of the post Newcomen engine. James Watt had considerable scientific knowledge and it was his knowledge of Black's theory of latent heat that led to his invention of the separate condenser.

One of the principal difficulties facing those who tried to develop steam power was the low quality of the materials they had to work with. As Dickinson said concerning the problems Newcomen would have faced "chains would break, pipes would burst, leather would tear away and incrustation would form in the boiler and on the interior of the cylinder."[51] Dickinson also attributes the failure of Savery's engine to imperfections of workmanship and unreliability of materials and in particular to the inability of tradesmen at that time to make boilers able to withstand a substantial amount of steam pressure.[52] It is much more likely that these sort of technical problems and their lack of knowledge of how to create a vacuum stopped the Romans developing a steam engine, than slavery did. By the end of the 18th century the situation had improved greatly with Watts engines having a much higher standard of workmanship in the making of its valve, valve gear and in the boring of its cylinders. Engine building had begun to move out of the hands of millwrights and into the hands of specialist manufacturers.[53] Dickinson considered new techniques for the boring of cylinders, introduced by John Wilkinson, were vital to the success of Watt's engine.[54] Equally the high pressure engines produced in the first half of the 19th century were dependent upon improved workmanship and materials to stop boiler explosions. It seems clear that progress in the metal working trades was a vital factor in the development of the steam engine without which the steam engine would not have been developed or would have remained a crude inefficient device restricted to pumping water out of mines and would not have become a key factor in the industrial revolution.

A further point that emerges from our study is the move from simplicity to complexity in engine development. When John Smeaton, who did so much to improve the Newcomen engine, first saw a Watt engine he considered it a pretty engine, but to complicated.[55] The move to greater complexity involved adding things to the engine such as the separate condenser, a second cylinder, expanding steam and rotary motion to improve its performance. Such progress had to be made one step at a time and in a particular order as the problems which were intended to be solved by adding to the complexity of the engine would only become apparent at an earlier stage of the engines development and the solutions were sometimes dependent upon newly acquired knowledge which arose only from practical experience in using the engines. Only when these problems became apparent was it possible to attempt to solve them, so the steam engine grew from a simple idea to a more complicated engine as people attempted to improve it.

The social conditions necessary for the development of the steam engine were a society where the free communication of ideas was allowed and encouraged. The steam engine was not invented by any one man and it was necessary for all those involved in its invention to be able to freely communicate their ideas and inventions. It was an invention that had its origins in antiquity and was actually developed over a period of about 100 years by a number of separate individuals. The steam engine was invented both due to individual brilliance on the part of those who contributed to it but also due to a considerable diffusion of knowledge between those contributors. Certainly James Watt developed his engine from a model of a Newcomen engine. Another improvement Watt made to the steam engine to give it rotary motion was a conical pendulum centrifugal governor which ensured the steady motion of the engine, even when the load on it varied. The same system was used in flour mills to regulate the speed of mill stones.[56]

A crucial point is that those who worked on steam engines published accounts of their work. That is how we know of their work and that is how they would have learnt of each others work. Giambattista della Porta published his work in his *Spiritali* in 1606, the Marquis of Worcester published his in his *A century of the names and scantlings of the Marquis of Worchester's inventions*; Sir Samuel Morland in a chapter of a book he wrote, the chapter being called *The principles of the new force of fire invented by Chevalier Morland ...*; Thomas Savery in a book called *The Miners Friend*; while Denis Papin published his work in *Philosophical Investigations*. The publishing of the work done by these men played a crucial role in the diffusion of knowledge of steam power and allowed each man to build on the work of his predecessors. It should be added that in many cases there was confirmation of the work of these men from other sources such as other people's books and diaries, British government state papers and the granting of Letters Patent. That the knowledge of the progress of steam power was reasonably widely known, at least within the circles of those interested in it, was shown by a poem written by Henry Beighton known as the *Prize Enigma* in which Beighton recites the history of the work done on steam power by the Marquis of Worcester, Savery and Newcomen. That Beighton apparently knew both Savery and Newcomen and knew of Worcester's work strongly suggests that Newcomen and Savery would have known of each other's work and that of the Marquis of Worcester. Further elements in the diffusion of knowledge of steam power was the presence of organizations such as the Royal Society and that the work was to some extent concentrated in particular areas such as London, where the Royal Society and the English court were located, and Devon.

There were a number of conditions necessary for the invention of the steam engine. A vital one was the presence of a need, initially that of how to get water out of mines and later how to drive the new machinery that was being produced as part of the industrial revolution. But needs are common and they are not always met. The reasons why those needs were met was due to the scientific progress that was going on in 16th and 17th century Europe concerning the knowledge of atmospheric pressure, how to create vacuums and of the properties of gases. Allied to this scientific progress was a belief in Europe at the time that progress could be made and problems could be solved. The inventors at the time such as Papin, Newcomen and Watt applied the scientific knowledge to solving the problems that existed and after long periods of trial and error were able to produce a working steam engine (Newcomen's) which was then improved to become Watt's engine. Crucial to the progress made by the inventors was the diffusion of scientific and engineering knowledge which enabled them to build on each other's work. The earlier development of printing was important to the diffusion process and the role of organizations such as the Royal Society was also important. Poor quality materials were a great difficulty for those trying to construct a workable steam engine. Once this difficulty became apparent at the start of the 18th century with the Savery and Newcomen engines, work was done to improve the materials so that Watt's engine, which required better materials, was able to be built.

History of Electricity

The magnetic properties of lodestone and the electric properties of amber had long been known and humans have always been aware of lightning. Lodestone, a magnetic oxide of iron, will align itself on a north-south axis if placed on a piece of wood floating in water. This enables it to act as a compass. Amber, if rubbed with certain materials, will attract light objects such as paper. The same phenomena can be observed with sulphur, glass, wax crystals and various gems. These natural manifestations of electricity and magnetism were first systematically studied by William Gilbert (1544-1603) in his book *On the Magnet*.

Electrical machines were made by Otto von Guericke in 1660 and Francis Hauksbee in the early 18th century. Guericke's involved a rotating sulfur ball which brushed against cloth and attracted various materials. Hauksbee's involved a rotating hollow glass ball which glowed when he touched it due to the friction between hand and ball. These types of machines were used by Stephen Gray (1666-1736) and Charles Du Fay (1698-1739) in experiments that showed how an electric charge could be moved along a stick and a thread and that there were two types of electricity, a positive charge and a negative charge and that similar types repel each other and opposite types attract each other. Gray and DuFay also showed that insulating material could stop electricity from being lost from charged objects and anything could be provided with an electric charge so long as it is properly insulated.

The machines used by von Guericke, Hauksbee, Gray and Du Fay could collect a limited electric charge but it wasn't until 1745 when the Leyden jar was invented that it was possible to store a substantial charge so it could be used in later experiments. The Leyden jar was a glass jar with metal foil on the inside and outside surfaces of the jar. The jar was charged by linking it to a charged body via a metal chain, causing the charged body to lose its charge to the Leyden jar. The charge when released could cause a substantial electric shock.

The next significant step in research on electricity was made by Benjamin Franklin (1706-90). Franklin conducted various experiments from which he decided all objects were normally in an electrically neutral state, but could by the use of friction gain electricity in which case they would be positively charged or lose it and be negatively charged. If an object was charged, whether positively or negatively, the electricity could be discharged to return the object to its electrically neutral state. This led Franklin to his idea of conservation of charge, that while electricity could be moved around, the amount of positive charge must be balanced by an equivalent negative charge, so that the total amount of electricity remains the same. Franklin was able to give Leyden jars both positive and negative charges and to show that the electricity was stored in the glass of the Leyden jar. Franklin's most well-known work with electricity was his experiments with lightning. His experiments, including the famous experiment of flying a kite during a thunderstorm, showed that lightning was a form of electricity.

The force between two electric charges is an inverse square law which was first discovered by Henry Cavendish in the 1760's. Cavendish however failed to publish his work which was eventually published by James Clerk Maxwell in 1879. The inverse square law is known as Coulomb's law and was first published in the 1780's when Coulomb carried out experiments on both the magnetic and electrical forces using a torsion balance. The torsion balance allowed Coulomb to measure the attraction and repulsion between two charged objects, suspended by an insulating fiber, so he could measure the force between the two charged objects.

The study of electricity in the 17th and 18th centuries had been limited to the study of static electricity. Static electricity involves a sudden rush of electricity like a flash of lightning. A steady flow of electricity along a wire is called an electric current. It was only with the invention by Alessandro Volta of the voltaic pile in 1800 that scientists were able to study electric currents. The voltaic pile came from a mistaken discovery of animal electricity proposed by Luigi Galvani. Galvani had noticed that dead frogs jerked when they touched an iron fence while hanging from a brass hook. Galvani thought the electricity came from the frog. Volta disagreed and conducted a series of experiments which suggested the electricity was caused by differences between the two metals.

All metals are reactive, in that they may lose or gain electrons, but some are more reactive than others. They can be considered as being in a reactive series running from the least reactive to the most reactive. When two metals far apart in the reactive series (one very reactive, one not very reactive) are placed on either side of a moist material, the more reactive metal, for example zinc, will lose electrons and the less reactive metal for example copper will gain electrons. This process constitutes an electric current which can flow through a wire connecting the two metals.

The voltaic pile consisted of a series of combinations of metals like copper and zinc, widely separated in the reactive series, and physically separated by a moist material, piled on top of each other and connected by a wire. It provided a flow of electric current that could be turned on and off at will. The current could be increased by adding more of the zinc, wet material, copper combinations or reduced by having fewer of these combinations within the pile. Volta's invention soon became a vital laboratory tool for studying electricity and for decomposing compounds and later when further progress was made it developed important applications outside the laboratory.

In 1820 Hans Oersted (1777-1851) discovered that an electric current affected the behavior of a compass needle. A compass needle is also affected by the force from a magnet so it seemed as though the electric current had the same effect as a magnet. This suggested electricity and magnetism were either the same force or closely related. Oersted's discovery also led to the invention of the electromagnet, a wire through which an electric current runs causing the wire to behave as a magnet.

In 1821 Michael Faraday was asked to write an article about Oersted's discovery and repeated his and others experiments. He also created his own experiments, one of which involved a vertical copper rod through which he ran an electric current. A magnet was placed near the bottom of the rod and the rod moved around the magnet. This involved the conversion of electrical energy into mechanical energy and was the world's first electric motor. In 1831 Faraday created an experiment which involved moving a magnet in and out of a circular coil of wire which created an electric current in the wire. This however only applied if the magnet was kept moving so the strength of the magnetic field near the wire was constantly changing. Oersted had discovered that an electric current that is moving electricity caused magnetism. Faraday had discovered that a moving magnet causes an electrical current. Faraday's discovery showed mechanical energy, such as motion, could be used to create electrical energy. This was the world's first electric generator or dynamo.

Faradays and others discoveries of the effects of magnets and electric currents on compass needles and on iron filings scattered around magnets or electric currents, which assume a pattern of concentric circles lead to ideas of lines of force. The lines of force were expressed mathematically by James Clerk Maxwell in 1873 in four equations which show how electric and magnetic fields behave in all circumstances. The equations applied equally to electrical phenomena as to magnetic phenomena so that the process of unifying electricity and magnetism into an electromagnetic theory was complete. A remarkable feature to come out of Maxwell's work was that the electromagnetic fields traveled at the same speed as the speed of light. The speed of light had recently been measured with some accuracy by Armand Fizeau (1819-1890) and Leon Foucault (1819-1868) and Maxwell's theory showed that light was a form of electromagnetic radiation.

Faraday's discovery of the electric generator and the electric motor were to change the world. The first practical electric motor was created by the American Thomas Davenport in 1835. Davenport’s motor used an electromagnet, a wire through which an electric current flows, to create a magnetic field. The electromagnet, like an ordinary magnet, has a north and south pole. However the poles of the electromagnet can be changed by changing the direction of the flow of the electric current. A needle placed between the poles will move in one direction and then when the current is reversed will swing all the way round in a circle, to point to the new pole. If the current of the electromagnet is constantly reversed at the right time the needle will continue going round and round. Davenport’s motor could drill holes in steel and was used to run a printing press. However, the motor was uneconomic due to its source of electric power being a battery.

It was not until the late 19th century when a number of developments came together that electric power and electric motors became economic. In 1873 Zenobe Gramme built a practical electric generator that allowed rotary power to be turned into electric current. Previous generators tended to overheat due to the action of the armature, the moving part of a generator. Gramme’s generator solved this problem, making it practical for power generation to begin. The world's first power generation station opened in England in 1881 supplying hydro-electric power to the local city council and to private premises. The station closed down in 1884 as it was unprofitable possibly because electric lighting cost more than gas and early light bulbs did not last long. In 1882 another power station using steam power opened in London and a power station using six generators was opened in New York.

A dispute arose as to whether electricity generation should use direct current or alternating current. Direct current had the advantage that it was cheaper for urban areas and rechargeable batteries could be used with it to ensure continuity of supply if generators broke down. Alternating current had the advantage that it was cheaper to transmit over long distances. This is because when electricity is transmitted over long distances, the transmission voltage is increased in order to reduce losses of electricity. When the electricity is sent to individual consumers transformers are needed to lower the voltage. Direct current transformers are complex and expensive, while alternating current transformers are simple and cheap. Thomas Edison backed direct current and George Westinghouse supported alternating current and Westinghouse was successful when he was able to supply the machinery for the Niagara falls hydro-electric station. Improvements in generators ensured fewer breaks down in supply and the benefits of economies of scale with large power stations requiring long distance transmission of electricity meant alternating current was always going to be the better system in the longer term.

Around the start of the 20th century the steam turbine, invented in 1883, began to be increasingly used in electric power stations. The steam turbine had fewer moving parts than a piston driven steam engine so it could run smoothly at greater speeds than a piston driven steam engine. Modern electric power stations may be fuelled by coal, oil or nuclear power or use the kinetic energy of falling water to make hydro-electric power.

A crucial development in the spread of electric power was the development of the electric light. When an electric current moves through a wire it creates heat. The heat is caused by the resistance of the wire which results in electrical energy being converted to heat energy. If the current is strong enough the wire will glow. An 1845 patent by J.W. Starr proposed the use of carbon filaments, as carbon is a good radiator of heat, inside a glass bulb containing no air so as to prevent the hot filament being burnt away. Joseph Swan tried to produce such a light but was not successful as it was not possible to create a perfect vacuum within the bulb with the vacuum pumps then available. Swan also decided the light needed a better source of electricity than batteries. The problem of creating the vacuum was solved when Herman Sprengel invented the mercury vacuum pump and in 1879 Swan produced a vacuum electric light. In the same year Thomas Edison produced his electric light and then both Swan and Edison began manufacturing the light bulbs, initially in competition, but later in partnership. The electric light was improved by the use of tungsten, which has a very high melting point in the filament when William Coolidge discovered how to create a tungsten filament, a difficult job given that tungsten is a hard and brittle material. Further improvements made were to fill the bulbs with a chemically inactive gas such as argon and to wind the filament into a coil, both improvements considerably extending the life of the light bulb.

The coming together of a reliable system of generating electricity, the invention of the electric light bulb and the invention of reliable electric motors was to result in the electrification of first world countries in the 20th century. In industry, electric motors were to drive all sorts of machines such as drills, grinders, lathes, rolling mills, conveyor belts and cranes. Steam engines which originally provided the power for factories, operated by means of belts and pulleys which is far less flexible than simply plugging a machine into a powerpoint, which became possible with electric power generation and electric motors. In transport electric trams and railways became commonplace. In the home electric motors powered by electricity generated far away and transmitted to the home were soon driving vacuum cleaners, washing machines, driers, waste disposal units, food mixers and dishwashers. Electricity in the home also powered lights, air conditioning, heaters, computers and televisions. Electric lighting allowed work and entertainment to take place at night and made the streets safer at night. Electricity also operates in communication systems such as the telegraph and the telephone. The use of electrically powered lifts has allowed the building of modern skyscrapers.

It is quite apparent if it was not possible to generate electricity many of the things we take for granted in modern life would not exist. Television, telephones, radio, computers, electric lighting and heating would not be possible without the discovery of economic electricity generation and how to control and manipulate electricity.

Electricity has had an enormous effect on the modern world. It has however only been possible due to the structure of the world we live in. Our world is largely made up of objects which are made up of atoms and all atoms contain electrons. It is only because electrons exist and because objects and atoms are able to lose and gain electrons that electricity is possible.

Electricity is simply the transfer of electrons from one object to another and this process can generate heat and light and by means of an electric motor can be turned into mechanical energy. It is these properties of electrons and electricity that have resulted in electricity having its revolutionary effect on human society and on human social and cultural history. If electrons did not exist or they were unable to move from atom to atom and from object to object electricity would not exist and the most convenient and possibly the only method available to humans of moving energy over long distances and using it efficiently in the home would not be available to us. If electricity did not exist there would be no telephones, television, computers, electric lighting and heating and the use of electricity in industry and transport.

The history of electricity reveals a series of discoveries with the simplest discoveries being made first and more complex discoveries being made later. Some discoveries could not be made without certain prior discoveries having been made. The earliest forms of electricity to be experienced by people were those that occur naturally such as lightning and those resulting from simple activities like rubbing an object which causes electricity by means of friction. Systematic experimentation concerning electricity began after the scientific revolution with the scientists constructing simple machines to create electricity and conducting simple experiments that showed electric charges could be positive or negative and that insulating material could stop an electric charge being lost from a charged object. This led to the invention of the Leyden jar which enabled electricity to be stored and used in later experiments. Franklin was able to do experiments using Leyden jars which further increased human knowledge of electricity. Up to this time only static electricity had been studied as it was much easier to create static electricity. Current electricity required the prior discovery of a battery such as the voltaic pile before the scientific study of current electricity could commence. It was only when current electricity could be produced and experimented with, was it possible to discover the connection between electricity and magnetism. Once current electricity could be produced it was soon discovered that an electric current affected the behavior of a compass needle leading to the invention of the electromagnet and eventually to Faraday's invention of the electric motor and the electric generator. When practical electric generators and motors were invented and the generation of electricity became economic, helped considerably by the invention of the electric light, the use of electricity began to spread throughout the first world. The order in which these discoveries were made was inevitable and given how valuable electricity is to human beings, it was also inevitable, that sooner or later in some society open to new ideas and technology, that electricity would be used to meet human needs.

Electric Telegraph

The use of electricity for the communication of messages began with the telegraph. An electric current is able to flow along a wire and arrive almost instantaneously at its destination. The first known suggestion for an electric telegraph was made in 1753 by Charles Morrison, a Scottish surgeon. It involved 26 wires, one for each letter of the alphabet, which could carry a discharge of electricity and attract a piece of paper representing a letter to an electrified ball where the message was to be received. In 1804 Francisco Salva in Spain, using an electric battery and a similar system of multiple wires as was proposed by Morrison, sent messages up to a kilometer, while in 1812 Dr Samuel von Sommering using improved batteries sent messages for 2 miles. In 1832 a system using one to six wires was created by Baron Pawel Schulling in Berlin with needles hanging over coils being moved by the current to indicate letters. William Cooke and Professor Charles Wheatstone received a patent in 1837 in England for a telegraph of 5 wires connected to needles which moved to indicate letters.

A system using a single wire was patented in 1838 by Samuel Morse. Morse realized a message could be sent by interrupting the current in such a way that the interruptions constituted a sign. The sign or a combination of signs could represent letters of the alphabet so a message could be sent. Samuel Morse was to invent this system known as Morse code but an instrument for sending and receiving the code was also needed.

The first instrument Morse came up with involved an electro-magnet with a pendulum. When Morse made or broke the current the electro-magnet moved the pendulum which was attached to a pencil which marked a paper tape. The marks on the tape were a series of zigzag lines which represented the letters of the alphabet. The instrument was eventually improved by replacing the pendulum with a clockwork mechanism and the zigzag lines were replaced with the dots and dashes of Morse code. The current from the system was initially too weak to send a message for any great distance so Morse invented the relay. This was an electro-magnet that could close the circuit on a new line and add more power to send the current further.

After some difficulties Morse was able to set up the first working telegraph in 1844 between Washington and Baltimore. By 1851 a submarine cable was placed across the English Channel and in 1866 a trans–Atlantic cable was laid. In America the Western Union Telegraph Company placed telegraph lines across America and soon every railway station had a telegraph office to control the movement of trains.

The telegraph continued to be improved when in 1855 Professor David Hughes invented a printing telegraph using a keyboard into which the letters were typed and the message was printed out where it was received. Duplex telegraphy which allowed two messages to be sent over the same line at the same time was invented by J. B. Stearns and patented in 1872. Transmission speeds were increased by the introduction of a punched tape system which allowed the transmission of 75-100 words per minute. Morse's original telegraph of 1838 could only transmit at up to 10 words per minute.

The telegraph was to vastly speed up the spread of news. In 1776 it took seven weeks for a sailing ship to bring word to London of America's Declaration of Independence. During the Crimean War word reached London by telegraph of the death of the Russian Czar in St Petersburg on the same day he died. The telegraph revolutionized diplomacy with governments being able to have continued and immediate contact with their diplomats abroad. Instructions could be sent and reports received without any time delay, although there was always the risk of interception by unauthorized persons. The Zimmerman telegram from the German government to Mexico was decoded by the British and published by the Americans and was one of the events leading to American involvement in World War I.

Public use of the telegraph system grew, so by 1870 it cost a shilling to send a 20 word message anywhere in Great Britain. Ninety million telegrams a year were being sent by the end of the 19th century. In America a telegraph office was available in almost every small town. However by 1918 the telegraph system was of declining importance due to the development of the telephone network.

The electric telegraph could never have existed at all but for the ability of an electric current to travel along a wire and our ability to send and control the current. If metallic wires were not able to conduct electricity or if we were unable to control the electricity through insulating materials, there would be no electric telegraph. The telegraph could not have been developed until after people learnt how to create and control an electric current. This required the invention of the battery and Morse's invention of the relay which allowed the current to travel long distances. The sending and receiving instruments required the earlier invention of the electro-magnet. Consequently, the telegraph could only have been invented after the invention of the battery, the relay and the electro-magnet all of which were dependent upon prior discoveries made by scientists engaged in research on electricity and magnetism. The telegraph provided an improved method for meeting the human need for long distance communication and its arrival in the mid 19th century was the culmination of a series of logical developments that led inevitably to the invention of the telegraph.

Telephone

The telephone works by converting acoustic energy into electrical energy. It turns the sound waves of the speaker's voice into a varying electric current which is sent along a wire and is then turned back into sound waves. The telephone consists of a transmitter and a receiver. The transmitter contains a metal box filled with carbon granules. The sound waves from the speaker's voice strike the side of the metal box, known as the diaphragm, causing the diaphragm to press against the carbon granules. This pressure varies as the sound waves vary and an electric current passing through the carbon granules is varied in accordance with the changes in the pressure on the carbon granules. The variable current is an electrical copy of the sound waves and is sent along a wire to the receiver. The receiver contains an electro-magnet and a thin disc capable of vibrating. The variable electric current enters the electro-magnet which varies in strength according to the variations in the current it receives. The variations in the strength of the electro-magnet cause the thin disc to vibrate producing sound waves which are a reproduction of the sound waves produced when the speaker talked into the transmitter. This enables the person listening on the receiver to hear what the person speaking into the transmitter has said. The invention of the telephone also involved the invention of the microphone and the loudspeaker. The microphone is the device in the transmitter which converts the speaker's sound waves into a variable electric current while the loudspeaker is contained in the receiver and converts the electrical current back into sound waves so the listener can hear what the speaker has said.

Alexander Graham Bell is usually credited with the invention of the telephone which was patented in 1876. However, at least ten men before him had the idea of the telephone and two of them produced a practical telephone. Philip Reis made a telephone in 1863, but did not take out a patent. Elisha Gray also invented a telephone but was beaten to the patent office by Bell by a few hours.

Bell's telephone used the same part of the phone for speaking and listening. Within a year Edison produced a phone which had a separate mouthpiece and earpiece and had a better transmitter. The first telephone exchange was created in 1878 in New Haven, Connecticut. It had 21 customers who were connected by operators sitting in front of a simple switchboard. Long distance calls began between Boston and New York in 1884 with copper wire replacing the iron wire which had previously been used. An automatic telephone exchange, which eliminated the need for operators to connect customers, was developed by Almon Strowger in the 1890's. It became widely adopted early in the 20th century. Long distance telephony was made difficult by electromagnetic induction between the telegraph and telephone systems. This problem was fixed by the use of twisted twin cable conductors and by "chokes" in telegraph circuits so as to allow telephone messages to be sent on telegraph wires. This allowed long distance telephone calls to be a reality.

The telephone system has had a dramatic effect on business and social life. It has assisted in the development of large scale national and international business and has kept families and friends in touch from all around the world. The number of telephones grew throughout the 20th century. In 1934 there were 33 million telephones in the world, by 1976 there were 380 million. By the 1970's the vast majority of homes in the western world had a telephone. The growth in the telephone system was accompanied by a substantial decline in the use of the electric telegraph system.

The telephone could not have existed without a means of converting sound waves into electrical waves and vice versa. The telephone was also dependent upon the ability to send an electric current down a wire. If nature did not make these things possible there would have been no telephone system. It is the particular properties of the carbon granules that they will conduct the electric current in proportion to how closely packed together they are and the closeness will vary in accordance to the pressure on them from the diaphragm. If neither carbon granules or any other material was able to cause the variation in the electric current then there would have been no telephones. The telephone would also never have existed but for the existence of electrons and that certain materials are easily able to lose electrons and conduct an electric current. Equally important is that other materials do not gain electrons and so are able to act as insulators to stop the current being lost to the air. If these materials were not available in nature we would not be able to send an electric current down a wire and there would be no telegraph, telephone or internet over telephone wires. It is the existence of these materials that makes possible the telegraph, telephone and internet so it can be said that the structure of nature, particularly the existence of electrons and conducting and insulating materials has a major effect on human social and cultural history.

The telephone could not have been invented until it had been discovered how to turn sound waves into electric vibrations and how to send an electric current down a wire. These discoveries were dependent upon a series of prior discoveries such as the electro-magnet and the battery which were themselves dependent upon earlier discoveries in the history of electricity. This means there was a definite time in human history when the telephone could have been invented. Only when the practice of systematic experimentation had been adopted at the time of the scientific revolution in 16th century Europe was the knowledge of electricity able to grow until it reached the stage of people being able to send a controlled electric current down a wire. The knowledge of how to convert sound waves into electrical vibrations also could not have been acquired until humans became aware of the properties of carbon granules and how to create and control an electric current. The invention of the microphone which converts sound waves into electric currents took place in the time in history when it was able to take place. It could not have taken place any earlier due to the necessity of prior discoveries and inventions.

The telephone was invented after the telegraph as the telegraph was an easier invention than the telephone. The telegraph involved sending and controlling an electric current, a simpler invention than turning sound waves into electrical vibrations. The invention of the telephone involved many of the same problems as the invention of the telegraph, but additionally required the invention of the microphone and the loudspeaker so it was always going to be invented after the telegraph. This meant there was always going to be a brief period when the telegraph dominated long distance communication before the telephone took over.

Radio

The existence of radio waves was first suggested by James Clerk Maxwell in 1864. He suggested radio waves were a form of electro-magnetic radiation with a long wavelength. Radio waves were shown to exist by Heinrich Hertz in 1888 when he conducted experiments that sent electrical waves, including radio waves, through space. Heinz used an oscillator and a resonator. The oscillator consisted of an electric battery that discharged electricity into a coil connected to two metal balls which were half an inch apart. When the discharge took place sparks jumped across the space between the two balls. The resonator was a circular piece of wire with a small gap in it and with two metal balls at the ends of the wire with the small gap separating the metal balls. When the electrical discharge occurred in the oscillator causing the spark to jump the half inch gap between the two balls in the oscillator, a spark jumped across the gap between the two balls in the resonator even though the resonator was not connected to the oscillator. The cause of the spark jumping across the gap in the resonator was electro-magnetic vibrations in space originating from the oscillator. Electro-magnetic waves of a very high frequency leave the oscillator and travel out in all directions and will create an electric current in any conductor they meet. These waves are radio waves, they are everywhere but are not normally detectable by people.

The detection of radio waves was made possible in 1890 when Professor Edouard Branley invented the coherer. The coherer was a glass tube containing metal filings and when radio waves hit the tube, it was able to detect the waves and turn them into usable electric current. The coherer was able to ring a bell and more usefully send Morse signals through the air.

Marconi began experimenting with radio waves in 1894 and by 1899 he had sent radio waves across the English Channel. In 1900 he invented the tuned circuit which allowed the sending of a constant series of waves of the same wavelength to which a receiver could be tuned to receive those waves. This reduced interference from other radio waves and allowed the waves to be sent over longer distances. In 1901 Marconi sent radio waves from Cornwall in England to Newfoundland in Canada showing that radio waves would follow the curvature of the earth rather than fly off into space, as other electro-magnetic waves such as light do.

An important development for radio was the invention by Lee De Forest of the triode or audion tube. The triode was a vacuum tube in which a filament, metal grid and a positively charged metal plate were placed inside a glass tube containing a vacuum. The heated filament sends a flow of negatively charged electrons toward the positively charged metal plate. However the metal grid with an alternating current changing from negative to positive many times per second, lies between the filament and the positively charged metal plate. When the grid is negative, electrons emerging from the filament are repelled. When the grid changes to a positive state a great force of electrons goes through the mesh of the grid towards the positively charged metal plate. This has the result that one moment few electrons reach the positively charged metal plate, while at the next moment there is an enormous rush of electrons towards the positively charged metal plate. This means a weak radio wave can be turned into a greatly increased electric current which can be further strengthened by putting it through other triode vacuum tubes. The ability to increase the strength of the current which can be done both when the current is sent and received allows radio waves to be sent and received over much greater distances.

The triode was invented in 1906 and in the same year the first public radio broadcast of speech and music was made. A microphone was used to turn the sound waves of the music and speech into electric waves. The first trans-Atlantic speech was made in 1915 from a U.S. Navy radio station in Virginia to Paris. The superhet radio receiver which was better able to pick up weak signals and could be tuned into radio stations more effectively was invented in 1918. Short wave radio began to be used for long distance communication in the first half of the 1920's.

The explanation for why radio waves followed the earth's surface was discovered in 1924. The earth is surrounded by the ionosphere, an area of electrified air about 80-200 miles above the earth's surface. The ionosphere reflects radio waves sent into the air back down to earth which reflects the radio waves back into the air. The radio waves bounce between the ionosphere and the surface of the earth as they travel around the world.

The transistor radio was developed in 1954 allowing much smaller and more portable radio receivers. It also resulted in radios appearing in cars and other motor vehicles.

Radio communication has had a wide variety of uses. A major early use involved ships at sea which could not communicate by telegraph. By 1903 50 merchant ships had radio telegraphy to keep in contact with shore stations. Navies adopted radio to enable more effective control of ships at sea and radio played a significant role in the Battle of Jutland. Many people were saved from the Titanic due to the radio operator on the ship informing nearby ships that the Titanic was sinking. The arrest of the famous murderer Dr Crippen was achieved when the captain of the ship Crippen was traveling on radioed his belief Crippen was on his ship to a shore based radio station. The station informed the police who arranged to be in port when Crippen's ship arrived in Canada.

Early radio was a form of wireless telegraphy using Morse code. Radio reached its full potential when speech and music were broadcast. Politicians as diverse as Franklin D Roosevelt and Adolf Hitler used radio to get elected. Voters heard the results of elections on the radio. Sporting events taking place thousands of miles away from listeners were broadcast directly to listeners. Radio stations were initially funded by the manufacturers of radios but later used advertising to make profits. Ultimately the broadcasting of music became the major function of radio, closely followed by radio-talkback which allowed the public to have their say on a wide variety of topics on radio. The transistor radio allowed people to have radio wherever they went and pop music stations kept the youth of western countries supplied with a constant stream of music.

Radio is only possible because the electro-magnetic spectrum includes waves of the radio wavelength. Without waves of that wave length radio communication could not have happened. Long distance radio communication was also dependent upon the existence of the ionosphere without which radio waves could not travel around the world. Long distance radio is also dependent upon the triode vacuum tube or some other form of amplifier. If there was no amplifier, radio could only have been used for reasonably short range communications.

Radio was also dependent upon the earlier invention of the microphone and loudspeaker. Without these inventions radio would have been limited to a form of wireless telegraphy using Morse code. There would have been no music, radio talk-back or sports commentaries. Radio was equally dependent on the human discovery of how to create and control electric currents and how to amplify them using valves and transistors. It could only appear after these discoveries had been made so it could not appear before the 20th century.

Television

Television has been evolving for over a hundred years. There were two main lines of development, the photo-mechanical and the electronic. The electronic was to prove the better system with the photo-mechanical falling into disuse in the 1930's.

The first idea for television was proposed by the American scientist G R Carey. Carey proposed a system for using electricity to send a picture from a camera to a receiver some distance away. Carey's system consisted of a number of light sensitive selenium cells each connected by wires to a battery and to an electric lamp. There was one electric lamp for each selenium cell. The cells reacted to the light and would send an electric current, which amounted to an electronic version of the picture from the camera, down the wire to the electric lamps. Carey envisaged a receiver consisting of the same number of lamps as there was for the selenium cells. The idea was not practical as to produce a clear image millions of electric lamps were needed.

Practical television required a single light cell connected to one lamp. Dr Paul Nipkow, a German physicist, invented the Nipkow disc, which was a fast turning circular metal disc with small holes in it. The rapidly spinning disc is placed between an object and a photo-electric cell and only a small section of the object is exposed to the photo-electric cell at a time through the holes in the disc. The cell converts the light into electrical signals which are transmitted to a lamp which would vary in brightness. The lamp would shine through the holes on a second revolving disc on to a screen on which the image is projected. The Nipkow disc was used to produce television in the 1920's but the discs could not be turned fast enough to produce good quality pictures.

The cathode ray tube was invented in 1887 by the German physicist Ferdinand Braun. Braun was studying the behavior of electrons when he discovered electrons hitting fluorescent paint caused the paint to glow. He placed paint on the end of a glass tube containing an electrode which produced a stream of electrons. He found he could control the electron stream by the use of electro-magnets and electrically charged plates.

The first television to work was created by Boris Rosing in St Petersburg in Russia. In 1906 he combined the cathode ray tube with Nipkow's disc system. Nipkow's disc was used to record a scene and turn it into an electrical signal. The cathode ray tube was used as a receiver but the picture produced was of very poor quality as at the time there was no way to amplify the electrical signal.

A fully electronic television system was suggested by A A Campbell Swinton in 1908. He suggested that the cathode ray tube could be used both as a television camera as well as a receiver. The camera could work with the end of the cathode ray tube being made of a photo-sensitive material and the scene being focused on that material. The cathode ray could scan the photo-sensitive material to produce an electrical signal which would match the strength of the light at each point on the end of the tube.

The first person to put Campbell-Swintons ideas into practice was Vladimir Zworykin, a Russian refugee working in America. Zworykin produced an electronic camera tube called the iconoscope and a receiver called the kinescope. The early versions produced very poor quality pictures. At the same time Philo Farnsworth in California produced an electronic television camera and receiver.

The first practical television broadcasts, however, were made by John Logie Baird and by Charles Francis Jenkins in the United States. Both Baird and Jenkins used Nipkow discs for both the camera and the receiver and they were able to produce crude television broadcasts due to improvements in electronic amplification and better photo-electric cells and electric lamps. However, picture quality using Nipkow disks could never be very good as it was impossible to rotate the disc fast enough to produce high quality pictures. The future lay with electronic television cameras and receivers as were invented by Zworykin and Farnsworth.

The television system invented by Zworykin and Farnsworth works by light reflected from the scene being sent through a lense and being focused on a metal screen. The metal screen is made up of millions of tiny specks of cesium oxide which act as tiny photo-electric cells emitting electrons when light falls upon them. The higher intensity of light falling on a given speck, the greater the volume of electrons emitted from the speck. The light sensitive screen becomes positively charged when it loses its electrons. The positively charged light sensitive screen is then scanned by a cathode ray controlled by electro-magnets which replaces the electrons lost when the light falls on the light sensitive screen making it electrically neutral. The electrons lost by the light sensitive screen flow to a positively charged plate, usually made of silver due to its high conductivity, where they constitute an electric current which is amplified before being sent as electro-magnetic waves to the television receiver.

The receiver contains a cathode ray tube and when the electrical signals reach the cathode ray tube it controls the electron stream. Electro-magnets in the receiver cause the cathode ray to move in a scanning motion identical to the scanning that occurs in the television camera. This causes the scene in front of the television camera to be reproduced on the screen of the television receiver. The receiver screen is made of glass coated with zinc oxide and other chemicals and contains millions of fluorescent molecules which vary in intensity when scanned in an identical manner to the variation in intensity to the light falling on the light sensitive screen in the television camera.

The extreme speed at which the cathode-ray works, and the accuracy of the electro-magnets controlling it, ensures that the picture has motion. It is not possible to send a whole picture so the picture is sent in tiny bits and the individual bits must be transmitted and received so quickly that the eye only sees the whole picture. Each scan of the television camera screen amounts to a single picture and each scan of the television receiver screen reproduces that picture. However, to create motion on the receiver screen it is necessary to send many pictures per second so that the eye gets the impression of continual movement. In Europe and Great Britain the pictures are sent at 25 frames per second due to alternating current running at 50 cycles per second and one frame is produced for each two cycles of alternating current. In the United States the alternating current runs at 60 cycles per second so that the pictures in the USA are sent at 30 frames per second. In addition, British and European pictures have 625 lines while American television pictures have only 525 lines. These differences in the speed at which pictures are sent and the number of lines per picture make the British and European television systems incompatible with the American system.

The technical problems of producing practical electronic television were largely solved in the 1930's but progress largely ceased during the Second World War as resources were diverted to the production of militarily useful electronic products such as radar. A number of attempts were made to introduce color television before and after the Second World War with a practical system being introduced by RCA in America in 1953. Color television involved the use of color sensitive mirrors that produced red, green and blue images on the camera tubes which convert the optical images into electric signals. The electrical signals are combined in an adding device to form a black and white electric signal which is then mixed with a coloring signal which is then broadcast. The color television receiver contains a box which has a decoder which changes the signal back into separate red, blue and green signals. Three electron guns in the receiver create electron beams for each of the red, blue and green colors which are directed around the screen by deflector coils. The screen has a coating of phosphor stripes of which some stripes are for the red beam and others for the blue or green electron beam. A shadow mask ensures only the appropriate red electron beam hits the red phosphor stripes. After passing through the phosphor stripes, the beams combine to produce a color picture on the receiver screen.

The 1960's saw the replacement of thermionic valves by transistors in television sets. This resulted in much smaller and more mobile television sets. The beginnings of the use of satellites for relaying television broadcasts all around the world also occurred at this time.

The social and cultural effects of television were immense. It has united the world, or at least the first world, into one enormous television audience. It has created what Marshall McLuhan called the "global village". It has brought entertainment, news, sports and information into nearly every home in the developed world and some homes in the developing world. Sporting events such as the Olympic Games are seen by hundreds of millions of people simultaneously all around the world. Political elections are broadcast nationwide and sometimes internationally so people instantaneously learn the results. Dramatic events such as moon landings and the September 11th terrorist attacks are broadcast around the world. Television has major political significance for example media images, shown on television, of the Vietnam War are often considered to have turned American public opinion against the war. Portrayals of events such as famine in third world countries can help encourage relief efforts. Extensive advertising takes place on television interrupting programs and annoying viewers. The effects of violence on television, especially on children, has been controversial. Television has been blamed for many things such as violence in society and obesity in children.

Television could not exist without the presence of some light sensitive materials capable of acting as photo-electric cells. Materials such as selenium or cesium oxide are able to lose electrons when light falls on the material and so create an electric current. Fluorescent molecules on the screen of the receiver are also necessary to turn the electric current back into pictures. The third requirement for television is the ability to control, amplify and transmit the electric current. If there was no light sensitive material capable of turning light into an electric current or no fluorescent material capable of turning an electric current into pictures or ability to create, control and amplify an electric current there would have been no television. It is only because nature allows light to be turned into electricity, electricity to be created, controlled and amplified and turned back into light that television is possible. This means the laws of nature and the properties of the materials in nature have had a significant effect on human social and cultural history.

The particular time in history that television appeared was inevitable. Once humans learnt how to create, control and amplify electricity and that certain materials would lose electrons when exposed to light and that other materials when struck by electrons would produce a picture, then the development of television was inevitable so long as there were appropriate social and cultural conditions for its invention and introduction. Those conditions existed in Europe and the United States in the first half of the 20th century.

Photography

The invention of photography required the understanding of two scientific ideas, one concerning physics and optics, the other involving chemistry. The optical idea was that of the *camera obscura*. The *camera obscura* (Latin for dark room) involves light reflected from an object passing through a pinhole in the side of a box or room and an upside down image of the object appearing on the far inside wall of the box or room. This effect is caused because light travels in a straight line and when some of the light rays reflected from an object go through a pinhole they cross and reform as an upside down picture of the object on the far wall of the box. The image is upside down because the light rays cross as the light rays coming from a lower point on the object will go to a high position on the far wall of the box and those coming from a higher point on the object will go to a lower point on the far wall of the box. The *camera obscura* had been known since classical times and had been used by Arab astronomers since the 9th century to look at the sun, stars and the moon. Roger Bacon in the 13th century and Leonardo da Vinci in the 15th century knew of the *camera obscura* and in the 16th century a lense was used in place of the pinhole. The image produced on the inside back wall of the box or room was of limited use as it could not be kept or reproduced.

It was to take a series of discoveries in chemistry before *camera obscura* images could be given permanence and could be reproduced. In the 18th century it was known that some compounds of silver, particularly silver nitrate and silver chloride, blackened on exposure to sunlight. In 1802 Thomas Wedgewood and Humphrey Davy took the first photographs by placing leaves on top of a piece of white paper, coated with a silver nitrate solution, in the sunlight. Where the leaves covered the paper, the paper remained white, while the rest of the paper blackened. This created the world's first negative but the photograph could not be fixed and the white areas soon darkened when exposed to light. Wedgewood and Davy also experimented with putting the treated paper in a *camera obscura* but were unable to find any way of making the images permanent.

It was a Frenchman Joseph Niepce who in 1827 discovered how to make an image permanent. He used a resin called bitumen of Judea to coat a glass plate and exposed it to an image in a *camera obscura*. Where the light hit the plate the resin hardened and turned white. The unhardened areas were darkened with iodine vapor to contrast with the white parts. Niepce produced the first permanent negative but his process has the grave limitation of an exposure time of many hours.

It was not until 1839 that a more practical method of producing a photograph with an exposure time of about 30 minutes was invented by Jacques Daguerre. Daguerre used a silver plate coated with a layer of silver iodine, a compound very sensitive to light. When the plate was exposed in the camera the picture appeared and was developed using mercury vapor and the image was fixed with sodium hyposulfite. The process soon became known as Daguerreotype.

A further photographic process was independently invented by William Talbot in 1839 and was improved in 1841. Talbot treated paper with silver nitrate, potassium iodine and gallo-nitrate of silver and exposed it to a *camera obscura*. He obtained a negative which he then laid in sunlight on top of paper treated with the same chemicals to produce a positive. Talbot found he could produce as many positives as he liked from a single negative and it was this advantage that resulted in his process, soon to be called Talbotype, being preferred over Daguerreotype, which produced only one photo for each exposure.

