Disease is ubiquitous. Disease afflicts humans. It afflicts animals. It afflicts plants. People refer to disease in their everyday conversation. Newspapers comment upon it. Parliament enacts legislation regarding it. Novelists write about it. Artists depict it. Physicians, veterinary surgeons and agriculturalists seek to combat it. Insurance companies offer reimbursement against it. Anthropologists study it. Philosophers debate its nature, and dictionaries define it. Disease looms large in human consciousness.

One might presume that, since disease is so important in daily life, human beings would know exactly what they mean by it. Most people seem to believe instinctively that they understand the nature of disease, and that their ideas about it coincide with other people’s ideas. The definition of disease therefore arouses little controversy in everyday conversation. People use the word *disease* as readily as they use the words *spade*, or *table* or *nose*. They suggest, when they joke that somebody *calls a spade a spade*, that the nature of the implement used to dig the garden is so obvious that it requires no further definition. Similarly with a table or a nose. They might debate how many legs a table must have, but—regardless of the answer—rarely deny that it is a table; whilst every human must surely know what a nose is. This high level of agreement
about so many commonly used terms perhaps creates an assumption that the meaning of disease is equally obvious and requires no further analysis. Is this, however, really the case?

Disease is a somewhat less concrete phenomenon than is a spade or a table or a nose. Its existence, most would agree, is incontrovertible, but its nature is less clear. It is something that seems to befall people and animals and plants. It rarely serves any useful purpose. It often carries dire implications. It is something that most of us would prefer not to have, but rarely succeed in avoiding. It commonly comes unannounced and at inconvenient times. It usually causes distress, but not always. It can have a fatal outcome. Some people appear more prone to it that others. It sometimes sweeps through whole populations producing social devastation, but its manifestations vary. Some diseases affect a person’s whole body, others merely a part of the body; some affect some parts of the body, others other parts. Some diseases only affect humans, whereas others affect both humans and animals. Some spread from animals to humans, others from humans to humans, and others still do not appear to spread at all. Some diseases affect plants, and few that affect plants seem to affect humans, but some humans can acquire diseases when they come into contact with plants that appear to have no diseases. Any reasonable analysis of the nature of disease must account for all these aspects and many others also.
The nature of disease is a topic that has attracted the attention of physicians, scientists and philosophers over millennia. The close association that existed between medicine and philosophy in the classical Egyptian, Palestinian and Greek eras ensured that scholars who flourished in those societies examined the nature of disease. Comparable developments occurred in classical Indian and Chinese civilizations. The natural philosophers of Renaissance and post-Renaissance Europe divided into competing schools of thought over the nature of disease. More recent years have witnessed an enormous flourishing of physicians, pathologists, and agriculturalists who study aspects of disease that relate to their individual disciplines. Most of these researchers have, however, examined ever-narrower aspects of specific diseases—such as manifestations, mechanisms and causes—rather than the generic nature of the phenomenon.

Some contemporary philosophers, on the other hand, have become interested in general aspects of the topic. They have proposed a number of novel ideas and reached some stimulating conclusions, although they can hardly yet claim to have reached a consensus. This lack of unanimity presumably implies that the issues involved require closer analysis if a formulation is to emerge that most of them can accept.
The object of the present thesis is to undertake such an analysis. It will start by outlining in this introduction the general background to the topic. It will then detail the more noteworthy of previously proposed theories about the nature of this phenomenon, classifying them according to their most prominent components, and assessing their several strengths and weaknesses. It will next discuss the specific philosophical issues of definition, causation, and explication in the biomedical context, before suggesting a comprehensive, but succinct, definition that acknowledges many older views about disease, encompasses current usage, and provides a theoretical base from which to work into the future. It will finally test the strengths and weaknesses of that definition to account for observed phenomena and to accommodate some former definitions.

* * * * * * * *

A detailed discussion of the nature of disease might reasonably start by examining the question of whether disease (or diseases) actually occur. Examination of this simple question immediately subdivides into examination of three inter-related questions. Firstly, do people believe that disease occurs (or diseases occur)? Secondly, do(es) disease(s) qua disease(s) indeed occur? And thirdly, if so, what explanation best describes the nature of that which occurs? Overwhelming evidence favours a positive answer to the first of these questions. That then implies,
but does not prove, the likelihood of a positive answer to the next question, but still leaves the answer to the third question as a major issue for discussion. This work focuses upon the latter two questions.

Substantial evidence points to the existence of a widespread belief in the occurrence of disease. Virtually every copy of every newspaper contains articles that mention it. Simple observation confirms that such reports reflect everyday conversation in the community. People speak about, and fear, disease. Parliamentarians, responding to the perception that diseases can threaten society, legislate to regulate the activities of those who might spread them. Writers and painters use disease as the central motif in many of their works. Disease drives millions of people to seek relief each year from physical or mental distress—help that is not economic, not military, not spiritual, not simply social, but uniquely medical. Disease thereby provides the *raison d’être* for the whole health care industry. Governments organize national health services to assist their citizens who succumb to disease, and commercial enterprises offer them insurance against the vicissitudes created by it.

The idea of disease is, furthermore, not confined to western civilization. Explorers, missionaries and colonizers in previous centuries often noted the existence of sophisticated belief systems among the local people whom
they encountered.\textsuperscript{1} Their contemporary counterparts, the professional anthropologists, find similar ideas in almost every society that they study.\textsuperscript{2} Even modern iconoclasts (such as Ivan Illich), who have decried the medicalization of the human experience and the development of a medico-political dictatorship over the whole of humanity, accept without demur the existence of the idea that we refer to as disease.\textsuperscript{3}

This unanimity of belief that disease occurs (or that diseases occur) nevertheless throws into stark contrast the profound disagreement that surrounds the crucially important associated issue of what people mean by disease. Fierce debates revolve around interpretations of the fundamental nature of the whole genre, and the limits of the phenomena that it includes.

These disagreements have huge practical implications. The categorization of a person as suffering from disease produces immediate consequences both for the individual and for others whom he or she contacts. Society

\textsuperscript{1} A huge literature exists on this subject that is beyond the scope of the present work to review. The Indian Ayurvedic system and the traditional Chinese system based upon the Yin-and-Yang theory are obvious examples, but the reviews by Carman (1976) of the topic in Kenya, Eiseman (1996) in Bali, Burton-Bradley (1990) in Papua-New Guinea, Brooker (1998) among the New Zealand Maori, and McCarthy (1957) on the Australian Aboriginals examine the prevalent ideas of disease in widely spread geographical regions uncontaminated intellectually by European views.

\textsuperscript{2} George Murdoch (1980) undertook a particularly comprehensive survey of this issue some years ago and concluded that every one of the 186 tribal groups that he studied from every continent and many islands of the world had some theory of other to explain disease.
grants a patient—a person who suffers from disease—relief from certain obligations that it demands of healthy people, but requires the patient to behave appropriately. This ‘sick role’ obliges the patient to try to overcome the disease, to seek appropriate professional assistance, and to avoid wilfully spreading contagious disease to others. The patient can then reasonably expect his or her family, and the state, to display solicitous emotional support during periods of incapacity; to provide material assistance, such as food and money, during those periods; and to offer relief from otherwise compulsory social and civic obligations—such as participating in certain family activities, and meeting certain onerous legal requirements. Consistent application of these rights and obligations inevitably requires that people agree about the limits and nature of disease. An agreed definition of disease becomes necessary since any inability to agree will create personal, administrative, and legal havoc.

The implications of disagreement about the limits and nature of disease are at least as severe for animals and plants as they are for humans. Allegations that certain animals carry particular diseases can trigger protective measures to prevent transmission of those diseases to other animals, and these can entail the slaughter of huge numbers of beasts. The economic implications of such actions are immense.\(^4\) The

\(^3\) Illich (1975), page 15.

\(^4\) Consider, for example, the economic effects on the farming community and the travel industry of ‘mad cow disease’ in the United Kingdom during
development of certain diseases in food crops or forests can cause environmental devastation and enormous anxiety among humans who find themselves exposed to products from the affected plants or trees.\(^5\)

Mental disease in humans is another area that tests the definitional limits of disease. This has for long generated unique challenges that carry immense social and political implications. People in positions of power have, for hundreds of years, sought to sustain their privileges by having those whom they wished to control labelled as insane. Nineteenth century examples of this occurred\(^6\), but, more recently, the leaders of certain totalitarian regimes recognized that burgeoning research into psychology and psychiatric medicine provided useful instruments for them to use against political dissidents. They employed psychiatrists to diagnose their opponents as suffering from psychiatric disease and to create disease categories into which to classify them. This thereby provided a superficially

\(^5\) An outstanding contemporary example of this occurs with genetically modified food crops: at the heart of international arguments that are occurring over these crops and that have huge economic implications is whether the artificially induced genetic modifications do or do not make them diseased, and so whether (if they do) the induced disease or diseases carry risks for humans and animals who eat them.

\(^6\) Two amusing, but sad, nineteenth century examples of this were the diseases labelled as *dрапетомания* and *дисаэстезия аэтиопис* by physicians in the Southern states of ante-bellum America. They diagnosed the former as an explanation of the actions of (and to legitimise reactions to) runaway slaves, and the latter to account for “hebetude of mind and of
respectable justification for the emotional isolation of such people and, not infrequently, their incarceration in political prisons posing as psychiatric institutions. External questioning of the legitimacy of these diagnoses and actions invariably revolved around the acceptability of classifying political dissidence as psychiatric disease, however disingenuous the label.  

A contrasting example of other difficulties that psychiatric disease poses arose in the latter part of the twentieth century when a group of psychiatrists and social activists associated with Thomas Szasz launched a vigorous attack on the whole idea of mental disease. They claimed that the diagnosis of mental disease was, even in open societies, generally a corrupt undertaking that merely provided some people with the opportunity of depriving innocent persons of liberty, and of providing guilty persons with the opportunity to escape retribution for their offensive actions by mounting a ‘successful insanity defence’. The consequent possibility of physically threatening and unpredictable people moving freely in the community inevitably alarmed the relatively defenceless members of society. Uncertainties over the definition of disease lay at the base of that whole issue. Such major disagreements over the nature of disease have the obtuse sensibility of body—a disease peculiar to negroes—called by overseers, ‘rascality’. (Cartwright, 1851).

7 See, for example, the paper by Harold Merskey and Bronislava Shafran in which they examined the application by authorities in the Soviet Union of the diagnosis of sluggish schizophrenia to activists whose views they found politically challenging.

8 Szasz (1976).
potential of creating intolerable stresses within legal and social systems, quite apart from potentially leading to inappropriate attempts at medical treatment on the one hand, or inappropriate denial of treatment on the other. They point to the need for a clear definition that can command universal or, at the very least, widespread respect.