The wet collodion process was developed in 1851 by Fredrick Archer and was soon to supersede both the Daguerreotype and Talbotype processes. The collodion process involved using a glass plate for the negative image, rather than paper. Liquid chemicals such as nitrocellulose and silver bromide were poured on the glass plate which was placed in the camera and exposed while the glass plate was still wet. The process was not very user-friendly with photographers often getting the chemicals over their hands, arms and clothes. The process also required that photographers carry substantial equipment around in order to do their photography. However exposure times were down to about ten seconds depending on the size of the plate and intensity of the light.

Experiments began with the use of dry plates from about 1853 as dry plates did not require immediate development and reduced the equipment photographers had to carry around. Early dry plates however had very long exposure times so that they were seldom used until after 1871 when Richard Maddox invented the gelatino-bromide dry plate which soon had an exposure time of one second. Between 1874-80 a cheap and fast way of making multiple prints from a single negative was invented. The process involved using a gelantine coated paper sensitized with silver bromide, a compound very sensitive to light. The process allowed prints to be made in the dark room without having to expose the print paper to sunlight.

Popular use of photography expanded enormously when in 1888 George Eastman introduced the Kodak camera. The camera used a flexible roll film made of paper coated with a light sensitive emulsion. A year later celluloid film was introduced which made the processing of the film cheaper and easier. Celluloid was first produced by Parkes in Great Britain in 1862 and then independently by Hyatt in the USA by compounding cellulose nitrate and camphor. Cellulose nitrate had earlier been produced by Braconnet in 1832 by treating cotton with nitric acid.

Color photography was experimented with in the late 19th century but the first practical system of color photography was introduced by Louis and Auguste Lumiere in 1907. Their autochrome process involved photographic plates on glass coated with red, green and blue granules that acted as color filters allowing some light rays through and reflecting others. Other color photographic processes were soon introduced and in the 1920’s and 1930’s smaller hand held cameras were introduced. In the late 1960's electronics were introduced into photography in the form of light sensitive photocells which determine the exposure time for photographs.

Photography has had a major effect on society in a number of areas. In science photographs were taken through microscopes and telescopes to provide permanent and accurate pictures of everything from magnified insects to stars and galaxies. Photographs could show movement that could not be seen with normal vision. Photographs were taken of galloping horses to show that at times they had all four hooves off the ground and of birds in flight.

Social reform was effected by photographs when photographs of sub-standard living conditions and of child labor hastened efforts to improve housing and to ban child labor. Photographs of war lead to greater understanding of the horrors of war which could never be shown by paintings which tended to glorify war. Photographs of the Vietnam War are often credited with undermining American public support for the war. Magazines began using photographs both in advertising products and in news stories from around the mid-19th century. Photography also became an art form, taking over the role of portrait painter and producing many other pictures previously produced by artists. This encouraged art to move towards more abstract images which could not be produced by photography until quite recently. Popular photography became widespread with the introduction of the Kodak camera. Pictures taken during holidays or of friends and family began to fill family photo albums.

Photography was only possible due to certain properties of light and of chemical compounds, such as silver nitrate and silver chloride. Light, when passing through a pinhole into a dark room or box, will produce an image of the scene outside the box or room on the far inside wall of the box or room. If this property of light did not exist there would have been no photography. The light sensitive nature of certain silver compounds was also vital to photography in order to allow the image provided by the *camera obscura* to be fixed and made permanent. If those light sensitive chemicals did not exist photography could not exist. The invention of photography could only take place after the discovery of the *camera obscura* effect and after the discovery of the light sensitive properties of the silver compounds. The *camera obscura* effect had been known since classical times while the light sensitive properties of the silver compounds became known after a series of experiments by European scientists such as Georg Fabricius, Angelo Sala, Wilhelm Homberg, Johann Schultze and Carl Scheele in the 16th, 17th and 18th centuries. Their discoveries were known to Wedgewood and Davy when they created the first photograph and to Niepce, Daguerre and Talbot when they discovered ways of fixing and making photographs permanent. Both the process of creating the photograph and fixing it were dependent on prior discoveries in chemistry made in the few hundred years before photography was invented in the 19th century. The existence of the *camera obscura* effect and of light sensitive chemicals which enabled an image to be fixed to produce photographs shows how the structure of the universe has a major effect on human social and cultural history.

The process by which photography developed in the 19th century followed a logical pattern with the easier discoveries being made before the more difficult discoveries. How to create the photographs was discovered, as it had to be, before how to fix them was discovered. Black and white photography was invented before color photography as color photography involves additional complications than there are with black and white photography. Throughout the 19th century exposure times for photography fell as new and improved techniques were developed. The whole process was improved until amateur photography became possible towards the end of the 19th century.

Motion Pictures

In 1824 the phenomena of persistence of vision was described by Dr Peter Roget. Human vision persists, for a second after a scene has disappeared, as a memory and Roget suggested that if successive pictures of a scene, with only slight differences between the pictures, are run before a person's eyes the memory of the previous picture will run over to the next picture and produce the appearance of continuous movement. Using Roget's idea, machines were invented that flashed pictures before an observer's eyes to create an impression of continuous motion. These machines such as the Phenakistoscope, the Zoetrope and the Thaumatrope relied upon rotating disks and drums to which a succession of pictures were attached and continuous motion was produced when the drum or disk was turned. The first moving picture to be shown on a screen was achieved by combining a Phenakistoscope with a magic lantern but the results were not very satisfactory.

Photography in the early and mid-19th century could not be used for moving pictures as moving pictures required many pictures per second while it took several seconds exposure to make a single photograph. Exposure times for making photographs declined throughout the 19th century and in the 1870's Eadweard Muybridge was able to set up a system using 12 to 24 cameras that enabled him to produce a moving picture of a horse at full gallop. Muybridge was later to take over a hundred thousand pictures of people and animals in motion and was able to show them as a moving picture on a kinetoscope, a machine invented by Edison Laboratories. The kinetoscope produced a good motion picture but could not project the pictures onto a screen. The kinetoscope contained a substantial part of the mechanism of a movie projector. It used 35mm celluloid film which ran at 46 frames per second for the duration of the film which was usually about 15 seconds.

Muybridge, in 1879, invented a projector called the Zoopraxiscope which enabled moving images to be projected onto a screen. He placed pictures onto a rotating glass disk which was connected to a slotted metal disk which was rotated in the opposite direction to create breaks between the pictures. A lantern using oxyhydrogen limelight was shone through the disks onto a screen to create a moving picture on the screen.

Improved projection machines were invented by C. Francis Jenkins and Thomas Armat in America and Louis and Auguste Lumiere in France. Jenkins and Armat's projector, known as a phantoscope, used an electric motor to turn a sprocket around which the film was wound. The turning sprocket unwound the film into a beam of light which put the pictures on the film on to the movie screen. The Lumiere brothers using a similar machine opened the world's first public cinema in 1895.

Movie cameras were developed in the 1890's by Etienne Marey in France and William Friese Greene and J.A.R. Rudge in England. Marey invented a "photographic gun" which took a series of photographs using sensitised glass disks. The camera could take 12 photographs per second with exposures of around 1/720 of a second and could show the movements of a bird in flight. Marey improved the camera by replacing the revolving disk with roll film wound on spools and then later replaced the paper rolls with celluloid which had good transparency so as to make projection onto a large screen easier. Marey's chronophotographic movie camera could take 50 pictures a second with exposures of 1/1000th of a second and was patented in 1890. Marey also invented a projector and discovered that if he took pictures of moving objects at 60 per second and then projected them at 10 per second he could see the pictures in slow motion.

Attempts were made to combine sound with moving pictures as soon as pictures began to be shown. The earliest attempts used Edison's phonograph on which recorded sound was played back in the cinema. It proved very difficult to synchronize the sound with the film, and without electric amplification, to produce enough sound to fill the cinema. *The Jazz Singer* was the first film to successfully combine sound and film in 1927. A much improved method of combining film and sound was developed by E.A. Lauste. This involved a photoelectric process in which the sound was recorded on the film as the picture was made. A microphone was used to capture the sound which modulated a light signal which fell on a strip on the edge of the film. When the film is shown, light goes through the strip onto a photoelectric cell causing a signal that activates a loudspeaker which produces the sound for the film. This process began to be used in the late 1920's and by 1930 only 5% of the major films produced were silent.

Color was added to the earliest films by coloring each frame by hand and later it was applied mechanically by using stencils. A process known as kinemacolor was invented by George Smith in 1909. This was a two color process in which black and white film was exposed at double the usual rate with red and green filters being used alternatively for succeeding frames. The film was projected through the same alternating filters. Kinemacolor was used for some years but had various problems in that special cameras and projectors were needed, good color requires three primary colors rather than two and moving objects usually had color fringes. The big breakthrough came with the development of Technicolor in the 1930's. It involved making three separate films in red, blue and green in a beam splitting camera. The three color negative films were used to produce three positive films and then the color was transferred to blank film. This enabled the production of a colored positive transparency able to be shown with ordinary projectors.

Motion pictures were to have a considerable effect on society. Their greatest effect was in the area of public entertainment, but they also played important roles in science, in education, in news delivery and in politics. Motion pictures were to become a vast industry dominated by Hollywood and making stars out of the actors appearing in the films. In the period after World War II before television became widespread 200 million people went to the movies every week. After television became common, it became the principal form of public entertainment, the number of cinemas in England, Scotland and Wales declined from 4,600 in 1950 to 1,600 in 1979. In science, motion pictures were used to examine the movement of people and animals and of the effect of explosions and bullets in flight and generally of the results of many scientific experiments. Politically, movies have been used to advance certain political movements such as Nazism in *The Triumph of the Will* and slightly more subtly the South's cause in the American Civil War in *The Birth of a Nation*. Many movies produced for public entertainment display subtle political values, in fact it is probably impossible to produce movies that do not display some political values. Public education was advanced by the making of documentaries on topics as diverse as wild life, history and science.

The development of motion pictures could not have taken place but for the existence of the phenomena of persistence of vision. If the persistence of vision did not exist all we would see was a series of individual pictures and there would have been no continuous motion. Motion pictures (of the type we are familiar with) were also dependent upon the prior development of photography. Little progress was made until the exposure times for photography fell so as to allow many pictures to be taken per second. Once this was achieved it was possible to invent a movie camera which could take the pictures capable of being used to create a motion picture. A means of projecting those pictures onto a screen was also needed and when that was achieved silent black and white motion pictures were able to be shown. The existence of persistence of vision and of the ability to invent photography, which depended on the *camera obscura* effect and of certain photosensitive chemicals allowed the production of motion pictures. This shows how the structure and laws of nature and the materials available in nature have a considerable effect in social and cultural history.

It was inevitable that silent black and white movies should appear before movies with sound and color. This is because the need for sound and color was not apparent until silent black and white movies had developed and what was needed to provide sound and color for motion pictures was not apparent until motion pictures had been developed. Only after black and white silent movies had been developed was it possible to work on how to add color and sound to those movies. Eventually in the 1930's high quality sound and color was able to be added to motion pictures.

Internal Combustion Engine

The steam engine was an external combustion engine as the fuel was burnt outside the engine. The first internal combustion engine where fuel is burnt inside the cylinder to force a piston to move was invented in 1856 by the Italians Barsanti and Matteucci. The principle behind the internal combustion engine was the same as that behind the steam engine, namely a piston being driven by alternative phases of expanding gas and vacuums. The first internal combustion engine to be produced in substantial numbers was a gas engine built by the Belgian Lenoir in 1860. The engine lacked power and consumed a considerable amount of fuel because the fuel and air mixture was not compressed before it was ignited. In 1862 Rocas, a French engineer patented a four stroke internal combustion engine which involved compression of the fuel and air mixture. The engine however was never built and in 1876 the four stroke engine was independently invented by Otto. The Otto engine produced more power and consumed considerably less fuel than the Lenoir engine.

The four stroke engine worked by the first downward stroke of the piston drawing the fuel and air mixture into the cylinder through an open inlet valve. The descending piston creates a partial vacuum in the cylinder and the valve in the cylinder closes and the piston rises, compressing the fuel and air mixture. The mixture is then ignited, causing the third stroke as the piston is forced downward. It is the third stroke that gives the engine its power. The fourth stroke occurs when an exhaust valve is opened and the rising piston forces the exhaust gases from the cylinder.

In 1883 Gottlieb Daimler, who had previously worked with Otto, designed a four stroke internal combustion engine that ran on petrol or gasoline. The engine ran faster than Otto's so that it produced more power for the weight of the engine. A carburetor was used to pass air over the top of petrol to mix the petrol vapor and air which was ignited to force the piston down in the third stroke. Further improvements by Karl Benz involved an electrical induction coil for ignition of the fuel mixture.

The effect of the internal combustion engine on society was immense. Its main advantage over the steam engine was its weight to power ratio. In 1880, the Otto gas internal combustion engine weighed 440 lbs. per unit of horsepower produced; by 1900 a petrol driven internal combustion engine weighed only 9 lbs. per unit of horsepower. The weight to power ratio allowed the engine to be used to drive motor vehicles, aircraft, tractors, submarines and tanks. During the 20th century motor vehicles were to replace railways as the principal means of land transport. The urban and rural environments of first world countries were to be crisscrossed by roads, highways and motorways built specifically for motor vehicles powered by the internal combustion engine. The ordinary citizens of first world countries enjoyed a new freedom of travel they had not previously possessed. Aircraft made considerable improvements in performance and safety during the 20th century. They became a new weapon of war but they also helped precipitate the enormous growth in international tourism that was to occur in the second half of the 20th century. Agricultural productivity improved greatly with the development of the tractor and other farm machinery powered by the internal combustion engine. The development of the engine also gave oil producing countries a wealth and an influence in world affairs that they had not previously possessed. These social and cultural consequences of the internal combustion engine were an inevitable result of the invention of the engine and the engine was only invented after certain earlier discoveries had been made.

The steps involved in the invention of the internal combustion engine followed in a logical and necessary order. The first step was the initial invention of the engine by Barsanti and Mattucci and its development by Lenoir. Only after the engine was invented was it possible to work out the best way to operate the engine which is by the four stroke cycle system that was invented by Rochas and Otto. The use of petrol in the engine was dependent upon the earlier developments of drilling for oil which began in the United States in the 1850's and by methods of refining crude oil by distilling or thermal cracking which was developed in the 1860's.

The internal combustion engine could only be invented because of certain properties of gases and vacuums. Gases expand when heated and that a piston will move to reduce a vacuum are properties of gases and vacuums which allowed the invention of the internal combustion engine. If gases and vacuums did not have these properties the internal combustion engine could not have been invented. A further requirement for an internal combustion engine is a suitable fuel which exists in nature in the form of oil deposits. This shows how the properties of the materials in nature have had a major influence on human social and cultural history.

Motor car

Motor driven carriages had been experimented with ever since the invention of the steam engine. The steam engine however was too heavy for the amount of power it produced to allow it to drive any sort of road vehicle. It was not until the invention of the petrol fuelled internal combustion engine that there was an engine light enough and powerful enough to drive a vehicle on the road.

The four stroke internal combustion engine was invented by Nikolas Otto, but his engine ran on gas. An internal combustion engine using petrol had been built by Jean Lenoir, in 1862, but it was too heavy and lacking in power to drive a road vehicle. The creation of a more powerful internal combustion engine fuelled by petrol was achieved by Gottlieb Daimler and Wilhelm Maybach, in 1883, which, because it turned faster than the gas engine, was more powerful for its weight. The first road vehicle powered by a petrol fuelled internal combustion engine was built by Karl Benz in 1885. The car could reach a speed of 8 mph with its engine which provided less than one horsepower. The car had a very unreliable electric ignition with a battery coil and spark plugs. The car also had a gearing system involving belts between pulleys of varying size so as to provide for different forward speeds. Belts were also used to transmit power from the engine to the wheels. The vehicle only had three wheels, two at the back and one at the front. The use of a single front wheel was designed to avoid problems with normal horse carriage steering, where turning was easy enough for horses but was very difficult for people.

The steering problem was solved in 1888 when Britain's first petrol fuelled car was built by Edward Butler. Butler used the Ackermann system which involved the front wheels being connected by a rod so that they turned about a common center. This avoided skidding when the vehicle turned, making turning safer and easier. After the introduction of the Ackermann system nearly all cars had four wheels and nearly all of them used the Ackermann system.

Daimler did not produce a car for sale until 1895 as he concentrated on the production of petrol fuelled internal combustion engines. A two cylinder engine built in 1889 providing three and a half horsepower and which ran at 800 rpm became the standard engine for early cars. Daimler did produce experimental cars that introduced the modern transmission system using a friction clutch and sliding pinion gears so as to allow a range of forward speeds. This system could transmit more power than the belts used in Benz's 1885 car. In the 1890's Benz began to produce improved four wheel cars using the Ackermann system.

The standard design for motor vehicles became gradually established in the last decade of the 19th century. The engine began to be placed in the front of the vehicle as it was found this provided greater stability than placing the engine in the center or rear of the car. Four wheels on cars with the Ackermann steering system became standard. The transmission system became standardized with the introduction of the propeller shaft which ran under the car and drove the rear axle. The most common gearing system used the manually operated sliding pinion gearbox, although some cars used an epicyclic which was the predecessor of automatic transmission. The suspension consisted of four leaf springs that connected the axles to the body of the car. Pneumatic tyres were first introduced in 1895, although solid rubber tyres remained in use for commercial vehicles until around 1930. The braking system used was improved when band brakes which applied pressure to the wheel hub were replaced by drum brakes which applied pressure to the inside of a drum revolving with the wheels.

The world's first mass produced car, the Model T Ford, was introduced in 1908. Between 1908 and 1927 when production ended, 15 million Model T's were built. When a conveyor belt was introduced into the manufacturing process, in 1913, the assembly time for the chassis fell from 12 hours to one and a half hours per car. The price of the Model T fell from $850 when manufacturing began to $260 per car.

Improvements to motor cars after World War I were limited and related mainly to improved engines and to better comfort and safety. Hand cranking was replaced with an electric starting system and the enclosed sedan began to replace open-top cars. All steel bodies became common after the 1920's. Hydraulic brakes on all four wheels became common and safety tyres with no inner tubes and instant self-sealing became common in the 1970's. Seat belts and air bags were also introduced to improve safety.

The effect of the motor car on society was immense. It gave the general public the freedom to travel when and where they liked, unrestricted by time tables and with a privacy not available on public transport. It involved the creation of a major new industry with millions of jobs. Motor cars also became a leading cause of death by accident in wealthy countries and a major cause of pollution.

The motor car could not be introduced without the prior invention of the internal combustion engine. Only that engine could provide enough power and was light enough to drive a road vehicle. Without the internal combustion engine there would have been no widespread motor car use in the 20th century. Once a reasonably efficient internal combustion engine had been invented the rest of what was required to produce a workable motor car was quickly put together. Transmission, gearing, braking, steering and suspension systems were already well understood and all that was required was to adapt them to the motor car. This is why the standard design for the motor car became established quite quickly, within about 20 years, after Benz's first car was built in 1885.

The motor car could not be invented without the internal combustion engine, which was only possible due to the properties of gases and vacuums and the existence of suitable fuels such as petrol and oil. This shows how the properties of matter and materials in nature have had a major effect on human social and cultural history. If the properties of matter were different, for example gases did not expand when heated or it was not possible to create a vacuum, then there would have been no internal combustion engine and no motor vehicles.

Aeroplanes

Human beings have always wanted to fly. The example of birds effortlessly soaring through the air had many human imitators both in myth and in real life. In myth Daedalus and Icarus are supposed to have made wings of feathers and wax but when Icarus flew to near the sun his wings melted and he fell to his death. The same result occurred in real life when people made themselves wings and jumped off buildings and cliffs flapping the wings as they plunged to their deaths. These bird imitation efforts, known as ornithopters, did not work as humans did not have the same muscle and bone structure as birds.

The kite was the first heavier than air machine to actually work. Kites had been flown in China and Japan over two thousand years ago and became known in Europe in the 16th century. Kites flew due to air pressure under the kite providing it with lift.

Gliders became the main focus of aeronautics research in the 19th century. Sir George Cayley built a 5 ft. model glider in 1804 and by 1849 he had a glider that flew for a few yards when towed and piloted by a young boy. In 1853 a glider piloted by Cayley's coachman flew 500 yards. Cayley's experiments lead to an increasing knowledge of the requirements for flight. Cayley studied bird flight and realized that birds could only stay in the air if they were in constant motion. If they stopped they would fall to the earth. Cayley also realized that the flapping of the bird's wings created air currents that were necessary to provide lift. He concluded that a heavier-than air machine had to keep moving to stay in the air and needed a means of creating air currents. Cayley investigated various engines such as steam engines and an internal combustion engine driven by exploding gunpowder to power an aircraft but he could not find an engine both light enough and powerful enough to make an aircraft fly. Cayley also investigated how to control an aircraft in flight and his gliders had a tail equipped with a rudder and elevator.

Otto Lilienthal began experimenting with gliders in the late 19th century. Like Cayley he studied the flight of birds and from 1891 to 1896 he made controlled flights of up to 750 feet in his gliders. He was killed in a glider accident in 1896 but his book *Bird flight as the basis for aviation* was very influential and assisted the Wright brothers although not all the information in the book was fully accurate.

The Wright brothers began experimenting with gliders in 1900 and by 1903 they had made more than a thousand flights. They learnt how to control the gliders in flight and in 1903 they built a 12 horsepower internal combustion engine with a propeller they designed themselves. They attached the engine to one of their gliders and in December 1903 they made the first powered flight at Kittyhawk, North Carolina. Some weeks before their flight Dr Samuel Langley attempted a powered flight that failed due to equipment failure. Langley's plane was later flown in 1914.

The Wright brother's aircraft began to be improved and redesigned with the engine in the nose, a long fuselage and a tail which helped in the control of the aircraft. With such an aircraft Louis Bleriot flew the English Channel in 1909. Better engines and other technical improvements lead to greater speeds, greater carrying capacity, longer ranges and increased reliability. These improvements occurred particularly rapidly in World War 1 and in 1919 John Alcock and Arthur Whitten-Brown made the first non-stop flight across the Atlantic. This feat was repeated in 1927 when Charles Lindbergh made the first solo flight from New York to Paris. Between the wars many long distance flights were made as aircraft performance improved.

Aeroplanes fly because of a scientific law known as Bernoulli's theorem named after Daniel Bernoulli (1700-1782). Bernoulli's theorem states that as air travels faster, it loses pressure. Aeroplane wings are designed to take advantage of Bernoulli's law. They are curved on top so that air traveling over the wing has further to go than air traveling under the wing. This causes air traveling over the wing to travel faster and as air speeds up its pressure falls so there is less air pressure on top of the wing, than under the wing, which forces the wing to rise. If the wing is properly attached to the rest of the aeroplane, the whole aeroplane will rise into the air.

In order for the aeroplane to rise off the ground it needs a flow of air over the wings. This is provided by the aircraft moving forward which results from the action of the propeller on the air. The propeller cuts through the air and forces it backwards which drags the plane forwards which creates the flow of air over the wings which causes the plane to rise off the ground.

The next major technological breakthrough was the invention of the jet engine. Work on producing the jet engine began before World War 2 and the first jet planes flew during World War 2. The jet engine was invented separately by Frank Whittle in Britain and by Hans von Ohain in Germany. The jet engine is based on Newton’s third law of motion which states that for every action there is an equal and opposite reaction. The jet engine operates by taking air into the front of the engine where the air hits a rapidly rotating fan which compresses the air and forces it into a combustion chamber where it is mixed with fuel. The mixture is then ignited and expands and shoots out of the rear of the engine under great pressure. The gas shooting out of the rear of the engine pushes the plane forward as the opposite reaction to the gas shooting backwards.

The use of the jet engine first in military aircraft and later in civilian aircraft was to result in still greater speeds, range and cargo carrying capacity for aircraft. Passenger transport which began after the First World War expanded greatly after the Second World War with the introduction of the jet engine. In 1937 two million people were carried by the world's airlines outside the Soviet Union. In 1947 twenty one million people were carried and in 1957 there were ninety million air passengers. Trans-Atlantic air travel multiplied forty times between 1950 and 1975 while the numbers traveling by ocean liners fell by 80%. This caused a massive reduction in the size of the ocean liner industry. Cabin pressurization increased the comfort of air travel and allowed aircraft to fly higher above bad weather. The development of the Concorde meant faster air travel and the introduction of the “jumbo jet” in 1970 led to more passengers and cheaper air travel. Mail was the first cargo regularly carried by aircraft, but as aircraft became larger and were able to carry greater cargo, almost any reasonably compact product can be transported by air.

Air travel is only possible due to air passing over an aircraft's curved wing having to travel further and faster than air passing under the wing so that there is less air pressure above the wing than below the wing so the wing and aircraft will lift off the ground. Without this feature of our physical environment and of the behavior of air and gases there would be no air travel. But air travel also requires some sort of engine to drive the aircraft forwards so as to produce a flow of air across the aircraft's wings. It was not until such an engine, the internal combustion engine, with its relatively high power and low weight became available that powered flight became possible. A further requirement of air travel is some way of controlling the aircraft. If it was not possible to control the aircraft any significant air travel would not be possible. It is only possible to control an aircraft due to the way in which rudders, elevators, and ailerons can alter air flows and because alteration of the air flows will cause the aircraft to turn, lift or dive. If air did not lose pressure when it travels faster, or if internal combustion or jet engines were not possible, for example if gases did not expand when heated, or if aircraft could not be controlled in the air, then air travel would not be possible. This shows how the properties of materials, such as gases, in our natural environment have had a major effect on human social and cultural history. If the properties of those materials were different, then human social and cultural history would be different.

While human beings have always wanted to fly, progress could only be made with the scientific study of flight done by people like Sir George Cayley, Otto Lilienthal and the Wright brothers. Prior to such scientific study, the efforts of the ornithopters were going nowhere. However the work of Cayley, Lilienthal and the Wright brothers on gliders could not achieve any useful result until the invention of an engine that was powerful enough to drive the aircraft forwards while at the same time being light enough to be carried by the aircraft. It was not until the internal combustion engine had been invented and improved that such an engine became available. Soon as the engine became available, powered flight became a reality.

The next major development in powered flight was the invention of the jet engine. The jet engine only works due to Newton's third law of motion that to every action there is an equal and opposite reaction. If Newton's third law did not work there would be no jet engine. The jet engine is also dependent on the behavior of the air-fuel mixture which expands when ignited. If the mixture did not expand when ignited there would be no jet engine. The jet engine only exists due to the third law of motion and the behavior of heated gases.

The History of Medicine

Illness and injury are as old as humankind. Stone age human remains show evidence of diseases such as arthritis, tuberculosis, inflammations, dental problems, leprosy bone tumors, scurvy, spinal tuberculosis, cleft spine, osteomyelitis, sinusitis and various congenital abnormalities and injuries. These diseases show in human skeletal remains and if more complete human remains were available, it is likely a much greater range of disease would be apparent. Given that human beings do not like pain, death and suffering there was a clear need to try and find a cure for diseases and injuries.

The curing and prevention of disease often involves an explanation of the cause of the disease. In the absence of knowledge of germs, bacteria and viruses and of human anatomy and physiology, stone age humans ascribed disease, injuries and death to supernatural forces, just as other inexplicable events such as storms, earthquakes and volcanic eruptions were considered to be caused by supernatural forces. This led to the need for a method of influencing the supernatural forces which required a person with knowledge of the supernatural world who could communicate with and placate the gods or spirits that caused the disease and injury. Priests, shamans, witch doctors and medicine men were often responsible for protecting the health of stone age humans by means of appropriate rituals and spells. A cave painting of what is considered to be a Stone Age medicine man dating from around 15,000 BCE is on the cave walls of the *Les Trois Freres* cave in the Pyrenees.

Stone Age medicine men would most likely have supplemented their spells and rituals with the use of various herbs, roots, leaves and animal parts and other medicines. Given the body's natural tendency to heal itself and placebo effects, it would have been difficult for prehistoric healers to work out whether their spells and herbs were actually working. Only in recent times with modern written records, statistical techniques and double blind studies involving control groups, can it be reasonably clear if a particular medicine is working.

The earliest clear example of a surgical operation is trepanning which involves boring a hole into the skull. This operation was first carried out in Neolithic times using stone tools. Some of the patients survived as shown by healing around the holes and some skulls even had several holes bored in them, indicating repeated operations. It is not clear why such a painful operation was carried out, but it may have been to allow evil spirits that were causing migraines, epilepsy or madness to escape from the patient's skull. It is also likely other surgical operations, such as the lancing of abscesses and the sewing up of wounds with bone or flint needles, were performed, but there is no clear evidence of this.

When nomadic hunter-gatherers first began to settle in permanent villages, which grew into towns and then cities, new health problems arose. Large numbers of people concentrated in small areas meant disease would quickly spread through populations. The domestication of animals resulted in many diseases spreading from animals to humans such as measles, smallpox and tuberculosis from cattle and flu from pigs and dogs. However a further result from living in cities was the development of writing which allowed a more organized medical profession and the possibility of accurate recording of symptoms and remedies.

Writing began in Mesopotamia before 3,000 BCE when it was invented by the ancient Sumerians. The Sumerians wrote on clay tablets and one such tablet contains lists of drugs, chemical substances and plants used for medical purposes. Magic and religion however played a major role in Mesopotamian medicine as injury and disease were considered to be caused by Gods, demons, evil spirits and witchcraft. Numerous magic spells, incantations and sacrifices were available to combat particular diseases and correct recitation was necessary for an effective cure. Whether a patient would survive or not could be divined by examining the liver of a sacrificed sheep or goat. The Code of Hammurabi, a law code made by a Babylonian King, sets out medical fees for various services and penalties for errors made by the doctor. Services referred to involved the opening of an abscess, the treatment of broken limbs, eyes and intestinal complaints.

Our knowledge of ancient Egyptian medicine comes from certain medical papyri and from the embalming of Egyptian dead. The papyri contain various descriptions of magic spells designed to drive out the demon causing a particular disease and of various prescriptions, including the dosage for particular diseases. Drugs used included castor oil, hartshorn, bile and fat from animals and copper sulphate. Treatment was prescribed for wounds and bruises and surgical instruments appear to have been used and broken bones were treated with splints.

The Egyptian practice of embalming and the favorable conditions of Egypt for the natural preservation of bodies shows us some of the diseases the Egyptians suffered from. Arthritis and inflammation of the periosteum and osteomyelitis were common. Spinal deformations and spinal tuberculosis, gout and virulent osteomas have been found in Egyptian mummies. Tooth decay was as common as in modern times and there is good evidence of kidney stones and gallstones, appendicitis and stomach and intestinal troubles. The lower classes in particular suffered from infectious diseases such as plague, smallpox, typhus, leprosy, malaria, amoebic dysentery and cholera and various parasitic diseases.

Egyptian physicians' knowledge of anatomy was not extensive despite the practice of embalming. This is because embalming was carried out by specialist technicians and not by physicians. Knowledge of internal organs was largely limited to an awareness of their outward appearance.

Chinese Medicine

The earliest Chinese medicine, in common with most other ancient civilizations, assumed disease and illness were caused by the gods or by demons. The correct remedies for illness involved ritual exorcisms and appeals to the Gods.

A more naturalistic explanation of illness developed with the belief in Yin and Yang. The Yin and Yang principles were considered to control everything and their interaction controlled the functioning of the human body. Yin was feminine, soft, cold, moist, receptive, dark and associated with water while Yang was masculine, dry, hot, creative, bright and associated with fire. Human health depended on a balance between Yin and Yang. Further factors affecting disease were wind, rain, twilight and brightness of day so there were a total of six disease making influences. Any of these six influences could upset the balance of Qi, which was a vital spirit similar to breath or air, which existed throughout the human body.

Chinese knowledge of anatomy was very limited due to a strict prohibition on the dissection of the human body. Chinese belief concerning the inner organs was largely erroneous. They believed there were five “firm” organs that acted as receiving organs and lay opposite five “hollow” organs who served the purpose of evacuation. The firm organs were the heart, spleen, lungs, liver and kidneys. The heart was considered to be the place of wisdom and judgement while the liver and the lungs were associated with the soul. The male’s right kidney was seen as the source of sperm and its connection with the passage of urine was not understood. The hollow organs were the bladder, gallbladder, colon, small intestine and the stomach.

Chinese doctors attempted to make a diagnosis by studying the state of the pulse. This practice known as sphygmology involved attempting to recognize some very subtle variations in the pulse. There were considered to be 51 different varieties of pulse which were to be taken in 11 different areas of the body. Chinese doctors were attempting to obtain far more information from the pulse, than it could possibly provide.

Acupuncture, aimed to restore the balance of Yin and Yang, and involved inserting needles into particular parts of the body. There were 388 areas of the body into which the needles could be inserted and they needed to be inserted at the correct time, based upon the weather, the time of day and the phases of the moon. The needles were left in anything from five to fifteen minutes. Acupuncture does appear to be effective for pain relief as the needles seem to make the body produce endorphins, the body's own natural painkillers. Claims have been made that acupuncture can cure many diseases including muscle, bone, respiratory and digestive disorders. A further Chinese treatment was Moxa which involved inflicting a slight burn on the skin. It was considered to be a treatment for a vast range of complaints such as diarrhea, abdominal pains, anemia, vertigo, nose bleeding, gout, toothaches and headaches.

Indian Medicine

Indian medicine began with the belief that illness was caused by the Gods or by demons and was a punishment for bad behavior. Over time however other beliefs arose such as that which considered good health required a balance being kept between the elements of air, bile and mucous.

India developed surgery to a higher standard than any of the other ancient civilizations. This was because the prohibition on human dissection which existed in Europe, China and the Arab world did not exist in India. This enabled the Indian physicians to obtain a good knowledge of human bones, muscles, blood vessels and joints. A wide variety of surgical operations were carried out, including cosmetic surgery on people who had been mutilated as part of a legal punishment. An adulterous wife could have her nose cut off as a punishment and Indian surgeons learnt how to repair the damage and replace the nose.

India is a land of many diseases and Indian doctors were familiar with 1,120 different diseases. They guessed the connection between malaria and mosquitoes, noticed that the plague was foreshadowed by the death of large numbers of rats and that flies could infect food causing intestinal disease. They were also aware that cleanliness could help in the prevention of disease.

Greco-Roman Medicine

Greek medicine derived its earliest beliefs and practices from Egypt and West Asia. Greek medicine later spread around the Mediterranean during Roman times and was to form the basis of the medical knowledge of Medieval Europe. Our knowledge of Greek medicine mainly comes from the Hippocratic writings and from Galen writing in the second century CE.

The earliest Greek medicine was based on religion. Asclepius, the son of Apollo, was able to cure disease and patients sleeping at his shrines would see the God in their dreams and receive advice on appropriate treatments. Around the sixth century BCE Greek medicine began to change with a greater emphasis on rational explanations of disease involving natural rather than supernatural causes. The Hippocratic writings, probably written by a number of authors, suggested liquids were the vital element in all living things. The human body contained four fluids or humors; phlegm, yellow bile, black bile and blood. Disease was caused by an imbalance of these fluids in the body. Such an imbalance could be caused by the weather or by extreme behavior such as over eating or excessive drinking. The medical practice of bleeding, which was to persist for several thousand years, originated from the belief there was an excess of blood which could be cured by releasing some blood from the body. Correct diet, bathing, exercise, sleep and sex would prevent illness. According to Hippocrates sex should be more frequent in winter and older men should have sex more frequently than younger men. He considered epilepsy to be caused by an excess of phlegm. Hippocrates however tells us little about infectious diseases and anatomy as the dissection of bodies was taboo as it was considered to be a violation of the sanctity of the human body.

The classical era taboo on human dissection led to some quite erroneous views of the human body. Aristotle considered the heart was where the soul was located and was the center of thought, sense perception and controlled bodily movements. He thought the brain cooled the heart and the blood. There was however a brief period in Alexandria where due to the ancient Egyptian practice of embalming and the more recent Platonic view that the soul and not the body, was sacred, human dissection was allowed. Herophilus and Erasistratus carried out dissections that led them to discover the nerves leading to the brain. They discovered there were two different types of nerves, one dealing with sense perception and the other with body movement. When studying the brain they discovered the cerebrum and the cerebellum and suggested the heavily folded human brain indicated humans higher intelligence compared to animals. They considered the lungs took in air that was then transferred to the arteries, the veins held blood and the heart worked like a bellows. After making significant discoveries that could only be made by human dissection, the taboo against dissection arose again delaying further progress until the 16th century. Until then, knowledge of the interior of the human body could only be guessed at from its external behavior or by comparison with animal anatomy.

Two further theories created by the ancient Greeks were the methodic theory and the pneumatic theory. The Methodic theory considered disease to be caused by a disturbance of atoms in the body and treatment involved manipulating the body by massage, bathing or exercise. The pneumatic theory considered breath to be a crucial factor in human health.

The high point of Greco-Roman medical knowledge came with Galen in the second century CE. Galen's two main areas of study were anatomy and physiology. As human dissection was illegal, his anatomical studies were based on dissections of animals, particularly the Barbary ape. He did, however, have the assistance of his study of gladiator's wounds, a human skeleton he had seen in Alexandria and of human bodies exhumed by natural events, such as floods. Galen's work on the bone structure and muscular system was a significant advance on anything else in antiquity. His belief in Aristotle's idea that everything had a purpose led him to assume every bone, muscle and organ had a particular function and he set out to describe each bone, muscle and organ and their particular function. He described the human skeleton and muscular system with some accuracy. He put an end to Aristotle's idea that the mind was located in the heart, locating it in the brain. Galen discovered seven pairs of cranial nerves, the sympathetic nervous system and he distinguished between the sensory and motor nerves. However he also found things that did not exist. The *rete mirabile* (wonderful network) is located under the brain of many hoofed animals but is not found in humans. Yet Galen's claim that it exists in humans was accepted for some thirteen centuries.

Galen's physiology, his concept of how the human body worked, began with a vital spirit, *pneuma*, taken into the body by breathing. The *pneuma* entered the lungs where it met some blood before passing into the left ventricle of the heart. The blood then flowed into the arteries and spread through the body feeding the flesh. When food enters the body it is converted into blood in the liver, some of the blood then enters the veins and spreads through the body and is fed into the flesh. Other blood flowed from the liver into the right ventricle of the heart from where some of the blood enters the lungs to absorb the *pneuma*. Some of the blood in the right ventricle, however, passed directly into the left ventricle and from there flowed into the arteries.

One problem for Galen was that he was unable to discover how blood moved from the right ventricle to the left ventricle, which were divided by a solid muscular wall. He eventually concluded there must be tiny holes in the wall, so small they could not be seen by the human eye. Galen's system correctly realized that the heart caused blood to flow through the body and that the arteries contain blood. Previously Erasistratus suggested the arteries only contained air, as the arteries of a dead body do not contain blood. Galen did not realize that the blood circulated and his suggestion of minute holes in the wall between the right and left ventricles of the heart was wrong.

Galen's pathology, his concept of illness, brought together Hippocrates theory of the four humors and Aristotle's idea of the four elements, air, fire, earth and water. Blood was considered to be warm and moist, yellow bile warm and dry, black bile cold and dry and phlegm cold and moist. Blood is associated with the heart, yellow bile with the liver, black bile with the spleen and phlegm with the brain. The following table shows how Galen brought the two ideas together.

Humor	Element	Organ	Qualities
Phlegm	Water	Brain	Cold & Wet
Blood	Air	Heart	Hot & Wet
Yellow bile	Fire	Liver	Hot & Dry
Black bile	Earth	Spleen	Dry & Cold

The table indicates the symptoms of the disease, the cause of the disease and the cure for the disease. If the patient has the symptom of being hot and perspiring, this is the quality of being hot and wet, this suggests there is an imbalance in the blood, so that bleeding is the cure. If they have a hot and dry fever, this suggests the yellow bile is out of balance, so that vomiting up the yellow bile is the cure. The humors could also affect a person's personality. An excess of phlegm would make one phlegmatic, of blood, one would be sanguine, of yellow bile, one would be choleric and of black bile, one would be melancholic.

An imbalance in the humors in particular organs could result in illness. Excessive phlegm in the bowels resulted in dysentery and an excess in the lungs caused tuberculosis. Cancer was caused by a massive imbalance in the humors. Stroke was caused by an excess of blood, jaundice by excessive yellow bile and depression by too much black bile.

Dark Ages

The fall of the Roman Empire marked the beginning of the Dark Ages in Europe. The later stages of the Roman Empire were a period of epidemic disease and population decline. The population of cities in particular was to fall and the city's paved roads, drains, aqueducts and public baths soon fell into disrepair. The decline of the cities was accompanied by a decline in classical learning which was opposed by the new Christian church. In 391 CE a Christian mob set fire to the great library of Alexandria and murdered the pagan philosopher Hypathia. The last pagan school of learning, the academy in Athens, was closed in 529 CE by order of the Emperor Justinian.

Medicine was not to escape the general decline of learning which accompanied the fall of the Roman Empire and the arrival of Christianity. There was a return to the belief that the cause of much illness was supernatural. Illness was a punishment from God for people's sins. The curing of such disease by medical practices was contrary to God's will. The only appropriate treatment was prayer and penitence. Diseases might also be caused by witchcraft, possession by demons or spells made by elves and pixies. Some of the old learning did survive, ironically in Christian monasteries where monks copied and translated classical writings. Their work mixed superstition and religion with classical learning and knowledge. Bede, (born 673 CE) an English monk famous for his *Ecclesiastical History of the English People* and one of the most learned men of the Dark Ages, also wrote on medical matters. He referred to Hippocrates and the theory of the four humors and prescribed bleeding as the appropriate treatment for hot fevers caused, as he believed, by an excess of blood. But he also considered magic incantations and the wearing of magic amulets as the way to deal with spells made by pixies. There are also stories of miraculous cures such as a leper sleeping where a saint died and being cured when waking the next morning.

Not much had changed by the 12th century CE when Hildegard of Bingen began to bring together classical medical beliefs with 12th century religious beliefs. She considered the imbalance of the four humors resulted from man's ejection from the Garden of Eden. The eating of the forbidden fruit destroyed the balance of the four humors in the human body. Sin was to cause the imbalance of the humors and was therefore the cause of disease. Some of her medical beliefs could not be regarded as scientific or rational. Her cure for jaundice was to tie a live bat, first to the patients back and then to the patient's stomach. Failing eyesight, caused by excessive lust, was to be cured by placing the skin of a fish’s bladder over the patient's eyes when he goes to sleep, but it had to be taken off by midnight.

Arab medicine

The Moslem prophet Mohammed was born in 570 CE and he and his successors were to conquer an empire extending from Spain to India. The early Moslems had a tolerant attitude to Christian and Jewish minorities who were allowed to freely practice their religions. The origins of Arabian medicine lay with a heretical Christian sect known as the Nestorians. The Nestorians under threat of persecution from orthodox Christians fled eastwards toward present day Iraq and Iran. They brought with them classical texts from a range of authors including Hippocrates, Aristotle and Galen which they proceeded to translate into Arabic. At this time the Arab world had a positive attitude to new ideas and was happy to adopt the ideas of classical scholars like Aristotle and Galen.

The first great Arab medical authority was Rhazes who was born in 854 CE. Rhazes believed illness had nothing to do with evil spirits or God and that classical authorities were not above criticism. He was in frequent disagreement with Galen. He considered Galen’s cure for asthma consisting of a mixture of owl's blood and wine did not work as he had tried it and found it to be useless. He questioned the belief that disease could be diagnosed by studying the patient's urine and was the first medical authority to understand the difference between measles and smallpox. Rhazes gave a full description of diseases he encountered, giving his diagnosis, prognosis and treatment. His understanding of the workings of the human body were however hindered by the Islamic prohibition on dissections of the human body.

Arabian medicine’s second great authority was Avicenna (980-1037) whose book the *Canon of Medicine* was to become the leading medical work in both Europe and the Middle East for some 600 years. Avicenna’s *Canon* includes many of the ideas of Hippocrates, Aristotle and Galen but also includes many of Avicenna’s own ideas. The *Canon* deals with a range of diseases and describes their diagnosis, prognosis and treatment. Avicenna accepted Hippocrates and Galen’s theory of the four humors. Treatments included bleeding, enemas and purges while diagnosis included examining the pulse and urine. Over 700 drugs were recognized by Avicenna and the *Canon* provided instructions on how they were to be prepared, which drugs should be used for which illness and their effects. Wounds were dealt with by cauterizing, a treatment that dates back to Ancient Egypt.

Surgery in the Arab world was not respected and surgeons were usually craftsmen. One exception to this is Albucasis (936-1013) who practiced in Cordoba in southern Spain. Alburasis wrote a book called *Tasrif* or the *Collection* which provided full accounts of surgery practiced at the time. The *Collection* was to become the standard book on surgery during medieval times. The book prescribes a range of surgical procedures including trepanning, dentistry, mastectomy and lithotomy and advocates cauterization as a treatment for a wide range of problems.

Medieval European medicine

European medicine began to move away from the supernatural explanations of disease with the founding of a medical school at Salerno. The school was probably founded in the ninth century and reached its greatest heights between the tenth and thirteenth centuries. Anatomy was taught at Salerno based on the dissection of pigs whose internal organs were thought to be similar to those of humans. *Passionarius*, a book written by Gariopontus, one of the teachers at the school, was based upon classical Greek learning while the arrival in Salerno of Constantine the African around 1075 with many Arab medical works was to greatly improve the medical knowledge at Salerno and eventually all of Europe. Constantine was to spend the remainder of his life translating the Arabic texts into Latin and so bring the classical Greek authors, upon whose work Arabic medicine was based, to Europe.

The translation of Arabic medical texts into Latin continued in early medieval times so that the works of Hippocrates, Aristotle, Galen, Rhazes, Avicenna and Alburasis became well known. They soon assumed a status of great authority and their initial impact was to help free medicine from supernatural and magical explanations and cures. Their status however was eventually to hold back the improvement of European medicine as new ideas contrary to those of the Greek and Arab writers had great difficulty in obtaining acceptance.

New medical schools at Montpellier, Bologna, Paris and Padua were founded that significantly increased medical knowledge. The knowledge of anatomy improved with the occasional human dissection being performed as post-mortem examinations for judicial purposes and with occasional dissections of the bodies of executed criminals. Anatomy was also improved by Mondino de Luzzi or Mundinus who taught at Bologna. His book *Anothomia* brought a new level of knowledge of anatomy, although he did repeat many of the errors of Galen. Mundinus however did most of his dissections himself, unlike other teachers who sat on a high chair somewhat above the body reading a book supposedly describing the dissection, but probably only loosely related to it. Guy de Chaulias, the leading surgeon of the fourteenth century was a pupil of Mundinus.

The most dramatic medical event of the 14th century in Europe was the arrival of the Black Death. It originated in China killing up to two thirds of the population and then spread along trade routes to Europe and the Arab world. It killed half the population of Cairo and between a quarter and a third of the population of Europe. The medical authorities in Europe had no solution to the Black Death. The idea of a contagious disease was beyond the understanding of medical knowledge in either the Arab or European world during the 14th century. The Arabs considered the Black Death was caused by evil spirits, the Europeans blamed everything from the Jews to God's punishment for human sins. Jews were accused of poisoning wells and entire Jewish communities were wiped out by vengeful Christians. Flagellants traveled around Europe whipping themselves for their sins hoping this would appease God. Conventional medicine of the time had no answers, bleeding, cauterizing and cleaning the air with incense were all tried and failed. Quarantining worked to some extent but the best advice was to run like the wind. The failure of conventional medicine during the Black Death led to a revival of supernatural explanations of disease.

The Renaissance

A revolution was to take place in medicine at the time of the Renaissance. It was to involve the breaking of the stranglehold classical and Arabic thought, especially Galen and Avicenna, had on medicine and its replacement by a belief in observation and experiment. One of the principal proponents of the new beliefs was Paracelsus who attacked academic learning, especially Galen and Avicenna and advocated learning from experience. His own ideas however were not much of an improvement on classical learning. He rejected the humoral theory, but considered everything was made out of sulphur, mercury and salt. Sulphur caused inflammability, mercury volatility and salt solidity in substances. He also believed in the "doctrine of signatures", the idea that assumed plants capable of healing visibly showed their healing qualities. Heart shaped lilac leaves would cure heart disease and yellow celandine would cure jaundice.

However, Paracelsus' interest in alchemy led him to some significant discoveries. He noticed the anaesthetic effects of ether and tincture of morphine which he called laudanum. He recognised that particular substances had their own individual qualities and that compounds including those substances often had some of those same qualities. He considered that each disease needed to be cured by its own remedy. The main value of Paracelsus's ideas were in his iconoclastic attack on classical medical learning, which was held in vastly excessive reverence in Paracelsus's time. After Paracelsus it became easier to criticise established medical learning and for new ideas to be accepted.

A contemporary of Paracelsus, Fracantorius, suggested contagious disease was caused by tiny seeds invading the human body. The seeds were too small to be seen with the human eye and could find their way into the body from the air, from bodily contact or from infected clothes or bed linen. Once they had entered the human body they could multiply causing people to fall ill. Fracantorius also considered each disease was caused by its own particular seed, leading Fracantorius to clearly distinguish between such contagious diseases as smallpox, measles, the plague, syphilis and typhus. Previously contagious diseases were sometimes considered to be versions of the same disease with varying degrees of intensity. Fracantorius's theory is virtually identical to the germ theory of disease but in the 16th century, without microscopes, he was unable to prove the theory. Physicians preferred other theories, such as the humoral theory, which while also unprovable at least had the support of tradition and ancient authority.

The study of anatomy was to undergo a revolution at the hands of Vesalius. Vesalius was able to dissect human corpses and this enabled him to provide a generally accurate picture of the human body. Previously anatomy had suffered from the prohibition on human dissection that extended back to classical times, so that knowledge of human anatomy was based on animal dissections. Before Vesalius the accepted authority was Galen whose anatomical studies were based on animal dissection and whose work had acquired such a status that to question it could involve accusations of heresy.

Versalius was able to obtain human corpses for dissection, as public authorities were prepared to allow the dissection of the corpses of executed criminals. Some physicians had previously dissected the corpses of criminals, but such was the reputation of Galen that they had not noticed or not dared to point out that the dissection of humans showed that much of what Galen had said was wrong. Versalius's strength was that he was prepared to rely on his observations and where these contradicted Galen he was prepared to say Galen was wrong.

Vesalius's great work was the *De Humani Corporis Fabrica*, usually called the *Fabrica*. It consisted of seven books, the first dealing with the skeleton, the second with the muscular system, the third with the veins and arteries, the forth with the nervous system, the fifth with the abdominal organs, the sixth with the heart and lungs and the seventh with the brain. The *Fabrica* especially books 1 and 2 were illustrated with high quality drawings showing the various human parts in considerable detail. In book 1 Vesalius emphasises that the bones supported the human body, played an important role in movement and provided protection for other parts of the body. The illustrations in book 2 show the muscles in the order in which a person dissecting a body would see them. The upper layer of muscles are shown then the layer below them and then the next layer and so on. Book 3 gives a good description of the arteries and veins and book 7 describes some of the structure of the brain for the first time.

The book corrected certain of Galen's errors. It questioned Galen's suggestion that blood flowed from the right ventricle of the heart to the left ventricle. Vesalius also showed that *the rete mirabile* did not exist, that the liver was not divided into five lobes, that the uterus had multiple chambers and that the pituitary was directly connected to the nose. Vesalius's exposure of such errors by Galen resulted in some criticism of Vesalius's work from physicians who considered any questioning of Galen to be outrageous.

Vesalius did make some errors. His descriptions of the visceral organs (the liver, the kidney and the uterus) were based upon those of pigs and dogs. He failed to notice the pancreas, the ovaries and the adrenal glands. His description of female organs was poor, probably due to there being fewer female bodies available for dissection. Nevertheless the book still represented an enormous advance in human knowledge of anatomy.

Circulation of the Blood

Classical physicians were aware of the existence of the heart, but had little idea of its function in the human body. They realised when the heart stopped beating, life would stop which led them to believe the heart had a significant role during and at the end of life. They considered the heart was where the soul was located when a person was living and the soul left the body when a person died.

Classical physicians had little understanding of the relationship between the heart and the blood. They did not know how blood got to the heart, how it got from the right ventricle to the left ventricle or what happened after it left the heart. They believed the heart provided a “vital spirit” to blood passing through the heart. They also believed the arteries did not contain blood as when a person or animal dies, the heart stops pumping blood into the arteries, which then contract and drive their blood into the veins. This only leaves air in the arteries of a dead person or animal and classical physicians only dissected dead bodies and so never discovered blood in the arteries. The veins in dead bodies are full of blood, especially the veins connected to the liver. This led classical physicians to believe that the liver created blood which was passed through the veins to the rest of the body. It was also believed that the body somehow absorbed the blood.

Galen, who had the opportunity to observe the internal organs of living human beings while acting as physician to injured gladiators, had a better understanding of the heart and blood. He understood the arteries contain blood in living people and that the heart was a pump which pushed blood from the right ventricle of the heart into the lungs which then flowed into the left ventricle and from there into the arteries. This circulation from the right ventricle to the lungs and then to the left ventricle was known as the pulmonary transit. Galen however still believed that the liver created the blood, but also that it pumped the blood to the rest of the body and that blood was passed directly from the right ventricle to the left ventricle of the heart. The irony is that Galen's work on the pulmonary transit, which was at least partly right was largely not noticed, while other work which was quite erroneous like the humoral theory was treated as holy writ.

The idea of the pulmonary transit was revived by the Arab physician Ibn al-Natis in the 13th century when he suggested that all the blood went from the right ventricle to the lungs and then to the left ventricle and none travelled directly from the right ventricle to the left ventricle.

In the 16th century the same idea was suggested by Michael Servetus and accepted by Realdo Colombo. Colombo also suggested the heart could act as a pump and discovered the presence of valves in the veins which ensured that the blood could move only in a single direction from the right ventricle to the lungs and then to the left ventricle.

The classical ideas concerning the heart and blood were beginning to be challenged in the 16th century. Ideas of the pulmonary transit, the heart acting as a pump and valves in the veins ensuring blood flowed only one way, questioned the classical orthodoxy still largely accepted in Renaissance Europe. Into this environment William Harvey proposed his ideas of the continuous circulation of the blood.

Harvey had been carrying out dissections on a wide range of living animals and it is from his observations of their living organs that he was able to understand how the blood circulates through the human body. His book *De Motu Cordis* begins by explaining the structure of the heart and what it does. The heart consists of two upper parts called the auricles and two lower parts called the ventricles. The left auricle and the left ventricle were separated from the right auricle and the right ventricle by an impenetrable muscular wall. The question of whether the auricles or the ventricles beat first was difficult to resolve as hearts would often beat too fast for normal observation to provide an answer. Harvey answered the question by observing the hearts of cold blooded animals like fish which beat slowly and then confirmed it by observing the slow beating hearts of dying warm blooded animals. He observed the auricles beat first, pushing blood into the ventricle which contracted, pushing blood out of the heart.

The classical theory considered that blood was made by the liver, flowed through the heart and was absorbed by the body. Harvey calculated the amount of blood that flowed through the heart of a dog. He calculated the number of heart beats per minute, which was the number of times the heart pumped blood out into the body. He also calculated the quantity of blood that was pumped with each heartbeat and concluded that the heart pumped blood weighing three times the weight of the whole body each hour. The question arose as to where all this blood came from and where did it all go. Blood equivalent to three times a person's body weight per hour could not come from food and drink consumed. No one could eat or drink that much per hour. Nor could that quantity of blood be absorbed by the body every hour. Veins, arteries and tissues would explode with that quantity of blood being poured into them every hour. Harvey suggested the solution to this problem was that blood was not being created by the liver or absorbed by the body, but that the same blood was constantly circulating around the body.

Galen had suggested that the blood moved in both directions in the veins and arteries. Harvey showed that valves in the veins ensured that blood moved in only one direction. He showed that blood in the veins always moved towards the heart, by pressing a vein, blood accumulated in the vein on the side of the compression away from the heart. The side of the compression close to the heart would be emptied of blood as the blood flowed to the heart and away from the compression point. When an artery was pressed the blood built up on the side of the compression closest to the heart. This indicated the blood flowed in a single direction, in the veins towards the heart, and in the arteries away from the heart.

The consequences of the blood all flowing in one direction and the same blood constantly being circulated, without blood being created by the liver or absorbed by the body, was a revolution in physiology. New ideas often receive considerable criticism and Harvey's idea of constantly circulating blood was attacked for daring to disagree with Galen. One rational criticism of Harvey's theory was that Harvey could not show how blood flowing out of the heart to the arteries could connect to the veins and flow back into the heart. Harvey suggested tiny connections, too small to be seen with the naked eye, linked the arteries and the veins but he could not prove their existence. This problem was solved by Marcello Malpighi, in 1661, when using a microscope he was able to observe the existence of capillaries linking the arteries and the veins which allowed blood to flow from the arteries to the veins so that the idea of the circulation of the blood was complete.

Jenner and vaccination

Smallpox goes back at least to Ancient Egypt and was in Greece in the classical period and was present in Ancient China and India. The symptoms of the disease were described by Al-Razi in 910 CE and involved blisters filled with pus appearing on the eyes, face, arms and legs. Twenty to forty percent of those who caught smallpox died from it and the survivors were covered with disfiguring scars. In London in the 17th and 18th centuries a third of the people had smallpox scars and the majority of cases of blindness were caused by smallpox.

It had been observed that people who survived smallpox did not usually catch it again. The idea developed that if a mild case of smallpox could be produced it would protect a person from future smallpox attacks. In the East dust from a smallpox scab was blown into the nose to induce a mild case of smallpox to create immunity from future attacks. In Ottoman Turkey smallpox material was rubbed into small cuts made in a person's arm. These methods of conferring immunity from smallpox were made known in England in the early 18th century but were ignored.

The practice of deliberately giving a person a mild case of smallpox began in England in the early 18th century with Lady Mary Montagu. The practice became known as variolation and Lady Montagu, who had learnt about the practice in Turkey, had her own daughter variolated in the presence of newspaper reporters, which ensured substantial publicity. Lady Montagu then persuaded the Prince and Princess of Wales to have their children variolated which ensured even more publicity. Variolation also took place in America where Zabdiel Boylston, a Boston physician, heard of variolation from an African slave and faced with a smallpox epidemic variolated 244 people of whom only 6 died. Surgeons however demanded patients go through a 6 week period of bleeding, purging and dieting before variolation which limited the popularity of the practice and resulted in patients being weakened before variolation took place. Variolation turned out to be quite dangerous with modern estimates that 12% of patients died, a lower death rate than the 20-40% who might die in a smallpox epidemic, but certainly not a perfect treatment for the problem of smallpox.

A better treatment was to come with Edward Jenner, who while training as a surgeon in 1768, heard that milkmaids who had contracted cowpox were immune from smallpox. Cowpox resulted in lesions on the milkmaid's hands, but had no other symptoms. Later Jenner met a Mr Frewster who in 1765 had presented a paper to the London medical society on the ability of cowpox to prevent future smallpox attacks. The paper was never published but reminded Jenner of what he had heard of cowpox from the milkmaids. Cowpox is part of a family of animal poxes, including horsepox, cowpox, swinepox and smallpox, all caused by the orthopox virus. All the animal pox diseases can infect humans and an infection from any of them will protect people from all the other animal poxes. In December 1789 Jenner began a series of experiments. He inoculated three people including his son with swinepox and later variolated them with smallpox and none of them produced the rash that usually came from variolation with smallpox. Swinepox seemed to protect them from smallpox. Later in 1796 Jenner put cowpox into a healthy 8 year old boy and after he developed normal cowpox symptoms variolated him with smallpox. The boy did not develop any of the symptoms that normally occurred with variolation with smallpox. Jenner then took fluid from the boy's cowpox pustule and used it to inoculate some more children and fluid from their cowpox pustules was used to inoculate some more children. Two of these were later variolated with smallpox, but did not develop any of the symptoms that normally occurred with variolation, confirming the initial experiment. The experiment showed that cowpox could provide protection against smallpox without any of the risks of variolation. The practice of cowpox inoculations, which began to be called vaccination, was soon done throughout the British Empire, the United States and Europe although there was some opposition to it. The opposition gradually disappeared and eventually late in the twentieth century smallpox was completely eliminated.

The discovery of anaesthesia

A vital component of modern surgical operations is the use of anesthesia. Without anesthesia operations would be excruciatingly painful and as a result many patients chose not to have operations. The pain of having limbs amputated could result in patients dying of shock and forced surgeons to perform operations with extreme speed. The best surgeons could amputate a limb in less than a minute. The state of mind of a person awaiting surgery would be similar to that of a person about to be tortured or executed. When London hospital was built in 1791, and was to act as a model for other hospitals, the design took into account the lack of effective anesthetics. The operating room was on the top floor, partly to allow sunlight through a skylight to illuminate the operation, but also so the patients' screams would not travel through the hospital and could be muffled by extra heavy doors. When an operation was to commence hospital staff would go to the top floor and assist in holding the patient down and if necessary in gagging the patient.

The problem with an effective aesthetic that will allow major surgery is that it must place the patient in a state where the central nervous system is depressed to an extent where painful stimuli cause no muscular or other reflexes. This is far beyond ordinary sleep as obviously performing surgery on a sleeping person will wake them. Effective surgical anaesthesia must place the patient in a state close to that of death.

In the past various attempts were made to reduce or eliminate pain during surgical procedures. Dioscorides, a Greek physician in the early Roman Empire, used drugs such as henbane and mandrake root to relieve pain. These drugs continued to be used into medieval times. Arab physicians seemed to have used drugs such as opium and hyoscyamus. Alcohol was often used but was probably more effective at making the patient easier to hold down than in relieving pain. Soporific sponges, involving the inhalation of drugs such as opium, mandragora and hyoscyamus, were used from around the ninth century. However, modern experiments with such sponges suggest they had no aesthetic effect at all. The use of soporific sponges was discontinued in the seventeenth century. It may well be due to the lack of effectiveness of pre-modern anaesthetics that their use was not widespread. Egyptian papyri and the Code of Hammurabi describe surgery without mention of anaesthetics. Only one Chinese surgeon, one Indian surgeon and a few Greek, Roman and Arab surgeons seem to have made any attempt to relieve pain during surgery. Pre-modern attempts to relieve pain during surgical operations seem to have been of little or no effect.

The first step in the development of modern anaesthetics was the discovery of ether. In 1275, the Spanish alchemist Raymundus Lullius produced ether by mixing alcohol with sulfuric acid. Paracelus used ether to relieve pain in 1605 in some of his medical patients but not in surgery as he was not a surgeon.

Nitrous oxide, soon to be known as laughing gas, was discovered by Joseph Priestly in 1772. Priestley however did not realise nitrous oxide could act as an anaesthetic. Others however soon discovered both nitrous oxide and ether had an intoxicating effect when inhaled and soon “ether frolics” and “laughing gas parties” became a popular source of amusement. It was soon observed that minor injuries such as bruises received at the frolics and parties were not accompanied by any pain. In addition, Humphrey Davy discovered that nitrous oxide relieved the pain of an inflamed gum and jaw and suggested nitrous oxide could be used in surgery. Similar observations concerning nitrous oxide were made by William Barton in the United States. In 1842 ether was used to painlessly extract a tooth, by a dentist, Dr Elija Pope, acting on the suggestion of William Clark, a chemistry student who had participated in ether frolics.

The first use of ether for surgical purposes was by Crawford Long in Georgia, USA in 1842. Long had attended ether frolics and had noticed bruises he had received while under the influence of ether had involved no pain. Realising that ether had stopped the pain, he used it in various surgical operations and in obstetrical procedures. He did not, however, publish his work until 1849.

A dentist, Horace Wells, while attending a nitrous oxide party in 1844 noticed a person injuring his legs without suffering any pain. Realising nitrous oxide could serve as a dental anaesthetic, Wells had one of his own decaying teeth removed by another dentist while he was under the influence of nitrous oxide. Wells experienced no pain and was soon performing dentistry using nitrous oxide on his own patients. However, when he attempted a public demonstration at Massachusetts General Hospital he used insufficient gas and the demonstration was not a success.

The public demonstration at Massachusetts General Hospital had been arranged by Wells' former dentistry partner William Morton. Morton, who had possibly seen Long operate in Georgia, became interested in ether as an anaesthetic and had discussed it with Charles Jackson, a doctor in Harvard's medical faculty and at Massachusetts General Hospital. Intending to patent the anaesthetic, Morton and Jackson disguised the ether by mixing it with aromatic oils and called it Letheon. They then arranged public demonstrations of the use of Letheon, in 1846, for pulling teeth and for an operation removing a tumour from a patient's jaw. Both the dentistry and the operation were carried out painlessly. Jackson and Morton however were forced to withdraw the patent for Letheon and reveal that Letheon was really ether by pressure from the surgeons involved in the operations. By the end of 1846 news of the use of ether as an anaesthetic had travelled across the Atlantic and in December 1846 it was used in an operation in London.

Jackson, Morton and Wells all claimed to be the discoverer of surgical anaesthesia and in 1847 the United States Congress became involved in trying to sort out who was the true discoverer of anaesthesia. Congress eventually dismissed Wells and Mortons claims and decided it was between Jackson and Long. The American Medical Association, in 1872, gave the credit to Wells, while in 1913 the electors of the New York University Hall of Fame named Morton as the discoverer of surgical anaesthesia. The American College of Surgeons, in 1921, decided Long should be credited with the discovery.

Attempts were soon made to use ether in obstetrics but it was found to be unsuitable. Ether often produced vomiting patients, irritated lungs and a bad smell. Chloroform had been discovered independently in 1831 by Samuel Gutherie in New York, by Eugene Soubeiran in Paris and by Liebig. Initially its anaesthetic quality was not recognised but Gutherie's daughter had become unconscious for several hours after tasting it. In 1847 Sir James Simpson, while looking for an anaesthetic to use in obstetrics, tried chloroform on himself and having found it to be an effective anaesthetic, began using it in surgical operations. Its use was soon extended to obstetrics, provoking considerable opposition from the Calvinist Church in Scotland on the grounds the Bible stated "In sorrow thou shalt bring forth children" showed women must suffer when giving birth. The Calvinist church opposition disappeared when Queen Victoria gave birth to her eighth child under the influence of chloroform. However chloroform was soon discovered to have its own problems as it could cause liver damage and five times as many people died under chloroform as died under ether.

The method of application of the anaesthetic developed over time. Long had simply poured ether into a towel for his patient to inhale. Morton used an inhaler made up of a round glass bottle with two holes and a mouth piece. Air passed through one hole into the bottle which contained a sponge soaked in the ether which was then inhaled by the patient through the mouthpiece which was attached to the other hole. Morton's inhaler did not allow the anaesthetist to have control over the amount of anaesthetic. Soon John Snow, who had provided the chloroform to Queen Victoria, created an improved inhaler which provided a 4% mix of chloroform in air. Joseph Clover produced a further improved inhaler in which the chloroform and air mixture was prepared in advance and held in an airtight bag. Sir Francis Shipway created an apparatus which allowed the anaesthetist to control a mixture of varying amounts of chloroform, ether and oxygen for inhalation by the patient.

A significant improvement in the provision of anaesthetics occurred with the introduction of the anaesthetic directly into the windpipe or trachea. This was first attempted by Frederick Trendelenburg, in 1869, who inserted the anaesthetic through a tube he inserted into a hole he had cut into the patient's windpipe. Sir Ian Macewan achieved the same result without cutting into the windpipe, in 1880, by inserting a metal pipe down the throat and into the windpipe. This allowed the development of endotracheal anaesthesia which was important for operations on the mouth and the jaw and for many modern cardiac and pulmonary operations. Endotracheal anaesthesia was further improved, in 1919, when Sir Ian Magill put tubes through the conscious patient's nose and mouth and down into the windpipe by anaesthetizing the throat with cocaine before inserting the tubes.

General anaesthetics were often not necessary for minor operations. A local anaesthetic which worked on a particular part of the body and avoided the small risk of death and several hours of recovery time involved with general anaesthetics was sought. Peruvian Indians knew about the anaesthetic qualities of the coca plants and in the nineteenth century cocaine was obtained from the plant. In 1872 Alexander Bennett observed that cocaine had anaesthetic properties and in the 1880's Carl Koller experimented with cocaine using it to anaesthetize frogs eyes. Soon cocaine began to be used as a local anaesthetic for eyes, the mouth, nose and throat and in the urethra. The use of cocaine was extended by injecting it into the nerves relating to the area to be operated on and eventually into the epidural space around the spinal cord which allowed a larger area to be anaesthetized. The use of cocaine as a local anaesthetic was discontinued with its replacement by novocaine which was synthesized as an aesthetic after 1905.

The germ theory of disease

The first person to see microorganisms was Anthony Leeuwenhook (1632-1723) a Dutch draper who was an expert maker of microscopes. His microscopes gave a degree of magnification which was not exceeded until the 19th century. He used his microscopes for observing a wide variety of phenomena. In 1675 and 1676 he looked at drops of rain water and found tiny animals within the water. Those animals would have included what we now call bacteria and other microorganisms. In 1683 Leeuwenhook looked at plaque from his own teeth and found it contained large numbers of small animals. Later samples of plaque did not contain the small animals, which Leeuwenhook suspected was because his drinking of hot coffee killed the little animals. Leeuwenhook also looked at scrapings from his tongue when he was sick and at the decay in the roots of a rotten tooth he had removed. In both cases he found vast numbers of the little animals. The presence of these animals in such great numbers in places of illness and decay raised the question as to whether the animals arose from the decay or whether they were attracted to it or whether they caused the decay. The question of whether the small animals were spontaneously generated from decaying materials or were attracted to it was the subject of much controversy. Francesco Redi (1626-1698) kept boiled meat in sealed containers and when maggots failed to appear suggested this showed there was no spontaneous generation. However in 1748 John Needham repeated the experiment and found small animals in the meat which he considered proved spontaneous generation. Lazzaro Spallanzoni suggested Needham had failed to seal his containers properly so that the small animals arrived on the meat through the air, rather than being spontaneously generated by the meat. Supporters of spontaneous generation argued that sealing the containers prevented some gaseous substance, necessary for spontaneous generation, from reaching the meat and so preventing the generation of the living organisms.

Whether micro-organisms caused the diseases they were so often found with was investigated by Agostino Bassi. In 1835 he showed that the silkworm disease, muscarine, was caused by bacteria. When he inoculated healthy silkworms with the bacteria, he produced the sickness in the silkworms. This suggested that other diseases may be caused by bacteria.

The question of spontaneous generation and whether micro-organisms played any role in causing disease were eventually settled by Louis Pasteur. He was to show that fermentation in wine, putrefaction of meat and infection in human disease all involved the same process and were all caused by the activities of microorganisms. The micro-organisms were generated not by decaying matter but were continually present in the air and when they were present in great numbers and were of unusual strength they could cause matter to decay and human beings to fall ill.

Pasteur began with fermentation in wine. At the time chemists such as Wohler and Justus von Liebig suggested fermentation was solely a chemical process with living organisms playing no role in the process. Fermentation in wine was a problem as sometimes the fermentation went wrong and soured the wine. Pasteur showed that fermentation was caused by microorganisms in yeast and that round yeast cells produced good wine, but long yeast cells created lactic acid which caused the wine to go sour. Pasteur showed that if the wine was heated it would kill the yeast and stop any of the wine going sour.

Pasteur next began to investigate putrefaction in meat with an experiment that allowed air to reach boiled meat via an undulating u-shaped tube. The meat did not putrefy and Pasteur considered this was because the dust particles containing the micro-organisms were caught on the low bend of the tube as they could not travel up the tube due to gravity. The micro-organisms did not reach the meat even though it was exposed to air so the meat did not putrefy. This showed it was not air that caused putrefaction, but microorganisms in the air.

Pasteur then began to investigate diseases in living organisms, first with silkworms and then anthrax which effects sheep and cattle and occasionally humans. Pasteur showed the disease killing silkworms were two different sorts of micro-organisms which caused two different diseases in the silkworms. In relation to anthrax it was already known that the blood of cattle who had died from anthrax contained micro-organisms and that these microorganisms were the cause of the disease. Robert Koch had discovered the anthrax bacteria, had cultured it and injected it into animals who had immediately died. He also found that anthrax micro-organisms could sometimes form spores, which were tiny organisms resistant to a range of environmental conditions. The spores were formed when the temperature was right and oxygen was present. Once the spores were formed they could survive for a considerable time and re-infect other animals making the disease difficult to control. Pasteur, with some difficulty, then produced an anthrax vaccine which he used to inoculate sheep who were later injected with the anthrax bacteria. The sheep did not develop anthrax and Pasteur had found a vaccine for anthrax.

Pasteur's last great achievement was to discover a vaccine for rabies. Rabies normally occurs in humans after they have been bitten by a rabid dog with the symptoms appearing between 10 days and several months after the dog bites took place. Pasteur studied the tissues of rabid dogs but could not find a microorganism that could have caused rabies. He decided the organism was too small to be detected with a microscope. Pasteur considered that the microorganism entered the body through the bite wound and over time moved to the brain, explaining the period of time between the bite and the arrival of symptoms. After some time Pasteur was able to produce a vaccine for rabies which was able to be injected in the period after the dog bite and before the onset of symptoms.

Pasteur's work had followed a logical path. He had first shown that fermentation was caused by microorganisms, that those micro-organisms originated in the air rather than from the fermenting matter and that microorganisms also caused putrefaction and infectious disease. He then showed how the diseases in both animals and people could be cured by vaccination. Pasteur's work established the germ theory of disease and put an end to other theories of disease such as the humoral theory.

Robert Koch, after isolating the anthrax bacteria, began using an improved microscope with a light condenser and an oil immersion lens. This enabled him to see bacteria that had previously been too small to be seen even with the best microscopes available. He also used new aniline dyes which helped him to distinguish between different types of bacteria. Koch also found a way of producing pure cultures of different types of bacteria by placing the bacteria on a solid culture medium, in place of the liquid culture medium then currently used, which only worked well with bacteria that moved in the bloodstream. With his improved microscope and better techniques for creating pure cultures of bacteria Koch began to search for tuberculosis bacteria in the tissue of humans who had died of tuberculosis. Using a microscope equipped with the oil immersion lens and condenser that was five times as powerful as Leeuwenhook's microscopes he was able to find a tiny bacteria which he called the tubercle bacillus. The tubercle bacillus was much smaller than the anthrax bacteria and was too small to be found without the use of his new improved microscope. To prove the tubercle bacillus caused tuberculosis, Koch needed to isolate it in a pure culture and to inject it into various animals. If it produced tuberculosis in those animals that would prove the tubercle bacillus was the cause of tuberculosis. After some difficulty he was able to produce a pure culture of the tubercle bacilli. He then injected this into animals which soon became sick and when he examined their diseased tissues he found they had tuberculosis. Koch had found the cause of tuberculosis, giving hope that a cure would eventually become possible.

If Pasteur established the germ theory of disease, it was Koch who was to turn bacteriology into a science. Koch formalized the methods for studying microorganisms and proving their relationship with particular diseases. To prove an organism was the cause of a disease Koch proposed the following criteria, which came to be known as Koch's postulates:

1. The organism must be present in every case of the disease.
2. It must be possible to prepare a pure culture, maintainable over repeated generations.
3. The disease must be reproduced in animals using the pure culture, several generations removed from the organism originally isolated.
4. The organism must be able to be recovered from the inoculated animal and be reproduced again in a pure culture.

Clearly the third and fourth postulates can only apply to diseases which apply to animals as well as humans and the postulates were not able to be applied to all micro-organisms for example viruses. Nevertheless the postulates provided a set of procedures for the investigation of diseases which were to establish the causes of a range of diseases which opened up the possibility of finding cures and treatments for the diseases. Between 1879 and 1906 the micro-organisms causing many diseases were discovered. The diseases involved included gonorrhoea (1879), typhoid fever (1880), suppuration (1881), glanders (1882), tuberculosis (1882), pneumonia (1882 and 1883), erysipelas (1883), cholera (1883), diphtheria (1883-4), tetanus (1884), cerebrospinal meningitis (1887), food poisoning (1888), soft chancre (1889), influenza (1892), gas-gangrene (1892), plague (1894), pseudo-tuberculosis of cattle (1895), botulism (1896), bacillary dysentery (1898), paratyphoid fever (1900) syphilis (1905), and whooping cough (1906). The discovery of the microorganism causing the disease did not always result in effective treatments.

Antiseptics

The increase in surgery produced by the use of anaesthetics simply highlighted another problem, the death of large numbers of patients due to infection. Patients dying from infection had long been a problem both in obstetrics and surgery. It was in obstetrics that the first understanding of the causes of infection arose, but it was in surgery that the solution to the problem was achieved.

Some doctors and surgeons sensed that a lack of cleanliness may be the cause of infection. Charles White in 1773 in Manchester suggested the cleaning of the surgery room, clothing and articles in contact with the patients but did not refer to cleansing of surgeons and others involved in operations. Alexander Gordon (1752-1799) suggested infection was carried from infected patients to uninfected patients. He suggested the cleansing of surgeons but did not realise that infected matter was involved in the spread of disease.

In the mid nineteenth century Ignaz Semmelweis was working at the maternity clinic at Vienna General Hospital. He noticed that the section of the hospital used for training medical students in obstetrics had a much higher rate of mortality, around 13% than the section used to train midwives, which was around 2-3%. Explanations considered for the variations in the mortality rates included that the poor single mothers and prostitutes in the hospital were less embarrassed when treated by women. Semmelweis noticed that the puerperal fever which killed many of the women immediately after they had given birth seemed to be the same disease that had killed the surgeon Jakob Kolletschka who died after cutting his finger in a post mortem. Later Semmelweis realised that medical students going to their section of the maternity clinic came from anatomy classes involving dissections and the handling of diseased body parts. Little attempt was made to clean up between the anatomy classes and the work done in the maternity clinic. Semmelweis suspected the students coming from the anatomy classes were bringing infection into the maternity clinic so he ordered students to wash and scrub in a chlorine solution before entering the maternity clinic. Within a month the mortality rate in the students section dropped to 2%, the same as for the midwives section. Despite his success Semmelweis became very unpopular with the medical students, his immediate superior and even the patients who felt he was suggesting they were dirty. Semmelweis left Vienna for a hospital in Budapest where he instituted similar hygienic reforms and again the mortality rate dropped dramatically. He published a paper on his discoveries, which was ignored, and then a book which was also ignored. Semmelweis then began to behave erratically, writing angry letters to those who criticised his work. He was soon induced or forced to enter a mental hospital and within two weeks was dead in circumstances that may have amounted to murder.

Joseph Lister was a surgeon in Glasgow who noticed that the mortality rate for compound bone fractures where the bone was exposed to the air was much higher than for broken bones where there was no exposure to the air. Broken bones exposed to the air often developed gangrene which was usually blamed on “miasma” or bad air. Lister did some experiments on frogs legs and concluded that gangrene was a form of rotting, involving the decomposition of organic material. He also read Pasteur’s work which suggested that putrefaction was the rotting of organic material caused by bacteria in the air. Lister accepted Pasteur’s idea that it was not the air that caused the gangrene but bacteria in the air.

The question was how to destroy the bacteria both in the air and in the wounds. Carbolic acid or phenol had been isolated in the 1830’s through coal tar distillation. It was used to clean sewers and after various experiments with crude carbolic, which killed tissue, Lister began to use carbolic acid. He would dress wounds in lint soaked with carbolic acid and spray the air in the operating room with carbolic acid. Lister published his work in 1867 in a paper entitled *On the Antiseptic Principle in the Practice of Surgery*. The mortality rates from Lister’s amputation operations fell from 45% to 15%, but despite this some doctors still refused to believe that bacteria existed or could cause infection. However, the results of using Lister’s methods soon became obvious and they began to be used throughout Europe. Over time he refined his procedures, getting rid of the carbolic spray and putting greater emphasis on using heat to sterilize dressings and instruments. There was also a move from anti-septic measures which destroyed germs in wounds to aseptic measures which ensured that everything that touches the wound such as instruments and the surgeon's hands are free from germs. Towards the end of the 19th century sterilized gowns, masks, caps and rubber gloves were introduced for surgical operations.

Antibiotics

Scientists experimenting with bacteria had on various occasions noticed that penicillin and other biological organisms could inhibit the growth of bacteria. In 1875 John Tyndall had observed penicillin had killed bacteria in some of his test tubes. In 1877 Pasteur had noted anthrax bacilli grew in sterile urine but the addition of “common bacteria” stopped the growth. In 1885 Arnaldo Canteri noted certain bacterial strains killed tubercle bacilli and reduced fever in the throat of a tubercular child. In 1896 a French medical student noted that animals inoculated with penicillin and a virulent bacteria did better than animals inoculated with the virulent bacteria only. In 1925 D A Gratia noted that penicillin could kill anthrax bacilli.

Alexander Fleming was experimenting with bacteria in 1928 when he observed bacteria in his petri dish had been killed by the *Penicillium* mould. Fleming began experimenting with the mould and soon isolated the substance that killed the bacteria. He called the substance penicillin and then tested its effectiveness against other bacteria. He found penicillin could kill a range of bacteria but there were some bacteria it did not affect. He injected it into animals and found that it did not do them any harm. Fleming then published his results in 1929 and then in a briefer report in 1932. Fleming's work was largely ignored and he then turned his research interests elsewhere. The prevailing scientific view at the time was that antibacterial drugs would not work against infectious disease and would be too toxic to use on humans. This belief was to change after 1935 when it was found that Prontosil could destroy streptococcal infection when given intravenously. Research on penicillin only began again in 1940, in Oxford, when Howard Florey and Ernest Chain discovered that penicillin was an unstable simple molecule. They were able to stabilize it by freeze drying it in a water solution. This produced a powder that was tested on mice and did not harm them and cured them of streptococci. It was also discovered that penicillin could travel through the body to attack infections wherever they were. Their results were published in August 1940 and Florey, Chain and their colleagues began to manufacture penicillin as fast as possible. The first human test of penicillin was on a badly ill policeman. The policeman improved until he seemed on the verge of total recovery when the supply of penicillin ran out and the policeman relapsed and died. More penicillin was manufactured and tested on humans and was found to regularly clean up infections. It was found to be effective against most forms of pus forming cocci and against tetanus, anthrax, syphilis and pneumonia. The manufacture of penicillin was greatly expanded when the United States began to produce it and new manufacturing techniques involving deep fermentation were developed. This involved submerging the mould below the surface of the culture medium. Eventually semisynthetic penicillins and penicillins that could be swallowed were produced.

Eventually a systematic search began for other antibiotics. Howard Florey outlined the procedure to be followed which involved the investigation of microorganisms to find out which ones produced an antibacterial substance, the isolation of that substance, testing the substance for toxicity, testing it in animal experiments and then testing it on people. The search for new antibiotics was to produce a substantial number of new antibiotics including streptomycin developed in 1944 which was effective against tuberculosis. Chloramphenicol, developed in 1949, was effective against typhoid fever. Antibiotics were eventually found that could act against every bacteria that causes diseases in humans. Some of those bacteria are now developing resistance to antibiotics and the development of new antibiotics is inhibited by the extreme cost, running into hundreds of millions of dollars, of obtaining United States government approval for the drugs. Nevertheless antibiotics have saved hundreds of millions of lives.

Medical Statistics

The use of statistics in medicine to determine the cause of disease or the success of a treatment has a relatively short history. In the past the causes of disease and the success of treatments were usually decided by physicians' personal experience with patients, which, assuming that physicians had similar experiences, lead to accepted beliefs as to the efficacy of treatments and the causes of disease. The beliefs would be recorded in authoritative medical texts and would in many cases become a sort of medical dogma. Disputing the dogma could involve accusations of unorthodox opinions that could lead to bad practices that could endanger patients' lives.

The idea of doing trials to test the effectiveness of medical treatments was suggested by the scientist, Johannes van Helmont and the philosopher George Berkeley. The first known trial to assess the cause of a disease seems to have been done by James Lind in an attempt to discover the cause of scurvy. Scurvy was killing large numbers of sailors on long sea voyages. Lind took 12 scurvy sufferers and divided them into 6 groups of 2 and each group was given a different dietary supplement. The two sailors given oranges and lemons rapidly recovered and the others did not. Lind eventually published his findings, and although there remained some confusion for sometime, eventually lemon juice became standard on long sea voyages.

One question, much debated in the 18th century, was whether smallpox inoculation was a good thing. In England inoculation was generally favoured, in France it was opposed. Various calculations were made as to the death rate from smallpox which was considered to be around one in ten, excluding fatalities of those under 2 years old. Other calculations were 1 in 12 and 1 in 7. This was compared to the death rate from inoculation which James Jurin, secretary of the Royal Society, calculated at 1 in 91. The Swiss mathematician, Daniel Bernoulli calculated that inoculation increased the average life expectancy by two years. A further problem was that people inoculated with smallpox could spread it to others and this was not taken into account in calculating death rates from inoculation. If people who were inoculated could be isolated for a period, then the figure might not be too high, but then if people who got smallpox naturally were isolated that would reduce the death rate from normal smallpox. An additional problem was that the rate of smallpox infection varied considerably from large cities where nearly everyone would, sooner or later get smallpox and the small towns and villages where most people in the 18th century lived, and many people could live their lives without getting smallpox. Modern estimates of the death rate from inoculation are as high as 12%, not much better than the death rate from normal smallpox infection.

The difficulty in calculating accurate death rates for inoculation and for normal smallpox infection, how to introduce into the figures people who caught smallpox from those who were inoculated and how to deal with the widely varying rates of smallpox infection between urban and rural areas gives some idea of the difficulty in working out whether inoculation was a good thing or a bad thing. The whole debate eventually became irrelevant when vaccination with cowpox, a quite safe form of immunization, became available at the end of the 18th century. A further illustration of the problem of accurate statistical analysis of medical treatments is contained in the work of Pierre Louis in the first half of the nineteenth century. Louis conducted several trials to test bloodletting as a treatment for various inflammatory diseases. He concluded from his trials that bleeding resulted in patients recovering earlier than if there was no bleeding and that if bleeding is done, patients who bleed earlier during the course of the disease recovered more quickly than those bleed later. However, the way Louis conducted the trial was not ideal. Those bleed earlier during the illness were on average 8 years and 5 months younger than those bleed later, which could explain the faster recovery. A further criticism of Louis's study was that the numbers involved in his trial were insufficient so there was a wide margin of error in his results so they were not reliable.