Physicians in particular have, for millennia, made attempts to explain and define disease. The intellectual range of their explanations has varied enormously, as has their success. None has attracted universal support, although traces of many of them persist in modern thinking. Some of the oldest theories postulated disease as being due to an imbalance of internal fluids or forces (for example, the classical Chinese theory of Yin and Yang, the classical Indian Ayurvedic *tridosha* theory, and the classical Greek theory of the four humours). Middle Eastern societies in Mesopotamia, Egypt, and pre-Christian Palestine, in contrast, favoured theories of punishment of sinful humans by an angry God or gods as the underlying reason for disease. Several sects of physicians who flourished in classical Rome (including the Dogmatists, Empiricists, Sceptics, Methodists and Pneumatists) held strongly competing views on the nature of disease. Competing schools also developed among physicians in Early Modern Europe, the most prominent of them being the iatrochemists and the iatromechanists. Others, such as Boerhaave, sought to amalgamate these

---

9 Szasz (1986).
chemical and mechanical explanations into composite explanations. Another set of views, popular especially in eighteenth and early nineteenth century Europe were those of the animists and vitalists. Many who advocated clinico-pathological views succeeded them. Later scientific advances encouraged groups of people to develop ideas that emphasized microbiological explanations or theories based on immunological, biochemical, endocrine, genetic psychosocial, or statistical components. Many of these explanations implied definitions of disease, and some of them led to formally stated definitions. All of the views are pertinent to any comprehensive assessment of the topic, especially as many of them contained suggestions that continue to resonate. None of them, however, succeeded in producing an explanation, or a definition, that has subsequently attracted universal support.

The hiatus that resulted from the failure of purely medical and scientific attempts to describe and define disease in persuasive terms created a challenge that has attracted the attention of contemporary philosophers. They too have produced a variety of propositions, but none of them seems as yet to have achieved much greater a degree of acceptance than the efforts of the physicians and scientists. The philosophical approaches adopted have ranged from brief, but incisive, reviews of the long medical and scientific history of discussion of the topic\(^\text{10}\) to a rapidly increasing

\(^{10}\) D. von Engelhardt (1995).
number of specific formulations.\textsuperscript{11} Many philosophers have adopted a strategy of striking a contrast between two or more theoretical positions as a prelude to deciding in favour of one of these.

A.P. Cawadias used that technique as early as 1937 when he contrasted the four categories of \textit{realist} views (disease as a separate thing), \textit{conceptualist} ones (disease as a concept that exists in an observer's mind), \textit{nominalist} ones (disease as a name for an artificial category of morbid phenomena), and \textit{Neohippocratic or constitutional} ones (a diseased organism is one that struggles against external evils).\textsuperscript{12} He favoured the last of these. Sir Henry Cohen (Lord Birkenhead) contrasted an \textit{ontological} view (disease as an entity, so \textit{realist}) with statistical deviation from the norm, and based his definition on the latter.\textsuperscript{13} Temkin\textsuperscript{14} and Whitbeck\textsuperscript{15} both contrasted \textit{ontological} ideas with \textit{physiological} ones (disease as something peculiar to the individual and differing from the normal),\textsuperscript{16} whereas Hudson distinguished \textit{realist} views from \textit{nominalist} ones (disease as merely an idea in the mind).\textsuperscript{17} Wulff, on the other hand, saw a tension between what he called \textit{essentialistic} views (disease exists independently)

\begin{footnotesize}
\textsuperscript{11} See H.T. Engelhardt (1995) and Hofmann (2001) for reviews.
\textsuperscript{12} Cawadias (1937).
\textsuperscript{13} Cohen (1955).
\textsuperscript{14} Temkin (1963).
\textsuperscript{15} Whitbeck (1977).
\textsuperscript{16} Temkin (1963).
\textsuperscript{17} Hudson (1966).
\end{footnotesize}
and others for which he used the term *nominalistic*—by which he meant that there are no diseases, but only sick people.\(^{18}\)

H. Tristram Engelhardt, Jr, proposed that six distinctive ideas exist about disease: namely as something wrong with an organism (what he called an *evaluative* view), disease as species atypicality, disease as a cause of suffering (a *weak normativist* view), disease as a lesion, disease as something that places a person in a sick role (a *social* view), and disease as something of community interest (a *democratic* view).\(^{19}\) This was not, however, his first discussion of the topic as he had earlier suggested that a disease is a theory that explains collections of observable phenomena in an organism that is experiencing unpleasant subjective manifestations.\(^{20}\)

Some other commentators, such as Barondess, tried to read significance into the variety of quasi-synonyms sometimes used in the English language in place of *disease*. He thus proposed that there is a distinction between *disease* as an objective event that involves disruption of structure or function, and *illness* as a subjective event characterized by discomfort and/or psychological dislocation.\(^{21}\) Goosens favoured a *normativist* approach (disease must be a threat to wellbeing) over a *normativist disvaluation* one, an *interventionist* one (disease must enjoin to action), and

---

\(^{18}\) Wulff (1979).
\(^{19}\) Engelhardt (1986), pages 157-201.
\(^{20}\) Engelhardt (1976).
\(^{21}\) Barondess (1979).
a neutralist one (diseases interfere with normal functioning). Brown similarly distinguished between normative views (disease as a structure or function of an organism that is bad) and objective ones (a structure or function of an organism deviates from normal). Hare believed that the major contrasts lay between disease as descriptive (statistically abnormal) and evaluative or normative (disease as bad). Gillon divided the theories into realist (disease as a thing), nominalist (disease as a difference from the norm that places an organism at a biological disadvantage), value-free (based merely on statistical abnormality), and disease as a malady (the consequence of an intrinsic evil). Fedoryka also saw the big contrast as being between value-free and value-laden concepts. Nordenfelt suggested that the principal distinction lay between a Biostatistical definition of disease as proposed by Christopher Boorse and a holistic definition as proposed by Georges Canguilhem (disease as something that interferes with a person’s ability to realize his or her vital goals).

Although reliance on contrasts between competing theories has provided many philosophers with a fertile method for debating the possible definitions of disease, others have adopted the alternative strategy of simply arguing for individual theories that they favour. Margolis, for

---

22 Goosens (1980).
24 Hare (1986).
26 Fedoryka (1997).
example, perceived disease as anything that disorders the body or mind relative to prudential function, which he listed as avoidance of death, prolongation of life, restriction of pain, gratification of desires, maintenance of security, and the like.  

King, Scadding, and Boorse followed Cohen in focussing on statistical deviation from the norm as the basis for their definitions. Redlich defined disease as what is treatable, whilst others saw a social construct in it without thereby implying that it was a mere myth or was false. Fulford argued for a bridge between biological and social theories of disease, but Clouser et al. took the quite different approach of skirting around the need to define disease by encompassing the whole collection of categories of disease, illness, injury, and sickness under the term malady. They then described this as a condition that could cause any of several evils to befall a person. Yet another radically different approach harked back towards a scientific interpretation when it noted that any disease process has four important aspects: causation, mechanism of development (or pathogenesis), structural alterations, and functional consequences.

---

28 Margolis (1976).
29 King (1954).
30 Scadding (1959).
31 Boorse (1977).
32 Redlich (1976).
33 Klerman (1977); Birch (1979).
34 Fulford (1993).
Many of these philosophers accepted that, despite the strength of their own individual arguments, no agreed definition of disease existed at the times when they were writing.\textsuperscript{37} An increasing level of scepticism, indeed, has developed among some writers on the topic in recent times. Temkin perhaps started the argument in 1963 when he suggested that the question ‘does disease exist or are there only sick people?’ lacks a meaningful answer.\textsuperscript{38} Ladd certainly focussed the argument a few years later when he claimed that disease is incapable of definition in the strict sense of the word.\textsuperscript{39} Hesslow, in 1993, went so far as to argue that we do not even need a definition of disease.\textsuperscript{40} Although Nordenfelt immediately disagreed with him\textsuperscript{41}, this did not prevent Worrall later from posing the provocative question: ‘Why ask “what is disease”?‘ and concluding that ‘There are many issues concerning medicine to which philosophers, especially philosophers of science, can make significant contributions…but I fear that the issue of “what is disease”? is not one of them’.\textsuperscript{42}

These many relatively recent philosophical discussions lead to some obvious conclusions. The different approaches mentioned above have

\textsuperscript{37} One can find explicit statements or comments along these lines in the writings of Temkin (1963), Hudson (1966), Hudson (1977), Hucklenbroich (1984), Brown (1985), Englehardt (1986), Hare (1986), Gillon (1986), Sadegh-Zadeh (2000), and Hofmann (2001).
\textsuperscript{38} Temkin (1963).
\textsuperscript{39} Ladd (1986).
\textsuperscript{40} Hesslow (1993).
\textsuperscript{41} Nordenfelt (1993), pages 15-26.
certainly produced new insights and some highly stimulating views.\textsuperscript{43} The intense disagreements and lack of consensus have, however, distinguished them. This has then left practical people, who need a working definition of disease that can command a reasonable degree of acceptance, totally dissatisfied. Their need for clarity of definition remains as urgent as it was before the philosophical debate started. The word \textit{disease} continues in everyday use and the ideas that it portrays continue to have grave social and legal implications. However tough the problem of defining disease might appear to scientists, physicians and philosophers, the world will not stand still whilst they conduct interminable discussions.

Considerations such as this have recently stimulated the Norwegian philosopher, Bjørn Hofmann, to analyse why the rival theoretical frameworks have failed to provide a useful conclusion.\textsuperscript{44} He has concluded that profound philosophical issues interfere with any attempt to define the term \textit{disease}. He has suggested that several of these are: whether instances of disease have anything in common; what the minimal attributes are of disease; whether the word \textit{disease} refers to a thing, or to some phenomenon other than a simple material entity; whether disease is best thought of as a deviation from normal; whether the idea of disease is a value-laden concept; and what the implications are of the several

\textsuperscript{42} Worrall (1999).
\textsuperscript{43} See Chapter 1.
\textsuperscript{44} Hofmann (2001).
alternative words that we sometimes use for disease. He has also noted the vast array of theoretical terms that various people have used to categorize the theories and definitions advanced\textsuperscript{45} and implied that some of the disagreements may have arisen from attempts to categorize ideas on a theoretical basis with confusion of the implications of the terms used, rather than by focussing on the attributes of disease that the various positions seek to demonstrate. He did not attempt to judge between the many existing philosophical positions, but despaired of reducing the concept to a particular perspective or to ‘a monistic conception’. He did not, nevertheless, deny the possibility of developing a concept and associated definition that might achieve widespread acceptance, whilst commenting that ‘the complexity of the debate reflects the complexity of the concept itself: disease is a complex concept, and does not easily lend itself to definition’.

Nevertheless, despite this claim, the practical and pressing problem of defining disease remains for communities that use the word regularly. The present research seeks to assist them. It arises out of a refusal to accept the invincibility of the nihilist position. The challenge for it is to rise above the disputes, often at cross-purposes, that have plagued the field, and to

\begin{itemize}
\item analytic, anti-realist, biological, causal, Cnidian, Coan, constructivist, deductive, descriptivist, deviation from the norm, empirical, entity concept, essentialist, Hippocratic, holistic, ideal, naturalist, nominalist, non-descriptivist, normative, ontological, particularist, Platonic,
\end{itemize}

\textsuperscript{45} They include: analytic, anti-realist, biological, causal, Cnidian, Coan, constructivist, deductive, descriptivist, deviation from the norm, empirical, entity concept, essentialist, Hippocratic, holistic, ideal, naturalist, nominalist, non-descriptivist, normative, ontological, particularist, Platonic,
address issues such as those that Hofmann noted. It uses two strategies, novel to this area, to achieve this. Firstly, whereas many previous philosophical attempts to define disease involved comparison between rival theories, with final adoption of one of these, the present research will seek to amalgamate many of the distinguishing characteristics that previous investigators have associated with the idea of disease into a comprehensive construct that can underpin a succinct definition. It will, secondly, set out to define disease by examining how this occurs in simple non-sensate organisms of the plant and animal worlds before moving to explanations applicable to the special circumstances of more complex sensate organisms such as humans, on the assumption that disease in the former will pose fewer problems to define than will disease in the latter. The presence of a mind in vertebrates such as humans adds complications that a comprehensive definition must eventually encounter and comprehend, but development of the explanation does not need to address the most difficult aspects before accounting for the more straightforward, despite the tendency of many previous investigators to adopt that approach. It will therefore address the former first and then later graft on to the resulting formulation explanations of the more complex issues. The most appropriate background, however, against which to develop a new definition would appear to be that provided by a detailed examination of previously proposed explanations and definitions of disease, and of some positivist, practical, rationalist, realist, reductionist, relativist, scientific,
weaknesses apparent in these. That, therefore, will form the first substantive chapter of this work.