A more successful use of statistics to discover the cause of disease occurred in the mid 19th century when John Snow discovered the cause of cholera. Cholera, like many infectious diseases, was assumed to be caused by miasma or bad air caused by putrefaction. Snow suspected that cholera could be transmitted by personal contact and through polluted water supplies. He examined the sources of the water supplies in London and compared it to mortality rates from cholera. Areas with clean water supplies, due to water being taken from the Thames above sewage outfalls, or with filtered water, or with water passed through settlement ponds, showed much lower rates of cholera than areas using unfiltered and unponded water taken from below sewage outlets. Areas with clean water had a death rate of 10 per 10,000 from cholera, areas with polluted water had a death rate of 110 per 10,000 from cholera.

Snow also investigated the cholera levels for households in the same areas, where the water supplies came from two separate companies, one of which supplied clean water to its customers and the other which supplied polluted water. Those customers obtaining clean water had 5 cholera deaths per 10,000, those obtaining polluted water had 71 cholera deaths per 10,000. The 5 cholera deaths per 10,000 could have been caused by visiting houses, pubs and cafes with polluted water and people who had fallen sick with cholera.

Snow's final study concerned a small area around Broad Street in London where 500 people died of cholera in ten days. Snow suspected a water pump supplying drinking water in the centre of the area could be responsible so he asked the local authority to remove the handle from the pump. This was done and the cholera outbreak ended. More particularly, Snow showed certain groups within the Broad Street area, people in a workhouse and those working in a brewery who did not use water from the pump, had an unusually low cholera death rate. He also showed that certain individuals from outside the Broad street area who drank water from the pump also died of cholera within the ten day period.

Snow's three studies provided powerful evidence that polluted water caused cholera but his findings were initially rejected. Two inquiries considered cholera still came from bad air and another study which concluded that the death rate from cholera rose as one moved from highlands to sea level also suggested bad air was to blame. Eventually when miasmic theories of disease lost credibility with the rise of the germ theory of disease, Snow's explanation of cholera was accepted.

The first truly scientific randomised control test was conducted on the drugs streptomycin and PAS as a treatment for tuberculosis. Tuberculosis in the mid twentieth century, was the most common fatal infectious disease in the western world. Its cause, the tubercle bacillus, had been identified by Robert Koch in 1885, but no effective treatment had been found for it. Antibiotics like penicillin did not work against it, as it had an impermeable waxy coat that protected it from antibiotics.

A new drug called streptomycin had been discovered in America in 1944 which seemed to work against tuberculosis germs. It inhibited the growth of tuberculosis bacillus on agar plates and was successful at curing tuberculosis in guinea pigs and when tried on a human patient with five courses of treatment between November 1944 and April 1945, cured the human patient. A second drug which showed promise as a tuberculosis treatment was PAS. It had been noted that Aspirin resulted in the tuberculosis bacilli absorbing increased amounts of oxygen and it was considered that a similar drug to Aspirin might block the supply of oxygen to the tubercle bacilli. PAS was tried and was shown to cause an improvement in the condition of tuberculosis patients.

Immediately after World War II Britain was short of money and could afford only a very small amount of streptomycin. The Tuberculosis Trial Committee, encouraged by one of its members Austin Bradford Hill, recognised there was not enough streptomycin to provide to all patients, decided to conduct a random control test with the streptomycin, providing streptomycin to one set of patients and comparing the results with another set of patients not receiving the drug. There was enough streptomycin to provide to 55 patients and the results of the treatment were compared with 52 patients who received the usual treatment provided for tuberculosis patients. Which patients received the streptomycin and which received the usual tuberculosis treatment was decided completely at random to avoid any conscious or unconscious bias in the allocation of patients to either group.

Six months after the trial had begun it was found that only four patients had died from the group given streptomycin while fourteen had died from the group receiving the conventional treatment. Streptomycin seemed to be an effective treatment with significantly fewer deaths in the group receiving the streptomycin. However a follow-up investigation, three years later, revealed 32 of the group using the streptomycin had died compared to 35 in the group not receiving the drug. After three years the group using the streptomycin was only slightly better off than the group not using it. What had happened was that over the period of treatment some of the tubercle bacilli had become resistant to the streptomycin and when this happened patients who initially seemed to be getting better, worsened and often died. The test revealed that not only did streptomycin not work in the longer term but that there was a problem of the bacilli becoming resistant to the streptomycin which, if it could be overcome could mean that streptomycin could still be an effective treatment for tuberculosis. If the drugs had simply been provided to doctors for treating patients it would have taken much longer to work out why it was not working.

A further trial was conducted which combined streptomycin with PAS with the aim of overcoming the problem of resistance from the tubercle bacilli. In the second trial resistance to streptomycin developed in only 5 patients compared to 33 in the first trial. The combination of the two drugs proved to be an effective treatment for tuberculosis and survival rates for tuberculosis patients went up to 80%. Eventually other drugs such as isoniazid and rifampicin were introduced and it was found that combining three drugs resulted in survival rates approaching 100%.

Random controlled trials were also found to be effective in proving the causes of certain diseases. After World War II the great majority of the adult population smoked and lung cancer deaths were rapidly increasing. Bradford Hill, Edward Kennaway, Percy Stock and Dr Richard Doll were asked to investigate whether smoking was a cause of the increasing number of lung cancer deaths. Smoking was only one possible explanation, others such as increased air pollution especially from motor vehicles were considered to be as likely or more likely the cause of increased lung cancer deaths, than smoking. The asphalting of roads was considered to be another possible cause of the escalating lung cancer deaths. Given that most adults smoked it was difficult to find a suitable control group of non-smokers. The investigation was conducted by creating a detailed questionnaire which patients suspected of having lung cancer completed. The questionnaire was also completed by patients who had other cancers and also by patients in hospital for reasons other than cancer to act as two control groups. It was found that 99.7% of the lung cancer patients smoked against 95.8% of the control group patients. This was not a great difference but it was also found that 4.9% of the lung cancer patients smoked 50 cigarettes a day as opposed to only 2% of the control group patients. The lung cancer rate amongst those smoking 50 cigarettes a day was over double for lung cancer patients than for the control group. The more people smoked the greater their chances of getting lung cancer.

The study conducted by Doll and Bradford Hill had looked at lung cancer patients and looked back in time at their smoking habits. They then decided to do a study of healthy people investigating their smoking habits and then observing how their health developed in the future. Doll and Bradford Hill decided to do the study on doctors, 40,000 of whom filled in and returned their questionnaires. Two and a half years later enough doctors had died for Doll and Bradford Hill to be able to show that the more the doctors smoked the greater the likelihood they had died of lung cancer. It was eventually found that doctors smoking 25 cigarettes per day were 25 times as likely to develop lung cancer compared to non-smokers.

The success of the random control tests on streptomycin and in showing that smoking causes lung cancer led to random control tests becoming standard practice to test new drugs and to identify the causes of disease. The testing has had its undesirable side with the testing costs running to hundreds of millions of dollars and so discouraging the production of new drugs and some studies of disease showing a relationship between environmental factors and the disease without giving any real indication of a cause and effect relationship.

Diagnostic Technology

The twentieth century has seen the development of a series of new technologies that have enabled physicians to see inside the human body. The technologies began with X-Rays and then CT scanners, PET scanners and MRI scanners were developed. These technologies all allowed physicians to see inside the body from the outside while other technologies such as endoscopy allowed physicians to invade the body with tiny cameras to observe the state of the interior of patients' bodies.

X-Rays were first discovered by Wilhelm Roentgen in 1895. Roentgen was experimenting with a Crookes tube, a glass tube with the air removed to create a vacuum and with electrodes to allow the production of an electric current within the tube. The electric current, consisting of a stream of electrons known as cathode rays, would cause phosphorescent material within the tube to glow. When experimenting with a Crookes tube, the German physicist Phillip Leonard has noticed that cathode rays could travel through an aluminium sheet he had placed over a window in the Crookes tube and turn slips of paper covered with barium platinocyanide salts, fluorescent. Lenard sent a Crookes tube to Roentgen for Roentgen to study the cathode rays. Roentgen repeated Lenard's experiments and found the cathode rays were escaping from the Crookes tube just as Lenard had found. Roentgen thought that the cathode rays might be passing through the walls of the Crookes tube as well as through the aluminium covered window in the tube. When conducting the experiment Roentgen noticed a screen coated with barium platinocyanide, a yard away from the Crookes tube, turned fluorescent. This could not be caused by cathode rays which only travel a few inches in the air. Roentgen moved the screen further away from the Crookes tube and the screen still turned fluorescent when he turned on the electric current in the Crookes tube. Roentgen placed objects like a book and a deck of cards between the Crookes tube and the screen and the screen still lit up when he turned on the current in the Crookes tube. Further experiments revealed that the ray causing the screen to light up, could penetrate a wide range of materials such as wood and flesh. Roentgen had no idea what the ray was so he called it an X-ray. When a human hand was placed in front of a photographic plate and exposed to X-rays, the plate showed the bones in the human hand. However the X-rays did not easily pass through metals and could not pass through lead at all.

X-rays were found to have a number of uses such as in crystallography, astronomy and in microscopic analysis, but their most important use has been in medicine. X-rays can provide a photograph of the inside of the human body. X-rays have a shorter wavelength than light so they can penetrate materials opaque to light. X-Rays can more easily penetrate materials of low density such as skin and muscle, but cannot penetrate materials of higher density, such as bone, bullets and kidney stones.

The use of x-rays in medicine was greatly extended by the employment of contrasting media such as barium salts and iodine solutions. Barium makes it possible to obtain x-rays of the large and small intestine and the stomach and the esophagus. Iodine allows an x-ray picture of the kidneys and bladder and also the carrying out of angiography. Angiography provides a view of the blood within the arteries and veins which will disclose blockages and other problems within the arteries and veins. The use of catheters allows contrast materials to be injected into the heart allowing x-rays of the internal structures of the heart. X-rays can be used to detect tumours, cancers and cysts.

A further enhancement of x-ray technology came with the development of CT or CAT scanners. The CT scanner uses x-rays, photon detectors and computers to create cross section images or tomograms of the human body. In 1963 Allan Cormack invented an improved x-ray machine using computers, an algorithm and tomograms. In 1972 Godfrey Hounsfield invented the CT or computerized tomography scanner. It allowed many x-rays to be taken, from multiple angles of thin slices of the human body and detectors opposite the x-ray tubes would collect the data, which was converted into digital data, which was then converted by an algorithm, a set of mathematical instructions, by a computer into x-ray pictures. The CT scanner could give three dimensional views of the body and provide much better resolution than ordinary x-ray images. It can show soft tissues and liquid parts of the brain and can show tumours as small as one or two millimetres in size. CT scanners have gone through a series of improvements involving various different generations of scanners. In the earlier scanners the x-ray beam lacked the width and the number of detectors to cover the complete area of interest, requiring multiple sweeps to produce a suitable image. In subsequent scanners, a wider x-ray beam and more detectors were used to shorten scanning times.

Endoscopy, also known as laparoscopy, involves inserting an instrument into the body either through the body's natural entrances or through a small hole surgically cut in the body. The instrument is used to observe the internal structures of the body and can also be used for surgery with tiny instruments at the end of the endoscope being manipulated by the surgeon through the endoscope.

Endoscopy goes back to the late nineteenth century but was not widely used as the views it provided of the interior of the body were too poor for practical use. Harold Hopkins, a physicist, heard about the problems with endoscopes and remembered that although light normally travelled in a straight line it could in certain circumstances be made to travel around corners by the use of curved glass. Hopkins considered that tens of thousands of flexible glass fibres operating together may be able to cause light to go around corners. He made an experimental endoscope and published his results in 1954. Basil Hirschowitz, a South African, working in the United States, read about Hopkins' ideas and created his own endoscope. Several hundred thousand fibres were wound together and to stop light jumping from one fibre to another which could cause the loss of the image, a technique of coating each fibre with a glass coating was developed. The endoscope allowed investigation of much of the interior of the body and some surgery on the interior of the body without having to make substantial incisions into the body.

Photography through an endoscope was not very satisfactory due to inadequate illumination and because the optical system was not good enough. Hopkins investigated the problem and found that an endoscope consisting of a glass tube containing thin lenses of air gave improved light transmission around eighty times stronger than conventional endoscopes made of an air tube containing thin lenses of glass. This allowed the taking of photographs through the endoscope and allowed greatly expanded surgical possibilities through the endoscope. Endoscopy can be used for surgery by instruments such as lasers or wire loop cautery devices attached to the head of the endoscope and controlled by the surgeon through the endoscope.

Modern Surgery

Surgery before the introduction of anaesthetics and antiseptic and aseptic practices was limited to a narrow range of operations, of which limb amputation was by far the most common. The quickest operations only were possible without anaesthetics and the mortality rates from infection were enormous before antiseptic practices were introduced. The introduction of gowns, masks, rubber gloves and the sterilization of instruments dramatically cut the death rate in surgery.

Abdominal surgery only became possible with anaesthetics and antiseptics. Christian Billroth (1829-94) pioneered operations in this area. Operations to remove the appendix and to close a perforated gastric ulcer began to be performed in the late 19th century. Brain surgery began with Sir William Macewan (1848-1924) in Glasgow and Macewan also developed operations to deal with bone diseases such as rickets.

Plastic surgery was to make great progress in the 20th century, two New Zealanders Harold Gillies and Archibald McIndoe leading the way. Plastic surgery dates back to ancient times and was practiced in pre-British India and Renaissance Europe when it was used to deal with the terrible damage caused by syphilis. During World War I Harold Gillies carried out plastic surgery on the badly disfigured faces of soldiers and sailors. He developed an operation whereby a skin flap was sliced from the upper arm, one end of the flap remaining attached to the arm and the other end was moulded over the nose and then sewn down. After several weeks the skin sewn to the face would take and the skin attached to the arm could be cut and sewn into place on to the face. When the injured had no facial skin at all Gillies took the flap of skin from the abdomen rolling it over the chest and sewing one end to the face. Holes would be cut in the skin for the nose, eyes and mouth. When that end had taken Gillies cut the end still attached to the abdomen and then sewed that into place on the face. This system involved two operations as if the skin was completely removed from the donor area before it had taken on the face it would die due to lack of blood supply. These techniques were further developed by Archibald McIndoe while operating on air force pilots injured in World War II.

Organ transplant had been first experimented with by Alexis Carrel early in the 20th century. He carried out various transplant operations on animals, discovering the problem of rejection where the transplanted organ was rejected by the receiving animal’s body. The problem of rejection was investigated by Peter Medawar when he observed skin grafts taken from a donor would last for ten days before rejection, while a subsequent skin draft from the same donor was instantly rejected. When the body suffers an infection from bacteria or viruses initially it takes time to identify the invading organism before the immune system attacks the invading organism. In the event of a subsequent attack by the same organism the organism is immediately attacked because the immune system recognises it as foreign material due to its previous contact with the virus or bacteria. The way in which the first rejection takes some time but a second rejection of the same material occurs immediately led Medawar to realise that it was the immune system rejecting the transplant in the same way as it attacked invading bacteria and viruses.

Organ transplant required a practical surgical technique which was developed by Joseph Murray who improved on techniques experimented with by Alexis Carrel on animals. The technique involved the sewing together of small blood vessels which allowed the attaching of the transplanted organs' blood supply to those of the recipient so that it could receive the recipient's blood. The first attempts at organ transplant were kidney transplants. This was because humans had two kidneys, but only needed one so living donors were readily available. Kidney transplants were also relatively straightforward operations, the main job being to connect the transplanted organs' blood supply to the recipient's blood supply.

Kidney transplants did however require the prior invention of the kidney dialysis machine. The dialysis machine was invented by Wilhelm Kolff, a Dutch physician in 1941. The dialysis machine performs the work of the kidneys when the kidneys fail. This mainly involves removing waste material from the blood. The dialysis machine is needed during transplants to keep people alive before the operation and for a period of time after the operation, often ten days or so, until the donated kidney begins to work.

A workable surgical technique and the dialysis machine allowed kidney transplants to be performed and the first operation was performed in 1954 by Joseph Murray on a patient whose identical twin supplied the donated kidney. The operation was a success with no rejection problems as the donated kidney came from an identical twin so that the recipient's immune system did not treat the donated kidney as foreign material. When however kidney transplants were attempted using close relatives as donors, the donated organs were rejected by the recipient's immune system resulting in the death of the recipient.

A drug known as 6-mp had been developed by George Hitchings and Gertrude Elion as a treatment for leukaemia. 6-mp worked by stopping the cancer cell from dividing by appearing to be a chemical necessary for the cancer cells division, but which was slightly different so that it stopped the cancer cell from dividing and so killed the cancer cell. 6-mp was used to stop the immune system rejecting transplanted organs by stopping the division of cells in the immune system. 6-mp was tried on rabbits and found to stop the rabbits immune system attacking foreign material, but leaving the rabbits immune system otherwise working. Hitchings and Elion also developed a new drug azathioprine that was an improved version of 6-mp. Azathioprine was tried on people but with poor results until high doses of steroids in short bursts were given to patients with the azathioprine. This had the desired effect of preventing the immune system attacking the transplanted organ while still leaving the immune system able to work against ordinary infections. Eventually another drug cyclosporine was developed which had the same effect and transplant operations for other organs such as the lungs, liver, bone marrow and hearts were developed.

Improvements in medicine and sanitation lead to people living longer and an increasing exposure to the diseases of old age. Arthritis became much more common in the twentieth century than previously. Arthritis of the hip was particularly a problem causing constant and serious pain to patients and greatly reducing mobility. The pain was caused by the rubbing of bone against bone in the hip due to the erosion of cartilage between the bones.

Some attempts had been made to provide artificial hips in the 1930's and 1940's but none had been particularly successful. A major difficulty was that the hip has to maintain the weight of the body as well as being completely mobile. John Charnley looked at the problem and came up with three innovations that were to lead to a practical artificial hip. He redesigned the socket, he cemented the artificial hip to the bones with acrylic cement and he lubricated the joint first with Teflon and then when that failed with polyethylene. Charnley's new artificial hip was an outstanding success and the hip replacement operation was to become a common operation in the late 20th century.

The heart is the most complex organ in the body and for the first half of the twentieth century surgeons did not touch it believing that to do so would kill their patient. In the 1930's and 1940's operations were carried out on the aorta and the pulmonary artery to ease symptoms caused by heart problems, but the heart itself was not touched. In the late 1940's surgeons began to widen heart valves through a hole cut in the wall of the heart while the heart was still working. However, much heart surgery, known as open-heart surgery, was only possible with the heart being stopped. If the heart was stopped some means of maintaining the blood supply to the body was necessary or the patient would die. John Gibbon and his wife Mary Hopkins began work on a machine that could perform the work of the heart and lungs in the 1930's. The machine needed to be able to add oxygen and remove carbon dioxide from the blood and to pump the blood through the body. The machine needed valves to ensure the blood all flowed in one direction and had to use glass tubes as plastic had yet to be invented. The Second World War delayed progress, but a heart-lung machine was created in the early 1950's. Early results were not promising but the machine was taken over and improved by the Mayo Clinic. Donald Melrose, in England, and Viking Bjork, in Sweden, also built similar machines to allow open heart surgery. The result was to be an effective heart-lung machine that could take over the functions of the heart and lungs during operations so as to allow surgery on the human heart.

Analysis of the order of discovery in the history of medicine

The question of the origin of infectious disease was in dispute for thousands of years, the matter not being settled until the late 19th century. The earliest cultures and civilizations considered the cause of diseases to be supernatural and the appropriate remedies to be appeals to the Gods and magical incantations. Such beliefs were perfectly reasonable based upon the knowledge available to our prehistoric ancestors and to early civilizations. They had no awareness of bacteria, viruses or other microscopic organisms. Given that beliefs in Gods were used to explain other mysterious events, such as earthquakes, storms and volcanic eruptions, the Gods were an obvious explanation of disease. Given also that diseases can kill human beings, it would be reasonable to assume they are caused by powerful beings, like Gods or powerful demons and evil spirits. As the body automatically tends to repair itself, due to the immune system, it must have appeared to our prehistoric ancestors that on occasions the magical incantations and appeals to the Gods had worked. When the patient died the death could be put down to the capriciousness of the Gods or the great power of the evil spirit, rather than there being anything wrong with the treatment used.

In the west, from the time of Hippocrates, natural causes of diseases, such as the four humors theory, were the favored explanation, although supernatural explanations continued to find acceptance. The same situation existed in China with natural causes of disease such as inadequate or imbalanced Qi and Yin and Yang being considered to be the causes of disease. A similar situation existed in India where a balance of the three elements, air, bile and phlegm was required for good health. The Greek, Chinese and Indian explanations of disease are quite similar, all involving imbalances in bodily substances and all acquired a status that made them impervious to criticism and a block on innovation.

The presence of blood, urine, vomit and diarrhea clearly shows the body has many internal fluids. Vomit and diarrhea particularly seem to be present at times of sickness and recovery often occurs after vomiting and diarrhea so that it would appear that getting rid of fluids from the body could cure sickness. Even bleeding was often followed by recovery from injury so that a limited loss of blood could be seen as promoting recovery. It is because the human body has these fluids and because getting rid of the fluids with vomiting, diarrhea and bleeding seemed to cure sickness and injury, ideas such as an imbalance of fluids caused ill health arose in Western, Chinese and Indian cultures. This gave rise to theories such as Hippocrates and Galen's four humors theory and to remedies such as bleeding and purging. The Chinese theory of an imbalance between Yin and Yang causing disease appears to be a more abstract version of the same idea. Given the knowledge of non-scientific societies these theories make good sense. A theory that microorganisms, invisible to the naked eye, cause disease is hardly credible for societies that have no evidence of the existence of the microorganisms. On the other hand bodily fluids plainly do exist and their removal from the human body seems to be associated with recovery from disease and injury.

The medicine of Hippocrates and Galen did not just relate to the four humors. It also dealt with qualities such as hot, cold, dry and wet. This is because many of the symptoms of disease relate to these qualities for example if a person has a temperature or fever, they are hot, if they are perspiring, they are wet. If they do not have a temperature they are cold, if they are not perspiring they are dry. Galen's theory was built up from the way the human body acts, both when it is sick and when it is healthy. If the human body functioned in a different way it would have led to a different type of medical theory. If for example the human body changed color when it was sick, rather than changing temperature, medical theory would likely involve explanations and treatments that involve colors with the aim of restoring the patient to his or her normal healthy color.

The traditional Chinese theory of medicine has considerable similarities to the classical theories of Galen. The western idea of *pneuma*, a vital spirit taken into the body by breathing, is similar to the Chinese concept of *Qi*. Galen's theory of the four humors considers much sickness is caused by an imbalance in the body fluids. The Chinese theory also deals with body fluids, known as *JinYe*. A healthy person will have the body fluids in balance, but if the body fluids are deficient, or if there is an accumulation of fluids, sickness can result. A further similarity between Galen's humoral theory and the Chinese theory is that the Chinese theory of Yin and Yang, like the humoral theory, considers sickness to be caused by imbalances within the body. The Chinese theory of blood also emphasizes that imbalances can cause sickness. Given that Yin and Yang, body fluids and blood should all be in balance to avoid sickness in Chinese medical theory, it has considerable similarities with Galen's humoral theory which considers sickness is caused by imbalances in the four humors. In both the humoral theory and traditional Chinese medicine the weather could cause imbalances in body fluids. A further similarity between Galen's theory and traditional Chinese medicine concerns the elements. Galen's theory uses the idea of the four Greek elements, air, fire, earth and water. Each element is associated with a particular organ, a particular humor and with the qualities of hot, cold, dry and wet. Water for example is associated with the organ, the brain, the humor phlegm and the qualities of cold and wet. Traditional Chinese medicine uses the Chinese elements of fire, earth, water, wood and metal. The elements are each associated with organs, one of which is a Yin organ and the other a Yang organ. Water for example is associated with the bladder and the kidney, while earth is associated with the stomach and the spleen. The elements are all interconnected so that if one of the organs and its element is in a state of imbalance, it will affect the other elements and their organs. This could affect the individual's facial color and emotional state as well as the functioning of the relevant organs. The Western and Chinese theories of medicine were so similar as each was derived from the same source. The source was the human body and the environment that could affect the human body. If the human body and the environment were different the theories would be different.

The naturalistic and supernatural explanations of disease co-existed for thousands of years, sometimes with one dominant and other times with the other being the more powerful.

Neither was more convincing than the other, in that both sometimes appeared to work and that both sometimes failed to work. When they failed to work, both the supernatural and naturalistic theories provided explanations for the failure. If the human body did not have an immune system, so that if a person got sick they inevitably died and the incantations to the Gods and the treatment provided by doctors never worked, then the supernatural and naturalistic explanations of disease and the treatments they gave rise to would never have existed. It is only because the human body fights against disease, often successfully, that the incantations to the Gods and doctors treatments often appeared to be successful which suggested that the explanations of disease were true and the treatments provided were sometimes working. Both the supernatural and naturalistic explanations of disease could have been proved wrong with modern double blind testing, but such testing was not done in the past because it required knowledge of sophisticated statistical techniques that only became available in the last 400 years. Even in the 18th century the English and French were unable to agree as to whether smallpox inoculation was desirable while in the first half of the 19th century Pierre Louis conducted trials which showed bleeding was a useful treatment. Even today, drug trials sometimes produce contradictory results. Even if testing had been done, the theories would probably have survived due to the lack of serious alternatives.

It was not until the late 19th century with the development of the germ theory of disease that the question of the origin of infectious disease was settled in favor of a naturalistic theory, but a theory completely different from any of the naturalistic theories previously accepted. When Fracantorius in the 16th century suggested contagious disease was caused by tiny seeds invading the human body, the theory was quite reasonably not accepted as there was no evidence of the existence of the tiny seeds or that they caused disease. Fracantorius theory was almost identical to the germ theory of disease and the germ theory was only accepted in the late 19th century with the work of Pasteur and Koch. Leeuwenhook had discovered microorganisms in the late 17th century but that did not mean that they caused disease. In fact the vast majority of microorganisms do not cause disease in humans. It was only with the more powerful 19th century microscopes that Pasteur and Koch were able to discover particular organisms which caused particular diseases in humans. They were able to show the organisms were the causes of the disease by isolating the organisms and by preparing a pure culture of the organism, which in the case of animals would then be injected into an animal causing the disease in the animal. This procedure known as Koch's postulate established the Germ theory of disease and was able to show which particular germs caused which disease.

The explanations of infectious disease were based upon the knowledge available to a society at a particular time. When that knowledge changed (the discovery of microorganisms and the discovery that some of them cause disease) the explanations of disease changed. Societies that considered the activities of supernatural beings as explaining otherwise inexplicable phenomena used supernatural explanations for the cause of infectious diseases. Supernatural explanations and naturalistic explanations of disease co-existed for thousands of years. Each was as convincing as the other until the germ theory of disease arose in the late 19th century. Naturalistic explanations of disease were based upon the natural world, and in particular, on the human body itself. Body fluids, organs and the elements of the natural world all had a prominent role in both Western and Chinese naturalistic explanations of disease. The Chinese and Western explanations of disease were similar because they had similar knowledge of the natural world and of the human body, so they developed similar theories to explain the origin of disease. If the natural world and the human body were different, then the theories explaining disease would have been different. When human knowledge of the natural world increased, with the discovery of microorganisms in the 17th century and the discovery in the late 19th century that some of those microorganisms caused disease in humans, the theories explaining the causes of disease changed. The germ theory of disease became the accepted explanation of infectious disease throughout the western world.

The practice of immunization (the modern name for vaccination, also known as inoculation) has been one of the most successful medical practices in history. It has been responsible for an enormous reduction in human suffering and has saved an enormous number of human lives. The injection of dead bacteria or their toxins, or dead or weakened viruses into the human body to create immunity against disease, has eliminated or controlled a considerable range of diseases. Immunization has been used successfully against anthrax, bubonic plague, chicken pox, cholera, diphtheria, Haemophilus influenzae type B, mumps, paratyphoid fever, pneumococcal pneumonia, poliomyelitis, rabies, rubella (German measles), Rocky Mountain spotted fever, smallpox, tetanus, typhoid, typhus, whooping cough and yellow fever.

Immunization works because the body's natural defences against infection are able to remember dangerous bacteria and viruses it has already had contact with and are able to react more quickly and more strongly to later infections from the same organism. When an infection occurs certain cells in the body respond by moving to destroy the invading bacteria or viruses. In order to destroy the invading bacteria or viruses the body’s immune system, a collection of free moving cells, has to recognise which materials in the body are foreign invaders and what is part of the body. It does this by matching the shape of receptors on the surface of defending cells to the shape of the surface of the invading organism and if they fit together the defending cells recognise an invading organism. Once recognition of an invader has taken place, other defending cells will attack and destroy the invading organisms. The defending cells can also produce memory cells which, in the event of a future invasion by the same organisms, are able to immediately clone large numbers of the appropriate defending cells to attack the invading organism, without having to go through the process of recognising the invading organism. This makes the immune system's response to invading organisms, which it has recognised before, much stronger, faster and more effective. This process known as the amplification of the response, is the basis for immunization. A dead or greatly weakened infectious organism is injected into the human body so that the defending cells will remember the organism, so that in a future attack the immune system does not have to go through the recognition process and can immediately attack the invading organisms with large numbers of cloned defending cells.

If the body did not work in this manner, for example if it did not produce memory cells which instantly recognise invading organisms, the process of immunization would not work. This would mean that the wide range of diseases immunization is effective against would still be killing vast numbers of people.

Smallpox was the first infectious disease to be treated with immunization, partly because it was one of the worst and most persistent diseases in history and partly because nature provided a ready made immunizing material, in the form of cowpox, which saved people from having to identify, isolate and produce a safe vaccine. The high mortality rate from smallpox and the observation that survivors were protected from future attacks, which could only be observed with a disease which was continually or often present made smallpox the obvious disease to immunize against. A disease which came and then disappeared often for centuries is a less urgent case to immunize against as it may well not come back for centuries making immunization unnecessary. Given that smallpox was often or continually around, it made sense to immunize against it. It also made it more easily observable that survivors were protected against future attacks. This was not so easily observable with diseases which involved major epidemics and then disappeared for long periods of time, so there were no future attacks from which the victims of earlier attacks could be shown to be immune. However early attempts at variolation were so dangerous, that it is not surprising that it never really caught on.

The reason why smallpox was the first disease effectively treated with immunization was because nature provided, in cowpox, a ready made vaccination material which was not dangerous to human beings. To produce effective vaccines for other diseases it was necessary to discover the bacteria or virus involved, to isolate it and to reproduce it. This process enunciated in Koch's postulates could only be done with better microscopes than was available in the 18th century. It also needed the understanding that germs cause infectious disease which was not established until late in the 19th century by Pasteur and Koch. This understanding was not needed for smallpox, where it could be empirically observed, even by milkmaids, that the natural vaccine, cowpox, prevented smallpox. With the other diseases it was necessary to understand the germ theory of disease and then to artificially produce a vaccine before it was possible to immunize against those diseases. The process of immunizing against smallpox was a lot simpler than the process of immunizing against other diseases, so immunization against smallpox occurred before immunization against the other diseases.

The taboo on human dissection applied in most human societies, except India, Ancient Egypt and Europe since the Renaissance. The result was substantially erroneous beliefs concerning human anatomy and physiology. Beliefs that the heart was the centre of thought, sense perception and controlled bodily movements, while the brain cooled the heart and blood held by Aristotle resulted from the taboo on human dissection. When the taboo was not present, such as in Alexandria during the Ptolemaic era, it was discovered that the brain dealt with sense perception and bodily movements. Further progress in anatomy and physiology was delayed until the Renaissance when some dissections of the corpses of executed criminals were allowed. This eventually resulted in the anatomical discoveries of Versalius and the circulation of the blood by Harvey. Many future developments in medicine, especially in surgery, were dependent upon the new knowledge of anatomy and physiology obtained from the lifting of the taboo on human dissection.

Progress in surgery was also dependent on the discovery of anaesthesia and antiseptic and aseptic practices. There were two main consequences from the discovery of anaesthesia. The first was that surgery became far more common as patients no longer tried to avoid it. The second was that surgical operations became a lot longer with emphasis being on precision and accuracy rather than on speed. With increasing time being spent on operations more intricate and complex operations could be performed which greatly widened the range of operations available. With much longer operations and the need for anaesthetics and anaesthetists the cost of operations went up as did the status of surgeons who were now able to do so much more for their patients. Surgery became a practical solution to many medical problems.

The idea that cleanliness was important to stop infections in surgery and obstetrics was only accepted after Pasteur had established the germ theory of disease which showed that bacteria in the air caused infections. Prior to the germ theory of disease being accepted, suggestions that cleanliness was important were ignored as there seemed to be no reason why cleanliness could stop infection or lack of cleanliness could cause infection. The discovery that infection was caused by bacteria in the air, led to the antiseptic idea of killing the bacteria to stop infection and then to the a-septic idea of sterilising everything that came in contact with the patient.

The ending of the taboo on human dissection resulted in vastly improved knowledge of anatomy and physiology; this, and the discovery of anaesthesia and the realisation of the importance of a-septics, formed the basis of modern surgery. Only when these developments came together, was it possible for modern surgery, with its sophisticated and intricate operations, to become a reality.

This led to new types of surgery which had never before been developed such as abdominal and brain surgery. Plastic surgery, which had been practiced crudely in the past, improved enormously and later led to cosmetic surgery. Hip replacement operations were developed after the invention of a practical artificial hip. Organ transplants began when surgical techniques were developed for joining small blood vessels and when the problem of rejection of donated organs was solved by the development of appropriate drugs. Kidney transplants developed rapidly after the invention of the kidney dialysis machine as it is a relatively simple operation and because there is a better supply of donated kidneys as human beings have two kidneys and only need one so as to allow transplants from living donors. Open heart surgery and heart transplants were developed after the invention of the heart-lung machine to keep the patient alive during surgery.

The use of antibiotics in medicine is only possible because nature provides such organisms that inhibit the growth of bacteria and allows the production of synthetic compounds that achieve the same result. If nature did not provide these organisms, or allow such compounds, there would have been no antibiotics used in medicine. Without antibiotics, medicine since the 1940's, would have been much less effective and hundreds of millions, who were cured of infections would have died. The discovery and use of antibiotics was impossible before the development of microscopes capable of observing bacteria. Only when such microscopes existed was it possible to observe that certain organisms were capable of killing or inhibiting bacteria. A number of such observations were made in the late 19th and early 20th century and eventually it was realised that penicillin, a substance taken from one of those bacteria killing organisms, could be used against infectious disease. When penicillin was proved to be effective, a systematic search was made for other antibiotics which resulted in the discovery of a number of other antibiotics. However it was only because nature has provided the antibiotics, that we have them and we have only had them since we acquired the knowledge of their existence and of how to use them.

The use of statistics in medicine has been of enormous use in showing the causes of disease and in assessing the effectiveness of treatments. Yet statistics are never able to provide a perfect answer to questions of drug effectiveness and the causation of disease. They may show a correlation between two variables, for example people living close to the sea have higher rates of cholera, than people further from the sea. This does not however mean that proximity to the sea causes cholera. Correlation does not prove causation as the correlated variable may be caused by a third factor, such as polluted river water which is more common closer to the sea. The third factor, often called a lurking variable, may well not be considered in the data so no effort is made to compare cholera rates among people drinking polluted water close to the sea with those drinking clean water close to the sea. If the comparison was made it would show that it was the quality of drinking water rather than proximity to the sea that was the important variable concerning cholera rates. When trying to discover the cause of increasing lung cancer after World War II, air pollution and asphalting of roads were considered likely causes as both were increasing at the time lung cancer rates were increasing. Working out which variable to study when trying to discover the causes of disease can be very difficult.

A further problem concerns trying to ensure the chosen sample is representative of the population which is being studied. Pierre Louis concluded bleeding was a useful treatment, but one of the groups he studied was substantially younger than another group. The sample must also be of sufficient size or simple co-incidence and high margins of error may provide misleading results. Pierre Louis' study of bleeding was criticised for having insufficient numbers in his sample.

Given the difficulties of doing good statistical studies it is not surprising that the causes of diseases and the effectiveness of treatments was never accurately assessed until recently. Modern statistical methods were only developed in the 17th, 18th and 19th centuries and arose from probability theory. It was only with the development of modern statistical methods that it has been possible to identify the causes of many diseases and to evaluate the effectiveness of treatments. Even with modern statistical methods the causes of some diseases, for example some cancers, are still difficult to pinpoint. Often different studies of the same phenomena will produce different results. In these circumstances it was impossible for people in the past to discover the effectiveness of treatments and the real causes of disease until the discovery of modern statistical analysis.

Modern diagnostic technology began with the discovery of X-rays. X-rays however could not be discovered until certain earlier discoveries had been made. X-rays were discovered through the use of a Crookes tube which required prior discoveries of an efficient air pump to create a near vacuum in the tube and the ability to send an electric current through the tube. Only when these discoveries had been made was it possible to discover X-rays. The use of X-rays was eventually improved and extended by the use of contrasting media and eventually by CT scanners after the invention of computers.

X-rays are a form of electro-magnetic energy and are useful due to their property of being able to pass through matter of low density but not matter of high density. This allows X-rays to be used to produce photographs of the interior of the human body, which is why X-rays are so useful in medicine. It is only because nature has provided a form of electro-magnetic energy that we have X-rays available to be used for medical diagnosis. If nature had not provided electro-magnetic radiation with that property we could not have the ability to see inside the human body for medical purposes by means of X-rays.

Endoscopy only became practical when Hopkins and Hirschowitz discovered a practical method to make light travel around corners. It was only because such a method exists that we are able to have modern endoscopy, and modern endoscopy could not exist until the discovery of how to make light travel around corners. Endoscopy was further enhanced when Hopkins discovered that thin lenses of air gave much greater light transmission than thin lenses of glass, so as to allow much better endoscope photography. If such lenses did not provide improved light transmission, then endoscope photography might still not be practical.

Our brief examination of the history of medicine has shown how the environment relevant to medicine has affected the history of medicine. The relevant environment includes the human body, how the human body works, the diseases that attack the human body, how the materials in the environment affect the human body and how the body reacts to disease and injury. If the human body was different then the history of medicine would have been different. If, for example, there was no immune system, then a lot of the confusion concerning the effectiveness of treatments used in the past would not have existed. When patients treated with prayers, incantations, herbs, medicines, moxibustion and bleeding recovered, it looked as though the treatment had worked. If patients died all the time as they would have if there was no immune system, it would have been clear all these treatments were failing and they would have been abandoned. If there was no immune system then modern treatments such as immunization would not work and would not be available. If the human body was different, the theories as to what went wrong with it when people got sick would have been different. Galen's humoral theory and traditional Chinese theories were based on the human body and how it behaved in sickness and in health. If the body was different then those theories would have been different.

Anaesthesia was only possible as materials in the human environment had the property of making people so unconscious that they could not feel pain. X-rays were only possible as electro-magnetic energy of a certain wavelength will pass through matter of low density but not matter of high density. Modern endoscopy is only possible because light can be made to travel around corners and thin lenses of air provide excellent light transmission. The use of antibiotics is only possible due to bacteria killing organisms existing in the human environment and the ability to create compounds that will kill bacteria. The properties of materials and matter and forms of energy in the environment determine what is possible in medicine.

When knowledge of the environment relevant to medicine changed, this resulted in new theories, such as the brain being the centre of thought and emotions rather than the heart, the circulation of the blood and the germ theory of disease. These ideas were the logical explanations of the new knowledge that human beings had acquired, just as the previous theories were the logical explanations of the knowledge humans possessed at those times. Increasing knowledge of the environment relevant to medicine also led to the development of new treatments such as anaesthetics and new drugs. The new theories and treatments inevitably had significant social and cultural consequences, such as greater life expectancy, reduced suffering and different attitudes concerning religious beliefs, all of which would themselves result in further social and cultural consequences.

Where taboos existed against the acquisition of new knowledge, such as the taboo on human dissection, then the acquisition of new knowledge will be delayed until the taboo is removed. This, in the case of medicine, meant erroneous ideas of human anatomy and physiology continued for as long as the taboo remained in place. Only after the taboo was lifted was it possible to make the anatomical discoveries of Versalius and for Harvey to discover the circulation of the blood.

The Discovery of the Periodic Table

This paper conducts a study of the history of chemistry from the phlogiston theory to the periodic table. It traces the changes and new discoveries made in chemistry from the 18th century to the second half of the 19th century. It examines the chemical revolution of the late 18th century, the discovery of new elements and theories, that eventually led to the development of the periodic table. It concludes by an analysis of the order of the discoveries made in chemistry during that period. It notes that many of the discoveries had to occur in a particular order and that the order of those discoveries was inevitable.

A new scientific attitude had begun to appear in Europe in the 16th and 17th centuries but this had little effect on chemistry until the 18th century. In the 18th century methods for the qualitative and quantitative analysis of minerals improved, resulting in the discovery of new compounds and elements. The blowpipe became a common laboratory tool while the practice of weighing precipitated salts was introduced by Torben Bergman when analyzing mineral waters. This practice was improved by Klaproth who heated the salts to dryness before weighing them, which produced more accurate results. Klaproth also began the practice of reporting the actual percentage composition as produced by his analysis regardless of whether it totaled 100% and this allowed the discovery of errors in analysis and the discovery of new elements in the materials analyzed.

The new laboratory methods lead to the discovery of new elements such as cobalt (1735), platinum (1740-1741), zinc (1746), nickel (1754), bismuth (1757), manganese (1774), molybdenum (1781) tellurium (1782), tungsten (1785) and chromium (1798). The oxides of zirconium, strontium, titanium and ytterbium were also discovered. Many of the new substances were metals and this led to the demise of the ancient doctrine of seven metals. In the second half of the 18th century Carl Scheele discovered hydrofluoric acid and the compounds hydrogen cyanide, lactic citric and malic acids and glycerol.

The phlogiston theory was introduced by Becker and Strahl in the late 17th and early 18th centuries. The theory considered all combustible substances contained phlogiston which was lost to the air during the process of combustion. A limited amount of air could only absorb a limited amount of phlogiston which explained why combustion ceased if only a limited amount of air was available. Combustion would also cease as soon as substances ran out of phlogiston. The phlogiston released into the air was absorbed by plants which were eaten by animals so that the phlogiston was recycled into known combustible materials.

The main problem with the phlogiston theory is that metals gained weight when burnt in air and the theory suggested phlogiston was lost, so one would expect substances to lose weight. This problem became acute when the gaseous state of matter began to be investigated in the mid-18th century.

It was at this time that gases were becoming much better understood and progress was made on distinguishing compounds from elements. In the early and mid-18th century air was considered to be an element. When scientists observed gases with unique properties, their difference from air was assumed to be caused by impurities. Boyle’s inverse pressure-volume law also convinced scientists that air was an element since the law applied to all gases. One difficulty with investigating and controlling gases was solved in the early 18th century by

Stephen Hales when he invented an apparatus for isolating gases so they could be studied separately. Hales' device, known as the pneumatic trough, allowed the collection of gases above water. A bent gun barrel, with its closed end containing various substances placed in a fire and with the open end in a vessel of water suspended upside down in a pail of water. The gases released from the substances in the closed end of the gun barrel would collect in the upside down container above the water and separate from the air. Hales apparatus lead to the identification of many gases such as carbon dioxide discovered by Joseph Black in 1755; hydrogen discovered by Henry Cavendish in 1766; nitrogen discovered by Daniel Rutherford in 1772; nitrous oxide discovered by Joseph Priestley in 1772 who in the years after that discovered ammonia, sulfur oxide and hydrogen chloride; oxygen was discovered in the 1770's independently by Carl Scheele, Joseph Priestley and Antoine Lavoisier. The ability to isolate, identify and handle gases soon led to the realization that these gases were forms of matter in the same sense that liquids and solids were.

The study of isolated gases soon showed they were different from each other and the differences resulted from differences in composition rather than from contamination by impurities. The idea of air as an element began to be replaced by the idea of gas as a state of matter.

The phlogiston theory was widely accepted by scientists by the middle of the 18th century. Despite the discovery of oxygen the phlogiston theory continued to be accepted until Lavoisier created a revolution in chemistry which destroyed the phlogiston theory, eliminated the four elements of antiquity and replaced them with the modern concept of elements as substances that could not be broken down and which were the fundamental substances of chemistry. Lavoisier was also involved with a reform of the nomenclature of chemistry, so that the names of compounds reflected the elements making up the compound.

Many experiments had been conducted by Priestley, Lavoisier and others that showed that metals and sulfur and phosphorus would increase in weight when burnt in air. It was also known that when the calx (oxide) produced when these substances were burnt in air, were themselves burnt using Hale's pneumatic trough, a variety of different airs (gases) would be produced. In particular, experiments were made involving the burning of mercury in air to produce mercury calx and then the burning of the mercury calx, using the pneumatic trough, to recreate the mercury and a gas in which candles burnt more brightly than in normal air and supported respiration in mice. The air in which the mercury was burnt was not able to support respiration in mice or combustion after the formation of the calx. Measurements made indicated that the weight of the original mercury and air absorbed on burning equaled that of the calx and also equaled the weight of the mercury and the gas produced when the calx was burnt.

According to the phlogiston theory, the gain in weight of the mercury as it was burnt in the air was caused by the release of phlogiston, which had a negative weight. This explanation was considered by Lavoisier and others as absurd. As the increase in weight of the mercury equaled the reduction in weight of the air in which the mercury was burnt Lavoisier concluded that part of the air had combined with the mercury to form the calx. Equally, as the air produced by burning the mercury calx was different from normal air and as the air left behind when the mercury was burnt did not support respiration or combustion, it seemed that a particular constituent part of the air that supported combustion had been removed from that air and had combined with the mercury to form the calx and had then been released by the calx into the pneumatic trough. This led Lavoisier to assume that air was not an element, but was composed of several parts, one of which supported combustion and respiration and one that did not. It also led Lavoisier to consider that combustion required the presence of the part of the air that combined with the mercury and did not involve any release of phlogiston from the substance being burnt. The new gas was eventually called oxygen by Lavoisier.

A further development involved the burning of hydrogen in air which produced a clear liquid which on analysis was shown to be water. Cavendish, Priestley and others as well as Lavoisier were involved in these experiments, but Lavoisier was the first to interpret them to mean that water was a compound of oxygen and hydrogen. Lavoisier's interpretation marked the end of the belief that existed from ancient Greece, that water was an element.

One of the results of the discovery of many different gases after the invention of the pneumatic trough was the understanding of gases as a distinct form of matter. Matter could be seen as changing from a solid to a liquid to a gaseous state by the application of various degrees of heat.

It also became clear the air and the gases it was made up of played a role in chemical reactions. Substances when heated would combine with various gases or would release gases into the atmosphere. In both cases new substances were created by these chemical processes.

A further feature of the discoveries was the confirmation of the law of conservation of mass. Tacitly assumed by many of the chemists, it was confirmed by many experiments dealing with gases combining with metals to form calx and then the calx, when burnt releasing the gases. When the quantity of gas absorbed by the metal and the amount released from the calx are measured they are found to be the same, confirming the scientists belief in the conservation of matter.

The measurement of substances involved in experiments assumed a much greater role in chemistry, than it had previously. The awareness that metals gained weight when burnt was an important element in the demise of the phlogiston theory. Chemistry was becoming a quantitative science. Once this occurred, the way was open for chemical equations and the calculations of the weight of elements, leading to Dalton's atomic theory in chemistry.

Some of Lavoisier's innovations did not survive. His belief that oxygen was a necessary part of all acids and his idea of caloric as an explanation of heat would soon be abandoned. However, by the end of the 18th century, his overall conception was largely adopted throughout Europe.

A debate arose between Berthollet and Proust in the early 19th century as to whether compounds were always formed from fixed proportions of their constituent elements or whether the proportions could vary. Their debate was resolved in favor of fixed proportions although there are now known to be some situations where the constituents of a compound can vary. However in many cases it became clear that compounds were made up of elements that combined in definite and fixed proportions. The question arose as to what lay behind those definite proportions. Dalton showed those proportions were not only fixed but related in a simple numerical manner.

This process was helped by the development of quantitative analysis in chemistry. Before the 19th century most work in chemistry was qualitative and concerned with the properties of substances and the courses of particular chemical reactions. By the late 18th century more emphasis was being given to the weight of substances entering into and resulting from chemical reactions.

Lavoisier's concept of an element provided the foundation for Dalton's atomism. Different elements had different atoms and this explained the different properties of the elements. The atomic theory was the outcome of the new quantitative work being done in chemistry, the discovery of fixed proportions in the elements making up compounds and the observation that the proportions were fixed in a particular numerical manner.

The discovery that air is a mixture made up of a number of gases rather than an element raised the question of why it was all mixed together rather than formed in layers with the heaviest gas at the bottom and the lighter gases higher up. Dalton's answer to this problem was the idea that if the particles of a particular kind of gas were self-repulsive but did not repel particles of a different kind of gas, then the formation of layers of gases would not occur. The cause of the repulsion was caloric, Lavoiser's explanation of heat, each particle of gas being surrounded by an atmosphere of caloric. As heat was known to flow from hot substances to colder ones, two equally hot substances would be mutually repellant. The problem remained that all particles of gas had the same repellant (caloric) wouldn't they still repel each other. Dalton considered that the particles of different gases were of different sizes and so would have varying amounts of heat so they would not repel each other. Only particles of the same gas would have the same amount of heat and would repel each other. This theory, although not correct was the best explanation for the mixing of gases in air before the kinetic theory of gases was developed in the middle of the 19th century. However the idea that the size of particles of different gases would vary, lead to the idea that the weight of the particles would vary.

This conclusion was also reached as a result of experiments concerning the solubility of gases in water. It had been observed that the mass of a gas dissolved by a liquid is proportional to the pressure. Elementary gases such as hydrogen and oxygen were less soluble while compound gases such as carbon dioxide were more soluble. Dalton considered the cause of the varying solubilities was the different size of the particles of the different gases. Again the varying size of the particles of different gases led to the idea that the weight of the particles would vary.

Dalton was to call the particles of gases, and of all substances, atoms. Elements were composed of simple atoms and compounds of compounded atoms. The elements varied one from the other, as the atoms making up different elements varied in weight. There were, however, difficulties in calculating atomic weights. It was impossible to weigh individual atoms so the system of atomic weights had to be based on a comparative system. Dalton chose hydrogen as a base for such a system and gave it an atomic weight of one. The atomic weights of the atoms of other elements were based on how much more they weighted in comparison with hydrogen. To calculate for example how much more oxygen weighed than hydrogen Dalton compared the weight of hydrogen and oxygen making up a quantity of water. He found the oxygen in water weighed 5.5 times as much as the hydrogen (the correct figure is 8) so he assigned an atomic weight of 5.5 to oxygen. Such a system would only work if the number of hydrogen and oxygen atoms in water was known, and in Dalton's time this was not known. To overcome this difficulty Dalton adopted his principle of simplicity when he assumed that if two elements formed only one compound, the compound would consist of one atom of each element. If there were two compounds formed of the same two elements, there would be two atoms of one element and one atom of the other element and so on. As water was the only known compound of hydrogen and oxygen it was assumed to consist of one hydrogen and one oxygen atom. Obviously the principle of simplicity was not a reliable guide to the chemical composition of compounds. The problem of accurate measurement of atomic weights and of accurately assessing the chemical make-up of compounds limited the usefulness and acceptance of the atomic theory. In addition, as the number of elements discovered increased in the early 19th century, it began to look as though there was an increasing number of fundamental particles. Many scientists considered the idea that there were a large number of fundamental particles was absurd.

The atomic theory did obtain support from Gay-Lussac in 1808. Gay-Lussac found that hydrogen combined with oxygen at a ratio of approximately two to one. In other experiments he discovered other gases combine among themselves in simple whole number ratios. This became known as the law of combination of gases. It suggested equal volumes of different gases contained the same number of particles.

However there were problems with the law of combination of gases by volume. Carbon monoxide, considered to contain one atom of oxygen and one atom of carbon, should be denser than oxygen. Yet, it was known to be less dense than oxygen. A further problem was that one volume of nitrogen combined with one volume of oxygen to give two volumes of nitric oxide rather than the one compound of nitric oxide.

There was considerable confusion over atomic weights, molecular weights and equivalents. This confusion made it impossible to write chemical formula with confidence. In a chemistry book written by Kekule he quoted nineteen different formulas that had been suggested for acetic acid. Atomic weights in the first half of the 19th century were decided by guess work and arbitrary rules. Gerhardt corrected Berzelius's atomic weight for sodium and silver by halving them, but he also halved the correct weights for zinc and calcium and so made them incorrect.

A resolution to these problems was offered by Amadeo Avogadro. He distinguished between an atom, as the smallest part of an element which can play a role in a chemical reaction and a molecule as the smallest part of a substance. He assumed that molecules of an element could consist of more than one atom of the element. This meant a molecule of hydrogen could contain two atoms of hydrogen. This also meant a molecule could split in two when involved in a chemical reaction. If this happened, then equal volumes of gases could contain the same number of particles. The reaction of nitrogen and oxygen could be explained by two atoms of oxygen joining two atoms of nitrogen to create two molecules of nitric oxide. Avogadro's theory provides for the calculation of molecular weights of substances in the gas or vapor state by determinations of the gas or vapor densities. The determination of densities is made on a relative basis with hydrogen, as the lightest gas, being chosen as the standard. All other substances are expressed as having a weight relative to that of hydrogen. However as the hydrogen molecule consists of two atoms, atomic weights should be related to the weight of half a molecule of hydrogen. Avogadro's theory provided clear information about the number of atoms in a molecule of a compound and provided a firm foundation for the writing of chemical formula.

However, Avogadro's theory was largely ignored. The terminology he used to explain his theory was difficult and many chemists refused to accept that the fundamental particles of elements could contain more than one atom. Avogadro's theory was only adopted after 1860 when Cannizzaro drew chemist's attention to it and explained how it could allow the correct calculation of molecular and atomic weights.

It was only after 1860 that the atomic theory gained considerable acceptance with the acceptance of Avogadro's theory which cleared up problems concerning the atomic weight of elements and the composition of compounds. This was followed by Mendeleev's periodic table and finally Einstein's 1905 explanation of Brownian motion as grain pollens being bumped about by the movement of atoms confirmed the atomic theory.

A new method of causing chemical decomposition became available around 1800. This involved the voltaic pile which allowed a continuous electric current to be passed through a substance causing decomposition. Before 1800 only static electricity had been available for chemical decomposition but the short term nature of the current limited its effectiveness in chemical experiments. The use of the voltaic pile was to allow the discovery of new elements and showed some substances previously considered to be elements were actually compounds. Sir Humphrey Davy was to isolate potassium, sodium, barium, strontium and magnesium by means of the voltaic pile while Gay-Lussac and Thenard discovered boron and Courtois discovered iodine. The isolation of potassium led to the discovery of other elements due to the chemical reactivity of potassium. The heating of various compounds with potassium resulted in the discovery of silicon and aluminum.

The use of the voltaic pile led to a further significant discovery. It was observed that when water was decomposed using the voltaic pile, the hydrogen and oxygen formed at different poles. It was then found that when an electric current was passed through solutions of salts, acids formed at the positive pole and bases at the negative pole. This observation led to the development of Berzelius's dualistic theory. Berzelius considered that atoms carried both a positive and negative charge, but only one of the charges was predominant. Metals were electro-positive as they were attracted to the negative pole in electrolysis. Oxygen was the most electro-negative element. As atoms possess both charges, an atom could be negative towards one element and positive towards another. Phosphorus for example was negative towards metals, but positive towards oxygen. This allows a series to be established from the most electro-positive element to the most electro-negative element. Chemical combination happened due to the attraction between opposite electrical charges. When such a combination occurred, the compound formed would be either positive or negative depending on the strength of the charges of the elements making up the compound. If the compound was positive it could then combine with negative compounds and elements and vice versa. Berzelius considered his theory explained the nature of chemical affinity. It would not, however, be accepted today.

An important new development occurred in 1860 with the development of the spectroscope which allowed the detection of new elements. The color spectrum had been known in Roman times and the refraction of light had been studied by the Arabs, Roger Bacon, Kepler and Descartes. Newton showed that a prism separated white light into its component colors and that another prism could turn the separated colors into white light. The spectrum was extended into the infrared in 1800 by William Herschel and into the ultraviolet in1801 by W H Williamson.

Dark lines in the spectrum of sunlight passed through a prism were observed by Joseph Fraunhofer in 1814. He studied and mapped the lines (eventually called Fraunhofer lines) and observed similar lines in the spectrum of light from the moon, planets and stars. Fraunhofer also discovered yellow lines in the spectrum of the flame he was using when studying the refractive index of samples of glass. Similar lines were observed in the flame of burning alcohol, oil and tallow when determining refractive indices. Such lines had also been observed in the spectra of many substances by many scientists. They had been observed in the spectra of metallic salts by Thomas Melvill in 1752. David Brewster had observed them in the spectra of "nitrous acid gas", sulphur, and iodine vapor and brown oxide of nitrogen. Similar studies were carried out on halogen vapors and other gases.

The identification of substances by means of the spectrum began when Andreas Marggrat used flame colors to distinguish sodium and potassium salt in 1758. John Herschel showed when the flame colors of boric acid and the chlorides of barium, calcium, strontium and copper were passed through a prism they showed certain lines which could be used to identify the substances. Brewster after observing sulphur vapor absorbed light from the violet end of the spectrum and iodine vapor absorbed it from the middle part, suggested "the discovery of a general principle of chemical analysis in which simple and compound bodies might be characterized by their action on definite parts of the spectrum".

This idea was put into practice by the invention of the spectroscope by Bunsen and Kirchoff in 1859. Bunsen used flame colors for the identification of salts in mineral water. Kirchoff suggested better results could be obtained if the light was passed through a glass prism and viewed as a spectrum. Kirchoff also outlined the reason for the bright and dark lines as being emission or absorption lines of light. He set out his laws of spectroscopy as:

1. An incandescent body gives off a continuous spectrum.
2. An excited body gives off a bright-line spectrum.
3. White light passed through a vapor has dark lines where the vapor ordinarily emits light.

The spectroscope provided chemists with an instrument of unprecedented sensitivity for the analysis of chemical substances. The spectroscope was able to map Fraunhofer lines with great accuracy and when only minute traces of an element were present.

The effects of spectroscopy soon became apparent. Bunsen and Kirchoff discovered cesium in 1860 and rubidium in 1861. Thallium was discovered in 1861 by Crookes and indium in 1863 by Reich and Richter. Spectroscopy was later involved in the discovery of gallium, the rare earths and the rare gases.

When Lavoisier provided the modern definition of an element as a substance that could not be broken down into simpler substances, he provided a list of 33 elements. These included several forms of energy and some substances later found to be compounds. However the attention Lavoisier drew to the elements and new analytical techniques such as the voltaic pile and reaction with potassium, once that element had been isolated by Davy, led to the discovery of many additional elements. Between 1790 and 1844 31 new elements were discovered but the number of elements remained limited to 58 from 1844 to 1860 as the unknown elements were generally present in minerals in to small quantities to be detected by the analytical techniques available at that time. It took the development of the spectroscope to allow the discovery of new elements to recommence.

The discovery of sufficient elements and the establishment of a reliable system of calculating atomic weights as provided by Avogadro were necessary before the next important development in chemistry. This was the system of classification of the elements known as the periodic table.

The first attempt at such a classification was made by Dobereiner in 1829. Dobereiner observed that it was possible to put the elements into groups of three with the atomic weight of one element in the group being the mathematical average of the other two elements. Dobereiner also observed that the members of the groups all had similar chemical properties.

A number of other attempts were made to discover some sort of relationship between the elements. The more significant of these were made by Beguyer de Chancourtois and John Newlands. Beguyer de Chancourtois in 1862 and 1863 with the benefit of the atomic weights that were accepted after the work of Cannizzaro persuaded chemists to accept Avogado's theory, arranged the elements in order of their increasing atomic weights around a cylinder. Beguyer de Chancourtois pointed out there were remarkable similarities in the elements on the same vertical line on the cylinder.

Newlands, also using Cannizzaro's atomic weights, arranged the elements in order of their atomic weights. He observed that similar elements would appear on a horizontal line if a new column was commenced with each eighth element. In some versions of his table Newlands used blank spaces for unknown elements, but in other versions he did not. Newlands called his arrangement "the law of octaves" but as with Beguyer de Chancourtois, Newlands work was largely ignored.

The idea of the periodic table was accepted due to the work of Mendeleev and Meyer. Mendeleev published his first periodic table in 1869 and a second version was published in 1871. Mendeleev arranged the elements in the order of increasing atomic weight and noted the properties of the elements recurred periodically in the table. Gaps were left in the table where Mendeleev considered there were elements yet to be discovered. Using his periodic table Mendeleev predicted the discovery of certain new elements. He stated the approximate atomic weights, valences, the sorts of compounds the element would be found in and other properties the undiscovered elements would have. When the new elements gallium, scandium and germanium were discovered and found to have properties extremely close to those predicted by Mendeleev, his periodic table became widely accepted.

Meyer's periodic table was broadly similar to Mendeleev's but tended to concentrate on the physical properties of the elements, while Mendeleev's concentrated on the chemical properties. Meyer produced a graph in which he plotted the atomic weights against atomic volumes of the elements. The graph showed similar elements appeared at similar places on the curve produced by the graph. Alkali metals appeared at the high points of the curve; nonmetals on the ascending sides and metals on the descending sides and in the low points on the curve. Meyer's graph helped to make the periodicity explicable and encouraged acceptance of the periodic table.

Analysis

The brief foregoing history of chemistry from the phlogiston theory to the development of the periodic table is provided to show certain aspects of the theory suggested in Part 1 of this book. One of those aspects concern human needs. The needs met by humankind's exploration of the nature of matter was simply the desire to know and understand the universe. Such needs seem to exist in all societies as for example all societies seem to have creation myths to explain how the universe came into existence. Equally, many societies have their own explanations as to the nature of matter. The Babylonians considered all matter was created from a first principle and that first principle was water. Such needs to explain the nature of matter would be covered by Maslow's cognitive needs, the need to know, to understand and to explain.

The foregoing history of chemistry is also intended to show how the discoveries made in chemistry were made in a particular order and had to be made in that order. One such discovery was Lavoisier's chemical revolution. This had a number of features such as elements being seen as substances that could not be broken down; air no longer seen as an element; air playing a role in chemical reactions; the concept of a gas being seen as a separate state of matter; confirmation of the law of conservation of mass; explanations of combustion and respiration and a theory of acidity. It seems very apparent that many features of this revolution were dependent upon the discovery of new gases in the mid and late 18th century and in particular upon the discovery of oxygen. Oxygen is a colorless, odorless and tasteless gas so that its existence is not obvious to the naked senses which resulted in its discovery not occurring until the later part of the 18th century. It is hardly conceivable that Lavoisier's explanation for combustion and respiration could have been made without the prior discovery of oxygen. Lavoisier's experiments with mercury calx showed that part of the air, the part we call oxygen could cause candles to burn more brightly, but that air without oxygen could not support combustion or respiration in mice. In order to make these discoveries it was necessary for oxygen to be isolated and identified as a particular component of air. Only then could its effect on combustion and respiration be studied and understood. The discovery of oxygen was itself dependent upon a means of isolating and controlling gases as was provided by the pneumatic trough. Without the pneumatic trough Lavoisier would have only been able to perform part of his experiment with mercury calx. He could have burnt the mercury in air to create the calx, an experiment that had been performed many times before. But without the pneumatic trough he could not have burnt the calx in a situation where the oxygen released by the combustion could be controlled and experimented with. The same or similar experiments had been performed by Priestly and Scheele but Lavoisier was the first to interpret the results with a new theory of combustion and respiration. Furthermore, Priestly and Scheele were as dependent on the pneumatic trough as Lavoisier was when it came to the discovery of oxygen. The consequence is that there was an order of discovery from the pneumatic trough to oxygen to Lavoisier's theory of combustion and respiration.

A similar order of discovery was involved in the discovery that air was not an element, that the gases in the air played a role in chemical reactions and the concept of a gas as a separate state of matter. The discovery of many gases using the pneumatic trough showed the air was made up of a number of gases and so was not an element. The role of gases or air in chemical reactions was shown by the gain in weight of metals when burnt in air to produce a calx and then the release of gases when the calx was burnt. The concept of a gas as a separate state of matter was shown when substances could be heated and could be shown to pass through solid, liquid and gaseous states. Again the pneumatic trough played a part in this as substances could be burnt and could be shown to produce particular gases. There was an order of discovery from the pneumatic trough to the discovery of gases, to the conclusion that air is not an element, that it plays a role in chemical reactions and that gases are a separate state of matter.

The proof of the law of conservation of matter was dependent upon both the pneumatic trough and upon the increasing use of quantitative studies in chemistry. When metals were burnt, it was known, before Lavoisier, that they increased in weight. However it was only with the use of the pneumatic trough that it was possible to isolate air to show that there was a decrease in the weight of the air that matched the increase in the weight of the metal. The accurate measurement of the decrease in the weight of the air and the increase in the weight of the metal were required before the law of conservation of matter could be confirmed.

The modern concept of an element as a substance that could not be broken down, was established by Lavoisier, when the traditional elements such as air and water were shown to be made up of simpler substances. Due to the use of the pneumatic trough and the discovery of a number of gases in the mid and late 18th century Lavoisier was able to provide a list of 33 elements, which replaced the traditional elements of earth, fire, air and water. This involved an order of discovery from the pneumatic trough, to the gases discovered through the use of the pneumatic trough, to Lavoisier's new concept of elements.

Lavoisier's theories of acids and caloric, while no longer considered correct, were dependent upon the prior discovery of oxygen, which itself was dependent upon the prior discovery of the pneumatic trough. This meant an order of discovery from the pneumatic trough to oxygen, to Lavoisier's theories of acid and caloric.

The next major development in chemistry, after Lavoisier's revolution, was Dalton's atomic theory. The atomic theory was based upon Lavoisier's concept of an element. A particular element would consist of a particular atom and different elements had different atoms with different weights. It was improvements in quantitative chemistry in the late 18th century that showed that different elements had different atomic weights. The law of definite and fixed proportions which suggested compounds were made up of elements in definite and fixed proportions and the proportions were related in a simple numerical way, provided support for the atomic theory. This was because the atomic theory explained the relationship between the elements as varying in weight in that simple numerical manner. The atomic theory was also based on Dalton's erroneous ideas as to why the different gases in the air did not form layers. Such ideas naturally could only be formed after it had been discovered that the air was made up of a number of different gases. The discovery that air was a mixture was dependent upon the experiments by Lavoisier and others which isolated oxygen, nitrogen and other gases which were dependent on the prior discovery of the pneumatic trough. This means a chain of discoveries runs from the pneumatic trough, to the isolation of oxygen and nitrogen, to the idea of air as a mixture. A further factor in the development of the atomic theory was experiments concerning the solubility of gases in water. These again were dependent on the isolation of various gases by the pneumatic trough, so an order of discovery from the pneumatic trough to the isolation of various gases to the atomic theory can be identified.

The atomic theory could, at the time of Dalton, and eventually did in the 1860's receive support from Gay-Lussac's law of combination of gases, which suggests equal volumes of different gases contained the same number of particles. Such a law could only have been proposed when a number of gases had been isolated which required the prior discovery of the pneumatic trough. Yet again an order of discovery can be identified from the pneumatic trough to the discovery of gases, to the law of combination of gases to the atomic theory.

The law of combination of gases naturally led to Avogadro's theory. Avogradro's theory was intended to explain the known behavior of gases during chemical reactions in a manner consistent with the law of combination of gases. This means both the experiments showing the chemical reactions and law of combination of gases were necessarily prior to Avogadro's theory. The acceptance of Avogadro's theory in the 1860's led to an accurate system for calculating atomic weights and to the eventual acceptance of the atomic theory.

The discovery of the voltaic pile lead to the discovery of new elements, including potassium, and the discovery of potassium lead to the discovery of further elements. The voltaic pile also led to the discovery of the electro-positive to the electro-negative series which in turn led to the development of the dualistic theory. A clear order of discovery runs from the voltaic pile to the discovery of new elements and to the dualistic theory.

The development of the periodic table was dependent upon certain prior discoveries. One was an accurate way of calculating atomic weights, another was the discovery of a sufficient number of elements to allow them to be organized in a coherent table and a further development was sufficient analysis of the properties of the elements to reveal the periodicity of the periodic table. A reliable method of calculating atomic weights was provided by Avogadro's theory. The discovery of the elements occurred throughout the 18th and 19th centuries assisted by new instruments for the investigation of chemical substances such as the pneumatic trough, the voltaic pile, potassium analysis and the spectroscope. It probably took the invention of the spectroscope in 1859 to allow the discovery of sufficient elements so a coherent periodic table could be created. The spectroscope itself could only be invented after the discovery of the spectrum and Fraunhofer lines, and that Fraunhofer lines allowed the identification of chemical substances. Only when a sufficient number of elements had been discovered and their properties analyzed was it possible to create the periodic table. Early attempts to create a coherent table of the elements by Dobereiner, Beguyer de Chancourtois, John Newlands and others had failed due to uncertain atomic weights and an insufficient number of elements being known. It was in 1869 that Mendeleev published his periodic table and 1870 when Meyer published his.

It is of course not surprising there is a specific order of discovery in chemistry from Lavoisier's revolution to the periodic table. Obviously the facts of chemistry were not all discovered at the same time. However it is suggested that the order in which the discoveries were made were, in many cases, inevitable and the discoveries in those cases could not have been made in any other order. In some cases this is obviously so, a good example being the use of the spectroscope to identify elements. It is simply not possible to invent the spectroscope without first discovering the spectrum, then discovering Fraunhofer lines within the spectrum and then the discovery that Fraunhofer lines showed the presence of particular chemical substances. An awareness of all these discoveries was a necessary ingredient to the invention of the spectroscope and all these discoveries *had* to take place in the order in which they did take place.

However, other discoveries will take place in a certain order without the later discoveries being dependent upon the earlier discoveries. The dualistic theory was necessarily dependent upon the prior discovery of the electro-positive and electro-negative series and the voltaic pile. It was not, however, dependent on Lavoisier's chemical revolution and yet it occurred after the chemical revolution. Could the dualistic theory have been created before the chemical revolution? It is possible the dualistic theory could have been invented before the chemical revolution. However such an event was unlikely because discoveries always vary in difficulty; some discoveries will be easier to make than others. The easier discoveries will tend to be made earlier than the more difficult discoveries. This will not always be the case if for example governments and corporations pour resources into a particular area and neglect other areas. Difficult discoveries in areas receiving the resources may well be made before easier discoveries in areas not receiving the resources. The invention of the atomic bomb during World War II and the discoveries made in the space race are such accelerated discoveries caused by governments pushing resources into particular areas. However such situations are unusual and normally easier discoveries will be made before later ones.

A further such example would be that Lavoisier's chemical revolution occurred over a hundred years later than Newton's establishment of classical physics. Could Lavoisier's revolution have occurred before Newtons? As Lavoisier's revolution was not dependent upon the earlier revolution in physics, it would have been possible for it to have occurred before Newtonian physics. However such a situation would be very unlikely if the discoveries required for Newton's revolution were easier than those required for Lavoisier's. This would seem to be the case as Newton's revolution was substantially dependable on directly observable phenomena, the only exception being an accurate understanding of planetary orbits which required the telescope. Lavoisier's revolution was dependent upon phenomena that could not be directly observed such as gases that were discovered by the use of the pneumatic trough. If however, the pneumatic trough had been invented earlier, the chemical revolution could have occurred earlier.

The final result is that there were certain discoveries in chemistry which could not have taken place without certain prior discoveries. These were cases in which the order of discovery was inevitable and no other order of discovery was possible. There were other cases where the order of discovery was not inevitable, but where there was a likely or probable order of discovery as certain discoveries were easier to make, than other discoveries, and so were likely to be made earlier than the other discoveries. The degree to which one discovery was easier than another would determine the likelihood of it being made before the other discovery.

The Discovery of the Atomic World and the Constituents of Matter

The idea that matter was made up of tiny particles called atoms goes back to Ancient Greece where Leucippus and Democritus claimed all matter consisted of atoms and empty space. This idea was not widely accepted and from the time of Ancient Greece until the late 18th century, the western world generally considered all matter to be made up of mixtures, in various proportions, of the basic elements of water, air, earth and fire. Traditional Chinese beliefs were similar and considered all matter was made up of water, earth, wood, metal and fire. The traditional view of the elements ended when scientists showed air was made up of other substances such as nitrogen and oxygen and water was made up of oxygen and hydrogen. The modern view of atomic theory began with John Dalton (1766-1844) who, building upon Lavoisier's idea of chemical elements, suggested that for each chemical element there was a particular atom. He suggested that each element was related to a particular atom with different elements varying from each other due to the particular weight of their atoms. The existence of atoms was widely, but not completely accepted by the late 19th century. It was only with the discovery of the constituents of the atoms and Einstein's explanation of Brownian motion in 1905 that the existence of atoms became completely accepted.

The first of the constituents of the atom to be discovered was the electron. During the 19th century physicists had learnt to create and control electric currents. But they had been unable to discover what electricity was as when electricity flows through a wire or gathers on materials that tend to pick up an electric charge, it is impossible to separate the properties of the electricity from those of the wire or other material. In order to understand electricity, it was necessary to separate electricity from the materials that hold the electricity. This meant studying the flow of electricity in a vacuum or a near vacuum.

This became possible with the invention of an improved air pump by Johann Geissler in 1858. Geissler's air pump made it possible to remove enough air from a glass tube to create a near vacuum in the tube. Physicists began to create experiments with a glass tube containing a near vacuum with a wire carrying an electric current at one end, known as a cathode, and a positively charged metal plate, known as an anode, at the other end. This apparatus was known as a Crookes tube. When a current was passed through the cathode a thin ray of light passed from the cathode to the anode. The ray of light became known as a cathode ray and was caused by the heating of the small quantity of gas remaining in the tube by something moving through the tube. It was observed that the cathode rays could be deflected by magnets and electric forces placed near the tube which indicated they had an electric charge. In 1897 J J Thompson measured the velocity and the amount of deflection of cathode rays. He found that the velocity of the cathode rays was about a tenth of the speed of light which meant that the cathode rays were not a form of electro-magnetic radiation. The deflection of the cathode rays showed they had a negative electric charge. Given that cathode rays were not electro-magnetic energy the rays had to consist of particles and given that the particles had a negative electric charge they could not be atoms which were normally electrically neutral. Thompson also tried to measure the mass and the charge of the particles but he was only able to measure the ratio of those properties. His experiments showed the ratio remained the same regardless of what gas was in the tube and of what metal was used for the electrode. Eventually in 1909 Robert Millikan was able to separately measure the charge of the electron. This enabled him to calculate the mass of the electron which showed it was a particle vastly smaller than the atom.

Electrons could only come from atoms and given there was a negatively charged particle within the atom and that atoms were usually electrically neutral, it seemed obvious that there must also be a positively charged part of the atom. A further question was how the electron and the positively charged material in the atom were arranged. One suggestion by J J Thompson was that the atom was structured as a plum pudding with positively charged matter making up the great bulk of the atom with negatively charged electrons embedded in the positively charged matter, just as raisins are embedded in a plum pudding. The discovery of the structure of the atom and of the positively charged material in the atom was made by the New Zealand physicist Ernest Rutherford in 1911. Rutherford set up an experiment in which he fired alpha particles at a thin sheet of gold foil and measured the effect on the alpha particles as they were scattered by deflection by the gold foil.

Alpha particles are helium ions, or helium atoms without electrons, so that they are positively charged helium nuclei. Alpha particles had been discovered as one of three types of rays emitted by radioactive elements. The other two rays were beta particles, later discovered to be electrons and gamma rays, a form of electro-magnetic energy of a particular wavelength. Radioactivity had been discovered by the French physicist Becquerel in 1896 and had been extensively studied by Pierre and Marie Curie. The Curies discovered and isolated a radioactive element, radium, which Rutherford used as the source of alpha particles in his experiments when the particles were fired at sheets of gold foil.

Rutherford, assisted by Hans Geiger and Ernest Marsden, began a study of the scattering of alpha particles when they were directed at thin metallic foils in 1907. Alpha particles emitted from radium passed through a slit in a lead screen and then hit a metal foil which resulted in some deflection of the alpha particles as they were affected by the atoms of the metal foil. The deflection of the alpha particles was measured when they hit a zinc sulphide screen on which they made a flash of light. It was initially discovered the alpha particles were deflected by no more than a few degrees and that as the angle increased, the number of deflected alpha particles declined. Then Rutherford casually suggested that Geiger and Marsden check to see if any alpha particles were deflected back by a large angle such as more than 90 degrees. To their surprise it was discovered that a few of the alpha particles were deflected backwards by more than 90 degrees. This could not be explained by J J Thompson's plum pudding model of the atom with the positively and negatively charged material spread throughout the atom as widely spread positively charged material could not cause the large angle backward deflection of alpha particles. Rutherford suggested that only the concentration of all the positive charges at a single point in the atom could cause alpha particles to bounce backwards. The few alpha particles which bounced back more than 90 degrees were explained by them bouncing back from the nucleus which occupied a tiny portion of the space of the atom. The positively charged particles making up the nucleus were soon called protons. The great majority of the atom was empty space which explained the large number of alpha particles deflected by a small angle. Those alpha particles simply went through the atom with only a slight deviation from a straight line caused by the positively charged alpha particles being deflected by the positive and negative material in the atom. Measurements were made that showed the nucleus contained nearly all the mass of the atom, but that it occupied only a tiny proportion of the space of the atom.

The discovery of the nucleus of the atom raised questions as to why the negatively charged electrons did not fall into the positively charged nucleus as opposite charges attract each other and why the positively charged nucleus in all atoms with two or more protons (every atom other than hydrogen) did not fly apart as same charged particles repel each other. Neils Bohr suggested that if the electrons were in constant motion around the nucleus in particular orbits then they would not fall into the nucleus due to centrifugal force. Electrons could jump from one orbit to another and could not occupy any position between the orbits. The reason why nuclei of two or more protons do not fly apart is due to the existence of the strong force which holds the nucleus together. The strong force results from protons exchanging special particles known as pions, a process capable of generating an attractive force between protons.

A further problem with the atom was that for atoms other than hydrogen, their atomic weight seemed to be at variance with their atomic number. The atomic number concerns the number of protons and electrons an atom has. Hydrogen has a single proton and electron so its atomic number is 1 and its atomic weight is 1. Helium, for example, has 2 protons and because it is normally electrically neutral, it normally has 2 electrons, so its atomic number is 2. However the atomic weight of helium is slightly more than 4 and as almost all the weight of an atom is contained in the nucleus its atomic weight would be expected to be two, rather than slightly more than 4 which it had been measured to be. Atoms of the same element had also been discovered by a process known as mass spectrograph to be of different weights. Ions of the same element, which are atoms which have either lost or gained electrons, so that they are electrically charged, were sent through a glass tube containing a vacuum. The stream of ions, being electrically charged, would be deflected by the presence of a magnet and it was observed that the stream of ions would break into several different streams, which suggested the ions had different weights. Lighter ions are deflected more than heavier ions suggesting the reason why the ion stream divided into several different streams was that the original stream consisted of ions of various different weights. Atoms of the same element with different weights came to be called isotopes. The question arose as to why atoms of the same element had different weights. The obvious explanation for these problems was that there was more matter in the nucleus than just the protons. This led Rutherford in 1920 to suggest there must be some as yet undiscovered electrically neutral matter in the nucleus to explain why the atomic weight of atoms, other than hydrogen, were more than double their atomic number and why atoms of the same element could have different weights.

Experiments had been carried out by W Bothe and H Becker involving the directing of very fast alpha particles from the radioactive element polonium at various light elements, such as beryllium. It had been observed that the beryllium when struck with the alpha particles emitted unusually highly penetrating radiation. This was initially considered to be electro-magnetic radiation, such as light, but when Irene and Frederick Joliot-Curie noted that the radiation from the beryllium hit a paraffin wax screen it knocked particles out of the paraffin wax at a speed higher than one would expect if the beryllium radiation was electro-magnetic radiation.

James Chadwick suspected the beryllium radiation might be the electrically neutral matter in the atomic nucleus guessed at by Rutherford and which would explain the difference between elements atomic numbers and atomic weights and how different atoms of the same element could be of different weights. In 1932 he investigated the beryllium rays and directed them at a range of materials. Chadwick discovered that, when he directed the rays at materials with increasing atomic weights, the velocity of the particles knocked out of the various materials by the beryllium rays, decreased. The result was consistent with the beryllium rays, being particles with a mass similar to that of protons, rather than electro-magnetic radiation. A further property of the particles was that they were electrically neutral, which was clear due to their considerable penetrating power. Charged particles deflected by electric fields within atoms have much less penetrating power than uncharged particles. The newly discovered electrically neutral particle was initially considered to be made up of a proton and an electron, but soon began to be accepted as another elementary particle. The particle came to be called the neutron and in the late 1930's there were considered to be three elementary particles, the proton, the electron and the neutron. The standard model of the atom at that time had the proton and neutron making up the nucleus with electrons orbiting the nucleus in particular orbits. The discovery of the neutron explained why the atomic weight of an atom could be at variance from its atomic number and how different isotopes of an atom could have different weights.

Striking the atomic nucleus with alpha particles had the disadvantage that alpha particles are repelled by the nucleus as both are positively charged. This reduces the power of the collision and limits the target nucleus to the lighter elements as the heavier elements, containing more protons, repel the alpha particles so strongly that the experiments were ineffective. It occurred to Enrico Fermi, in Rome, that using the newly discovered neutrons could result in more effective bombardment experiments. He systematically began to bombard each element with neutrons, starting with the lightest elements first. When he reached fluorine he discovered that it became radioactive when bombarded with neutrons and that the same effect was observed for many of the heavier elements. The same artificial radioactive effect had been observed by Frederic and Irene Joliot-Curie and was then reproduced by John Cockcroft with his particle accelerator and in

America with cyclotrons. It meant that radioactivity which had previously been limited to a few elements could now be produced by many different elements.

Fermi came up with another interesting result when he discovered that hydrogen atoms could slow down neutrons and that slower neutrons generate more activity than faster neutrons. Substances such as water, which contain hydrogen atoms, could slow down or moderate neutron velocities, a discovery that was of importance to nuclear reactors.

Fermi's bombardment of uranium with neutrons and similar experiments carried out by Otto Hahn and Lise Meitner in Berlin, had produced a number of substances which were believed to be new elements but were actually uranium isotopes produced through nuclear fission. Similar experiments, with uranium, carried out by Irene Joliot-Curie and Paule Savitch in Paris produced a substance that looked like the element lanthanum, but which Joliot-Curie and Savitch thought must be actinium, an element close to uranium on the periodic table. Otto Hahn carried out similar experiments and believed he had produced radium, when he had actually produced barium, which chemically was similar to radium.

The reason scientists all over Europe were misinterpreting the results of experiments involving the bombardment of uranium with neutrons was that all the previous bombardment experiments had only knocked a few protons and neutrons out of the nucleus. To turn uranium into a uranium isotope or into actinium which had only three protons less than uranium was no surprise, but to produce barium with only 56 protons or lanthanum with 57 protons to uranium's 92 did not seem possible. Eventually testing of the materials resulting from the bombardment of uranium confirmed the materials were barium and lanthanum and that the bombardment had caused the uranium nucleus to split apart. The results of the experiments were published in early 1939 and the process now called fission was soon being repeated around the world.

It was immediately realised that fission would not only split the nucleus, it would also release additional neutrons which could hit additional uranium nuclei breaking them up to release even more neutrons to hit even more uranium nuclei. It was quickly realised such a process called a chain reaction would release a tremendous amount of energy, and act either as a bomb or, if appropriately controlled, for peaceful purposes. Scientists had long realised there was enormous energy in the atomic nucleus, but they had no way of releasing it until the development of nuclear fission.

The beginning of the war in Europe in September 1939 sped up research into the possibility of producing an atomic bomb. Scientists in Britain, France and Germany studied how an atomic bomb might be built, although those in Germany showed little enthusiasm for the job. Serious efforts to produce an atomic bomb began in the United States with a letter sent by Albert Einstein to President Roosevelt in September 1939 stating that recent discoveries had made possible the building of a new weapon capable of enormous destruction. The American attempt to build the atomic bomb came to be called the Manhattan Project and was under way by mid 1941 and became increasingly urgent after 7 December 1941 when America entered the war.

Whether a chain reaction was really possible depended on whether splitting the uranium nucleus would release additional neutrons and if so how many. If only a few additional neutrons were released then the chain reaction would gradually reduce and stop as some neutrons are lost during the process. In order to obtain a continual chain reaction it was necessary for the production of additional neutrons to reach what is known as a critical size where the new neutrons being released was equal to the number of neutrons being lost. It had been found that slower neutrons were more effective at breaking up the uranium nucleus and to slow the neutrons down a moderator, such as water or paraffin wax, containing hydrogen atoms was used. It was soon discovered that those substances tended to absorb neutrons making it more difficult to achieve a chain reaction. Eventually it was realised that heavy water or graphite were better moderators as they slowed neutrons down without absorbing them. The French were the first to achieve a chain reaction in 1939 but the chain reaction faded without reaching critical size.

A further problem that needed to be sorted out was the separation of the uranium isotopes U-238 and U-235. U-238 was by far the most common form of uranium but U-235 was required for making a bomb. The two isotopes of uranium were usually mixed together with U-235 amounting to only 0.71% of the mixture. The separation of U-235 from U-238 was achieved by an electro-magnetic separation plant using a two stage process and by a gaseous diffusion plant. Eventually 60 kilograms of U-235 was produced and used in the bomb dropped on Hiroshima.

An alternative to the use of U-235 was the use of Pu-239. This required the building of a nuclear reactor which would turn uranium into Pu-239 which is an isotope of plutonium. The nuclear reactor was built by Fermi with graphite acting as the moderator, with uranium inserted into the graphite reactor in a lattice structure. Fermi's reactor went critical on 2 December 1942 and established a self-sustaining chain reaction. Once the reactor had produced the plutonium it had to be removed from the uranium. Eventually enough plutonium was produced to enable a test bomb to be exploded and for the bomb dropped on Nagasaki.