* * * * * * * *

social, subjectivist, universalist, value-free, and value-laden.
1. PREVIOUS SCIENTIFIC IDEAS ABOUT DISEASE

Any serious attempt to explain and define the idea of disease must acknowledge previous accounts of the subject. Suggested explanations and definitions lie at the heart of a huge medical and scientific literature that has focussed upon human diseases, with less frequent references to diseases in animals and plants. The present chapter surveys some prominent pre-scientific and scientific suggestions made during the past four millennia that have culminated in a recent flurry of philosophical discussions. It also briefly assesses some of their more obvious strengths and weaknesses. The literature on this topic is huge, with a labyrinthine collection of interlocking, but competing, theories and contributing ideas. One cannot hope to do more than provide a brief overview of the vast subject, but one must make some assessment of its major themes in order to understand the many strands that are woven into the present fabric of the idea of disease. Various approaches are possible in undertaking a survey of it: I shall address the major issues in an approximately geographical and chronological manner for the eras before the 19th Century, and then proceed topic by topic thereafter.

§1.1 ANCIENT IDEAS
The Yellow Emperor’s Canon of Internal Medicine is probably the oldest surviving written account of disease.\(^1\) It contains responses to enquiries that a Chinese emperor made in about 2650 BC, when he sought to relate medical to religious and philosophical thought in his dominions, and it continues to have a strong influence in East Asia. It used the pre-existing Chinese philosophical tenets of a yin-yang duality and the Tao to explain disease by starting from two assumptions: that matter consists of five interacting elements (metal, water, wood, fire, and earth); and that impersonal celestial forces create cycles of night following day, winter following summer, rain following drought, and so on. The yin-yang idea then postulated many opposites as existing in the universe—such as man (the yang force) as the opposite of woman (the yin), active of passive, positive of negative, heat of cold, beauty of ugliness, and good of evil—with each partner in any pair requiring the other to exist to balance it. So disease, it supposed, is the opposite of health and wisdom equates to knowing the Tao, or ‘Right Way’, to achieve the equilibrium that is health. Success in that exercise would promote a flourishing life, whereas a waning yin would lead to yang predominance and manifest as aging and decay.

This underlying theory had some attraction, such that Chinese physicians then used it to integrate the animate and inanimate worlds by speculating

\(^1\) Veith (1992).
about a system of numbers that revolved around the figure five (from the five elements). They believed that this caused biological structure and function to remain in yin-yang equilibrium; and that a complex set of conventions would help them diagnose, treat and prognosticate about disease. They identified the site of disease by relating the body’s organs to the seasons of the year (to the obvious four of which they had to add a mythical fifth) and other factors. Characteristics of the pulse and respiration suggested certain organs, but differed depending upon whether the physician examined the pulse at the right wrist or the left. Diagnosis described a yin-yang imbalance (in terms of the seasons, the weather, and disregard of the Tao) rather than any named disease. A human could prevent disease by attaining balance in all actions; and treat it by adjusting the five items of spirits, nutrition, bowel movements, medications, and acupuncture according to a therapeutic numerology. Treating with the opposite to the character of the disease would promote balance; so cool climate diseases would require hot treatments. Insertion of hollow acupuncture needles into the twelve important channels of the body (that corresponded to the twelve channels of the Yangtze River) would drain accumulated toxins to reimpose balance. Application of moxibustion cones to the skin was similarly supposed to drain toxins into blisters. Cosmic relations such as the position of the planets, the season, and the weather would identify the best time to start treatment, with the respiratory cycle fine-tuning this.
The two great strengths of the classical Chinese explanation of disease were its accord with a general philosophy of the universe, and its description of the relationship between health and disease as an imbalance. The latter idea was to recur in several later explanations. Its contemporary weaknesses—that still remain—were the speculative nature of its atomic theory (has anyone ever had good grounds to claim that matter consists of non-subdivisible elements?), its numerological theory, its interpretation of many physical observations, and the treatments that it promoted. The idea of disease as imbalance and disequilibrium nevertheless provided an intuitively persuasive explanation for its observed association with various extreme circumstances, both physical and emotional. The production of balance within a complex physico-chemical system furthermore implied the presence of auto-regulatory feedback mechanisms, an idea that continues to reappear in contemporary studies of endocrinology, molecular biology and physiology (see §1.4 below).

In classical India, unlike China, several successive theories of disease developed. The Harappã civilization (c. 3000-1500 BC) came first, followed by the Vedic (c. 1500-400 BC), the Buddhist (after c. 500 BC),
and the Ayurvedic (after 200 BC) cultures. The Harappā people emphasized personal hygiene to avoid disease. The Vedic people, whilst recognizing injuries and empirical causes of some diseases, attributed others to sorcery or sin that provoked intruding demons such as takmán. They tried to gain protection for themselves and their crops (and damage for their enemies) by chanting to the demons, wearing devices such as amulets, performing rituals, and offering sacrifices. The Ayurvedic system, in contrast, resembled the Chinese ideas by invoking imbalance as the mechanism of disease. Its instigators, the philosophers Charada and Susruta, based their tridosa theory on an atomic and numerological concept of matter that the earlier proto-scientist, Kanada, had proposed. They perceived the universe as having six categories, with one (substance) subdividing into nine entities that included space, air, fire, water, and earth. Atoms, they believed, combined chemically to produce new substances. Food converted by chemical reactions in the body into eight dhātus, seven of which were anabolic, whereas the eighth (the malas) comprised the excrements (faeces, urine, and sweat) and the dosas (three forces known as vāyu, pitta, and kapha—reflecting perhaps

---

4 The hymns included the Rig, Sāma, Yagur, and Atharva Vedas, with the associated Brāhmanas, Āranyakas, and Upanisada, for details of which see Basham (1988), pages 234–235. See also the Atharva Veda 5.22, quoted by Zysk (1985), pages 40–41.
water, air, and fire\textsuperscript{5}. Disease (\textit{vyadhi}) occurred when disturbance of the \textit{dosas} destroyed an organism’s healthy equilibrium of structure and function.\textsuperscript{6} The type of disequilibrium determined the nature of the disease. Specific external stimuli (such as abnormal weather, physical activity, emotional stress, or food) could produce excess of one or other \textit{dosa}. Similar mechanisms would cause disease in animals as in humans.\textsuperscript{7}

The Ayurvedic system gained great popularity and continues to attract followers in India and elsewhere, despite attracting criticisms similar to those of the classical Chinese theory—especially a lack of objective identification of such essential components as the \textit{dhātus}. Its positive aspect is, however, also similar—namely the presence during health of a self-regulating physiological equilibrium that is lost in disease. The pervasiveness at a cross-cultural level of this idea of imbalance in disease presumably therefore makes it one that any successful theory should seek to accommodate.

Foreign ideas emanating from China (carried mainly by itinerant Buddhist monks\textsuperscript{8}), Mesopotamia, Greece (especially after the invasion of India by Alexander the Great in 326 BC\textsuperscript{9}), and the Arab world, also influenced

\textsuperscript{5} Jagi, \textit{Indian System of Medicine} (1973), pages 7-8 and 111-120.
\textsuperscript{6} Dubey (1978).
\textsuperscript{7} Basham (1988), page 502.
\textsuperscript{8} Basham (1988), page 62.
\textsuperscript{9} Basham (1988), pages 58-61.
Indian ideas about disease. The Buddha’s views on the subject derived from a general belief that the causes of, and solutions to, each person’s problems lie in that individual’s own actions. He held that life and the body are impermanent, that sickness and death are inevitable, that death is only the end of the physical body, and that enlightenment exists forever in the truth of the Dharma.\textsuperscript{10} Causation is crucial: everything that happens comes about because of actions over which humans can seek control. The world of the Fourfold Noble Truth is full of suffering caused by ‘the thirsts of the physical body’ and ‘the illusions of worldly passion’. He taught that elimination of desire, that underlies all human passion, would remove suffering. The Buddha entreated people to follow a Noble Path of eight stages to achieve this, and thereby to deal with disease.\textsuperscript{11}

Buddhist teachings provide an obviously attractive means of avoiding and treating many diseases, but have the major weakness of a strongly anthropocentric nature and an assumption that organisms ultimately create and control all their own vicissitudes. Butterflies and amoebae and plants are as susceptible to disease as are humans, and it is difficult to see how they could avoid it by eliminating illusions of worldly passion. A human, similarly, could hardly be considered to indulge in objectionable passion by continuing to breathe, even though that person might thereby inhale invisible bacteria and so develop tuberculosis. Despite these

\textsuperscript{10}Buddha (1988), page 24.
criticisms, the Buddhist view has strengths with regard to many human
diseases that any new explication would do well to accommodate. Many
things that go wrong in people’s bodies occur as a result of conscious, but
unwise, decisions that the individuals themselves make, so the wisdom
and prophylactic implications of the Buddha’s assertions contain an
obvious ring of truth. Any successful theory of disease should presumably
try to accommodate them in much the same way that it accommodates the
idea of imbalance.

Interactions between the classical Indian and Mesopotamian civilisations
created some similarities in their ideas also about disease. The
Mesopotamians used empirical treatments such as herbal medicines and
enemas for skin and bowel diseases, and for snake and spider bites, but
their world during the period from 3000 to 538 BC contained a complex
nexus between religion and medicine.\textsuperscript{12} Their belief that various gods,
goddesses, devils, and sorcerers could torment people by inserting
demons into their bodies had an obvious similarity to Vedic ideas. Specific
gods purportedly caused specific diseases: Adad, for example, would
smite a person on the neck, whereas Istar would cause pain in the chest
of someone who angered him. A person who infringed society’s many
behavioural prohibitions risked punishment by an enraged god who
governed issues of personal hygiene and human interactions. The gods

\textsuperscript{11} Buddha (1988), page 74.
might punish those who touched, or sat on the bed of, or ate the food of, or even talked to, someone identified as 'unclean' due to disease. Cure required exorcism of the demon by a sorcerer-priest-physician who would name many evil spirits in the hope of diagnosing the mediator of the disease by mentioning its name, providing magic medicines, and conducting rituals. These might include sacrificing an animal, making a statuette of the patient into which the demon could migrate, placing coloured threads on the affected part of the body with magical numbers of knots tied in them, or advising the patient to recite prayers whilst walking a certain path in a certain way.

These Mesopotamian views had the advantage of encouraging people to prevent disease by adopting lifestyles that had (or, at least in modern terms, would be described as having) a sound public health rationale. They suffered, however, from an important weakness that was common to every theologically based theory of disease: namely the speculative nature of the various gods who were alleged to control disease and the absence of information about the ways in which the gods (if they existed) mediated it. The proposal of a divine explanation for the otherwise inexplicable had an obvious attraction for people living in a society in which religion loomed large as it directed their actions into paths that had a strong empirical

---

12 Thompson (1994).
wisdom, but it then begged as many questions as it answered since it merely clouded the issue of why these were wise paths to follow.

Egypt, under its unified kingdom that existed from 3000 to 332 BC, communicated with Mesopotamia and experienced an even greater interaction between religious and empirical ideas of disease.\textsuperscript{13} The religious construct suggested that disease resulted from the intrusion of spirits. Separate gods existed for everything, including each of the 36 parts of the human body, sickness, healing (the goddess Sekhomet), and death. Sekhomet had priests, physicians, and magicians as followers, all of whom could diagnose disease. A symptom was a disease. Ordinary people believed that it arose from ‘something entering from outside’, ‘the breath of an outside god or death’.\textsuperscript{14} All but the most obvious injuries had supernatural causes. Reciting incantations, performing rites, and wearing amulets to honour the spirits were important therapeutic manoeuvres.

Egyptians sought relief from their diseases by praying to great physicians who had achieved deification—people such as Imhotep during the IIIrd

\textsuperscript{13} Information about disease comes especially from various papyri, of which the principal ones of medical importance are the \textit{Gynaecological Papyrus of Kahoun} and the \textit{Veterinary Papyrus of Kahoun} (both XIIIth Dynasty, about 1850 BC); the \textit{Papyrus Ebers}, the \textit{Papyrus Berlin 3027}, the \textit{Papyrus Hearst}, and the \textit{Papyrus Edwin Smith} (all approximately early XVIIIth Dynasty, about 1550 BC); the \textit{Papyrus Berlin 3038} and the \textit{Papyrus Chester Beatty No. 6} (XIXth Dynasty, about 1300 BC); and the \textit{Papyrus Carlsberg No. VIII} (XIXth or XXth Dynasty about 1200 BC). See Sigerist (1951).
Dynasty. These men usually built their reputations as exponents of an alternative tradition that emphasized material mechanisms rather than theistic causes. Many of them believed that both health and disease travel through the body in a series of channels (the *metw*) that resemble canals in irrigated farmland.\(^{15}\) Obstructions could cause flooding of some parts of the body and droughts elsewhere. Diseases, they thought, often entered through the mouth and departed through the bowel. Words such as áaã and *whdw* described certain important physical items, the translations of which remain uncertain, although the latter possibly denoted an inflammatory substance in the faeces that could enter the blood to produce clotting and pus, infect open wounds, and respond to treatment with enemas.\(^{16}\) That interpretation, if correct, perceived *whdw* as a mediator of disease rather than as disease in general or as any disease in particular. The implication was that it represented an internal substance or force that connected external initiating factors of disease in a particular organism with the manifestations of the disease in that organism. This was a novel idea and if it represents an accurate interpretation of the word was an early suggestion of a proposal that often recurred thereafter as people sought to conceptualise disease. It expresses an idea that nowadays looms large in scientific understanding of disease processes and which this thesis will later examine in considerable detail.

---

\(^{14}\) Wilson (1962).  
\(^{15}\) These had a certain analogy to the land of the Nile Delta.  
\(^{16}\) Steuer (1948).
Biblical Palestine, like Mesopotamia, had close relationships with Egypt. The Old Testament (1800-400 BC) provides insight into the Israelites’ ideas about disease that also included both empirical and divine explanations of causes and manifestations. Their divine views revolved around sin and God’s wrath. Transgressions that his chosen people made against his codes of conduct attracted punishment with disease or death of themselves, their animals, and their crops—perceptions that still remain important in some Middle Eastern societies and for followers of some religious codes derived from them.\(^\text{17}\) Moses wrote what was perhaps the most succinct summary of this view when he stated that doing evil and rejecting God would bring overwhelming disaster, confusion, trouble and destruction\(^\text{18}\):

\[
\text{He will send disease after disease on you... The Lord will strike you with infectious diseases, with swelling and fever; he will send drought and scorching winds to destroy your crops. These disasters will be with you until you die.}
\]

Some other Old Testament passages, however, conveyed empirical notions. The Israelites recognized the contagious nature of many skin diseases, for which they isolated both the patient and the fomites.\(^\text{19}\) They noted symptoms such as lameness, blindness, hunchback, dwarfism,

\(^{17}\) Deuteronomy 32, 39, Psalm 91, 5-8. Joseph gained control of Egypt by interpreting the pharaoh’s dream that seven years of famine would follow the destruction of the crops (Genesis 41, 1-57).

\(^{18}\) See Deuteronomy 28, 20-22; Exodus 9,3; 12, 21-28; 1 Kings 8, 37; 2 Chronicles 6, 28; Amos 4, 9; and Hagar 2, 17.
pain, and wasting that they did not immediately attribute to divine wrath.\textsuperscript{20} Their empirical viewpoint waned somewhat during the inter-testamentary period (400-4 BC) when Babylonian, Egyptian, and Persian ideas encouraged the Jews to perceive the world as a battleground between God and his traditional opponent Satan, with the latter sending unclean spirits of the underworld to disturb people’s behaviour.\textsuperscript{21} Satan particularly attacked people whom God forsook\textsuperscript{22}, although obedience to divine law would provide protection.\textsuperscript{23}

The perceptions about disease of the early Christians, as displayed in the New Testament, differed from traditional Jewish ideas. Jesus promoted reform by preaching a message of love rather than of retribution, claiming to cherish all regardless of their sins, and aiming to cure disease rather than to cause it. He suggested, in response to a question asked by a rabbi as to whether a congenitally blind man’s own sins or his parents’ sins had caused his affliction, that neither had sinned and that the man could regain his sight by taking appropriate personal steps that included having faith in the possibility of cure.\textsuperscript{24} Most New Testament authors adopted a similarly empirical approach to disease, focussing on its manifestations and cure rather than speculating about its causes or mechanisms. The complaint, to

\textsuperscript{19} They often described these as leprosy; see Leviticus 13; and 15,2. \textsuperscript{20} Leviticus 21, 16-20; Job 33, 19-22; Psalm 25; Psalm 48, 6; Isaiah 21, 3-4. \textsuperscript{21} Matthew 12, 22-30. \textsuperscript{22} Satan, for example, made sores break out all over Job’s body (Job 2, 7). \textsuperscript{23} Psalm 91, 5-8.
them, was the disease. A display of faith might assist in its cure, without necessarily implying a mechanism of causation. This marked change in views that occurred between the Old and New Testaments has attracted remarkably little previous comment, but the striking reduction in emphasis on divine causation perhaps marked a turning point in conceptualisation. This point was, however seemingly lost on some later Christian theologians, such as St Benedict, who espoused positions on disease that more closely resembled those of the Old than the New Testament when they attributed most cases to sin, especially if it afflicted their enemies. Others however, such as the Venerable Bede, never adopted that viewpoint.25

This overview suggests that empirical and divine ideations about disease coexisted in each of the civilizations of classical India, Mesopotamia,

24 John 9, 1-41.
25 Kroll and Bachrach (1986). Much subsequent Christian thought has continued to reflect ambivalence on the issue, with the empirical viewpoint clearly regaining ascendency in recent times. Thus the Church of England’s Book of Common Prayer indicated in 1662 that sickness was ‘certainly’ God’s visitation for which sufferers had to repent of their sins, display patience, give ‘humble thanks for this heavenly visitation’, and take the Lord’s chastisement in good part (pages 297-298); whereas the late 20th Century revision of that publication proposed prayers that sought assistance for those who suffer from diseases and those who work to discover their causes and cures, without particular emphasis on the role of sin (pages 573-574). Different Christian denominations moved in this area at different speeds, and even at different speeds within the same denomination. The Roman Catholic Church exemplified this when it promulgated prayers for use in Australia in 1986 that resembled the contemporary Australian Anglican prayers mentioned above, whereas it
Egypt, and Palestine. The relative importance of the two views varied from civilization to civilization. The presence of an obvious physical cause favoured an empirical explanation, but the means whereby the cause linked to the manifestations usually remained obscure. The absence of an obvious physical cause favoured supernatural explanations, but these then tended to obscure rather than illuminate the ways in which disease came about. They failed furthermore to address the reason for the obviously diverse manifestations that diseases could assume. Their strength, nevertheless, lay in securing omnipotent support by all believers for behaviour that predisposed to longevity and fulfilment.

Issues similar to those mentioned above stimulated interest in disease in classical Greece and Rome, with which the Israelite and early Christian worlds interacted closely. Many religious cults flourished there, including that of Asclepius (a semi-divine physician and putative son of the god Apollo) that focussed upon disease. The disciples of Asclepius, led by Euphyron at Cnidus on the coast of Asia Minor, and by Hippocrates (c. 460-375 BC) on the nearby island of Cos, had by 400 BC erected beautifully colonnaded temples in tranquil surroundings where the sick could bathe, exercise, diet, and take incubation treatment—sleeping

---

continued to state that ‘illness is mysteriously linked to sin and evil’ in its 1994 revision of its catechism produced for international use (page 375)
peacefully with the aid of hallucinogenic drugs while their dreams and the murmured suggestions of priests soothed their emotions.\textsuperscript{26}

Egyptian ideas about disease influenced the Cnidian physicians who developed rather mechanical views about it.\textsuperscript{27} They did however take a huge step forward from previous ideation when they started to identify distinct disease types distinguished by symptoms such as pain, fever, colour of the urine, sweating, vomiting, thirst, cough, and menstrual irregularities. The physicians at Cos, in contrast, favoured an idea that disease arose from imbalances of forces in the body. Their views therefore had fundamental similarities to the classical Chinese and Indian Ayurvedic views, although they differed importantly in that the Chinese theory relied upon imbalance of two forces, the Indian upon imbalance of three forces, and the Greek upon imbalance of four forces. The four forces of the Coans were the \textit{humours} of blood, phlegm (a cold slimy substance), yellow bile, and black bile (a putative thick black fluid supposedly secreted

\textsuperscript{26} See Sakula (1984); Phillips (1973), pages 197-201; and Adams (1939). Their other important temples were at Epidaurus near Corinth and near Pergamum in Asia Minor. Although the term \textit{Asclepiads} eventually referred only to physicians, these men were as much priests as physicians. The claim that medicine was divorced from religion during the Hippocratic era derives more from the ideological programme of the 18\textsuperscript{th} Century German philosopher, Johann Heinrich Schulze, than from empirical evidence. Hippocrates had long held an unimpeachable reputation as the father-figure of Western medicine, whilst Schulze had a distinctly anti-clericalist agenda: by distancing Hippocrates from religion and identifying him as a 'rational' physician, Schulze strengthened his own case, relatively scanty as was his supporting evidence (see Lonie, 1978, pages 54-55 and page 70 footnote 17).
by the adrenal glands). An intellectual ancestor of this idea was perhaps the earlier proposal by Alcmaeon of Croton (a Greek colony in South-Eastern Italy) that health reflects an equal balance of the powers of moisture and dryness, coldness and heat, bitterness and sweetness, ‘and the rest’, whereas supremacy of any one—often induced by environmental causes—results in disease.