Once the U-235 and Pu-239 had been produced it was necessary to turn the material into a bomb. This required some sort of initiator to set the explosion of. It was decided to use the conventional gun method to explode the uranium bomb but this method could not be used for the plutonium bomb. This is because of spontaneous fission caused by the presence of the Pu-240 isotope within the plutonium which could result in pre-detonation. A new method of initiating the explosion, called implosion, was developed for the plutonium bomb. This involved the bomb having a hollow core, surrounded by plutonium which was surrounded by high explosive. The detonation of the high explosive forced the plutonium into the hollow centre of the bomb which creates a supercritical explosive mass. When the bomb is in that state a burst of neutrons obtained by mixing polonium and beryllium starts the chain reaction. The implosion method of exploding the bomb was used when the plutonium bomb was tested on 16 July 1945 producing the largest man made explosion the world had yet seen. Then on 6 August 1945 a uranium bomb was dropped on Hiroshima and on 9 August 1945 a plutonium bomb was dropped on Nagasaki.

The atomic bomb was followed by the hydrogen bomb. The idea of the hydrogen bomb arose among the scientists working on the atomic bomb in the early 1940's. The scientists realised that energy could be produced not just by the fission of heavy nuclei, but also by the fusion of light nuclei, such as hydrogen. Fusion required extreme temperatures such as could be provided by an atomic explosion, so that the atom bomb could provide the heat required for a hydrogen bomb. Work began seriously on the hydrogen bomb and the United States exploded an effective hydrogen bomb in 1954. The Soviet Union exploded an atomic bomb in 1949 and then a small hydrogen bomb in 1953 and a somewhat larger one in 1955. Great Britain soon developed nuclear weapons and was followed by France, China, India, Pakistan and almost certainly Israel. Nuclear weapons proliferation has become a major concern of modern international relations with countries such as Iraq, Libya, North Korea and Iran being suspected of a desire to obtain nuclear weapons. The possibility of terrorist groups obtaining access to nuclear weapons is also of considerable concern.

The discovery of the atomic world did not just lead to the development of weapons of war. Peaceful uses of nuclear fission were developed in nuclear power stations which used nuclear power to create heat, which is used to produce steam, which is used to generate electricity. Nuclear power generated in power stations is made by bombarding uranium with neutrons to produce a chain reaction in the same fashion as atomic bombs work, but the system is controlled by control rods made of materials such as boron which absorb neutrons. The control rods enable the chain reaction to be started, stopped and to be increased and decreased as desired. The chain reaction produces an extreme amount of heat. The heat is removed from the reactor by pumping heat absorbing materials such as air or water through the reactor. The heat is then used to generate steam which is used to power steam turbines which generate electricity.

Accelerators and nuclear reactors produce various radioactive materials which have important uses in medicine, industry, agriculture and scientific research. In medicine radioactive materials assist diagnosis due to the tendency of some elements to concentrate in particular areas of the body. Potassium concentrates in the muscles, iodine in the thyroid and phosphorus in the bones. Cameras can take photos of those organs when a patient is injected with the appropriate radioactive elements. Radioactive elements can be used to treat disease for example radioactive iodine can be used to treat Graves disease, a thyroid disease and thyroid cancer. Radioactive elements cesium-137 and copper-67 are also used to treat cancers. Cobalt-60 is used to sterilize surgical instruments and to treat cancer.

In industry, radioactive substances are used to study the density of materials, to inspect goods for flaws and to track substances passing through piping systems. This is possible because radioactive materials lose energy as they travel through substances and it is possible to detect the energy loss disclosing where the radioactive material is. This is particularly useful where it is not possible to have direct contact with the substance being studied. Radioactive materials are used to check the quality of metal used in cars, to locate and assess oil and mineral deposits and to inspect jet engines and pipelines.

In agriculture radioactive materials have been used in the study of plants and animals. This has produced plants and animals that are more disease resistant and plants that have a greater nutritional value. Improved knowledge of fertiliser use has resulted from the use of radioactive materials to show how plants absorb fertiliser. Radioactive materials have been used to control insect pests such as Mediterranean fruit flies.

Radioactive dating of materials has caused a revolution in both geology and archaeology. In geology radioactive dating has enabled the dating of rocks which has led to calculations of the age of fossils and to rates of evolutionary change and of the age of the earth. Radioactive dating is based on the fact that some isotopes are unstable and decay by the emission of alpha and beta particles and gamma rays. While the rate of decay of an individual isotope is unpredictable, the rate of decay of a large quantity of atoms of a particular isotope is predictable. The rate of decay is usually called the isotope's half life, which is the length of time it takes for half of a given quantity of atoms of the particular isotope to decay. The rate of decay is constant over time so that the quantity of the isotope in some materials will disclose the age of the material. If the half life of the isotope is one million years and half the isotope has decayed, the material containing the isotope is a million years old.

Radioactive dating is usually done on a mass spectrometer. The mass spectrometer operates by producing a stream of ions from the material which is to be dated. The ions pass through a magnetic field which deflects them based on their mass and charge. This indicates the quantity of the isotope in the material or the degree of decay of the isotope which allows a calculation of the age of the material.

There are a number of unstable isotopes which decay at a predictable rate. U–238 has a half life of 4.5 billion years, U-235 has a half life of 700 million years and K-40 has a half life of 13 billion years. If a given rock has several different unstable isotopes in it, then the dating can be done on both isotopes providing a useful cross check on the dating process. Using radioactive dating techniques the age of the earth was calculated in 1953 at 4,550 million years, a figure still accepted today.

Radioactive dating of archaeological samples is usually done using the carbon-14 isotope which is why it is often known as radiocarbon dating. Carbon-14 has a half life of 5,730 years which makes it useful for dating materials back to around 60,000 years old. Carbon-14 is created by the collision of cosmic rays with nitrogen and oxygen atoms which release neutrons which collide with nitrogen atoms to create carbon-14. The amount of carbon-14 in the atmosphere is very small and the carbon combines with oxygen to form carbon dioxide. Plants absorb the carbon dioxide through photosynthesis and animals acquire the carbon dioxide through eating the plants. When an organism dies, its intake of carbon-14 ceases and the existing carbon-14 in the organism begins to decay at its standard rate. The amount of carbon-14 left in an organism when it is examined indicates how long ago the organism died. If there is plenty of carbon-14, the organism died more recently, if most of the carbon-14 has decayed, then the organism died long ago. A reasonably exact date can be calculated based on the quantity of carbon-14 left in the sample examined. Materials that can be dated include all plant remains and human and animal remains, such as bones. Radiocarbon dating has meant dates have been established for the end of the Neanderthals, the domestication of various animals and plants, the beginning of cities and civilization, the development of writing and many other major prehistoric events.

The only means of studying the atom open to scientists early in the 20th century was through the use of particles, usually alpha particles, generated by natural radiation from radioactive materials such as radium. This method had its limitations due to the limited energy levels of such particles. If the atom could be hit by particles with much greater energy levels a much greater range of effects could be observed. During the 1930's three different methods to deal with this problem were developed. In England John Cockcroft and Ernest Walton developed the voltage doubler accelerator, while in America Robert Van de Graff invented the Van de Graff generator and Ernest Lawrence created the cyclotron. Cockcroft and Walton, using their particle accelerator, directed a high speed proton beam at lithium and observed alpha particles being emitted from the lithium. This was the first nuclear reaction created by artificial means and its effect was a million times greater than could be produced by alpha particles. However it was the cyclotron that was to become the primary particle accelerator as it could provide a much higher voltage and give much greater energy to particles than the other two machines.

The cyclotron works by continuously sending particles around expanding circular paths and periodically accelerating them to greater and greater speeds. The particles are accelerated by means of electric fields which push and pull charged particles in ever expanding circular paths until the path reaches the borders of the machine. The cyclotron was, however, limited by accelerating increasingly fast particles in a constant magnetic field. The synchrotron was then developed to apply an increasing magnetic field to the accelerating particles. A synchrotron usually consists of a number of magnets laid out as a hollow ring which accelerates the particles in short bursts by increasing the power of the magnetic field as the particles go by. Synchrotrons are usually used for accelerating protons but are unsuitable for accelerating electrons as electrons travelling around a ring will emit photons and lose energy. The solution to this problem was to accelerate electrons in straight lines in linear accelerators. Once these particle accelerators were invented further improvements were made to them for example by using superconducting magnets and designing collider particle accelerators. Superconducting magnets are made from materials capable of being kept at close to absolute zero temperatures and are able to conduct electricity without loss of energy. Collider accelerators involve directing a high energy particle beam, not against a stationary target, but against another accelerated high energy particle beam to create collisions involving much higher energy than there is with a particle beam hitting a stationary target.

A new means of detecting particles, known as the cloud chamber, was invented in 1911 by Charles Wilson and began to be used to detect the results of the collisions of cosmic rays with particles in the atmosphere. Cosmic rays are particles emitted from stars, some of which obtain high energies and penetrate the earth's atmosphere where they will collide with atomic nuclei.

The collisions produce a cascade of particles which collide with other nuclei sending even more particles flying round, some of which reach the ground. The cosmic ray collisions result in the production of numerous particles, the presence of which can be observed in cloud chambers.

Cosmic rays were first discovered in 1912 by Victor Hess but were initially believed to be gamma rays and electrons. It wasn't until the early 1930's that cosmic rays were correctly identified as being positively charged particles, mainly protons. When this was understood scientists began to carry out a series of experiments, using cloud chambers, which showed the existence of a range of new particles. The positron was discovered in 1932 by Carl Anderson and in 1937 he and Seth Neddermeyer discovered the muon. The pion, (or pi-meson) was discovered in 1948 and the lambda in 1947. In 1953 the Sigma particle, which comes in three types, one positive, one neutral and one negative was discovered. In 1954 the cascade particle, which decays into a lambda and a pion and which comes in two types, one with a negative charge and the other with no charge, was discovered.

Cloud chambers work by showing the track of a charged particle passing through the chamber, in the form of a line of droplets. The cloud chamber contains a vapour which condenses into droplets around ions produced when the charged particle, passing through the cloud chamber, removes electrons from atoms in the cloud chamber. The cloud chamber can be placed between the poles of a magnet creating an electric field within the cloud chamber. The charged particle will move in a curved path through the cloud chamber, the curve showing whether the charged particle has a positive or negative charge. The degree of curvature and the number of droplets indicates the mass and velocity of the particle.

The cloud chamber was eventually superseded by the bubble chamber in the 1950's. The bubble chamber was much better at detecting high energy particles, such as those produced by particle accelerators, than the cloud chamber. The bubble chamber works in a similar way to the cloud chamber, but is filled with a liquid, usually liquid hydrogen, held under pressure and at a temperature just below its boiling point. When particles move through the liquid they create a trail of ions, the pressure on the liquid is then reduced which causes bubbles to form around the ions disclosing the path of the particles. The chamber may be in a magnetic field which discloses the charge of particles and their momentum. The particles are photographed by a number of cameras to show a three dimensional picture of the particle track.

Bubble chambers had some limitations which led to their replacement by the 1980's with more useful particle detectors. Bubble chambers produced photographic rather than electronic results which meant people, rather than computers, had to examine the results. This is a problem particularly with experiments that need to be repeated and analysed a number of times. The liquid in the bubble chamber is both detector and target so they cannot be used with accelerators producing particle beam collisions. Bubble chambers lack the size to show the results of high energy collisions and to allow accurate estimates of momentum in such small chambers. These sorts of problems resulted in bubble chambers being superseded by a variety of particle detectors such as spark chambers, wire chambers, drift chambers, silicon detectors, ionisation chambers and other detectors often combined together to produce the best results.

The results of cosmic ray experiments and experiments involving increasingly powerful particle accelerators was the discovery of a vast range of new sub-atomic particles. Many of those particles are very short lived as they quickly decay into other particles but may play important roles in the structure of the atom. By the middle of the 1970’s hundreds of particles had been identified and studied. This vast proliferation of particles came to be known as the particle zoo. Attempts were soon being made to classify the particles into various groups based on the properties of the particles. Certain particles such as electrons, muons and neutrinos are only involved in the weak interaction, which involves radioactive decay, and are known as leptons. Most other particles, such as protons and neutrons are involved in the strong force that holds the atomic nucleus together and are known as hadrons. Particles which decay into other particles including protons are known as baryons, while those whose decay produces photons (particles of light) are called mesons. Particles who decay slowly are called strange, while those that decay more rapidly are classified as non-strange. Numerical values can be assigned to strange particles. A strange particle decaying into a normal particle, such as a proton, would have a strangeness value of –1. If it decayed into an antiproton, its strangeness value was +1. If it slowly decayed into another strange particle it had a strangeness value of –2 but if it decayed quickly it had the strangeness value of the particle it had decayed into. Particles may also be classified by angular momentum, which is a property known as spin. Particle spin may be measured in any ½ number of whole numbers, for example ½, 1, 1½, 2, 2½ and so on. Spin must also be in particular directions such as up, down or sideways. Particles can also be classified by a quality known as isotopic spin. Isotopic spin is based on a particle's electric charge and involves treating particles that are identical, apart from their charge, as being the same type of particles. Particles such as pions, which come with either positive, negative or no charge are treated as the same type of particle but with different isotopic spins.

The classification of the hundreds of subatomic particles, constituting the particle zoo, made possible a rationalization that came to be known as the eightfold way. Murray Gell-Mann and Yuval Ne’eman independently came up with this rationalization which was similar to the periodic table of the elements. They created a graph with the vertical axis measuring a particle's strangeness and the horizontal axis measuring isospin. Particles are represented on the graph by a point. If the eight lightest baryons are put on the graph at points where their strangeness and isospin values intersect the shape produced by the positions on the graph of points representing the baryons strangeness and isospin values forms a hexagon with two particle points being in the centre of the hexagon. The same result is achieved if the eight lightest mesons are positioned on the graph in accordance with their strangeness and isospin values. Why the graphical representation of the particles should produce a hexagon with two particles in the centre was unknown. Gell-Mann later learned of the mathematics of Lie groups, which was developed in the

19th century. One of the Lie groups called SU(3) dealt with octuplet representations and asserted that the different particles in the octuplets were related in particular ways.

The particle Lie group SU(3) that provided for the octet representation also allowed other representations such as those with ten particles. This was fortunate as not all the groups of particles fit the octet model. The four delta particles could not fit the octet model but could fit a system of ten particles. The delta particles and five others produced an upside down triangle on the graph with one particle missing at the bottom of the triangle. Gell-Mann, using the mathematics of the SU(3) Lie group predicted the discovery of the new particle and its properties. The new particle was discovered in 1963 showing that the SU(3) symmetry worked and was the right way to classify particles.

The question then arose as to why the eightfold way or the SU(3) system worked for the classification of hadrons. The baryons would only form multiplets in accordance with SU(3) mathematics if there was some underlying reason, just as atomic structure provides an underlying reason for the structure of the periodic table. Gell-Mann and George Zweig independently suggested the SU(3) symmetry would make sense if the hadrons were made up of constituents called quarks by Gell-Mann and aces by Zweig. Gell-Mann's terminology was eventually accepted. Quarks were considered to be particles that had electric charges that were only fractions of the charges possessed by electrons and protons. All particles previously identified had electric charges of one (the charges of protons and electrons) or multiples of one. No particle had ever been observed with a charge that was a fraction of one. This was eventually explained by the idea of confinement, whereby quarks, while being the constituents of many particles, could not be knocked free of the particles by accelerator experiments. They cannot be knocked free from particles because as quarks break free from a particle, they form a new particle known as a meson which contains a quark and an antiquark.

The properties of quarks could be deduced as their properties had to explain the properties of the particles they made up. Baryons such as protons and neutrons were made up of three quarks, consisting of two quarks known as up quarks and one down quark for a proton, and one up quark and two down quarks for a neutron. Up quarks had an electric charge of +2/3 and down quarks had an electric charge of –1/3. Two up and one down quakes for a proton equals 2/3 + 2/3 + -1/3 =+1 being the electric charge of the proton. One up and two down quarks for a neutron equals 2/3 + -1/3 + -1/3 = 0 being the electric charge of the neutron. Leptons, such as electrons, were not made up of quarks at all.

When quarks were first suggested in the early 1960's only the up, down and strange quarks and their antiparticle equivalents were proposed as the constituents of hadrons and mesons. Evidence of the existence of these quarks was discovered in 1967 in experiments where electrons were fired at protons and the electron scattering was measured and suggested protons were made up of other particles. It had been suggested that quarks should come in pairs and there should be an additional quark called the charm quark to partner the strange quark. The J/psi particle was discovered in 1974 and was considered to be made up of a charm quark and its antiquark. The discovery of the charm quark quickly led to general acceptance of the quark model. In 1977 a further quark, the bottom quark, was discovered and its partner, the top quark, was discovered in 1995.

The quark model had a number of problems that delayed its acceptance. The electron scattering experiments had suggested that the constituent particles of protons were weakly bound and nearly non interacting within the proton. This suggested it should be easy to knock a quark out of a proton, but no one had been able to do this. The quark model also conflicted with Pauli's exclusion principle that states you can not have two identical fermions in the same quantum state. Quarks are fermions, yet in some particles, such as the omega minus particle, three identical quarks have the same spin. A further problem was that there needed to be some explanation of why quarks in the same particle with the same charges did not repel each other and destroy the particle they made up.

The solution to these problems was the idea that each quark had to be of one of three colors. Each quark in addition to its other properties had to be either red, blue or green. These colors are not real colors, they are just labels for a particular property quarks possess. The property could just as easily have been called small, medium and large or breakfast, lunch or tea. Each quark is considered to be of one of the three colors so that up quarks for example will be either a red up quark, a blue up quark or a green up quark. Equally every down quark will have color and be either a red down quark, a blue down quark or a green down quark. This applies for all the different types of quarks.

The quark color idea explains why quarks do not exist outside of the particles they make up. This is because the quarks are held together within the particle by the exchange of particles known as gluons. Gluons are able to carry color and create a color field between the quarks when gluons are exchanged. When quakes move apart and threaten to break out of the particle, the color field between the quarks becomes stronger and holds the quark within the particle. Should the color field be overcome the quarks will combine to form new particles containing the quarks so that the quarks are never seen independently but are always constituents of a larger particle.

Quark colors also mean that there is no conflict with Pauli's exclusion principle which states that two identical fermions cannot be in the same quantum state. Quarks in the same particle are not in the same quantum state if they have different colors so the exclusion principle can still apply.

Color has never been detected because the various colors of the quarks within a particle cancel each other out to make the particle containing the quark colorless. Quarks can only exist in combinations that make the particle they are contained in colorless.

There are six types or flavours of quarks. These are the up, down, strange, charm, top and bottom quarks. Each quark comes in three colors giving a total of 18 types of quark or 36 if their antiquarks are included. The quarks plus leptons, such as electrons, constitute the basic constituents of nature. These particles and the various particles involved in the transmission of force, such as gluons, photons and vector bosons constitute the standard model of elementary particles accepted at the current time.

Analysis

The change from the traditional Western and Chinese view of the elements involving materials such as water, air, earth, wood, metal and fire, to the chemical elements making up the periodic table, to atoms, to particles such as protons, neutrons and electrons and then to quarks was inevitable. The order of discovery of these ideas of the ultimate constituents of matter was necessary, in that they could not have been discovered in any other order. The traditional Western and Chinese views of the elements was always going to be the earliest explanation of the structure of the material world. This was because the traditional Western and Chinese view of the constituents of matter were based upon elements visible to the unaided senses, so inevitably they were discovered before the chemical elements found in the periodic table most of which are not directly visible to the naked eye and are often mixed up with other substances. It is hardly surprising that pre-scientific societies considered air, water and earth as basic constituents of matter as the air, water and earth are all around us and they seem to represent each of the various forms of matter, gas, liquid and solid. All matter is to some extent like water, if it is liquid, air, if it is a gas and earth if it is solid. The Chinese elements wood and metal are normally solids and are important to human beings and in the case of wood is an extremely widespread material. Fire is something widespread in human societies and it is of vital importance in a wide range of areas such as cooking, warmth, light, metallurgy and pottery. In general the elements in both the pre-scientific West and China were all around us, reflected the states of matter, solid, liquid and gas and were of great importance to us so it is to be expected that they were considered to be the basic constituents of all matter.

When some of the traditional Western and Chinese elements were broken down, air into a mixture of nitrogen, oxygen and other gases, and water into hydrogen and oxygen, it became clear that the traditional elements were not the basic constituents of matter. The chemical elements discovered by scientists in the late 18th and 19th centuries and which became incorporated into the periodic table became the new basic constituents of matter. Dalton's atomic theory was created to explain the differences between the new elements, each element having its own particular atom and being of a different weight from the other atoms. The atomic theory was largely accepted by the end of the 19th century and was confirmed by Einstein's explanation of Brownian motion and the discovery of the constituents of the atom.

The first constituent of the atom to be discovered was the electron. This was because the electron has the property of being easily separated from the atom and because it was the carrier of electricity. Electricity is a force visible to unaided human senses and the electron was discovered through studies made of electricity. Experiments investigating electricity in a near vacuum in a Crookes tube lead to the discovery of the electron. Those experiments with electricity in the form of cathode rays showed the cathode rays were made up of particles with an electric charge. Given that the particles were not a form of electro-magnetic radiation due to their low velocity, they could only come from the atoms which constituted matter.

The discovery of the electron was dependent upon the prior discovery of how to control electricity and of how to create vacuums which led to the creation of instruments such as the Crookes tube. The electron might also have been discovered by investigations of beta decay which consists of electrons or from the bombardment of atoms with alpha particles where the scattering of the alpha particles could have revealed the existence of small negatively charged particles or by the photo-electric effect where electrons are knocked out of metals. Electrons were discovered through the use of Crookes tubes because the Crookes tube was the earliest method for the discovery of electrons to become available. However whatever method was used it is quite clear that the electron was only going to be discovered in the late 19th or early 20th century. It was not going to be discovered by the Romans or in Medieval Europe and Asia. It could only be discovered after certain prior discoveries had been made and those discoveries had not been made during Roman and Medieval times.

The proton was the next particle discovered because it was a charged particle and the atom required a positively charged particle to balance the negatively charged electron, so that the atom could be electrically neutral. The discovery of the proton and the atomic nucleus was made by Rutherford's bombardment of gold foil by alpha particles and the analysis of the alpha particle scattering. This was only possible because of the prior discovery of radioactivity by Becquerel in 1896 and alpha particles which quickly followed from Becquerel's discovery. Rutherford's experiment was only possible because nature had provided radioactivity, a form of decay of unstable elements, which involves the emission of alpha particles which could be used to bombard other substances. If radioactivity and alpha particles did not exist, protons would not have been discovered when they were. Their discovery would have been delayed until either the development of particle accelerators, which first occurred in the 1930's or cosmic ray experiments which also began in the 1930's.

The neutron was discovered in 1932 and was the last of the constituents of the atom to be discovered. The discovery occurred later than that of the other atomic constituents due to the particular properties of the neutron. Unlike the proton and the electron the neutron carries no electric charge, so its existence will not be disclosed by experiments involving the scattering of charged particles fired at the atomic nucleus. If the neutron had a charge it would have an effect on particles fired at the atomic nucleus, either attracting them or repelling them and so showing its existence. As it has no charge, it has no effect on such particles, so it was quite hard to detect the neutron. The neutron, unlike the proton and the electron does not survive outside the atomic nucleus, as when it is outside the nucleus it decays into a proton, an electron and an antineutrino. The neutron has an important property, its mass and this property was what allowed its discovery. The neutron's mass explained why the elements other than hydrogen had atomic weights in excess of their atomic numbers and why different isotopes of the same atom could exist. It was the mass of the neutron that enabled the neutron to knock particles out of the paraffin wax in experiments involving beryllium rays. It was the properties of the neutron, such as its lack of charge and inability to survive outside the nucleus that were the reasons for the neutron being the last constituent of the atom to be discovered. It was, however, its property of having mass, that led to its discovery. In general it was the properties of the constituents of the atom that determined their order of discovery. The particles with charges and which survive outside the atom were discovered before the particle with no charge and which decays into other particles when outside the atom.

The simple system of the proton, neutron and electron as the basic constituents of matter was not to last. The invention of a series of new instruments such as particle accelerators and particle detection methods such as cloud chambers, bubble chambers and other detectors were to result in the discovery of a vast range of new particles. The particles were to be discovered using both particle accelerator and cosmic ray experiments. These particles which were to collectively be called the particle zoo could not have been discovered without the invention of particle accelerators or without the existence of cosmic rays. If nature did not provide cosmic rays or if particle accelerators could not be built for example if it was not possible to accelerate particles due to the electro-magnetic force not existing, then it most likely would not have been possible to discover those particles. If it was not possible to detect particles by means of instruments such as cloud and bubble chambers we would not have known of the existence of many of the particles that make up the particle zoo. If nature had been different then the particle zoo may never have been discovered so human history would have been different.

The particle zoo inevitably led to a study of those particles and that led to a classification scheme which resulted in the eightfold way analysis of the particles independently developed by Murray Gell-Mann and Yuval Neeman. The eightfold way, which was based on SU(3) Lie group mathematics, revealed a pattern among the particles which indicated some underlying situation that gave rise to the pattern. Using the pattern provided by the eightfold way Murray Gell-Mann and George Zeweig were able to predict that hadrons, which include protons and neutrons, were made up of particles that were eventually called quarks. Acceptance of the quark model took some time but eventually the discovery of quarks and the idea of colors which explained why quarks did not exist outside of hadrons and why quarks did not contradict Pauli's exclusion principle led to the acceptance of the quark model.

The idea of quarks could only be developed after the discovery of the particle zoo. It was the particles making up the particle zoo, and more particularly their properties that provided Gell-Mann and Neeman with the information that led to their suggestion that hadrons were made up of particles that were to be called quarks. The existence of these particles and their properties could explain the properties of the particles they made up. This idea based upon various graphs representing particles and showing a pattern among the particles that could be explained by some underlying factor such as hadrons being made up of other particles which are now called quarks. The properties of quarks such as charge, spin and isotopic spin could be calculated from the properties of the hadrons they made up. This meant that quarks could not be conceived of until after the discovery of the particle zoo and the eightfold way analysis of the particle zoo.

There is a hierarchy in the discovery of the constituents of matter from the original Western and Chinese traditional elements, to the chemical elements that make up the periodic table, to atoms, to protons, neutrons and electrons, which are the constituents of atoms, to quarks which are the constituents of protons and neutrons. Human ideas of the constituents of matter had to follow that order of discovery, with each particle being discovered before its constituents, as it was the order of discovery of that which is closest to us to that which is furthest from us. The order of discovery of particles is also affected by the properties of the particles. The charges of particles, their mass and ability to survive outside the particles they make up and other properties will make a particle harder or easier to discover.

Nuclear weapons are only possible due to nuclear fission. Nuclear fission is only possible due to the existence of very heavy elements such as uranium and plutonium and to the existence of the neutron. If the very heavy elements like uranium and plutonium did not exist then fission would not be possible as it is only those elements that will break apart when hit by neutrons and release their own neutrons to start a chain reaction. Only very heavy elements contain sufficient neutrons to create a sustainable nuclear reaction.

The second requirement for nuclear fission is the existence of neutrons. If neutrons did not exist then nuclear fission and nuclear bombs and power would not exist. Neutrons are used to start the fission process, but if neutrons did not exist then possibly some other particles could be used for that purpose. However for the continuation of the process, the creation of a chain reaction neutrons are necessary as the only other constituent of the nucleus, the proton would be repelled by other uranium or plutonium nuclei as they are both positively charged. If nature did not provide both heavy elements such as uranium and plutonium or did not provide the neutron then nuclear fission would not be possible. If nuclear fission was not possible then both nuclear bombs and nuclear power would not be possible.

Nuclear bombs and power could not be invented until after uranium and plutonium had been discovered. Uranium was first discovered in 1789 and was isolated in 1841. Plutonium was first produced in a cyclotron in 1941. The important discovery was that of the neutron by James Chadwick in 1932. As soon as the neutron had been discovered scientists began to use it to bombard various elements as its lack of an electrical charge made it a more effective bombardment tool than the positively charged alpha particles. The bombardment of uranium resulted in nuclear fission and the awareness among physicists that the breaking up of the uranium nucleus would release enormous amounts of energy that could be turned into a bomb. The Second World War caused research into the making of a nuclear bomb to be sped up so that both plutonium and uranium bombs were produced by 1945. If there was no Second World War nuclear weapons would have taken somewhat longer to develop, but even without the war nuclear weapons would have been created probably in the 1950's.

The development of nuclear weapons led to the development of nuclear power when the world's first nuclear reactor was built in 1942 and created the world's first self sustaining chain reaction. Nuclear reactors for the generation of electricity began to be built in the 1950's and 1960's in many countries.

Nuclear weapons and nuclear power are only possible because nature provides the materials for nuclear weapons and power in the form of heavy elements such as uranium and allows the production of plutonium and because nature provides neutrons which allow the creation of a sustained chain reaction within reactors. Without uranium and plutonium and neutrons history would be greatly different as there would have been no nuclear weapons or power in the 20th century. Given that nature provided uranium and neutrons and allows the production of plutonium, human curiosity about the world around us, would inevitably, sooner or later have led to the development of nuclear weapons and power. Nuclear weapons and power were developed when they were in the 20th century as they could not have been developed until a vast number of prior discoveries had been made.

The use of radioactive materials in medicine, industry, agriculture and science only occurred after the discovery of radioactivity and knowledge of how it could be created in accelerators and nuclear reactors and the realisation that such materials could be useful in medicine, industry, agriculture and science. Radioactivity was discovered in 1896 and the realisation that it could be created in cyclotrons and reactors occurred quickly after the invention of those machines. Studies of radioactive materials lead to the realisation of their uses in medicine, industry, agriculture and science. In particular, their use in radioactive dating led to a revolution in ideas of the age of the earth and in the dating of prehistoric and historic organic materials. Previous ideas of the age of the earth and of organic remains had no objective basis and were just unscientific guesses. Only with the advent of radioactive dating were reliable dates able to be established in archaeology and geology.

It is only because nature provides radioactivity that we have radioactive dating. If nature did not provide radioactivity we would still have no real idea of the age of the earth or of archaeological remains. Nature also allows the creation of radioactive materials in cyclotrons and nuclear reactors which provide materials for medical, industrial, agricultural and scientific uses. If nature did not allow this artificial creation of radioactive materials, their use in medicine, industry, agriculture and science would be vastly reduced. We would only be able to use naturally occurring radioactive materials and if nature did not provide them we would have no sources of radioactive materials for use in medicine, industry, agriculture and science.

The analysis of the discovery of the atomic world and the constituents of matter shows how the discovery proceeded through a particular path and that it had to proceed through that path. Particles were discovered in a logical order and when they were discovered the social and cultural effects of the discovery appeared in human history. When the neutron was discovered and combined with heavy elements such as uranium and plutonium, nuclear weapons and power were quickly invented. The discovery of radiation and the artificial creation of radiation in accelerators and nuclear reactors was quickly followed by the use of radioactive materials in medicine, industry, agriculture and science and for radioactive dating. If however the structure of matter was different, for example there were no neutrons or radioactivity, then there would have been no nuclear weapons or power and no use of radioactivity in medicine, industry, agriculture and science and for the dating of ancient materials. If however the structure of matter was different it may have provided humans with other means for meeting human needs. If nuclear fusion was possible at normal temperatures we may have had an abundant, clean and safe energy source which may have ended human reliance on fossil fuels and the concerns about the greenhouse effect. If the structure of matter was different then certainly the course of human social and cultural history would have been different.

Conclusion

These case studies cover a number of the most significant events in human history. They show that these events occurred only because of the characteristics of certain materials and matter that exist within the human environment. The development of stone tools was only possible because certain rocks, when hit with another rock, would break in a conchoidal fracture to provide a flake with a sharp edge. The discovery of how to use fire was only possible because organic materials would ignite and burn when subjected to sufficient heat. The domestication of plants and animals was only possible because a small number of plants and animals had a particular genetic make-up that gave them certain characteristics which made it possible for humans to domesticate them. The invention of pottery was dependent upon the particular qualities of wet clay, that it can be molded into a variety of shapes which it will retain when dried. Metallurgy became possible due to the ability to smelt metals from their ores and to melt the metal to make alloys and to allow casting of the metals. Smelting and melting were only possible due to the particular temperatures at which ores would smelt and metals would melt. Writing was only possible due to the ability of humans to give symbolic representation to the things they see and the sounds they make in their spoken languages. Glass could only be made due to the presence of materials such as silica, soda and lime in the human environment and to their particular characteristics that when they were heated and mixed together they produced the solid transparent substance we call glass. Microscopes and telescopes were dependent upon the prior invention of glass and upon the way in which light changes direction as it moves through one medium to another. Different shaped lenses using the law of refraction enable the light to be focused at different points allowing magnification of the object being observed. The steam engine and the internal combustion engine are only possible due to certain characteristics of gases and vacuums. Gases expand when heated and vacuums in a cylinder will cause a piston to move to reduce the size of the vacuum. Aeroplane travel was only possible due to the internal combustion engine (and later the jet engine) and because an appropriately shaped wing will cause the air pressure under the wing to be greater than air pressure above the wing causing wing and aeroplane to rise. Electricity is only possible because certain materials reasonably easily lose electrons to allow an electric current to be created. Electricity generators and electric motors were able to be developed due to the ability to turn motion into electricity, by means of a moving electro-magnet, and the ability to turn electricity into mechanical energy. Photography was possible due to the camera obscura effect and because certain chemicals could make an image permanent. Motion pictures were only possible due to the persistence of vision and the ability to produce photographs with very brief exposure times. The explanations given for the origin of infectious diseases in both classical western and traditional Chinese medicine were similar and based on supernatural explanations or naked sense observations of the behavior of the human body. If the human body had behaved differently, different theories would have arisen to explain illness. The theories that arose were logical and based on the best information available at the time and theories of that type were inevitable as they were the only plausible explanations of disease available at that time. Immunization as a remedy for disease was only possible due to the body having an immune system, which can be made to quickly attack invading organisms, if it has previously been introduced to those organisms. If the human body was different then the treatments available would be different and the history of medicine would have been different. Modern surgery is only possible because materials exist in the human environment that can be turned into anesthetics. If those materials did not exist then surgery would still be carried out as it was in the past, quickly, painfully, rarely and limited to only a small range of operations. Nuclear fission, which is the process used for nuclear power and weapons is only possible due to the existence of neutrons and very heavy elements such as uranium and plutonium. If neutrons or the very heavy elements did not exist, there would be no nuclear power or weapons. Modern scientific knowledge of the age of the earth and of organic materials is only possible because certain elements are radioactive and decay at a set rate allowing calculations of the age of the materials. If there were no radioactive elements then we would still have a very limited idea of the age of the earth and of a vast range of archaeological materials. What could happen in human social and cultural history is controlled by the structure of nature and the laws of the natural sciences. If nature was structured in a different way or the laws of physics, chemistry and biology were different then human social and cultural history would have been different.

The properties and characteristics of the materials in the human environment do not just allow human beings the opportunity to do certain things like making tools, pottery, glass, engines and scientific instruments, they also have a great effect on the course of human history, in that the order in which such discoveries and inventions are made is affected by how easy or how difficult it is for humans to make the inventions or discoveries. It was a lot easier to invent stone tools than metal tools, so inevitably stone tools were invented and used long before metal tools were invented or used. It was a lot easier to use and control fire, than to learn how to make it, so people learnt how to use fire, long before they learnt how to make it. Some discoveries and inventions cannot be made without certain prior discoveries or inventions being made. Glass and most pottery and metallurgy required the prior discovery of how to use and control fire. There is a considerable range of inventions and discoveries that either could not or would not have taken place without the development of widespread sedentism which was dependent upon the domestication of plants. Those inventions and discoveries include mathematics, writing, substantial permanent buildings, metallurgy, pottery and glass manufacture. Pottery and substantial permanent buildings could have been made by hunter-gatherers but were not, as it would have been uneconomic for nomadic people. Printing could not have been developed without the prior development of writing and printing with moveable type could not have been developed without the invention of the alphabet as it would not have been economic without an alphabetic writing system. Telescopes and microscopes would not have been possible without the prior discovery of glass making and some knowledge of the law of refraction. Without the microscope we would have no knowledge of micro-organisms and the cause of many diseases. Without the telescope we would probably still believe the sun and planets orbited an unmoving earth. The change from an earth centered universe, the common sense theory for societies with unassisted vision, to a sun centered theory such as the Newtonian system, and then to general relativity was inevitable. This was because the knowledge provided by unassisted vision, naturally lead to an earth centered universe, the knowledge available from 17th to 19th century telescopes and mathematics lead naturally to a sun centered system such as the Newtonian system and better telescopes and new mathematics such as non-Euclidean geometry lead naturally to a theory such as general relativity. The steam engine could not have been invented without prior scientific discoveries concerning the behavior of gases and vacuums. The petrol-driven internal combustion engine required the same knowledge of gases and vacuums as the steam engine and also the prior discovery of oil exploration. The internal combustion engine was a necessary prior invention for the invention of the motor car and the aeroplane. The electric motor and the electric generator could not have been invented without the prior discovery of the electro-magnet, which was dependent upon the prior discovery of the voltaic pile, which itself was dependent upon previous scientific investigations of electricity. The history of chemistry reveals many examples of discoveries that could only have been made after certain earlier discoveries had been made. The modern idea of the elements in chemistry could only have been developed after traditional elements such as air and water had been shown to have been composed of other substances. The periodic table could only be discovered after the modern concept of elements had been made and a significant number of elements had been identified. The identification of the elements depended upon prior inventions such as the pneumatic trough, the voltaic pile, potassium analysis and spectroscopy. The development of the periodic table was also dependent upon a means of calculating atomic weights, which was provided by Avogadro's theory. Photography was dependent upon the prior discovery of the *camera obscura* and the appropriate chemicals that would allow a picture to be preserved. Motion pictures were dependent upon the persistence of vision and photography with very low exposure times allowing the taking of many photographs per second. The development of improved microscopes, in the 19th century, led to the discovery of specific microorganisms, which caused specific disease. This resulted in the end of traditional explanations for disease and the development of the germ theory of disease. The germ theory could not have arisen until after it had been clearly shown that germs cause disease and that required the prior invention of sufficiently powerful microscopes. Modern surgery could not have been developed until after the discovery of anesthetics and the importance of anti and a-septics in the 19th century. Anesthetics required prior discoveries in chemistry and anti and a-septic practices required the understanding that germs cause infection. Modern immunization practices require the prior discovery of which microorganism causes which disease so that a weakened strain of the organism could be used for immunization. It also required the understanding that giving the patient a weakened dose of the disease could protect the patient from future infections. The order of discovery of the constituents of matter from the traditional western and Chinese elements, which are visible to the naked eye, to the chemical elements of the periodic table, to atoms, to protons, neutrons and electrons and then to quarks was an inevitable order of discovery as each discovery involved a breaking down of the previous constituent which first had to be discovered before it could be broken down. The development of nuclear power and weapons could only take place after the discovery of the neutron and the very heavy elements. Those discoveries were dependent upon a series of discoveries such as radiation and alpha particles and the splitting of the atom. The order of discovery concerning the materials in the human environment and of technology that resulted from such discoveries was not haphazard or accidental. The order of discovery followed a logical order and an order that it had to follow. The easier discoveries were made before the harder discoveries; discoveries that were dependent upon prior discoveries being made, were only made after those discoveries; and inventions that were not economic or did not meet human needs were not made until they made economic sense or until a need arose. The course of human social and cultural history is written into the structure of the universe.

The structure and characteristics or properties of the materials in the human environment provide human beings with the opportunity to do certain things, like making stone or metal tools, pottery, glass, or various engines, or domesticating various plants and animals. Whether human beings take advantage of these opportunities depends upon the characteristics of individual human societies. Every human society will contain a wide variety of human personalities, some of them open to new and better ways of doing things. Other human personalities are by nature conservative and are inclined to cling to the established way of doing things. In societies where the conservatives are politically or culturally dominant, new ideas and techniques might be banned or simply not be able to be established in that society. In societies where innovators are politically and culturally dominant new ideas and techniques may well, if they are good enough, be adopted by the society. History provides many examples of conservative and innovative societies. Conservative societies would include Tokugawa Japan (1612-1868), Ming and Manchu China, Ottoman Turkey and 17th and 18th century Spain and Portugal. Innovative societies would include Great Britain since the 17th century, the United States through its history, Japan since the Meiji Restoration (1868-to the present), Egypt under Mehemet Ali and Russia under Peter the Great. Some societies can be innovative in some ways and conservative in others. Europe in medieval times was conservative when dealing with ideas, especially those concerning religion, but was reasonably innovative concerning technology. Some societies may be conservative at some times in their history and innovative at other times. What is clear is that at one time or another there will always be some societies which are innovative. It is also clear that over the long term the innovative societies are likely to be more successful and dominant than the conservative societies. European dominance of the world since the Renaissance is largely because European societies or many of them were highly innovative. Ottoman Turkey's decline into the "sick man of Europe" is due to it being a conservative state which was resistant to new ideas and technologies. The ideas and technologies were easily available to the Turks, being produced in abundance by northern Europeans. If the Turks had been as open minded to new ideas as were the Japanese after the Meiji Restoration the decline of the Ottoman Empire may never have happened.

The difference between European states and society since the Renaissance and Ming and Manchu China in attitudes to new ideas and technology was immense. Etienne Balazs in an essay *Significant aspects of Chinese Society* published in *Chinese Civilisation and Bureaucracy* states:

"Chinese ingenuity and inventiveness ... would probably have brought it to the threshold of the industrial age, if they had not been stifled by state control. It was the state that killed technological invention in China."[57]

A crucial factor for the development of new ideas and new technology was the feeling that progress was possible; the belief that inventors held that if they thought about things enough and tried this and that, they might eventually be able to work out an answer to the problem. Such a belief would not exist to anything like the same extent in Confucian China as it did in Europe since the Renaissance. Etienne Balazs considered the intellectual climate of Confucian orthodoxy was not favorable to any form of trial and experiment, to any sort of innovations or to the free play of the mind. The imperial bureaucracy was quite satisfied by the traditional techniques which satisfied its traditional needs. The Chinese mandarins had little interest in science, commerce and utility. Their principal field of study was ancient Chinese authors. A late seventeenth century Jesuit traveler noted that educated Chinese were more attracted to antiquities than modern things. He observed this directly countered the Europeans' love of novelty for its own sake.[58] Chinese culture, compared to Europe, was static and conservative and lacked the sense of progress so strongly present in European culture.

One reason why a society may be conservative and opposed to innovation may be due to the activities of special or vested interests. In *The Writing Systems of the World* Coulmas suggested Egyptian hieroglyphics were:

“hard to learn and the privilege of an elite group. Naturally this group, the clerks and priests of the royal household of the Pharaoh had no desire to endanger their status. … It was in the best interest of those few to guard their privilege and make sure that writing was complicated and not readily available for everybody. … Indirectly the social privilege of writing may well have contributed to the stability and conservatism of the Egypt script whose development stopped short of the alphabet.”[59]

Special interests opposing change can often delay technological and other changes in any society or culture.