The simple contrast of ideas about disease created by the early Greek priestly physicians developed during the later stages of the Greco-Roman civilization into a bewildering series of speculative theories. These profoundly influenced European medicine for the following 1800 years. An understanding of them is pertinent to any subsequent attempt to explain and define disease since strands of several of them have persisted through until the present time in Western (and so in international) thinking. Any contemporary explication and definition of disease cannot therefore divorce themselves entirely from that inheritance.

Perhaps the most prominent of the several sects of physician-philosophers of the Greco-Roman world were the Dogmatists who, as direct successors of Hippocrates, were the strongest proponents of the

---

29 Kirk and Raven (1975), page 234.
The physicians of Alexandria—who both accepted the Dogmatists’ views, and practiced dissection and vivisection—caused the humoral theory of pathophysiology to gain a close association with anatomical dissection. They contrasted with the *Empiricists*, founded by Akron of Agrigentum (c. 440 BC), who believed that experience was everything, and therefore only accepted findings that they could perceive with their own senses. They denied the possibility of knowledge of the insensible (since experience of the insensible is impossible), of developing general rules, or of discovering obscure causes. Symptoms, to them, were diseases. They accepted at most that *syndromes*—collections of symptoms that occurred commonly in association with each other—were diseases. Repetitive patterns of symptoms might guide the physician’s opinions about prognosis and treatment, but could do no more. They prescribed remedies that experience suggested might relieve symptoms, but their refusal to generalize or to form theories caused them to approach patients in an inherently disorganized and individualistic manner. Their approach, nevertheless, attracted many adherents until the Middle Ages, with strong overtones of it clearly persisting to the present day.

---

The views of a third sect, the Sceptics, resembled those of extreme Empirics.\textsuperscript{31} Their most famous representative, Sextus Empiricus (c. AD 280), believed that humans know nothing, can search for the truth and theoretically gain knowledge of it, but can never be sure of having found it. They perceived every patient as a different organism, observable only for the moment, remaining uncertain even that their observations were true. They could neither prognosticate nor treat.\textsuperscript{32} They faced the practical problem that their philosophy prevented them from providing the relief that sick people craved, so they eventually became mere spectators of disease.\textsuperscript{33} Their patients often drifted away from them to practitioners who belonged to another sect, the Methodists. Although their leaders, Thermiston of Laodicea (c. 50 BC)\textsuperscript{34} and Thessalos (c. AD 50)\textsuperscript{35}, suspected that some unknown thing might cause microscopic pores in the tissues to contract and relax so that atoms could pass through these to redistribute bodily fluids and make the body excessively wet or dry, most of them did not pretend to understand disease: they just treated it. Unlike the Dogmatists, they discounted the four humours; unlike the Empiricists, they claimed to rely on existing knowledge rather than immediate observation; unlike the Sceptics, they claimed to have an effective method; but, like the Sceptics, they saw each sick person as an individual

\textsuperscript{31} Allbutt (1921), pages 163-175; Philips (1973), pages 156-160; Edelstein (1967), pages 195-251.
\textsuperscript{32} Edelstein (1967), pages 195-203.
\textsuperscript{33} Allbutt (1921), pages 163-164.
\textsuperscript{34} Neuburger (1910), page 208.
without seeking to accumulate knowledge. Their treatments were gentle
and mostly harmless but pro-active (for example, counteracting heat by
applying cold), and so they flourished in Rome especially in the first two
centuries of the Christian era.\textsuperscript{36}

The \textit{Pneumatists} were yet another Roman sect. They followed the highly
developed system of Stoic logic, physics, and ethics founded by Zeno of
Citium (336-264 BC). They believed that \textit{pneuma}, a fire-like breath,
permeated all living things, maintaining life, but also causing disease.\textsuperscript{37}

Fevers resulted from the inhalation of polluted air. Air that humans and
animals swallowed with food caused abdominal diseases. The blood
carried air from the intestine to cause distant manifestations, known as
dyscrasias (a term that persists in everyday use in modern descriptions of
disease). These then unbalanced the four humours that the Pneumatists
accepted as secondary phenomena. Their explanations covered each
manifestation—with epilepsy, for example, attributed to breaths that
passed rapidly through the body.\textsuperscript{38} Their physical theory made them
materialists, but the spiritual implications of the pneuma placed them
among the vitalists who believed in the importance of the soul. Their
treatments resembled those of the Methodists.

\textsuperscript{35} Edelstein (1967), pages 173-179.
\textsuperscript{36} Edelstein (1967), pages 189-191.
\textsuperscript{37} Ricken (1991), page 195.
The Pneumatists soon fragmented as a coherent sect, although the materialist-vitalist dichotomy in their ideas has caused their influence to persist until the present time and as such forms a view that any contemporary explanation of disease must also attempt to accommodate. Some of them, however, became Eclectics, a group of people who hardly warranted the title ‘sect’ as they accepted that all the competing dogmas had merit, and used whichever suited their purposes at any given time. The outstanding Eclectic physician was Galen (AD 130-200)\(^\text{39}\), a Dogmatist who admired Hippocrates, dissected animals, believed that diseases arose from imbalance of the four humours, and classified them by disturbance of the humour that seemed primarily responsible. He was an Empiricist in that he relied on experience and observations, but considered that Empiricism ultimately produced intellectual sterility in medicine. He was a Pneumatist in that he accepted the existence of a pervasive life-spirit. He was a Methodist in that he adopted a pragmatic therapeutic approach, basing his treatment on attempts to counteract one tendency with its opposite. He resembled, also, the Sceptics in decrying the futility of many other Roman physicians whom he considered as mere flatterers and charlatans.\(^\text{40}\) Galen’s principal importance to any conceptualisation of disease was that his writings remained the major resource for ideas in the Christian and Islamic worlds for the next 1500 years.

\(^{38}\) Phillips (1973), pages 54-55.
\(^{39}\) Siegel (1968), pages 4-18.
\(^{40}\) Neuburger (1910), pages 245-263.
years. It was not until 1661 when Robert Boyle (Oxford and London, 1627-1691) published The Sceptical Chymist that they came under real threat.

The strength of the Greco-Roman debate about the nature of disease was that it proposed many alternative theories and provided a milieu in which to compare the merits and weaknesses of each. It thereby provided a model for any new examination of the topic. The strengths of some theories, for example of maintaining equilibrium as proposed by the theory of the four humours, resembled the strengths of comparable theories in classical China and India. Others had the merit of variously emphasizing aspects of manifestations, of causes, and of mechanisms. The weaknesses of most arose from the rudimentary nature of the empirical knowledge upon which they relied and (except perhaps for the eclectic views) the confrontations that they created between the various investigators of disease. Their greatest problems were a lack of factual information upon which to base reasonable judgements about the nature of disease, and consequent speculation that conflated with factual information that did later emerge.

One might conclude, at first glance, that the many ideas of disease that held sway in the ancient civilizations have little relevance for modern

This is, however, far from the case. The vast array of complex terminology that contemporary society now uses to refer to the idea of disease, not only in the English-speaking world but also in many other languages, is just one example of their persisting influence. Modern international ideas of disease have developed from those of the Western world and these have developed ineluctably from those of the Greco-Roman world. They, in turn, developed from the contemporary and earlier civilizations with which they had direct or indirect contact. These intellectual developments provide the provenance of modern ideas. Disregard of them makes modern ideas mere intellectual orphans.

Thinkers during the 16th to 18th Centuries struggled with the diversity of available theories, modifying these in terms consistent with the prevailing philosophical insights of their times, eliminating some unsustainable concepts, and developing a basis for their successors during the 19th and 20th Centuries to mould into the contemporary concepts from which any new explication and definition must derive.

§1.2 EARLY MODERN IDEAS: ALCHEMY, PHYSICS AND VITALISM

One of the first people to launch a serious attack on the Dogmatists’ theory of the four humours was Philippus Aureolus Theophrastus Bombastus von Hohenheim (‘Paracelsus’, 1493-1541)\(^43\), a radical Swiss physician and professional antagonist who gained as much attention for

\(^{42}\) Boyle (1661).
his unconventional behaviour as for his unconventional theories. He realized that a good explanation of disease needed a chemical component, but found the existing humoral theory and Galen’s eclecticism unpersuasive. He turned instead to the ancient Egyptian theory of alchemy that held all substances to be reducible to one essence, and that valueless and impure materials (like lead) would change into valuable ones (like gold) if decomposed to their elemental origins then reconstituted.

Paracelsus used this idea to suggest that health and disease, through their dependence on chemistry, also arose from the same root. He suggested that they eventually separated, with one waxing while the other waned, but were essentially interchangeable because of their origins. Health was good and pure (analogous to the alchemists’ gold), disease evil and impure (like lead). The soul, which represented to him health, was good and survived; whereas the body, being subject to disease, was rotten and eventually died. God ruled over all, creating new life from the persisting soul by maintaining the good of health and discarding the evil of disease. Paracelsus distinguished between different diseases, rather than postulating a single disease state with protean manifestations. The cause, rather than the symptoms, revealed the nature and force of any given disease, the precursors of which often existed before birth. Climate was

---

important, but God often produced diseases that warned individuals about sin, and equated to a type of purgatory. Physicians could not presume to judge or modify purgatory, so successful treatment merely indicated their choice of the opportune time to intervene when God was about to relieve the disease (‘Only when the hour of recovery strikes for the patient does God send him to the physician’).  

Paracelsus assembled these ideas into a composite theory in which two types of disease occurred in living creatures: those of the soul, and those of the material body. He proposed that five entia (essential beings) influenced humans, each able to produce distinctive diseases: ens Dei (God), ens astorum (the stars, or climate), ens veneni (poisons), ens naturale (the natural constitution of the body), and ens spirituale (the spirits, or psyche). His complex ideation effectively proposed several simple, but novel, advances: that Galen was wrong about the humours, that disease often has a chemical basis, that causation plays an important role in disease, that diseases have individual identities, and that disease is bad. The second of these marked him out as the founder of the iatrochemical school of medical philosophy. His belief that disease is bad, while hardly novel, identified a theme that would stimulate an important 20th Century philosophical debate (see §2.1 below).

The proposals made by Paracelsus had at least two serious weaknesses: their reliance on the controversial alchemical view of physical substance, and their proposal of the ill-defined theory of entia whose implied physical existence was as illusive as that of the humours that they replaced. He had, however, several successors in the iatrochemical tradition, who tried to improve upon his theories by making distinctive proposals regarding physiology and disease.⁴⁶

Daniel Sennert (Wittenberg, 1572-1637) sought to preserve Aristotle’s four elements and amalgamate them with three other chemical principles (the former explaining aspects that were hot or cold, moist or dry; the latter explaining the colour, odour, and taste of substances), all of which, he believed, had a role in disease.⁴⁷ Jean Baptiste van Helmont (Louvain, 1577-1644), in contrast, rejected Aristotle’s elements and alchemical principles, and proposed instead a complex physiological system based on chemical interactions.⁴⁸ He postulated that each material object had an archeus (a vital force or ferment) and that matter had three phases of life—first life (the germ), middle life (growth and maturity), and last life (decline and death). Interaction between the middle lives of two objects would produce a battle between their archei. The archeus of a person or animal who ate food would engage the archeus of the food in battle; the

⁴⁶ King (1978), pages 64-94.
food would then enter the middle life of the consumer, who would assume some of the characteristics of the food if the consumer's archeus prevailed. Health was the consequence of an individual's archeus prevailing over invading archei, and disease the consequence of it succumbing. This situation had arisen, van Helmont believed, only after the unfortunate events in the Garden of Eden, since the archei of humans had previously reigned supreme in the perfect state of the naïve world.