Conservative societies or groups within societies can sometimes delay change, but they do not seem to be able to stop it completely. This is because competition from more innovative societies may force a conservative society to be more innovative. When Commodore Perry’s ships arrived in Japan in 1853 the Japanese realized their weakness and began a process of modernization. The loss of almost all Ottoman territory in Europe to Austria, Russia and newly emerging Balkan states caused Turkey to modernize in the early 20th century. The weakness of China in the face of European power in the 19th century and Japanese power in the 20th century caused China to begin modernization in the 20th century. Economic competition within a society may also force change as new technologies are developed. Businesses that fail to adopt new technologies may well find themselves uncompetitive in the market place and liable to go bankrupt and to be replaced in the market place by more innovative businesses. When states pass laws banning innovation such laws will often be broken by those who do not accept such laws or who simply hope to profit from breaking them. Even if a society was totally isolated from all other societies there would be periods when those who support innovation would be in power so that conservatives would tend to delay rather than stop change. In the longer term good ideas are almost certain to be adopted.

Appendix 1

A Problem with some Theories of History,

Social Change And Cultural Evolution

By Rochelle Forrester

ABSTRACT

Marx, Durkheim, White, Johnson and Earle and Sanderson produced linear theories of history, social change and cultural evolution but their theories have a common deficiency. None of them provide an ultimate explanation for social, cultural and historical change. This failure was rectified by J. S. Mill, who suggested increasing human knowledge was the ultimate cause of social, cultural and historical change. However, even Mill did not ask what caused the increasing human knowledge and why the knowledge had to be acquired in a particular order and how this could affect human history.

Attempts to understand and observe a pattern in the broad sweep of history are usually known as substantive or speculative theories of history or as macro-histories. A considerable number of such theories of history, social change and cultural evolution have been proposed by various philosophers, historians, sociologists and anthropologists. Twenty such theories are offered in *Macrohistory and Macrohistorians* [60](ed) by Johan Galtung and Sohail Inayatullah and thirteen in Part I of *Theories of History* [61](ed) by Patrick Gardiner and more are offered in *Philosophies of History* [62]by Rolf Grunger and *Philosophy of History* [63]by Alan and Barbara Donagen. Not all deal with the same subject matter, as some tend to deal with the rise and fall of civilizations, empires, cultures and religions. Others tend to deal with economic, technological or scientific changes in history. Such theories of history, social change and cultural evolution usually attempt to discern a pattern or meaning to history. Sometimes a linear pattern showing history as moving towards a particular end or result is proposed. A cyclical pattern is sometimes proposed involving history in some way periodically repeating itself. Sometimes a mixture of the two is proposed involving repetition in history accompanied by progress towards some end.

Speculative theories of history, social change and cultural evolution often involve a mechanism or an explanation as to how change takes place in history. They may also propose a purpose or justification of history.

This paper will examine six speculative theories of history, social change and cultural evolution all of which propose a linear pattern to history, and will point out a deficiency common to five of them. This deficiency is the lack of an ultimate driving force for historical change, a deficiency which is met by the idea that increasing human knowledge is the ultimate driving force for history. This view is stated by Mill in the fourth philosophy of history examined in this paper, but he does not say what causes the increasing human knowledge and why it comes to us in a particular order.

Marxism

There are many interpretations of Marxism and the interpretation proposed below is one close to the theories suggested in this book. Marxism proposes a linear development of history from a state of primitive communism, to ancient or slave society, to feudalism, capitalism and then to socialism. Marx begins by noting that humans work to meet their needs. We work to meet our basic needs of food, shelter, clothing and reproduction and also to meet our higher, intellectual, imaginative and aesthetic needs.

The production and reproduction of human society requires the use of productive forces. Productive forces consist of human labour power and the means of production. The means of production consists of the instruments of production and the raw materials that labour power works on.

The process of production requires humans to be in some relationship with the means of production and with each other. These relations are work relations required during the productive process and ownership relations which concern control of the means of production. The ownership relations of production determine who owns the means of production. The people who own the means of production form a particular class. Those who have no ownership rights in the means of production will form another class. Throughout history (except for a period of primitive communism) there has always been one class who owned the means of production and another class who did not. The class that owned the means of production tended to be economically privileged and were the ruling class in any society. The ideology of the ruling class tended to be the ideology of society as a whole. Society was divided into an infrastructure which consisted of the productive, economic part of society and a superstructure, which consists of the political, legal, religious and other non-economic aspects of society. The nature of and changes in the superstructure will usually be caused by the nature of and changes in the infrastructure, although the superstructure may have some limited effect on the infrastructure.

The relations of production and with it the class structure and ideology of a society may change as a result of a change in the productive forces. Marx is not particularly clear on this point, but it seems hard to see how the relations of production could change without change in the productive forces.[64] The relations of production change due to an increase in societies productive capacity, which requires a change in the forces of production.[65] Marx does not however say what causes changes in the productive forces.[66] Often the changes in the productive forces can be accommodated within the existing relations of production, so no change in the relations of production is required. However, on occasion the improvement in the productive forces will require changes in the relations of production in order to ensure the full productive potentiality of the change in the productive forces will be utilised. This situation may result in changes in both the work relations of production and the ownership relations of production. A change in the ownership relations of production will result in a new class owning the means of production and such a change is likely to occur only after a period of class war between the class owning the means of production and the class about to assume ownership of the means of production. A change in the ownership relations of production will result in a new mode of production as for example when primitive communism was replaced by the slave society, which was in turn replaced by feudalism, which was replaced by capitalism. These changes in the mode of production involved in ancient society, the slave owner owning the means of production, in feudal society, the lord owning the means of production, and under capitalism, the capitalist owning the means of production. In each of these modes of production there was a class that did not own the means of production. In ancient society this class was the slaves, in feudalism it was the serfs and under capitalism it was the workers.

The above description of Marxism constitutes a model of historical change. Marx however also attempted to give an actual description of the historical changes that led to the capitalist mode of production. Primitive communism begins with hunting and gathering, but develops into agriculture and pastoralism, but with the land still being owned communally. This society was classless but was soon to be replaced by societies based on slavery.

Slavery began due to war, but was further encouraged by the development of agriculture and animal rearing and by crafts such as metal working and weaving. Slaves were initially owned communally but private property in slaves developed, encouraged by trade and by an increasing economic surplus. Eventually private property in land develops due to individual's eagerness to own the land they possess. Private property developed due to the increasing productive capacity of society as economic surpluses were better utilised privately rather than communally. However as private property developed, inequality grew, as private owners could lose their property through usury, mortgages or trade in a way in which communally owned property could not be lost.[67] This lead to the establishment of a wealthy aristocracy and a further growth of slavery. The mode of production of primitive communism gave way to a slave society, ruled by a state and with a new superstructure and divided into slave owners, free men and slaves.

The earliest form of slave society is the Asiatic mode of production. Asiatic societies are ruled over by despots who organise large-scale public irrigation projects to assist their society's agricultural production. The people live in villages and engage in agriculture and small handicraft manufacture. The economic surplus produced is consumed by the government leaving the villagers with little incentive to improve their productive forces. The productive forces of these societies, for example, pre-European India, fail to grow as private property in land and slaves does not develop, as everything is under the control of the despot.

The ancient mode of production, which involves both classical Greece and Rome, was dominated by slave labour. Private property in both land and slaves developed and the rich and powerful dispossess the poor. Large land estates worked by slaves dominated Roman agriculture. Initially slavery leads to significant productive progress and ensures the full utilisation of the productive forces. However slavery eventually retards development of the productive forces as the supply of slaves dries up, slaves require considerable supervision, slaves are reckless with the instruments they use and slavery causes material production to be stigmatised. While Rome had many of the prerequisites for the development of capitalism, such as widespread commerce, money and free labourers, Marx considered it could not have developed capitalism due its insufficiently advanced productive forces. Class conflict leads to the destruction of the classical world. While this is happening slavery on the large estates becomes uneconomic and gives way to hereditary tenancies.

The feudal mode of production begins when serfdom develops as peasants seek protection due to widespread lawlessness, wars and Viking invasions. The serf is tied to the land and the surplus the serf produces goes to his lord. The serf's responsibilities to the lord are usually fixed, but his production on his own land is not fixed, so there is an incentive for economic development. This means there is encouragement for the improvement in productive forces and allows an increase in production for exchange. Under feudalism the towns grew in population and the guild system developed to protect and enhance craft industries. The craftwork improved in quality and efficiency and tools were improved and skills developed. But ultimately the guild system began to hamper further development of handicraft industries by stopping large-scale technical production, with an increased division of labour, which was needed to allow improvements in the productive forces. The guilds limited the master's capital and the number of workers the master could employ.

Marx considered that feudalism decayed before the beginning of capitalism. Feudal production relations dissolved, private property spread and money relations grew. This eventually allowed a primitive accumulation of the conditions necessary for the beginning of capitalism. These conditions were a population of free labourers, free of feudal encumbrances, and a means of production that could allow their independent subsistence and capital adequate for financing industrial production. The free labour force is produced by the expulsion of the peasantry from their land and the capital is derived from the colonial system, commercial wars, over taxation and protectionism. Capital was also derived from usury and the capital of merchants. Merchantry also encouraged the production of goods for trade, a necessary element in capitalism. Capitalist attitudes such as the worship of private property and the pursuit of profit are present in usury and merchantry.

Merchants would hire labour to manufacture the goods they intended to sell and would become capitalists. Producers who buy their own raw materials, rather than from a merchant, and who produce for the world market, would also become capitalists. Some artisans and even wage labourers would also be able to turn themselves into small capitalists. Under the pressure of these developments the guilds collapse and the capitalists are able to hire any number of workers they like.

The final requirement for capitalism is the development of productive forces able to support capitalism. The productive forces need work relations that can get the maximum production from the productive forces. Capitalism provides this by organising production more efficiently and increasing the division of labour. Capitalism develops, as it is inevitable that people will attempt to make money by hiring free labour.

Durkheim

Durkheim, a sociologist, was interested in the relationship between the individual and society. He considered this relationship changed over time and this led him to produce a theory of long-term social change.

Durkheim produced this theory in his book *The Division of Labour in Society*.[68] In this work he suggested early societies had a form of social solidarity he called mechanical solidarity. Such societies were characterised by a very low level of division of labour, so that all members experienced the same conditions of existence and carried out one of a limited number of roles within society. Social organization was simple and local and takes the form of an aggregation of individuals. Such a society, because it is an aggregate, rather than a collection of mutually dependant parts, may lose a part of itself and can continue to function. This situation is analogous to simple organisms that can divide to form new organisms. The parts of such a society are held together by mechanical solidarity. This solidarity is derived from commonly shared beliefs that exist because members of the group share the same conditions of existence. Property is owned in common and such a society has a low level of individualism. The commonly shared beliefs are called the conscience collective or the collective conscience. Religion is a typical form of the conscience collective in early societies. It tends to be local and concrete in its ideas and deals with beings that are connected to natural phenomena such as animals, trees and storms.

The best way to understand the moral codes and the collective conscience of simple societies held together by mechanical solidarity is to observe their legal codes. Moral beliefs are not easily observed but law and the sanctions provided for breaches of law provide an external index allowing us to objectively assess the state of a society's moral beliefs. An investigation of the sanctions prescribed by codes of law will indicate what type of moral code a society has.

There are two main types of sanctions that may be provided by legal codes. Repressive sanctions are those that involve inflicting some sort of suffering or loss of liberty or even loss of life on transgressors. Infractions are usually severely punished, as they are a threat to the solidarity of the society. As religion is such an important part of the collective conscience and moral beliefs of society, breaches of law tend to be breaches of religious law. The function of repressive sanctions is to reaffirm the collective conscience so as to reinforce social solidarity. Repressive sanctions are typical of simple societies whose cohesion is maintained by mechanical solidarity. An alternative type of sanction are restitutive sanctions, which are common in areas of commercial and civil law. Restitutive sanctions involve restoring the state of affairs that existed before the breach of law. Restitutive sanctions are common in complex modern societies.

Small, simple societies with little division of labour held together by mechanical solidarity begin to change as population density and volume increases. Increasing population, improvements in transport and communications and the growth of cities all bring about increasing social interaction. This results in increasing competition and conflict over scarce resources and in the beginnings of a process of disintegration of the societies based on mechanical solidarity. An increased division of labour is a possible solution to the conflict caused by increasing social interaction. Increasing division of labour will mean a society will become more complex and made up of parts that are mutually dependent upon each other. It will become more organic like complex biological systems, which are made up of a number of independent parts, none of which can survive without the others. Durkheim considered such a society to be based on organic solidarity, rather than the mechanical solidarity, societies with little division of labour were based upon. Organic solidarity involved the interdependence of people in systematic relations of exchange with each other. Organic solidarity results not from the similarity of individuals, which is the basis of mechanical solidarity, but from differences between them. A society based on organic solidarity will still have a collective conscience, but it will be of a more secular nature than the collective conscience of a society based on mechanical solidarity. In particular organic solidarity allows an individualism that could not exist under mechanical solidarity.

The increasing population and social interaction which resulted in an increased division of labour will lead to the gradual replacement of repressive legal sanctions with more restitutive sanctions. The breach of religious rules cease to be regarded as criminal acts, although repressive sanctions remain in certain areas for offences against persons and property and for offences against the dignity and authority of the state. The change from repressive to restitutive sanctions reflected the change in the collective conscience that resulted from a change in society based on mechanical solidarity to one based on organic solidarity.

However the process of change from mechanical solidarity to organic solidarity produces strains and tensions within society. This takes the form of class and sectional conflict and social and psychological pressure on individuals. This is because social evolution takes place imperfectly and the de-regulation of the old moral order is not immediately replaced by a new moral order. This creates a situation Durkheim calls anomie, which involves the absence of regulation by either shared moral rules or formal legal rules. This situation is made worse by inequality caused by the inheritance of wealth and factors that stopped individuals from entering the occupations most suitable for them. The existence of anomie showed that the line of development of the division of labour had taken an abnormal or pathological course. Equality of opportunity was needed in societies with an advanced division of labour in order to produce organic solidarity.

Durkheim's solution to these problems is a system of regulation covering conditions of employment and creating institutes which would administer codes of conduct binding on all those engaged in particular occupations. Such regulation would create a normal form of the division of labour allowing organic solidarity in a society with considerable social differentiation, but with full equality of opportunity.

White

White proposed a theory of the evolution of culture based upon humankind's control of increasing quantities of energy in his book *The Science of Culture*.[69] Human culture can be divided into three subsystems of culture, the technological, the sociological and the ideological.

The technological consists of the material, mechanical, physical and chemical instruments and techniques used by humankind to survive in nature. It includes the tools and materials of production, subsistence, shelter and war. The sociological consists of interpersonal relations between individuals and groups which are expressed in patterns of behaviour. This includes the social, kinship, economic, ethical, political, military, religious, occupational and recreational systems that exist within a culture. The ideological consists of the ideas, beliefs, knowledge, myths, theology, legends, literature, philosophy, science, folk wisdom and common sense that exist within a culture.

Each of these sub-cultural systems influences and is influenced by the others. However, technological has a much greater effect on the other two than they have on the technological. When technological systems change, the social system will change with it. Technological systems determine social systems, the technological system is the independent variable, and the social system is the dependent variable. In a similar fashion each technological system will tend to have an associated ideological system that will change as the technological system changes. However the ideological system is also affected by the sociological system. All of these systems influence each other, but the technological system is much more powerful than the other two and it determines what sort of sociological and ideological systems exist within a culture.

All biological systems absorb energy in order to maintain themselves and to grow and develop. The same applies to cultural systems which must harness and control energy to meet human needs. The means by which they do this is by the technological instruments available within that culture. The efficiency of these technological instruments varies. The productive output of a culture depends upon the efficiency of the technological means by which energy is put to work. The degree of cultural development in terms of productive output is determined by the amount of energy harnessed per capita and by the efficiency of the technological means by which it is put to work. This can be expressed in the formula E ´ T ® C, where C represents the degree of cultural development, E is the amount of energy harnessed per capita and T is the efficiency of the technology used in the expenditure of the energy. This means culture will evolve as the amount of energy harnessed per capita increases or as the efficiency of the technological means by which the energy is put to work increases.

The earliest source of energy exploited by human beings was human energy. This form of energy is very limited so that the cultural development that can take place using this source of energy was also very limited. Improving the technological means of putting energy to work developed these cultures to some extent, but these cultures tend to be simple, meagre and crude. Fire, wind and water could be used as sources of energy, but only to a very limited extent in the earliest cultures. This is because they lacked the technology to use fire, wind and water as a substitute for human muscle power.

The first great increase in the amount of energy available for cultural development came from the domestication of plants and animals. The yield of food and other plant materials was much greater per unit of human labour from agriculture than could be obtained by the gathering of wild plants. The yield of food and other animal products, per unit of human labour, obtainable from domesticated animals was much higher than could be obtained from wild animals. Some domesticated animals could also be used to carry goods or to pull ploughs or vehicles.

The result of the great increase in the amount of energy, controlled by human beings, brought about by the domestication of plants and animals, was the great civilizations of antiquity, in both the old and the new worlds. Great cities arose, great engineering projects were built, ceramics, textiles and metallurgy were developed, astronomy, writing and mathematics began and great works of art were made. All aspects of culture saw great progress and development. However after a period of considerable progress, the cultural development plateaued and progress continued only at a very slow pace.

Cultural development only began to re-occur at a substantial rate when a new means of harnessing energy was developed. This new means of harnessing energy was the use of the steam and internal combustion engines to produce energy from fuels such as coal, oil and gas. This resulted in great increases in population and in wealth, bigger cities and a rapid development in the art and sciences. This increase in cultural development continues today and may be enhanced by the harnessing of energy from the atom. On the other hand the whole process may cease if atomic energy is used in a full-scale nuclear war.

The amount of energy harnessed by a culture is not the only determinant of cultural development. Tools and machines are required to put energy to work and the efficiency of those tools affects the amount of energy harnessed and the amount of cultural development that can take place. A more efficient bronze or iron axe will chop a tree with fewer strokes than a stone axe so that less energy is expended to achieve a given task. More energy is then available for other tasks, so that with more efficient tools more cultural development can take place, than with less efficient tools.

However there is a limit to how much tools can be improved. When these limits have been reached little cultural development can take place unless there is an increase in the amount of energy harnessed. There is no limit to the amount of energy that can potentially be harnessed, but there is a limit to the efficiency of the tools used to harness it. It is the amount of energy that can be harnessed that is the principal factor in cultural development.

A changing technological system will affect the type of social system within a culture. Societies based upon human energy tend to be relatively small and have little structural differentiation and specialization of function. Societies based on the early stages of agriculture and pastoralism also have only minimal social differentiation and specialization. They have a high degree of social equality, have free access to the resources of nature for all and are based on kinship ties.

When agriculture and pastoralism reached a certain level it became possible for part of the population to produce food for all. This enabled part of the population to work at activities other than food production. This resulted in society becoming divided along occupational lines and becoming structurally differentiated. As the population increased, kinship relations were replaced by a society based on property relations, states were formed and society was divided into two major classes. One class was a small powerful, wealthy ruling class and the other a large exploited class of peasants, serfs or slaves. These trends were encouraged by the development of organised warfare and by commercial practices such as money lending.

The social system created by agriculture and pastoralism had the effect of reducing technological progress, to such an extent that cultural development nearly ceased. This is because the ruling class had ample for its needs so did not feel any need to increase production by increasing the efficiency of its technology. The exploited class did not feel any need to make the technology more efficient, because if it did, the increased production would be appropriated by the ruling class. As neither class would receive any benefit from improved technology, technological improvements became very rare in societies based upon agriculture and pastoralism. This situation continued until the fuel revolution caused technological and cultural development to recommence.

The fuel revolution brought with it a great increase in population and a process of urbanisation that resulted in the great majority of people living in cities. A capitalist industrial economy and parliamentary political system replaced European feudalism. The social structure became even more differentiated and functions more specialised. A two class system remains but the ruling class consists of industrial and financial lords and the exploited class are an industrial and urban proletariat.

Mill

Mill's theory of history is contained in his book *A System of Logic*.[70] Mill begins by defining states of society by which he means the simultaneous state of all the greater social facts or phenomena. This includes the degree of knowledge, of intellectual and moral culture, the state of industry, the class structure, the form of government and law and the beliefs of society. Mill notes that the different elements that make up the state of society will usually have what he calls a uniformity of co-existence. Where certain elements exist, certain other elements will usually co-exist with them. Particular economic states tend to be associated with particular forms of government, law and religious and other beliefs. However the state of society at any one time is caused by the state of society preceding it, so the fundamental problem is to find laws by which a state of society causes the state of society that succeeds it.

This problem is made more complex as the character of human beings is caused by the circumstances in which they live, but also humans affect the circumstances in which they live. The effects, human character, react back to the causes, the circumstances in which humans live. This causes people and the circumstances in which they live to change over time and this change is of a linear or progressive character. It may be possible by examining the order of succession of the different states of society, to discover a law explaining and predicting this linear progression, but such a law would only be an empirical law and not a scientific law. Such an empirical law could not be used to predict future events unless it is connected to the psychological and ethological laws that control the action of circumstances on people. Only then will it become a scientific law. A scientific law of history would require not only a study of history, but must incorporate laws of human nature which are influenced by the state of society at any given time and which change over time and consequently become unpredictable over any considerable length of time. In addition the causal links between human nature and the empirical laws revealed by human history are too complicated for us to understand.

There are two kinds of empirical laws of society. The first called social statics deals with the coexisting uniformities that exist within society at any one time. Certain social phenomena will usually coexist with certain other social phenomena within a given state of society. By comparing one state of society with its co-existing social phenomena with other states of society with their social phenomena, it may be possible to reach certain laws of social statics. The second kind of empirical law, called social dynamics, deals with the succession of states of society; the change from one type of society to another. Social dynamics attempts to explain the sequence of states of society. Social dynamics may involve observing various trends in history, but the observation of trends does not tell us whether those trends will continue or not. In order to produce better empirical laws, it is necessary to combine social statics with social dynamics. This enables us to observe not only the changes in the different elements of society, but also the relation of one element with the other elements of society at a given time. This may allow us to produce a scientific law of the development of human society.

This study would be greatly assisted if there were one element in society that was the principal cause of social change. When that element changed then all the other elements would make a corresponding change, to create a particular order for change within society. There is such a social element; it is the state of the speculative faculties of humankind. This involves the knowledge and beliefs of humankind. Mill considered every considerable historical change in the material conditions of humankind was preceded by a change in the state of human knowledge. The progress of industry must follow and depend on the progress of knowledge. The beliefs of humankind will also determine the moral and political state of humankind. The order of progression in human society depends on the order of progression in the knowledge and beliefs of humankind. Certain truths cannot be discovered or inventions made, until certain others have been made first and certain social improvements can only follow others. The order of human progress may to a certain extent have definite laws assigned to it. However the rate of progress or whether progress takes place at all for a period is not something that can be made subject to any law. In the longer term progress must occur, as societies can be certain of eventually producing leaders and thinkers through whose efforts progress takes place.

Johnson & Earle

A number of modern theories of history, social change and cultural evolution place a considerable emphasis on population growth and technology. In the *Evolution of Human Societies: From Foraging Group to Agrarian State*[71] Allen Johnson and Timothy Earle propose an evolutionary process the driving force of which is a positive feedback between population growth and technological development. In their theory Johnson and Earle distinguish between the subsistence economy and the political economy. The subsistence economy is the household economy and is designed to meet human needs at the household level. It produces no surplus other than a security margin which is required for times of shortages. The political economy concerns the exchange of goods and services in an integrated society of interconnected families. All societies have a political economy but the process of social evolution makes the political economy larger and more complex. More sophisticated political economies seek to obtain a surplus from the subsistence economy to finance political, social and religious institutions and are controlled by elites. As the feedback between population and technology intensifies, problems arise due to the threat of overpopulation and the solution to the problem will normally involve the creation or improvement of the institutions of the political economy.

Increasing population means that the subsistence economy needs to be intensified to feed increasing numbers of people from the same resources. Intensification can involve four problems being production risk, warfare, technological needs and resource deficiencies. The solutions to these problems usually involve strengthening the powers of leaders and increasing the economic integration of communities.

Production risk is the risk that insufficient food may be produced for the expanding population. The problem may be solved by measures such as community food storage or agreements with other groups for reciprocal visiting and feasting in lean times. Such arrangements will support a larger population but require political leadership and support.

The problem of warfare arises as intensification makes certain territory more productive so that the benefit of seizing the territory increases relative to the cost of seizing the territory. This means warfare will become more common and the solution to this problem involves the formation of alliances with other groups and more effective defence. These measures however will require more effective political leadership and control.

Intensification may result in a problem of inefficient resource use which may be solved by the development of costly new technologies. The development of technologies such as irrigation systems may require considerable organization and could lead to greater political organization and control.

The problem of resource deficiencies caused by population growth can increase the need for goods not capable of being produced locally. These goods must be obtained by trade and may involve food imported to cover local production shortfalls or tools which can not be produced locally due to an absence of local raw materials. Such trade will help feed an increasing population on the same resource base. Trade however requires leaders empowered to make decisions on behalf of the local community which increases control over the local community. The various methods used to solve the problems of intensification all involve the surrender of political control by the community to leaders resulting in greater power for certain individuals and less freedom for the great majority.

Johnson and Earle's theory of social evolution was designed to explain the change from foraging groups to agrarian states. However they consider that the techno demographic engine they propose has also operated since the industrial revolution in the modern industrial world. A major difference between agrarian states and the societies that preceded them on one hand and the modern industrial world on the other hand is the much greater role both governments and the self-regulating free market plays in the modern industrial world.

The increase in population in the industrial world required an intensification of production just as in previous societies. The intensification process would involve the same problems of production risks, warfare, inefficient resource use and resource deficiencies as was involved with agrarian states and the societies that preceded them. However the increased role of governments and free markets in the industrial world would mean that these problems were solved somewhat differently in the industrial world.

Production risk is reduced by the ability of people to use bank savings and insurance to reduce risk and the rapid movement of commodities from seller to purchaser reduces loss caused by spoilage and allows food to reach people affected by natural disasters. However this comes at the loss of family and traditional security and when the market fails for one reason or another people become dependent upon the state which leaves them subject to state control.

Warfare is encouraged by the increasing value of the land and resources due to improved technology and population growth, making it more worthwhile to violently seize the land and resources. Free markets discourage warfare as trade increases the value of peace. However, sometimes warfare is used to forcibly bring communities within the free market system. The control of violence within a group or state allows greater intensification of production, but also allows elites to strengthen their political control of the group or state.

The problem of insufficient resource use can be solved by the use of substantial amounts of capital available in free markets. The accumulation of capital results in capital acquiring a sanctity which strengthens the power of the owners of capital. Large amounts of capital enable an ever increasing portion of the world's resources to be brought within the free market. The free market brings an intensification of production throughout the world creating economic integration and increasing stratified decision making over the world's production.

The problem of resource deficiencies is solved by the free market moving resources to wherever the demand for them is greatest. This enables the population to grow without being hindered by insufficient resources. However the market is managed by elites who use capital and political and military resources to protect their own interests.

It appears to be a basic rule of social evolution that the expansion of the political economy, while solving problems in the subsistence economy, involves opportunities for elites to increase their control over society. Increased intensification of production and integration of economic communities leads to increased stratification. Only political controls can restrict the power and wealth of elites and protect the environment from damage caused by free markets and population increase.

The evolution of human society has involved a loss of freedom. The problems caused by technological change and population growth can only be solved by creating a compromise between individual freedom and community-based political controls. The global economic integration taking place in the modern world is an example of the intensification, integration and stratification processes that have always occurred in social evolution. Intensification in the modern world takes place through the process of free markets and integration in the modern world is primarily in the form of increasing involvement in free markets. Stratification in the modern world means elites have great wealth enabling them to protect their interests by political means. The mechanism of the feedback between technology and population growth leading to intensification of resource use requiring increased stratification and political controls applies equally to modern societies as it does to the evolution from foraging to agrarian states.

Sanderson & Evolutionary Materialism

Stephen Sanderson in his book *Social Transformations: A general theory of historical development*[72] proposes a model for social evolution. Sanderson calls his model evolutionary materialism and he considers evolutionary materialism to be a theoretical strategy which is an "abstract set of assumptions, concepts and principles designed to serve as a broad theoretical guide to explaining empirical reality." It is an orientating device for creating and assessing theories rather than a theory itself.

Sanderson outlines a number of propositions which constitute the theoretical strategy of evolutionary materialism. The first set of propositions dealing with the nature of world history state that "world history reveals social transformations and directional trends of sufficient generality such that typologies of social forms can be fruitfully constructed. These directional sequences of change constitute the bulk of what is known as social evolution. Social evolutionists concentrate on general and repeatable patterns of social evolution … but also show due respect for the unique and nonrecurrent in world history." Social statis (continuity in the social patterns of a social system), devolution (retrogression to an earlier evolutionary stage) and extinction (the elimination of the basic patterns of a social system) are basic facts of world history, but do not undermine an evolutionary interpretation of world history. World history does not involve a predetermined pattern, but represents the aggregation of the actions of individuals and groups responding to biological, psychological and social needs. Social evolution is to be explained by use of the same causal explanations that are used in all the sciences.

Sanderson's second set of propositions concern the nature of world history. Social evolution occurs at all levels within social systems from societies to social classes to kinship groups. It is studied mainly at the macrosociological level but applies also at the simplest microsociological level. Social evolution often involves increasing social complexity or differentiation but also involves transformations that involve reduced complexity. There are some similarities and some differences between social evolution and biological evolution and the differences are enough for social evolution to be studied in its own terms and not along the lines on which biological evolution is studied.

Sanderson's third set of propositions deal with the principal causal factors in social evolution. Sanderson considers the principal causal factors involve the material conditions of human existence. These factors involve-

-Technology which involves all the knowledge, tools and techniques available to a society.
-Demography which involves variations in human populations and particularly the increasing pressure of population on limited resources.
-Ecology which involves all aspects of the natural environment, particularly those that interact with technology and demography.

-Economic factors which involve the forms of social organization within which goods and services are produced, distributed and exchanged, including the ownership of the means of production.

The causal factors apply in the long run and in the majority of cases but do not completely determine the course of social evolution. Non-material factors play a role in social evolution but in a quite secondary way. The material factors are important as they concern basic human needs for subsistence and the reproduction of human life. Human needs for subsistence and reproduction are a priority in human life and this leads to a casual priority in social evolution. Which material conditions or combination of conditions are casually important varies from one period to another and can only be identified by empirical study. There is no universal cause of social evolution and the driving engines of social evolution are different in different historical periods.

Sanderson's fourth set of propositions deals with adaption. Adaption is the process by which people originate social patterns which are devoted to meeting their needs and wants. It concerns the origin or persistence of social patterns. Adaption relates only to individuals and not to any social group larger than the individual. This is because only individuals can have needs and wants. Adaption can be in response to either or both of the physical or social environments. Sanderson considers that much of what social evolution concerns comes from adaptional processes.

Sanderson's fifth set of propositions concern the role of agency and structure in social evolution. He considers that human individuals acting in their own interests create social systems and structure. The systems and structures often develop in ways people never intended due to their actions having unintended consequences. The social systems and structure reflect back on individuals in that they create constraints within which human action takes place. Social evolution represents the effects of the interplay between human agency and social structure. Human agency does not occur freely in that human action is constrained by the biopsychological nature of human individuals and by the social structures that surround them.

Sanderson's sixth set of propositions concern the units of social evolution. He considers the units of social evolution to be social groups, structures and systems and not individuals. Individuals are the units of adaption but they do not evolve in social evolution. Social evolution can occur both due to forces within a society and as a result of forces external to a society.

Sanderson's seventh proposition concerns the pace of social evolution and he considers the pace of social evolution varies from one time to another. However he considers that social evolution was much slower in earlier periods and is faster in recent times.

Sanderson's last propositions concern the methods of studying social evolution. He considers the comparative method, which involves ordering synchronic data into typologies that are treated as reflecting historical transitions from one evolutionary stage to another, is an important tool of evolutionary analysis. The use of the comparative method is justified to the extent it could be independently corroborated by other data. Diachronic or historical and prehistoric data is to be preferred to synchronic data. Social evolutionary analysis involves the acquisition and synthesis of data from archaeological, historical, ethnographic and sociological sources. All of these contribute to the development of evolutionary theories.

Comment

All of these theories have their critics and some of the criticisms may be justified. Nevertheless there would seem to be some truth in the theories. But the theories of Marx, Durkheim, White, Johnson & Earle and Sanderson all have one failing in common; they fail to provide an ultimate explanation of social, cultural and historical change. Marx tells us that class warfare is the driving force of history, but class warfare is the means by which an alteration in the productive forces causes a change in the ownership relations of production. No mechanism, however, is offered for the changes in the productive forces.[73] Shaw in *Marx's Theory of History* suggests a technological determinist theory as an explanation for changes in the productive forces. Such an explanation is often considered controversial, but some sort of explanation is needed for the change in the productive forces. Karl Federn in *The Materialist Conception of History* suggests human intelligence could determine changes in the productive forces.[74] This idea is dismissed by Shaw on the grounds that human knowledge and productive intelligence is already built into the concept of productive forces.[75] However just because human intelligence is built into the concept of productive forces is not a reason for it being unable to be used as an explanation for the development of the productive forces. It would simply mean that productive forces are able to generate their own momentum, rather than relying on outside forces, but the question still remains how can they do this? In order to explain this, it is necessary to explicitly state that an element within the concept of productive forces, drives the productive forces forward and to explain what this element is and how it is able to produce change in the productive forces. Marx has failed to do this.

Shaw's technological determinism is also a suitable candidate to explain the change in the productive forces, but it just begs the question as to what causes the level of technology available to a mode of production? and what causes changes in the level of technology? It still does not provide an ultimate cause for historical change.

Durkheim’s theory has the same problem. His driving force for historical change is increasing social density, caused by population increases, improved transport and communications and the growth in cities. However we are not told what causes the population increases, growth in cities and improved transport and communications. Population increases in pre-industrial societies were always limited by the ability of the environment to support an increased population within that societies mode of production. While population will tend to increase, it is usually constrained by limited food supplies, disease, war or other factors. There tends to be a stable population level for a particular environment in a particular mode of production. Durkheim fails to tell us how population can increase in a particular mode of production or if the mode of production changes, as they obviously do, what causes the mode of production to change. He also fails to tell us what causes transport or communications to improve and what causes the growth of cities. Like Marx, Durkheim fails to give us an ultimate cause of historical change.

White is the same. He tells us that cultures evolve as the amount of energy harnessed per capita increases or as the efficiency of the technological means by which energy is put to work increases. What White does not tell us is what causes the amount of energy harnessed per capita to increase. Nor does he tell us what causes the efficiency of the technological means by which energy is put to work, to increase. One suspects White might suggest improved technology, but even this would just raise the question of what causes the technology to improve. White, just like Marx and Durkheim, has failed to provide us with an ultimate explanation of historical change.

Johnson and Earle consider a positive feedback between technology and population is the driving force of history. The problems associated with increasing population are solved by increasing the powers of leaders and elites. However the theory does not explain which of population and technology begins the process. If population increases first it is likely to be constrained by limited food supplies and disease and other factors. If technology improves that may allow population growth but no explanation is given for why and how technology improves. As with Marx, Durkheim and White no ultimate cause of historical change is provided.

Sanderson’s evolutionary materialism provides a theoretical strategy for social evolution rather than a theory. He does suggest certain causal factors as the driving force for social evolution being technology, demography, ecology and economic factors. Again, while appreciating that Sanderson is providing overall guidelines rather than a specific theory, it is hard to see how any of these factors could be an ultimate driving force for historical change. No suggestion is provided for how and why technology changes, how population growth can occur given limited food supplies unless improved technology allows increased population. If this is the case then there needs to be an explanation for the improved technology. Ecology will vary over time but by itself could not be the ultimate driving force for history as changes in ecology do not match changes in human history. Even if the ecology does not change, social evolution may well take place. Economic factors involving changes in the social organization by which goods and services are produced, distributed and exchanged are the results of changes in human social and cultural history. Again, no ultimate cause of human historical development is provided by Sanderson’s evolutionary materialism.

Some progress towards such an ultimate explanation is provided by Mill when he suggests that changes in the state of human knowledge have always preceded and caused changes in the material conditions of humankind. The progress of human society depends on the order of progression in the knowledge and beliefs of humankind. Increasing human knowledge could explain the change in Marx’s productive forces, Durkheim’s increase in population density, White's increasing energy consumption per capita, Johnson and Earle's and Sanderson's changes in technology and population.

However, Mill has left us with an unanswered question. He does not tell us what determines the state of human knowledge at any given time and what determines the order in which knowledge becomes available to us. The answer to this, as stated in the first part of this book is the nature of the environment, which we inhabit and the structure and properties of nature and their relationship to human beings. Human beings can only discover the facts concerning the properties and structure of nature in a particular order so we move through states of knowledge in a particular order. That order is determined by how close particular facts concerning nature are to us. We discover the closer facts before we discover the facts which are further away from us.

This however, is as far as we can push the questions back. What determines the structure and properties of the universe is a question that can not be answered scientifically. Such a question belongs to the realms of theology and metaphysics and we are not able to come up with definite answers to such a question.

Appendix 2

The Discovery of Agriculture

By Rochelle Forrester

The domestication of plants and animals has been a much discussed event in prehistory and anthropology. It has however been much troubled by a lack of any firm knowledge of how the process took place. Most attention has focused on trying to identify when, where and in what circumstances agriculture first emerged. Why agriculture emerged has usually been explained by its offering significant economic advantages to human populations over that which would be provided by the hunting and gathering lifestyle. This has been called into question by recent studies of modern hunter-gatherers which suggest hunter-gathering may be a better lifestyle than previously imagined. However if this is true and if it is possible to use studies of modern hunter-gatherers to assess the living conditions of hunter-gatherers before the agricultural revolution, then it is necessary to explain why humans took to agriculture and why they did it when they did.

A further point that needs to be explained is why most of humanity took to agriculture at the same time. Anatomically modern humans, *Homo sapiens sapiens* emerged around 200,000 years ago in Africa. For approximately 190,000 of those years they obtained their food by hunting and gathering. Then within a period of about 8,000 years the great majority of humanity were making their living by farming. Why such a long wait, followed by the spread of agriculture across a large part of the land inhabited by humans? Obviously diffusion of agricultural knowledge is an explanation for its rapid spread in this 8,000 year period, but it seems clear that agriculture was independently invented in a number of areas and most certainly in the new world.

Various explanations have been put forward for the development of agriculture. One involves plant mutations such as mutant maize, but such mutations would have been available many times before agriculture was developed, but were ignored. When agriculture developed, a wide variety of different crops were domesticated, and it is hardly likely that they all developed convenient mutations at approximately the same time without those mutations occurring many times previously. We need to explain why the human population took advantage of the mutations, if that was how agriculture developed, when they did and why they had previously ignored the mutations.

Another explanation is that the right conditions for agriculture developed due to climate change that preceded the development of agriculture. However this explanation has the problem that many different climates would have existed on earth during the 190,000 years that *Homo sapiens sapiens* inhabited the earth before the development of agriculture. Many of these climates would have been just as suitable for the development of agriculture as the climates in which agriculture eventually developed. Yet agriculture did not develop until some 10,000 years ago despite the presence of suitable climates for the development of agriculture in the preceding 190,000 years of *Homo sapiens sapiens* occupation of the planet.

An alternative explanation for the development of agriculture is that it was forced by population pressure. The problem with this view is that it does not explain how humans learnt to engage in agriculture and why there was population pressure at that particular time 10,000 years ago but not at other times in human prehistory. The human population through most of this time was able to expand into new lands, such as America, but the population in Africa, Europe and the Middle East were not able to expand into new lands in the way that the North East Asian population was able to expand. Local population pressures would have developed many times in prehistory but did not give rise to the development of agriculture. Before humans began to move into America and Australia, they had for several hundreds of thousands of years occupied all of Africa, Europe and Asia and despite population pressure, never developed agriculture. The population theory says that agriculture developed in the Middle East because humankind ran out of room to expand in South America, as though the people of the Middle East felt population pressure 10,000 years ago due to humankind running out of room to expand in South America. It is hardly likely the people of the Middle East would have felt population pressure due to events in South America. In the modern world, with its advanced transport and communications, some countries such as Japan are arguably over populated yet it does not have much effect on other countries. It would seem likely due to excessive migration into fertile areas, or due to once fertile areas becoming less fertile, there would be excessive population pressure on the land at many times in prehistory, but there is no evidence that this ever led to the development of agriculture.

Many of the proposed explanations for the development of agriculture have the common defect of not being able to explain why agriculture developed when it did, and not before, as the proposed explanations involve conditions which almost certainly existed many times before agriculture was actually developed. The only plausible explanations of the development of agriculture are those that are able to answer the question of why agriculture did not develop before 10,000 year ago.

One explanation that does not suffer from this problem is that suggested by L H Morgan (1877) and V Gordon Childe (1955) and others that agriculture developed as part of a natural process of cultural evolution when a certain level of knowledge and technology had developed. This view has been much criticized in the last twenty or thirty years due to research into modern hunter-gatherer societies. This research suggests that the knowledge that plants grow from seeds was available to hunter-gatherers in prehistory.[76] Evidence cited in support of this position is that modern hunter-gatherers understand agriculture and that hunter-gatherers must inevitably have a considerable knowledge of the plants and animals they live off. It is claimed there is no significant difference between the knowledge hunter-gatherers have of the plants and animals they needed for their survival and the knowledge of plants and animals required for agriculture and domestication.[77] There is a problem with this as obviously modern hunter-gatherers could and would have learnt plants are grown from seeds from 10,000 years of contact with agrarian societies.

The claim that modern hunter-gatherers having a knowledge of agriculture, shows that prehistoric hunter-gatherers knew about agriculture, is a logical error. It is literally a *non-sequitur* in that it does not follow that modern hunter-gatherers knowing of agriculture, means prehistoric hunter-gatherers knew how to engage in agriculture. This is because modern hunter-gatherers would have inevitably known of agriculture from thousands of years of contact with agrarian societies. This means there is no evidence at all for the belief that prehistoric hunter-gatherers knew how to practice agriculture.

More specifically the logical error is that of the fallacy of composition. The fallacy claims that because a part of the whole has a particular characteristic, then all parts of the whole have that characteristic. The claim is that as some hunter-gatherers (i.e. modern hunter-gatherers) know about agriculture, then all hunter-gatherers know about agriculture. This obviously does not follow as there is no reason to believe, just because some hunter-gatherers, know about agriculture, all will.

It is quite difficult to find hunter-gather groups that have had no contact with agrarian societies. Where there are such hunter-gatherers they do not seem to understand that plants grow from seeds. The Australian Aborigines were quite familiar with the seeds of various grasses, but they seemed to be unaware that the grasses and other plants grow from seeds.[78] An analogous situation between seeds and plants is between sex and giving birth. The Australian Aborigines believed a woman became pregnant when a spirit being enters her body and before contact with Indonesians and Europeans seemed to have little understanding of the relationship between sex and pregnancy.[79] They do not seem to be alone in this; the Trobriand Islanders studied by Malinowski seemed to be in the same position. If hunter-gatherers are unable to work out the relationship between sex and giving birth, both matters they were closely involved with; it seems unlikely they would understand the relationship between seeds and plants, things which while they have some familiarity with, they would not be as familiar with as they would be with sex and child-birth.

A similar situation exists with the belief from the time of the ancient Greeks to the mid-19th century in the spontaneous generation of life forms from non-living matter. Certain life forms such as maggots, bees, mice and others were considered to arise spontaneously from other matter such as hay or decaying plant or animal matter. Spontaneous generation was eventually only disproved by experiments by Pasteur and the development of powerful microscopes in the mid-19th century. If a literate society, well acquainted with the rules of logic, continued to believe in spontaneous generation some hundreds of years after the start of modern science, then it is very likely that prehistoric hunter-gatherers would have been unlikely to work out that plants come from seeds. The most probable and plausible belief for prehistoric hunter-gatherers as to the source of plants, given their knowledge at the time, was spontaneous generation from the earth. Alternatively, prehistoric hunter-gatherers may have believed plants come from the gods or some other supernatural cause.

A further point is that if it was true that the knowledge of agriculture and domestication was known to hunter-gatherers before agriculture and domestication became common, then one would expect to find some evidence of agriculture and domestication long before 10,000 years ago. It is hardly likely that the conditions (whatever they were) that lead to the development of agriculture some 10,000 years ago; never occurred in the previous 190,000 years *Homo sapiens sapiens* has been on this planet. One would expect to find evidence that where the conditions were right, agriculture was practiced and then if the conditions later turned against agriculture it would be abandoned. Such evidence exists with "lost cities" in America and Zimbabwe, but these cities were obviously built long after the discovery of agriculture some 10,000 years ago. It seems clear that agriculture only developed 10,000 years ago and then by both diffusion and independent invention was adopted by the great majority of human beings. This hardly supports the idea that the knowledge required for agriculture was widely known amongst hunter-gatherers prior to 10,000 years ago.

A further problem for the idea that early hunter-gatherers had knowledge sufficient for agriculture is that they could, to borrow a phrase from Thomas Kuhn and *The Structure of Scientific Revolutions*, be considered to be living in a different paradigm from the people who practice agriculture. Hunter-gatherers are interested in where the food is and how to get it. Farmers however are interested in how to make plants grow. They need to know about the planting of seeds, the creation of clearings, which plants grow best in which soils, the enrichment of soils and the watering of their gardens, the importance of removing weeds, conservation measures such as are involved with shifting agriculture and how plants can be improved by a process of selection. These sorts of measures, necessary for successful agriculture, will not be obvious to hunter-gathers. Plants growing wild, the only plants known to pre-agriculture hunter-gatherers, grow without being in specially cleared areas. Which soil plants grow best in is of no interest to hunter-gatherers; they are looking for plants not soil types. That plants grow better when the soil is enriched and weeds are removed would not be obvious to hunter-gatherers. That nutrients in the soil get exhausted after a few crops and it is necessary to plant additional crops at a new location, or to let the land lie fallow, would not be obvious to hunter-gatherers. Knowledge of these things could only be developed by trial and error, not by simple observation of wild plants. It could only be developed by the actual practice of agriculture. The most hunter-gatherers could learn simply by observation would be that plants require water and that plants grow well in areas cleared by fire. As much of what is needed for successful agriculture can only be learnt by trial and error and not by the observation of wild plants it seems that pre-agricultural hunter-gatherers could not have had the knowledge required for agriculture. Present day hunter-gatherers may well have that knowledge but it is obvious they could and would have learnt that knowledge from contact with agrarian peoples over thousands of years. That of course is one of the ways the diffusion of agriculture occurred. However, when the environment of the hunter-gatherers, who had learnt of agriculture, was unsuitable for agriculture, then the hunter-gatherer lifestyle continued.

A final problem for the idea that hunter-gatherers in prehistoric times knew plants grow from seeds is that this idea is far from obvious. Seeds look quite unlike plants, so there is no reason to believe they will eventually grow into plants. There is also a significant time period before seeds turn into plants so that it is not obvious the seeds will become plants. Finally, in many cases seeds will not grow into plants, due to factors such as poor soils, a lack of water or too many weeds.

A similar situation exists with the domestication of animals. The earliest domestic animals are believed to be dogs which were domesticated in South-West Asia 12-14,000 years ago. Dogs would be useful assets to hunter-gatherers being capable of acting both as guard dogs and also as playing a role in hunting as they do today, for example when hunting pigs. Yet they were only domesticated after 190,000 years of *Homo sapiens sapiens* existence. It seems likely the domestication of animals took so long because for a long period of modern human existence they were simply unaware of the usefulness of dogs and other domestic animals and of how to domesticate them. If prehistoric hunter-gatherers did know how to domesticate dogs surely they would have done so.

The view that agriculture was adopted because it offered economic advantages in comparison with hunter-gathering has been questioned recently. Studies of modern hunter-gatherers have suggested they obtain ample calories and protein and consume a wide variety of food. Their lifestyles are usually preferred to those of farmers and they obtain their food supplies with less labor than is required of farmers. Many studies suggest the hunter-gatherer lifestyle is simply overall superior to that of farmers.[80]

There are, however, problems with these studies. There are a limited number of them; labor costs are measured in a variety of ways; how does one compare the costs and benefits of sedentism? how does one assess the fact that farmers normally produce a surplus and the costs of storage? Cohen suggests there is probably no method of fairly comparing agriculture with hunter-gathering.[81] It has been suggested by Hill and Hurtado[82] that the results of studies of modern hunter-gatherers are so variable that no group could be considered to be typical and could be used as an analogue for studying our ancestors. Considerable attention has been directed towards the !Kung San who seem to be an unusually prosperous group of hunter-gatherers.

There is however a much greater problem. It is quite uncertain as to whether studies of modern hunter-gatherers gives any real indication of what life was like for prehistoric hunter-gatherers. Some suggest that as hunter-gatherers only occupy marginal environments in recent times, while before the development of agriculture they would have occupied better lands, they would have been better off in earlier times. However, whether a group is prosperous or not depends not just on the fruitfulness of the land but also on the size of the population on that land. Poor quality land may support a small population in some affluence while a larger population on better land may not live very well at all. The prosperous !Kung San actually live in a desert but live well presumably due to a low population density on the land.

It is also suggested that the presence of agricultural people would interfere with the ability of hunter-gatherers to move at will and so reduce their economic opportunities and their standard of living. However it is not at all clear that before agriculture hunter-gatherers were able to move at will. Hunter-gatherers tend to have territories and to wander into another band's territory could produce conflict. So it is not necessarily the case that hunter-gatherers in prehistory could wander at will, so whether their choices of movement were any more restricted after the development of agriculture than before, is somewhat doubtful.

A more significant matter is that modern hunter-gatherers have a number of benefits not available to prehistoric hunter-gatherers. The first is that modern hunter-gatherers have access to goods and tools that prehistoric hunter-gatherers did not have, due to trade with modern agrarian and industrial societies. Most modern hunter-gatherers have access to iron, making hunting, digging for food and cutting down trees considerably easier. Other products such as pottery, rope and modern medicines might well make the lives of modern hunter-gatherers more comfortable than their prehistoric counterparts. Some modern hunter-gatherers actually hunt with shotguns. One effect of this is that it is likely to give modern hunter-gatherers the edge when it comes to confronting large carnivores. Prehistoric hunter-gatherers armed with flint, bone or ivory tipped spears or arrows may not necessarily have been the top predator in environments containing lions, tigers, leopards, bears, wolves and other fast and well equipped predators. Bears were hunted by the Tlinguit Indians of the north- west coast of America and men were sometimes killed in these hunts. Nowadays the Tlingit use powerful steel traps when hunting bears.[83] Snake bites and attacks by jaguars represent a significant proportion of deaths among the Ache in eastern Paraguay.[84] A further benefit modern hunter-gatherers have over their predecessors is that of a higher authority to control and keep order between them. In the event of a dispute between two hunter-gatherer bands there is a much more powerful authority, the government of whatever state the hunter-gatherers live in which will usually prevent them from slaughtering each other. There is no such authority to enforce law and order for prehistoric hunter-gatherers. Disputes may end up being settled by force to the benefit of the strongest or most numerous. Hill & Hurtado note that among the Ache warfare and accidents account for 73% of adult deaths. The equivalent figures for the Hiwi are 39% and for the !Kung-San 11%.

Yet a further advantage modern hunter-gatherers have over their prehistoric counterparts is that modern hunter-gatherers may well receive support in bad years from modern governments. Climates vary and most areas will occasionally suffer from drought which will cause the destruction of the plants and animals hunter-gatherers live on. For prehistoric hunter-gatherers this would mean famine unless they were able to move towards more fertile areas. This would not be easy if the drought covered a large area and because prehistoric hunter-gatherers would not necessarily know where the better areas are. Migration to other areas may well involve conflict with other hunter-gatherer bands. Modern hunter-gatherers may well be protected from such disasters but such protection was not available to their prehistoric counterparts. The true test of how people live is not their average or good years but how well they survive in their bad years, as there is little value in having a number of good or average years if they are followed by a single bad year that causes half the band to die of starvation. In these circumstances it seems hardly likely that studies of modern hunter-gatherers will give much idea as to how prehistoric hunter-gatherers lived.

The most convincing explanation of the development of agriculture is that by Robert and Linda Braidwood. They emphasize cultural rather than environmental, plant mutation or population explanations for the development of agriculture. All those explanations have the problem that they cannot explain why agriculture suddenly developed when it did after such a long period of hunter-gathering. The Braidwoods argue that it was improvements in human technology and human knowledge of the environment over time, that lead to the development of agriculture.[85] It is of course impossible to trace the growth in human knowledge in prehistoric people but improvements in human technology are to some extent traceable.

This can be shown in a number of ways. Brian Fagan shows how over time stone tool makers learnt how to make better and better use of a pound of flint to produce successively greater volumes of cutting edge as shown in the case study of Stone Tools.[86] A similar process can be seen in technological changes that occurred after about 30,000 bp. These included improved techniques for the working of raw materials. Before this time technology largely involved the use of only four techniques, those of percussion, whittling, scraping and cutting all of which required only a limited range of hand motion. After 30,000 bp, new techniques were added including pressure flaking, drilling, twisting grinding and others, which involved different motor abilities than those used previously. Secondly, in the earlier period the main raw materials used were stone, wood and skin. Later on bone, ivory and antler and less importantly shell and clay were added to the original materials. Thirdly, the number of components in composite tools expanded considerably after 30,000 bp, increasing the complexity of the tools used. Fourthly, the number of stages involved in manufacturing artifacts significantly increased after 30,000 bp. Before 30,000 bp manufacturing involved only a short series of single stage operations, while later there were often several stages of manufacture to produce the final product. The number of processes and techniques had increased as had the degree of conceptualization required to manufacture the product.[87]

In the period between the Middle and Upper Palaeolithic there were substantial improvements in the artifacts available to people. Hunting equipment improved by the use of narrow bone or ivory points for spears which had greater penetrating power than earlier flint tipped spears. Spear-throwers and the bow and arrow were also introduced, allowing prey to be killed from a greater distance. Cooking was made more effective through the use of cobble-lined hearths which allowed heat to be retained longer and at a more even temperature. Improvements in clothing seem to have been made between the Middle and Upper Palaeolithic, providing humans with much better protection against the elements. Eyed needles seem to have been invented around this time. Housing became more sophisticated in the Upper Paleolithic with many structures being made of mammoth bones suggesting that some sort of sophisticated transport device such as sledges were used to move the bones. Art, which played little role in the Middle Paleolithic, became much more extensive in the Upper Paleolithic. Cave paintings appeared in Europe, Australia and North and South Africa. Many artifacts such as bone needles, ivory beads, spear throwers and bows had engravings or carvings performed on them. Artistic objects such as Venus figurines were traded over considerable distances suggesting the Upper Paleolithic had much improved trade and communications than the Middle Paleolithic.[88] Technology developed by hunter-gatherers in the Middle East, to utilize wild cereals, such as stone sickles and underground storage pits were useful to early cereal farmers in the Middle East.

The substantial improvements in the tools, clothing, art and general culture of humankind between the Lower and Upper Paleolithic could only have taken place with a gradually increasing knowledge of how to make better and better use of the materials in the environment. It seems likely that the increased knowledge of the human environment shown by archaeological finds of tools, art and other paleolithic objects would have been matched by a gradually increasing knowledge of the plants and animals humans live off. Hunter-gatherers are known to have a very great knowledge of the plants and animals in their immediate environment, but that does not mean they always had such knowledge. In particular knowledge not directly related to the hunter-gatherers survival, such as how to make plants grow and how to tame animals would not necessarily be immediately known to hunter-gatherers and might only be learnt after a long period of gradually increasing knowledge. As noted earlier, such knowledge was irrelevant to the hunter-gatherer lifestyle, and so may have taken some time to become part of the culture of humankind.

There is very little in the way of hard facts known about the domestication of plants and animals. Most theories as to how this came about contain a fair amount of guesswork. Nevertheless the best theory would seem to be that the knowledge required for the domestication of plants and animals gradually increased over time until enough was acquired to allow the domestications to take place. Theories involving climate change, fortuitous mutations and population pressure causing the domestications all have the problem that such factors could have occurred many times before the agricultural and pastoral revolutions without agriculture and pastoralism being introduced. This strongly suggests that before the agricultural and pastoral revolutions human beings simply did not know how to successfully grow plants and how to domesticate animals.

Appendix 3

Guttman Scale Analysis

By Rochelle Forrester

Guttman scale analysis is a method used in the social sciences, of dealing with binary information, that is information with a yes or no answer, where that information can be assembled in a particular order. An example of information that can be assembled in a particular order might be I can tolerate cats, I like cats, I would like to own a cat. Agreement with the last item implies agreement with the earlier items. A further example would be I know what numbers are, I can add numbers and I can do quadrilateral equations. Anyone who can do quadrilateral equations, must be able to add numbers and will know what numbers are. Equally anyone who knows how to add numbers must know what numbers are. This situation, where person A can say yes to the third proposition will also be able to say yes to the first and second propositions and person B who says yes to the second proposition will be able to say yes to the first proposition, while person C can only answer yes to the first proposition, can be arranged into a table. Such a table with plus signs representing a yes and minus signs representing a no could be as below.

Table A

adding numbers	+	+	-
understanding quadrilateral equations	+	-	-
understanding numbers	+	+	+
	Person A	Person B	Person C

This table shows no particular pattern but can be arranged without changing the data by putting the most common proposition, understanding numbers in the bottom row, with the next most common proposition, adding numbers as the next row and the least common proposition understanding quadrilateral equations as the top row. The people can also be arranged with the least knowledgeable being listed first and with the one with the greatest knowledge last.

Table B

understanding quadrilateral equations	-	-	+
adding numbers	-	+	+
understanding numbers	+	+	+
	Person C	Person B	Person A

This will produce a table like that above known as a scalogram. The scalogram has a stair step look known as a perfect scale with the number of pluses increasing as one moves from left to right through the people. The scaling effect is not caused by manipulating the data, it must be present within the data for it to appear. If one for example simply tossed a coin (heads for pluses and tails for minuses) and inserted the results from the coin tosses into a table, no matter how much the table was rearranged you would not get a stair step profile as is obtained above. This can be seen from the attached table with 3 sets of 6 coin tosses.

Table C

1	+	-	+
2	+	+	-
3	-	+	+
4	-	-	+
5	+	-	-
6	-	+	+
	First set of tosses	Second set of tosses	Third set of tosses

Obviously there is no stair step profile here. Nor can a stair step profile be produced by moving the first, second or third set of tosses around or by changing the order in which the tosses are recorded in the left hand column. The actual results of the coin tosses cannot be changed as they represent the real data produced by tossing the coin and are equivalent to the data of the mathematical knowledge being shown in Tables A and B.

Why do we get the regular stair step result for the pluses in Table B when we could not get such a result from a random process, such as coin tosses? The stair step profile is caused by the data itself which is not random but which involves a process of accumulation. A person who knows how to do quadrilateral equations must also know how to add numbers and what numbers are. A person who knows how to add numbers must also know what numbers are. The person who knows what numbers are will not necessarily know how to add them or how to do quadrilateral equations. The different levels of knowledge is reflected in the number of pluses in the table which can be arranged in the stair step scalogram pattern. The process of accumulation is not present in the data produced by the coin tosses. Each coin toss is a separate act unrelated to the other coin tosses.

Guttman scale analysis has been used by anthropologist Robert Carneiro to show both the complexity or degree of evolution of a society and the sequences by which societies develop certain traits. The type of traits Carneiro investigated were the development of stone tools, copper, bronze and iron metallurgy, the use of pottery, the domestication of plants and animals, the development of writing and numerous other traits. In his article *Scale Analysis, Evolutionary Sequences and the Rating of Cultures* Carneiro actually deals with as many as 618 cultural traits all involving pre-industrial societies.

Guttmam scale analysis involves listing the societies chosen for the analysis along the bottom of a sheet of graph paper and the cultural traits along the side of the graph paper. As many or as few societies or traits as desired may be used. Societies and traits can be listed in an arbitrary order. If a particular trait is present in a society it is indicated by a plus sign (+) on the graph paper and if it is absent it is indicated with a minus sign (-).

A simple such analysis will produce a table such as is shown below.

Table D

settlements of 100 +	+	-	+	+
food surplus	+	-	+	+
social stratification	+	-	+	-

iron tools	+	-	-	-
	Romans	Tasmanians	Inca	Iroquois

Such a table can be rearranged with the most common traits being listed at the bottom of the traits and the least common at the top. The societies can also be rearranged with the societies with the fewest traits being listed first and the one with the greatest number listed last. This will produce a scalogram like that below.

Table E

iron tools	-	-	-	+
social stratification	-	-	+	+
food surplus	-	+	+	+
settlements of 100+	-	+	+	+
	Tasmanians	Iroquois	Inca	Romans

There is something about the societies and traits which gives this particular stair step pattern. The pattern is derived due to the order in which the societies have derived the cultural traits. The traits in the lower part of table E were derived earlier than those on the top part of the table and due to this more societies have those traits. If one examines the traits from bottom to top that is the approximate order in which the traits are accumulated in societies. Settlements of 100+ and food surpluses (both with 3 pluses) began about 10,000 years ago, social stratification began soon after food surpluses and iron tools developed last (around 1200 BCE with the invention of iron smelting by the Hittites). We arrange the table to have the most common traits at the bottom and the least common at the top and this coincides with the actual historical order societies acquired the traits. This must be because the traits occurred in the order in which they are shown in the above table. Any sample of societies and traits that have actually existed should show the stair step profile regardless of whether the samples are selected or are random. The only restrictions on this are that the traits selected should be retained in the societies over the long term and should arise in approximately the same order in different societies in which they exist. If they arose in different orders, scaling would fail and the fact that scaling is usually successful indicates that the traits tend to arise in various societies in very similar order.

If scaling occurs then certain things can be said about the societies and traits involved. Societies in the right hand column have all the traits that other societies have and some additional ones as well. If a particular trait is present in a society, we can predict that it will have certain other traits. A society with iron tools can also be predicted as having social stratification, food surpluses and settlements of over 100 people. If a trait is absent from a society, then we can predict other traits will also be absent. If a society does not have a food surplus, it will not have social stratification or iron tools. If we know a society's highest and lowest traits we can predict whether a society has any other trait. The highest trait for the Inca is “social stratification” and the lowest is “settlements of 100+”. This means we can also say the Inca will also have food surpluses but not iron tools. If we know the number of traits a society has we can say what they will be. If a society has just two traits they will be “settlements of more than 100 people” and “food surpluses”.

The reason why scaling works in cultural evolution is because traits will usually accumulate within a society over time leading to societies developing greater complexity over time. If traits accumulate over time, then over time societies acquire more traits. Over the same period of time, society A may develop 1 trait, society B 3 traits, society C 5 traits and society D 6 traits represented by 1,3,5 and 6 pluses as shown on the table below.

Table F

6	-	-	-	+
5	-	-	+	+
4	-	-	+	+
3	-	+	+	+
2	-	+	+	+
1	+	+	+	+
	Society A	Society B	Society C	Society D

When societies develop a trait it is usually not lost so societies with trait 6 will normally have the previous 5 traits as well as trait 6. If societies did not retain traits then a random pattern would be produced much like with coin tosses which can not be arranged in a stair step pattern.

Perfect scaling is rarely achieved and various means have been developed to measure the degree of scaling. The most common such method is known as the coefficient of reproducibility. The coefficient of reproducibility measures the degree to which we can predict which items a society will have if we know the number of traits it has. To measure the coefficient of reproducibility, we total the number of traits whose presence or absence would have been wrongly predicted from each society's scaling. The total number of these errors is divided by the product of the total number of traits and societies in the scalogram. This will produce a decimal fraction which when subtracted from one gives the coefficient of reproducibility.

The formula for the coefficient of reproducibility is

$$1 - \frac{\text{number of errors}}{\text{traits x societies}}$$

The product of traits x societies is simply the number of results from the yes or no question as to whether a society has a particular trait or not. It is the number of pluses and minuses contained in the table and the coefficient of reproducibility involves a comparison of the number of errors against the total number of pluses and minuses in the table.

An example of the calculation of the coefficient of reproducibility can be seen from the table below.

Table G

6	-	-	-	-	-	+
5	-	-	-	-	+	+
4	-	-	-	-	+	+
3	-	-	-	+	-	+
2	-	-	+	-	+	+
1	-	+	+	+	+	+
	A	B	C	D	E	F

The number of errors for society D is 2 as it does not have trait 2 and it has trait 3 when it only has a total of 2 traits. Society E also has 2 errors as it does not have trait 3 and it has trait 5 when it only has 4 traits. This gives 4 errors in total, which becomes the numerator while the denominator is the product of traits x societies as shown below.

$$1 - \frac{4}{6 \times 6}$$

Perfect scaling produces a coefficient of 1.00 while no scaling at all produces a coefficient of 0. Depending on what societies and traits are used, scaling seems to be typically above .90 while if traits arose in random order in societies the scaling would be 0. It is not necessary for scaling to be 1.00 to indicate there is something in the data that needs explaining; anything above 0 indicates a pattern for which there must be some sort of causal factor. The causal factor for scaling above 0 is that societies do actually acquire traits in a similar order. The reasons societies acquire traits in a similar order is because they have similar problems and similar resources at their disposal to solve those problems. They discover how to develop and use those resources in a similar order of discovery.

There may be a number of reasons why perfect scaling with a coefficient of 1.00 does not always occur. Societies like the Aztecs and Maya of Central America did not have large domesticable animals available to them so they could never develop traits such as plough agriculture or wheeled transport. Societies in areas with no copper, tin or iron deposits could not develop copper, bronze or iron metallurgy. Agriculture was never going to be developed by the Inuit, Laplanders or by desert dwellers. Lack of large domesticable animals, plough agriculture, wheeled transport and metallurgy will certainly ensure that the Central American civilizations will develop traits in a different order from Old World civilizations.

A further reason for traits being developed in different orders in different societies concerns the diffusion of traits. Traits will spread from one society to another so the order in which they are acquired may vary greatly depending upon whether or not diffusion takes place. A society open to the diffusion of traits such as Japan after the Meiji Restoration or Russia from the time of Peter the Great will acquire traits in a different order from societies that are resistant to diffusion such as Ottoman Turkey and Tokugawa Japan. Some societies, such as some Islamic societies, may be open to receiving some traits, for example those involving technology, but may be reluctant to accept other traits such as those involving political systems or social organization.

How traits are expressed will also affect scaling. The trait "tool use" scales effectively while a trait of "use of stone tools" does not scale as it is lost when stone tools are superseded by more efficient metal tools. Absolute monarchy will only scale to such a time as when it has not been superseded by democratic institutions. Where traits are superseded rather than accumulated, they will not scale well. However, the new traits can appear on the scalogram in replacement of the superseded traits.

The accuracy of the description of traits can have an effect on the accuracy of scaling. It may be somewhat uncertain as to whether a particular society has a trait or not. Do modern Scandinavian societies have religion? A few people in those societies do, but most do not. Does one say a society has religion when 1 person does and millions do not? Some societies will be in a state of transition from not having a trait to acquiring the trait or from having the trait to losing the trait. The same problem can arise from what is a society? Does a society have agriculture when it imports all its food. No society is truly separate from other societies, yet we treat them as separate societies when doing scale analysis.

If factors such as lack of particular resources, diffusion, how traits are expressed and the accuracy of trait description are taken into account then it may be possible to produce perfect scaling with a coefficient of 1.00. The exclusion of traits that do not scale well could provide valuable information as to what extent human social and cultural developments are necessarily unilateral and to what extent it is multilateral. Traits which do scale effectively would indicate unilateral development as they are acquired in the same order in many or all societies, while those that do not scale well are acquired in different sequences in different societies indicating multilateralism.

Bibliography

General

Alexander, William & Street, Arthur (1962) *Metals in the Service of Man*, Pelican Books, Harmondsworth, England

Balazs, Etienne (1964) *Chinese Civilisation and Bureaucracy*, Yale University Press: New Haven and London

Basalla, George (1988) *The Evolution of Technology*, Cambridge University Press, Cambridge

Braudel Fernand (1967) *Capitalism and Material Life* 1400-1800, Fontana, London

Cardwell, Donald (1994) *The Fontana History of Technology*, Fontana Press, London

Crump, Thomas (2001) *A Brief History of Science*, Robinson, London

Daumas, Maurice (ed) (1964) *A History of Technology and Invention*, Presses Universitaires de France

Dennell, Robin (1983) *European Economic Prehistory*, Academic Press, London

Diamond, Jared (1998) *Guns, Germs and Steel*, Vintage, London

Donagen, Alan & Barbara (1965) *Philosophy of History* New York

Durkheim, Emile (1984) *The Division of Labour in Society* Basingstoke

Fagan, Brian (1995) *People of the Earth: An Introduction to World Prehistory*, Harper Collins, New York

Federn, Karl (1939) *The Materialist Conception of History* London

Galtung, Johan & Inayatullah, Sohail (ed) (1977) *Macrohistory and Macrohistorians* Westport, Conn

Gardiner, Patrick (ed) (1959) *Theories of History* Glencoe

Gribben, John (2002) *Science: A History, 1543-2001*, Penguin Books, London

Grunger, Rolf (1985) *Philosophies of History* Aldershot, Hants

Harris M (1987) Anthropology and the Study of Culture in *Cultural Anthropology*, New York

Johnson, Allen & Earle, Timothy (2000) *Evolution of Human Societies: From Foraging Group to Agrarian State* Stanford University Press, Stanford

Lenski, G & J (1970) *Human Societies*, McGraw-Hill Book Co, New York

Lilley, S Technological Progress and the Industrial Revolution 1700-1914 in *The Fontana Economic History Of Europe* ed C Cipolla, (1973)

Mertons, R (1973) *The Sociology of Science*, Chicago

Meyer, Jerome (1956) *World Book of Great Inventions*, The World Publishing Company, New York

Mill, J S (1872) *A System of Logic* London

Mokyr, Joel (1990) *The Lever of Riches: technological creativity and economic progress*, New York
Ogburn, W F (1922) *Social Change*, New York
Ogburn, W F & Thomas, D S, Are Inventions Inevitable in *Political Science Quarterly* 37 (March 1922)
Popper, Karl (1957) *The Poverty of Historicism*, London
Price T. D. & Gebauer (1995) *Last Hunters-First Farmers*, Santa Fe, New Mexico
Rambo, A Terry, The Study of Cultural Evolution in *Profiles of Cultural Evolution* (1991) Ann Arbour, Michigan
Renfrew, Colin (1973) *Before Civilization*, Penguin Books, Harmondsworth, England
Sanderson, Stephen (1999) *Social Transformations: A general theory of historical development* Rowman & Littlefield Publishers, Lanham
Shaw, W F. (1978) *Marx's Theory of History* Stanford, California
Silver, Brian L (1998) *The Ascent of Science*, Oxford University Press, Oxford
Taylor, Gordon Rattray (1983) *The Inventions that Changed the World*, Reader's Digest, London
Usher, Abbott Payson (1954) *A History of Mechanical Inventions*, Harvard University Press, Harvard
White, Leslie (1970) *The Science of Culture*, Toronto
Williams, Trevor (1982) *A Short History of Twentieth Century Technology*, Clarendon Press, Oxford
Zohary, Daniel and Hopf, Maria (2000) *Domestication of Plants in the Old World*, Oxford University Press, Oxford

Discovery of Steam Power

Burstall, Aubrey F. (1963) *A History of Mechanical Engineering,* Faber & Faber: London Buxton,
Dickinson, H W (1963) *A Short History of the Steam Engine*, Cass: London
Lewis, Brian (1971) *Coal Mining in the Eighteenth and Nineteenth Centuries*, Longman Group Ltd: London
Lord, John (1923) *Capital and Steam Power*, P S King & Son Ltd: London
Neil (1978) *The Economic Development of the British Coal Industry*, Batsford Academic: London
Roll, Sir Eric (1930) *An Early Experiment in Industrial Organisation*, Augustus M Kelly Publishers: New York
Rolt, L T C (1963) *Thomas Newcomen*, David and Charles: Dawlish, London
Thirring, Hans (1958) *Energy for Man*, Indiana University Press: Bloomington
Tunzelmann, G N (1978) *Steam Power and British Industrialisation to 1860*, Clarendon Press: Oxford

Watkins, George (1978) *The Steam Engine in Industry Vols 1 & 2*, Moorland Publishing: Derbyshire, England

History of Chemistry

Hudson, J (1992) *The History of Chemistry*: Chapman & Hall New York
Findlay, A (1965) *A Hundred Years of Chemistry* (3rd ed): Methuen, London
Ihde, A J (1964) *The Development of Modern Chemistry*: Dover, New York
Leicester, H M (1956) *The Historical Background of Chemistry*: Dover, New York
Partington, J R (1961-70) *A History of Chemistry*: Macmillan, London

History of Writing

Coulmas Florian (1989) *The Writing Systems of the World*, Basil Blackwell, Oxford
Fischer, Steven Roger (2001) *A History of Writing*, Reaktion Books, London
Sampson, Geoffrey (1985) *Writing Systems*, Hutchison, London
Senner, Wayne (ed) (1989) *The Origins of Writing*, University of Nebraska Press, Lincoln

History of Medicine:

Coleman, Vernon (1985) *The Story of Medicine*, Robert Hale, London
Friedman, Meyer & Friedland, Gerald (1998) *Medicine's 10 Greatest Discoveries*, Yale University Press, New Haven & London
Glasser, Ronald (1997) *The Light in the Skull*, Faber & Faber, Boston
Gordon, Richard (1993) *The Alarming History of Medicine*, Sinclair-Stevenson Ltd, London
Gutherie, Douglas (1958) *A History of Medicine*, Thomas Nelson & Sons Ltd, London
Hicks, Angela (1996) *Principles of Chinese Medicine*, Thorsons, London
Porter, Roy (1999) *The Greatest Benefit of Mankind*, Fontana,
Reid, Daniel (1996) *The Shambhala Guide to Traditional Chinese Medicine*, Shambhala, Boston
Rhodes, Philip (1985) *An Outline History of Medicine*, Butterworths, London
Rubin, Stanley (1974) *Medieval English Medicine*, David & Charles, London
Shortwer, Edward (1987) *The Health Century*, Doubleday, New York
Singer, Charles & Underwood, E Ashworth (1962) *A Short History of Medicine*, Clarendon Press, Oxford
Siraisi, Nancy G (1990) *Medieval & Early Renaissance Medicine*, The University of Chicago Press, Chicago
Strathern, Paul (2005) *A Brief History of Medicine*, Constable and Robinson Ltd, London
Venzmer, Gerhard (1968) *5000 Years of Medicine*, MacDonald & Co (Publishers) Ltd
Williams, Guy (1975) *The Age of Agony*, Constable, London

Williams, Tom (2002) *Chinese Medicine*, Vega, London
Wootton, David (2006) *Bad Medicine*, Oxford University Press, Oxford

The Discovery of the Atomic World and the Constituents of Matter

Davies, P C W (1979) *The Forces of Nature*, Cambridge University Press, Cambridge
Kane, Gordon (1995) *The Particle Garden: Our Universe as Understood by Particle Physicists*, Addison-Wesley Publishing Co, Reading, Massachusetts
McKay, Alwyn (1984) *The Making of the Atomic Age*, Oxford University Press, Oxford, New York
Oerter, Robert (2006) *The Theory of Almost Everything*, Pi Press, New York
Squires, Euan (1985) *To Acknowledge the Wonder*, Adam Hilger Ltd, Bristol and Boston
Trefil, James (1994) *From Atoms to Quarks, An Introduction to the Strange World of Particle Physics*, Doubleday, New York
Weinberg, Steven (2003) *The Discovery of Subatomic Particles*, Cambridge University Press, Cambridge

The Discovery of Agriculture

Abbie, A A (1970) *The Original Australians*, A H & A W Reed: Sydney
Cohen, M. N. (1977) *The Food Crisis in Prehistory*, Yale University Press; New Haven & London
Dennell, R (1983) *European Economic Prehistory*, Academic Press: London
Elkin, A P (1974) *The Australian Aborigines*, Angus & Robertson Publishers: London
Fagan, B (1995) *People of the Earth: An Introduction to World Prehistory*, Harper Collins, New York 111-115
Megaw, J. V. S. (ed) (1977) *Hunters, gatherers and first farmer beyond Europe*, Leicester University Press: Leicester
Oberg, K (1973) *The Social Economy of the Tlingit Indians*, University of Washington Press
Price, T. D. & Gebauer, A. B. (ed) (1995) *Last Hunters-First Farmers*, School of American Research Press: Santa Fe, New Mexico
Rindos, D (1984) *The Origins of Agriculture*, Academic Press: Orlando
Sahlins, M (1972) *Stone Age Economics*, Aldine Atherton: Chicago

Guttman Scale Analysis

Carneiro, Robert (1968) Ascertaining, Testing and Interpreting Sequences of Cultural Development, *Southwestern Journal of Anthropology* 24:354-374

Carneiro, Robert (1969) The Measurement of Cultural Development in the Ancient Near East and in Anglo-Saxon England, *Transactions of the New York Academy of Sciences* 31: 1013-1023
Carneiro, Robert (1962) Scale Analysis as an Instrument for the Study of Cultural Evolution, *Southwestern Journal of Anthropology* 18:149-169
Carneiro, Robert (1970) Scale Analysis, Evolution Sequences, and the Rating of Cultures in *A Handbook of Method in Cultural Anthropology*, ed Raoul Naroll and Ronald Cohen, 834-871
Peregrine, Peter N, Ember, Carol R and Ember, Melvin (2004) Universal Patterns in Cultural Evolution: An Empirical Analysis Using Guttman Scaling, *American Anthropologist* 106(1):145-149

End Notes

Notes for Part I

[1] Diamond, J (1998) *Guns, Germs & Steel*, London, 408
[2] Rambo, A Terry (1991) The Study of Cultural Evolution in *Profiles of Cultural Evolution*, Ann Arbour, Michigan, 27-28.
[3] G & J Lenski, (1970) *Human Societies* New York, 26-31.
[4] George Basalla (1988) *The Evolution of Technology*, Cambridge, 212
[5] S Lilley, (1973) "Technological Progress and the Industrial Revolution 1700-1914" in *The Fontana Economic History of Europe* (ed) C Cipolla 192-195
[6] Diamond, J (1998) *Guns, Germs and Steel,* London, 362-363
[7] Ogburn, W. F. & Thomas, D. S. (March 1922) "Are inventions inevitable?" *Political Science Quarterly* 37 83-98; Ogburn, W. F. (1922), *Social Change*, New York 90-122.
[8] Price, T. D & Gebauer, A. B. (1995) *Last Hunters-First Farmers* Santa Fe, New Mexico, 3
[9] Mertons, R (1973) *The Sociology of Science* Chicago, 364.
[10] Ibid., 371.
[11] Universal Patterns in Cultural Evolution: An Empirical Analysis using Guttman Scaling, *American Anthropologist* 106(1):145-149
[12] Carneiro, Robert (1968) Ascertaining, Testing and Interpreting Sequences of Cultural Development in *Southwestern Journal of Anthropology* 24:354-374

[13] Caineiro, Robert Scale Analysis, Evolutionary Sequences and the Rating of Cultural in *A Handbook of Method in Cultural Anthropology* (1970) ed by Raoul Naroll and Ronald Cohen, Columbia University Press, New York and London p 834-871 on pages 839-841
[14] Caineiro, Robert Scale Analysis, Evolutionary Sequences and the Rating of Cultural in *A Handbook of Method in Cultural Anthropology* (1970) ed by Raoul Naroll and Ronald Cohen, Columbia University Press, New York and London p 834-871 on pages 843-845
[15] Mokyr, Joel, (1990) *The Lever of Riches* New York, 13.
[16] Ibid., 29-30.
[17] White, Leslie (1970) *The Science of Culture* Toronto
[18] Ibid., 366
[19] Harris, M (1987) "Anthropology and the Study of Culture" in *Cultural Anthropology* (2 ed) New York, 17.
[20] Popper, Karl (1957) *The Poverty of Historicism*: London, 108-109
[21] Ibid., 110-111.
[22] Ibid., 128-129.
[23] Ibid., 129.

Notes for Part II

[24] Dennell, Robin (1983) *European Economic Prehistory*, Academic Press, London, 81-87
[25] Ibid., 87-96.
[26] Fagan, Brian (1995) *People of the Earth: An Introduction to World Prehistory*, Harper Collins, New York 111-115
[27] Braudel Fernand (1967) *Capitalism and Material Life* 1400-1800, Fontana, London 111
[28] Diamond, J (1998) *Guns, Germs & Steel* London, 160-162
[29] Renfrew, Colin (1973) *Before Civilization*, Penguin Books, Harmondsworth, England 188-192
[30] Coulmas, Florian (1989) *The Writing Systems of the World*, Basil Blackwell, Oxford 34
[31] Ibid., 41
[32] Ibid., 47
[33] Dickinson, H W (1963) *A Short History of the Steam Engine*, Cass: London
185-186
[34] Rolt, L T C (1963) *Thomas Newcomen*, David and Charles: Dawlish, London, 20-21
[35] Ibid., 23
[36] Dickinson, H W (1963) *A Short History of the Steam Engine*, Cass: London
16
[37] Rolt, L T C (1963) *Thomas Newcomen*, David and Charles: Dawlish, London, 24-25.
[38] Dickinson, H W (1963) *A Short History of the Steam Engine*, Cass: London 29-30

[39] Burstall, Aubrey F. (1963) *A History of Mechanical Engineering,* Faber & Faber: London Buxton 191
[40] Rolt, L T C (1963) *Thomas Newcomen*, David and Charles: Dawlish, London 65
[41] Dickinson, H W (1963) *A Short History of the Steam Engine*, Cass: London 51
[42] Ibid., 29
[43] Ibid., 74
[44] Ibid., 64-65
[45] Ibid., 80
[46] Thirring, Hans (1958) *Energy for Man*, Indiana University Press: Bloomington 51
[47] Ibid., 51-52.
[48] Burstall, Aubrey F. (1963) *A History of Mechanical Engineering,* Faber & Faber: London Buxton, 279
[49] Lewis, Brian (1971) *Coal Mining in the Eighteenth and Nineteenth Centuries*, Longman Group Ltd: London 10.
[50] Neil (1978) *The Economic Development of the British Coal Industry*, Batsford Academic: London, 11
[51] Dickinson, H W (1963) *A Short History of the Steam Engine*, Cass: London, 35
[52] Ibid., 35.
[53] Rolt, L T C (1963) *Thomas Newcomen*, David and Charles: Dawlish, London 135
[54] Dickinson, H W (1963) *A Short History of the Steam Engine*, Cass: London 74.
[55] Rolt, L T C (1963) *Thomas Newcomen*, David and Charles: Dawlish, London 134
[56] Dickinson, H W (1963) *A Short History of the Steam Engine*, Cass: London 83.
[57] Balazs, Etienne (1964) *Chinese Civilisation and Bureaucracy*, Yale University Press: New Haven and London 11.
[58] Basalla, George (1988) *The Evolution of Technology*, Cambridge University Press, Cambridge 175
[59] Coulmas, Florian (1989) *The Writing Systems of the World*, Basil Blackwell, Oxford 70

Notes for Part III

[60] Galtung, Johan & Inayatullah, Sohail (ed) (1977) *Macrohistory and Macrohistorians,* Westport, Conn
[61] Gardiner, Patrick (ed) (1959) *Theories of History,* Glencoe
[62] Grunger, Rolf (1985) *Philosophies of History,* Aldershot, Hants
[63] Donagen, Alan & Barbara (1965) *Philosophy of History,* New York
[64] Shaw, W F. (1978) *Marx's Theory of History,* Stanford, California, 60-64.
[65] Ibid., 76.
[66] Ibid., 65.
[67] Ibid., 121-122.

[68] Durkheim, Emile (1984) *The Division of Labour in Society,* Basingstoke
[69] White, Leslie (1969) *The Science of Culture,* New York
[70] Mill, J S (1872) *A System of Logic,* London
[71] Johnson, Allen & Earle, Timothy (2000) *Evolution of Human Societies: From Foraging Group to Agrarian State,* Stanford University Press, Stanford
[72] Sanderson, Stephen (1999) *Social Transformations: A general theory of historical development,* Rowman & Littlefield Publishers, Lanham,
[73] Shaw, W F. (1978), *Marx's Theory of History* Stanford, California, 54, 65, 124.
[74] Federn, Karl (1939) *The Materialist Conception of History,* London 14, 16.
[75] Shaw, W F. (1978) *Marx's Theory of History,* Stanford, California, 65.
[76] Cohen, M. N. (1977) *The Food Crisis in Prehistory*, Yale University Press; New Haven & London, 19
[77] Cohen, M. N. (1977) *The Food Crisis in Prehistory*, Yale University Press; New Haven & London, 20-23
[78] Elkin, A P (1974) *The Australian Aborigines*, Angus & Robinson Publishers: London, 51
[79] Abbie, A A (1970), *The Original Australians*, A H & A W Reed: Sydney 200-202
[80] Sahlins,M (1972) *Stone Age Economics*, Aldine Atherton: Chicago
[81] Cohen, M. N. (1977) *The Food Crisis in Prehistory*, Yale University Press; New Haven & London, 34
[82] Hill & Hurtado, 1989, *American Scientist*, 77(5),436,442
[83] Oberg, K (1973), *The Social Economy of the Tlingit Indians*, University of Washington Press, 67-68
[84] Hill & Hurtado, 1989, *American Scientist*, 77(5) 442
[85] (Price & Gebauer, 1995, 25). (Braidwood, LS & RJ Prehistoric Village Archaeology in South East Turkey, Oxford: *British Archaeological Reports* S 138), (Braidwood RJ The Agricultural Revolution, *Scientific American*, 203, 130-141). (Braidwood RJ Prehistoric Men, Glenview, Illinois, Scott Foreman & Co)
[86] Fagan, B (1995), *People of the Earth: An Introduction to World Prehistory*, Harper Collins, New York 111-115
[87] Dennell,R (1983), *European Economic Prehistory*, Academic Press: London 81-87
[88] Dennell,R (1983), *European Economic Prehistory*, Academic Press: London 81-87