Franciscus de la Böe (also known as ‘Sylvius’; Leiden and Amsterdam, 1614-1672) claimed to rely only on experience.49 He condemned ancient and modern dogmatism, reasoning back from observations to causes, but considering causes less reliable than observations. He failed, however, to apply his own worthy precepts when he based faulty reasoning on faulty observations. His physiological interpretation held that the blood brought excessive acid (as lymph and chyle) via the superior vena cava, and alkali (as bile) via the inferior vena cava, to the heart where they mixed, releasing particles of fire to produce heat. Excessive release resulted in febrile diseases with consequent expansion of the blood volume. This made the heart speed.50 Digestion, he thought, relied upon fermentation that resulted from mixing food with water, air, and heat; and from the interaction of ‘acid’ pancreatic juice with ‘alkaline’ bile. This led to

49 King (1970), pages 93-112.
absorption of the fluid components as chyle, and excretion of solids in the faeces. Disturbances of these proposed chemical mechanisms led to gastrointestinal diseases.

Thomas Willis (Oxford and London, 1621-1675) was an iatrochemist who sympathized with atomic theory (although he worried about a lack of empirical evidence to support it), and promoted a physiological and pathological role for fermentation. He identified the latter as the motion of particles within a body that changed the body to a final and manifested (‘perfected’) form. He suggested that the wind carried inanimate particles that could enter a body and become ferments. Fermentation occurred especially in the intestine, and had a close relation to putrefaction. Disease, he believed, resulted from fermentation, and death from immobility of particles when fermentation ceased. He denied Aristotle's explanations, claiming that fire degraded matter into five ‘elements’ (mercury or ‘spirit’, sulphur, salt, earth, and water: that is, the four elements of the Aristotoleans and the three principles of the alchemists).

The profound weakness of all the iatrochemical theories of disease was that each of them contained components that lacked empirical justification.

50 Sylvius claimed that the ventricle of the heart was the hottest part of the body and so must contain fire.
51 King (1970), pages 93-112.
This then immediately opened them to criticisms similar to those directed at classical Dogmatism. They had to compete, furthermore, with an alternative view that had greater empirical justification and that was arising simultaneously as a result of the revival of anatomical dissection that occurred in early Renaissance Italy. Physicians in Salerno, Bologna, Venice, and Montpellier had for some years performed autopsies, but the publication in 1543 by Andreas Versalius of an anatomical atlas based on dissections he had undertaken at Padua revolutionized understanding of human anatomy. Its brilliant illustrations emphasized the machine-like integration of bones and muscles. That insight correlated with contemporary developments in mathematics, astronomy, and physics—several of them also associated with Padua. The thrust of this huge body of work was to characterize the whole universe as a system that functioned according to physico-mathematical rules. Several physicians

55 These included Copernicus’s theory that the earth moves around the sun (1543); Kepler’s laws of planetary motion (1596 and 1619); Harvey’s theory of the circulation of the blood (1628); Galileo’s observations on celestial movement (1632); Descartes’ linkage of physics, mathematics, philosophy, and physiology (1630-49); and Newton’s theories of calculus and gravity (1687). Copernicus was a Pole who had studied in Padua; Kepler, a German, dedicated his book to King James I of England, to whom Harvey had recently become physician; Harvey had studied in
(who became known as *iatromechanists*, or *iatrophysicists*, or *iatromathematicians*) perceived a correlation between the universe and each living body within it that led them to liken each organism’s body to a machine in which disease resembled mechanical failure.\(^{56}\) The supreme iatromechanist was a Prussian, Friedrich Hoffmann (Halle, 1660-1742). His *Fundamenta Mediciniae* of 1695 expounded his mechanical views, but simultaneously demonstrated his inability to provide a comprehensive and persuasive account of disease on physical principals alone.\(^{57}\) He believed that:

1. Life and health consist in the proper functioning of the vital actions; disease in the distorted or diminished functioning; and death, in its total destruction.
2. When the vital actions, involved in the movement, sensation, and nutrition of the machine, are injured, then disease is said to be present.

He claimed that the ‘principal and efficient cause of disease’ was damage to the solid parts of an organism through impaired mixing of its fluid particles by, for example, blockage of tubules or pores in the viscera. Disproportions could manifest as warmth with salinity and sharpness, or as coldness with acidity and melancholy. Some diseases were contagious, others not; some malignant, others benign; some acute, others chronic; some ‘essential’, others symptomatic; some endemic, others epidemic;

---

\(^{56}\) They included Giovanni Borelli (1608-1679), Lorenzo Bellini (1643-1704), Giorgio Baglivi (1668-1707), and Archibald Pitcairne (1652-1713). See King (1978), pages 95-124.
some recurrent and others not. An increase in the motion of the fluids would cause an acute illness; a decrease would cause a chronic illness. Malignant illness would result from effects on the whole organism, whereas uneven circulation of the blood would produce local involvement. Viscid or irritant particles, violent motion, and spasmodic contractions could all block vessels, resulting in increased flow elsewhere and inflammation. ‘Corruption’ of stagnating fluids (‘humours’) could produce plethora, whereas disproportions of fluids could cause further obstructions. Heredity, age, sex, climate, winds, air, seasons of the year, foods, drink, sleep, exercise, and emotional disturbances all predisposed to various diseases.

Hoffmann’s beliefs about fevers reflected a view that ‘all contagious disease is propagated through the air’. Obstruction of the vessels and pores of the skin by viscid material carried in the blood would prevent the secretion of sweat and so cause fever. Blood would then divert to the centre of the body, producing a rapid pulse. Poorly digested food could adhere to the intestinal coat, corrupt it, enter the chyle, travel through the vessels, obstruct those in the skin, and cause fever. ‘In our own microcosm, as in all of nature, all things take place by a definite number, weight, measure and time’, so fevers could be continuous, quotidian, tertian or quartan. He attributed smallpox to the presence of an impure

caustic salt of a particularly sharp type. He believed that these impurities circulated in the blood, stimulating the heart and arteries to violent motion, and 'when a particular ethereal air flows in, nature strives to liberate itself by a certain motion similar to fermentation'.

Hoffmann—despite alluding to mathematics and describing the body as a machine—required help from questionable Galenic ideology and hypothetical properties of fluids and chemistry in his attempt to explain disease. This staunchest iatromechanist could not, therefore, make his theory of disease survive without a chemical component, yet the problem for iatrochemical theories was their reliance on speculative humoral or alchemical allusions.

Robert Boyle had, in fact, already identified a possible escape route from this impasse faced by the iatromechanists in *The Sceptical Chymist* that he published in 1661, although his views failed to gain immediate acceptance. He had noted that no one had ever created 'elements' from the substances that the alchemists claimed contained them, and he denied the inevitability of such a discovery. He had promoted, instead, careful chemical analysis to elucidate normal and abnormal states of organisms, accepting the merits of both mechanical and chemical ideation, provided that they rested firmly on observation. He had thus sought to harmonize the opposing chemical and mechanical theories.

---

58 Hoffmann (1971), pages 53-54.
Herman Boerhaave (Leiden, 1668-1738) was an outstanding person who followed Boyle’s lead, by seeking to fuse the competing views. His efforts met with a reasonable degree of success. Boerhaave was influenced both by the iatriochemist, Sylvius, and the iatromechanist, Pitcairne; and he sympathized with the confusion created for medical students by the conflicting Greek, Arabic, chemical, and mechanical explanations of disease. He recommended studying all the theories; performing experiments on living animals; applying the principles of hydraulics, mechanics, and chemistry to biology; and recognizing the weaknesses in Descartes’ influential physical analyses (persuasive as were the mathematical ones).  

He also commented acerbically that ‘One cannot act on an archeus or principle of thought’. He believed instead that the principles underlying disease are usually obscure, but that humans can observe these and learn about them from their senses. He considered chemical processes as important, the existing theories of iatrochemistry as incredulous and unsubstantiated, and the investigation of chemical processes to deserve as rigid criteria as did the investigation of physical ones. He commented about health and disease:

When the body is able to exercise its functions with ease, pleasure and a certain regularity, we speak of health; if it does not, we speak of disease. So disease is a condition of the living body, in which one or more of its functions is not

---

60 Lindeboom (1986).
61 Lindeboom (1986), page 271.
62 Lindeboom (1968), page 278.
properly exercised. Essentially disease is disturbed movement.

This idea of functional failure was one that was to recur through into many late 20th Century attempts to define disease. He argued, however, that whereas weakness or stiffness due to age in the ‘fibres’ of the body causes some diseases, other diseases result from disturbances of fluids (humorum morbi). He speculated that microscopic substances dissolved in the blood became abnormal in disease, perhaps because poor nutrition made the blood acid or alkaline, whereas inflammation was ‘a process in which solids and fluids are both involved’.  

Boerhaave’s views failed, nevertheless, to achieve universal acceptance, facing in particular a challenge from the idea of vitalism—of damage to a vital force that supposedly suffused organisms with a life-giving spirit. A waning spirit would cause disease, loss of it death. Its intellectual origins lay in the Pneumatist sect of classical Rome, and in Galen’s eclecticism that had included a proposal that the liver imbued the blood with natural spirits, the left ventricle of the heart created vital spirits that the brain converted into animal spirits, and a pneuma pervaded the whole organism. Francis Glisson (London, 1597-1677) had similarly argued that all living

---

63 Lindeboom (1968), page 280.
tissues have the property of irritability\textsuperscript{64}, an idea that had also appealed to van Helmont. Georg Stahl (Halle and Berlin, 1669-1734), a contemporary of Boerhaave, suggested that \textit{the soul}, akin to van Helmont’s chief archeus, imbued a person’s body and struggled against a hypothetical substance, \textit{phlogiston}, to prevent disease, producing alterations in vascular tone that manifested as symptoms.\textsuperscript{65} H.G. Gaubius (1705-1780), who succeeded Boerhaave in Leiden, proposed somewhat similarly, that each disease was a distinct entity that arose from a hypothetical seed.\textsuperscript{66} The seeds resembled the \textit{archei}, albeit with a more strongly ontological tone that avoided some of the mysticism inherent in van Helmont’s iatrochemistry, but similarly lacked objective evidential support. Another successor of Boerhaave was the Leiden-trained Swiss aristocrat, Albrecht von Haller (1708-1777) who promoted nerves, muscles, and movement as key factors in health; and muscular irritability in disease. He cited in support, however, some inconclusive experiments and eventually merged his views with those of the vitalists.\textsuperscript{67}

\textsuperscript{64} Glisson was President of the Royal College of Physicians of London and a founder of The Royal Society. He stated his arguments in \textit{De Ventriculo et Intestinis}, published in 1677.
\textsuperscript{65} King (1978), pages 129-130.
\textsuperscript{66} Long (1928), pages 121-122.
\textsuperscript{67} King (1978), pages 175-177. Another group that arose about this time, the animists, closely resembled the vitalists in believing that some factor acts blindly and unconsciously in the bodies of healthy organisms and is damaged in disease.
Several other people incorporated vitalist overtones into their theories of disease at about that time. Théophile de Bordeu (Montpellier and Paris, 1722-1776) speculated that the brain, heart and stomach (the Tripod of Life) are crucial to survival, with each area in the brain producing a vita propria that flowed through the nerves to the glands triggering these to secrete substances into the blood that integrated the functions of other organs. Diseases (cachexias) developed when the secretions initiated a process of irritability followed by coction and later by crisis. The cachexias would then purportedly produce manifestations in the external secretions (such as urine, sweat, bile, and mucus) and the pulse.68 Paul-Joseph Barthez (Montpellier and Paris, 1722-1776) also promoted this system and introduced the term vitalis agens (vital principle).

Marie-François-Xavier Bichat (Paris, 1771-1802) proposed, in contrast, that human organs contain twenty-one types of tissues, each with a separate vital principle that gave it contractility and sensibility.69 Disease, followed by death, occurred when the vital principles failed. His follower, Johannes Müller (Coblenz and Berlin, 1801-1858), a mystic, a sceptic, and ‘the founder of German physiology’, later claimed that vital processes exist, but are beyond the ability of humans ever to understand.70 That situation typified the weaknesses of vitalism: its anti-intellectualism that

70 Garrison (1929), pages 451-452.
diminished its attraction to anybody who genuinely sought knowledge, and its belief in forces that by definition nobody could understand. Vitalist theories all benefited from the inability of materialists to provide persuasive explanations of disease, although this alone hardly entitled their explanation to prevail by default.\textsuperscript{71} The major criticism that they faced was the obvious one that nobody had ever succeeded in identifying the vital spirit, either chemically or physically. Some of its disciples, furthermore, did little to advance its cause.\textsuperscript{72}

\textsuperscript{71} A point made cogently by J.S. Haldane (1921), pages 17-18.
\textsuperscript{72} The 18\textsuperscript{th} Century Scot, John Brown (see Brown, 1795), was an outstanding example of these. A man who antagonized many who met him, Brown charmed with his ideas others who merely read his writings. He consequently had few followers in Britain, but many in Germany, Austria, France, and North America. Benjamin Rush, in the newly independent United States, was one staunch acolyte. Brown’s ideas, furthermore, so attracted the students of Göttingen that they rioted for two days in 1802 in support of him until the authorities summoned the cavalry to re-establish order. Brown, despite following philosophically the tenets of Asclepidean Methodism (see Garrison, 1929, pages 314-315) gained little credibility from its venerable position because of the practical consequences of his theory. He claimed that life depended on excitability. An organism that had this in moderation was healthy, whereas excess or deficiency of it produced disease. He consequently labelled diseases as sthenic or asthenic. Death occurred either when stimulation became so great that the organism’s excitability increased to levels of violence, or became so weak that it had exhausted its excitability. Therapeutic principles followed from these physiologic principles: when weakness occurred, stimulation should help; when violence developed, sedation should assist. He proposed alcohol and opium as useful medications with which to produce these respective results. His opponents noted that he so believed in his system that he treated himself liberally with the therapies that it suggested, and perhaps demonstrated by his early death its possible implications. His system gained little support from subsequent commentators (see Long, 1928, pages 122-123). The reason for its attraction to the students of Göttingen was, however, clear.
Vitalism was, indeed, merely one of the many theories of physiology and disease that arose in Europe between the 16th and 18th Centuries, each of them proposed by individuals who sought to understand the ways in which living organisms functioned and the ways in which their structures and functions failed. Curious as many of the theories now appear, their instigators were without exception conscientious people who made genuine efforts to understand and explain the phenomena that they observed. The manifestations of disease were obvious to them, although the causes generally remained obscure. They mostly directed their efforts towards speculating about processes within the organisms—usually humans—that they studied. They occasionally drew conclusions regarding causes, but they more often conjectured about processes within the organisms that they perceived as sufficient in themselves to explain the manifestations. The invariable weakness of their conjectures was that they lacked objective information upon which to base conclusions, thereby losing them the necessary breadth of support that would have fostered general acceptance. This in turn led to rejection and widespread scepticism. The immediate consequence was that by the early 19th Century the exercise of proposing general theories of disease had become an unattractive one. It was not indeed until relatively recently, as many more data have become available, that an interest has arisen in venturing back into this exercise. The present thesis joins that movement by advancing a theory to explicate and define disease, and in doing so it
attempts to embrace some objectively founded implications of the 16\textsuperscript{th} to 18\textsuperscript{th} Century experience.

\section*{§1.3 THE 19\textsuperscript{th} CENTURY: STRUCTURE AND CAUSATION}

Four movements that had their genesis during the late 18\textsuperscript{th} and early 19\textsuperscript{th} Centuries profoundly affected ideas about disease. These were the systematic classification of knowledge, the development of compound microscopes, the emergence of microbiology, and the development of chemical analysis.

The systematic classification of general knowledge was a topic of particular interest to several French and Scottish encyclopedists during the 18\textsuperscript{th} Century. They and others gave detailed attention to the subjects of botany, geology, chemistry, and medicine, with the result that classifications of disease soon started to appear. Historians usually regard Thomas Sydenham (1624-1698) as the harbinger of medical classification and nosology. A British empiricist, he emphasized practical experience and personal observation as the basis of medical understanding. He described several distinct diseases, principally fevers.\footnote{See Swan (1763). Sydenham described such fevers as intercurrent fever, scarlet fever, pleurisy, ‘bastard peripneumony’, erysipelatous fever,} Many clinicians in London and elsewhere followed his example of observing patients during life, performing autopsies on them after death, correlating their clinical and
pathological findings, and writing descriptions that avoided theoretical speculation. They often also named disease entities based upon their clinico-pathological observations. Disease, as a result, changed from being a single undifferentiated phenomenon distinguished only by variable external symptoms into diseases, a series of discrete patterns with distinctive names that affected groups of people or animals or plants.

François Boissier de Sauvages (Montpellier, 1706-1767) was, in 1763, the first person to publish a comprehensive classification of diseases. He based his ten main classes, with divisions and subdivisions of each, entirely on symptoms. Several other authors, all of whom relied on symptoms, promptly followed, until Mason Good proposed an anatomical basis for classification in 1817. His approach, however, outlived its

quinsy, cholera, measles, and small pox; and such other conditions as gout, rheumatism, ‘stubborn itch’, St Vitus dance, dropsy, and apoplexy.

They included John Fothergill (1712-1780, who described migraine, trigeminal neuralgia, and diphtheria), William Heberden (1710-1801, who described angina pectoris, rheumatoid nodules, and chicken pox), William Hunter (1718-1783, who described congenital heart disease, and retroversion of the uterus), William Charles Wells (1757-1817; rheumatic heart disease, and kidney disease), R.-T.-H. Laennec (Paris, 1781-1826, tuberculosis), Richard Bright (1789-1858; Bright’s disease or glomerulonephritis), Thomas Addison (1793-1860; Addison’s disease or adrenal insufficiency), Robert Graves (Dublin, 1796-1853; Graves’ disease or thyrotoxicosis), and Thomas Hodgkin (1798-1866; Hodgkin’s disease or lymphoma).

Boissier de Sauvagnes (1763).

His classes were superficial affections, inflammations, fevers, convulsions, breathlessness, paralyses, painful diseases, madness, discharges, and wasting diseases.

They included the Swedish naturalist, Carl Linné (Linnaeus) in 1763, William Cullen of Edinburgh in 1769, Vogel (1772), MacBride (1772),
usefulness in the mid-19th Century when an International Statistical Congress resolved to develop an improved and uniform system. A version devised by William Farr of London eventually gained general assent, based upon groupings of epidemic, constitutional, local (arranged anatomically), developmental, and traumatic diseases.\textsuperscript{78} The Royal College of Physicians of London started the regular publication of a \textit{Nomenclature of Diseases} in 1857, moving from a symptomatic basis of classification to an anatomical one by 1885, with some aetiological entries. The cumulative effect of these developments was to reinforce the idea of diseases as physical entities. An increased understanding of causation of types of disease that thereafter emerged resulted in increasing aetiological emphasis, a trend that the American \textit{Standard Classified Nomenclature of Diseases} and the World Health Organization’s \textit{International Classification of Diseases} also reflected.\textsuperscript{79}

The advent of compound microscopy in the 1830s was the reason for the dramatically increasing importance of anatomical description. Matthias Schleiden (Berlin, 1804-1881) had suggested in 1831 that plants are aggregations of microscopic cells, a proposal that Theodore Schwann (Berlin, 1810-1882) extended by identifying the cell as the basic unit of all

---

Sagar (1776), Pinel (1789), Erasmus Darwin (1796), Crichton (1804), Parr (1809), Swediaur (1812), Young (1813), Mason Good (1817), and Hosack (1821).

\textsuperscript{78} World Health Organization (1957), volume 1, pages vii-xli.

\textsuperscript{79} This commenced publication in 1932.
biological tissues upon the survival of which an organism’s life depended.\textsuperscript{80} Johannes Müller then started the real move to histology in 1838 when he described the microscopy of tumours, whilst Jacob Henle (Berlin, 1809-1885) not only classified tissues, but also demonstrated many abnormalities in them. Rudolf Virchow (Berlin, 1821-1902), who had launched his career by successfully attacking a humoral theory of disease proposed by the great Viennese pathologist, Carl Rokitansky\textsuperscript{81}, identified the cellular basis of inflammation, distinguished benign from malignant tumours, and popularised several terms that soon became commonplace in biology.\textsuperscript{82} He expounded a purely clinico-pathological (or anatomical) viewpoint of disease, and regarded diseases of plants and animals as entirely analogous to those of humans. Disease to him was abnormality of cells and the clinical manifestations that this generated. Despite lacking interest in causation, linking his theory of inflammation to vitalism, rejecting Darwin’s theory of evolution, denying that bacteria play any role

\textsuperscript{80} Schwann (1839 and 1847).
\textsuperscript{81} Rokitansky (1804-1878) was, as a doyen of Viennese medicine, an ardent clinico-pathologist who personally performed 30 000 autopsies and had access to the reports on 40 000 more. He lost all credibility, however, when he speculated about unverifiable dyscrasias as a mechanism of disease and failed to defend his views successfully against a mocking attack by the youthful Virchow. See Rokitansky (1846 and 1849-1854) and Miciotto (1979).
\textsuperscript{82} These included parenchymatous inflammation, thrombosis, embolus, leukaemia, and amyloidosis. See Virchow (1860).
in disease, and claiming that microbiology would provide no help in understanding it, he made huge intellectual advances.\textsuperscript{83}

Virchow's scepticism failed to prevent microbiology from emerging as a powerful factor in the understanding of disease in the late 19\textsuperscript{th} Century. People recognized that any persuasive account must explain the long history of observations of contagion and epidemics that dated from pre-historic times. Boccacio and others had described the plague that swept through 14\textsuperscript{th} Century Europe, leading Frascatorius to suggest in 1546 that *seminaria* ('seeds' or 'germs') of disease travel in the air and in fomites.\textsuperscript{84} He had noted that some infectious diseases cause immunity, and often only attack particular organs, speculating that the germs of consumption (tuberculosis) lodge preferentially in the lungs.\textsuperscript{85} The observation, using a simple microscope, by Antony van Leeuwenhoek (Delft, 1632-1723) in 1683 of motile *animalicules* in washings from human teeth was probably the first sighting of micro-organisms, and prompted the suggestion in 1723 that similar creatures might cause tuberculosis.\textsuperscript{86} Agostino Bassi (Lodi, Italy; 1773-1856) was, however, the first person to confirm the association of a particular organism with a specific disease when, in 1835-36, he

\textsuperscript{83} Perhaps there was an element of German nationalism—a powerful factor in the Teutonic consciousness of the mid-19\textsuperscript{th} Century—in his strong views: vitalism had a long history of appeal in the Germanic world, Darwin was an Englishman, and French and Italian scientists were particularly involved in developing microbiology.

\textsuperscript{84} He entitled his book *De conagione et contagiosis*.

\textsuperscript{85} Bullock (1960), pages 3-19.
showed that a microscopic fungus caused muscardino disease in silkworms, that it passed from animal to animal by physical contact and on food, that fungal seeds disseminated the disease, and that they killed the host before maturing.\textsuperscript{87}

Bassi’s observations, occurring at much the same time as the development of the compound microscope, stimulated increasing interest in microbiology and the role of germs in human, animal, and plant diseases. Investigators soon identified yeast as a micro-organism. They also associated bacteria with putrefaction and with puerperal fever, vibrios with human syphilis, Trichomonas parasites with some genital discharges, and ringworms with skin lesions. Louis Pasteur (Paris, 1822-1895) demonstrated that bacteria caused fermentation, thereby overthrowing previous theories that this was a purely chemical process. He went on in 1866 to show that other strains could release toxic chemicals that spoiled wine and beer. The observation that unfiltered discharges from suppurating gunshot wounds were infectious, but that appropriately filtered discharges were not infectious, explained the microbiological basis of septicaemia—a condition with obvious human counterparts that investigators could induce by injecting putrid substances into animals.\textsuperscript{88}

\textsuperscript{86} Marten (1723), pages 40-41.
\textsuperscript{87} Bassi (1835-1836).
\textsuperscript{88} Pierre Piorry (Brussels, 1794-1879) named \textit{septicaemia} in 1837; Bernard Gaspard (Paris, 1788-1871) first induced it in animals; and Edwin
Robert Koch (Göttingen and elsewhere, 1843-1910) finally codified the germ theory of infection in 1877 by describing simple experiments that he had performed on animals infected with anthrax. He and others identified many other bacteria that they associated causally with various human and animal diseases.

The bacterial genesis of many febrile illnesses became apparent during the late 19th Century; however, certain other contagious and febrile diseases such as smallpox, measles and haemorrhagic fevers remained enigmatic. A satisfactory explanation had nevertheless to account for these and the Russian botanist, Dmitri Ivanovski (St Petersburg, 1864-1920), made the crucial observation in 1892 that elucidated them. He noted that he could not identify bacteria or fungi on plants that suffered from tobacco mosaic disease, but that inoculation of a cell-free filtrate made from the leaves of diseased plants would induce disease in healthy plants. He concluded that ‘a filterable virus’ must cause the condition.

Persuasive evidence that viruses could also infect animals came with their

---

Klebs (Würzburg, 1834-1913) described the infectivity of the unfiltered discharges in 1872.

89 Koch (1877). He used tissue stains and a compound microscope with oil immersion lens to gain insights previously unavailable with weaker equipment.

90 Ivanovski (1892). People had previously indiscriminately used the term *virus* (derived from the Latin word for *poison*) before the late 19th century to denote any micro-organism, but it developed special implications after Ivanovski introduced the phrase ‘filterable virus’.
demonstration in cases of foot-and-mouth disease in 1898\textsuperscript{91} and in humans with yellow fever in 1902.\textsuperscript{92} The discovery by Frederick Twort (London, 1877-1950) in 1915 of bacteriophages—viruses that could parasitise and damage bacteria—led to a realisation that organisms responsible for diseases in one species could themselves succumb to diseases caused by a third species.\textsuperscript{93} This then created yet a further intellectual dimension with which any general theory of disease had to wrestle, especially as viruses remained invisible until the development of electron microscopy in the mid-20\textsuperscript{th} Century. Microbes, it seemed, could cause many diseases in hosts of many kinds, but these microbes were themselves susceptible to disease. An accommodation of this observation in explicatory theory of disease had become necessary.

The development of clinical chemistry was another circumstance that ranked in importance with the development of microbiology as a contributor to the understanding of disease in the 19\textsuperscript{th} Century. The invention by French scientists during the Revolutionary and Napoleonic eras of a sophisticated weighing balance, the graduated measuring cylinder, the burette, and the metric system of weights and measures enabled chemists to undertake systematic analyses of the chemical content of urinary stones and thereby to start replacing the mysteries of

\textsuperscript{91} Loeffler and Frosch (1898).
\textsuperscript{92} Carroll (1902).
\textsuperscript{93} Twort (1915).
alchemy with some objective certainty. They moved on to analyse the chemical content of many bodily fluids and tissues. Physicians soon noted that the urine of some patients who suffered from generalized oedema precipitated upon heating, indicating the presence of protein, and that at autopsy such people had evidence of kidney disease.\textsuperscript{94} These observations shed light on the processes that underlay the developing clinico-pathological ideas about disease. Advancing methods of analysis provided a basis for understanding the inorganic chemistry of living organisms, whilst organic chemistry developed later and revealed the nature of even more complex biological compounds.\textsuperscript{95} A spectacular series of technical developments particularly aided these advances.\textsuperscript{96} The accurate analysis of chemicals led to the development of ideas about failure of function in various organs. Detailed understanding of this crucial component of many disease processes required more refined methods than the mere analysis of the plasma level of some individual chemical. Dynamic assessments of renal function, liver function, and lung function, all based on chemical analyses, but enhanced by physical methods and correlated with clinico-pathological findings, led eventually to the identification and improved understanding of such conditions as uraemia,

\begin{footnotesize}
\textsuperscript{94} Wells (1812).
\textsuperscript{95} The inorganic substances included glucose, urea, creatinine, uric acid, carbon dioxide, oxygen, calcium, iron, chloride, sodium, and potassium; whilst the organic ones included proteins, cholesterol, and haemoglobin.
\textsuperscript{96} These included the introduction of plungers for hypodermic syringes, colorimetric analysis, centrifugation, quantitative micro-methods, spectro-
\end{footnotesize}
diabetes, and cirrhosis. A persuasive definition of the nature of disease had to account for all these aspects.

§1.4 THE 20th CENTURY: PATHOGENETIC MECHANISMS

The scientific developments that occurred during and shortly after the 19th Century set the scene for the remarkable burgeoning of research into particular diseases that was to occur during the 20th Century. Medical science, veterinary science, and agricultural science all developed numerous sub-specialties that flourished as they developed new insights. Five of these aspects in particular deserve detailed examination because of their implications for subsequent attempts, including the present one, to explain and define the general idea of disease. These were immunology, molecular biology, endocrinology, genetics, and psychiatry.

People had been aware of immunity since classical times. Edward Jenner (Berkeley, England; 1749-1823) stimulated a great surge of interest in the subject of immunity when he developed vaccination against smallpox in 1798 by infecting potential victims with the mild, but immunologically protective disease of cowpox. Louis Pasteur later noted that certain physical insults to bacteria weakened their ability to cause

analysis, the photoelectric colorimeter, ion exchange chromatography, electrophoresis, and radioimmunoassay.

97 See, for example, George (2002).
98 Thucydides (c. 460-c.400 BC) had mentioned it in his history of the Peloponnesian War.
disease, even though they maintained their ability to induce immunity. He found that injections of these *attenuated* organisms provided protection against rabies. The discovery in the late 19th Century that bacteria can produce substances (later called *exotoxins* and *endotoxins*) that will induce the production of neutralizing factors (*antitoxins*) in animals into which they are injected led to the development of the idea of a *humoral mechanism of immunity* that enabled the production of effective antitoxins against diphtheria and tetanus.\textsuperscript{99} This contrasted with the *cellular mechanism of immunity*, first recognized in 1882, in which white blood cells attack and phagocytose foreign particles that gain entry into a living organism.\textsuperscript{100}

Advocates of the two mechanisms of immune protection against disease, humoral and cellular, initially saw themselves as rivals. The discovery in 1894 of a circulating chemical factor, *complement*, that contributed to interactions between *antigens* and *antibodies* advanced the claims of the humoralists; but the demonstration in 1903 that another serum factor, *opsonin*, enhanced the efficacy of cellular mechanisms of immunity

\textsuperscript{99} Emil Behring (Berlin, 1854-1917) headed the team of scientists responsible for much of this work (see Behring and Kitasato, 1890; and Behring, 1890). The term *humoral* struck an obvious verbal association with the Dogmatist theory of classical Rome.

\textsuperscript{100} The eccentric Russian zoologist, Ilya Metchnikoff (Messina, Italy; 1845-1916) observed in 1882 that cells in an avascular transparent starfish larva attacked a rose-thorn inserted into it. He also recognised that humans use the same mechanism for protection against bacteria. See
resolved the tension between the two schools. It rapidly became clear that any persuasive explanation of disease would have to encompass both of their sets of ideas. The demonstration by chemical analyses performed in the 1930s and 1940s that antibodies are immunoglobulins produced by plasma cells\(^{101}\) encouraged immunologists to examine the molecular mechanisms that underlay such diseases as allergy, anaphylaxis, and serum sickness. The perplexing problem remained, however, of explaining why individuals fail to produce destructive antibodies against their own antigens: in effect, why everyone does not succumb promptly to immunologically generated diseases. Frank Macfarlane Burnet (Melbourne, 1899-1985) successfully explained this in 1957 when he realized that antibody-producing lymphoid cells and their progeny (‘clones’) cannot subsequently sense antigens that they had met during their foetal life; a person’s own proteins are such antigens, so people develop tolerance to their own constituent proteins.\(^{102}\) Advances such as this emphasized the importance of the immune system in those diseases where tolerance does not exist. They also focussed attention upon the protection that it potentially offers against other diseases. Its overwhelming importance in maintaining the viability of humans and animals then reinforces awareness of the need to accommodate its complex ramifications in any general explication and definition of disease.

---

\(^{101}\) Bing and Plum (1937); Fagraeus (1948).

\(^{102}\) Metchnikoff (1884). Robert Koch identified the phagocytosis of bacteria by white blood cells.
A detailed exploration of the mechanisms underlying immune function had to await the arrival of improved methods in molecular biology. This did not occur until the final two decades of the 20th Century. The discovery then of the cytokines, a unique series of circulating chemicals produced under genetic control that carry signals between cells and regulate their response to injuries such as infection, was one of the great achievements of the molecular biological revolution.\(^{103}\) These substances fall into three major categories.\(^{104}\) The haematopoietic cytokines (interleukins) act on progenitor cells in human and animal bone marrow to stimulate production of red cells, white cells and platelets. Deficiencies of them can cause diseases such as anaemia to develop. The pro-inflammatory cytokines protect an organism against invading pathogens by enhancing cellular and humoral immune responses. Excessive production of them can simultaneously provide such protection, but do so only at the price of creating overt manifestations of disease such as fever, lethargy, aching muscles, and hypotension. Deficiency of them can expose an organism to the dangers of invasion. The anti-inflammatory cytokines (interferons) activate feedback loops that control the production of pro-inflammatory cytokines, so their role in disease is to assist in maintaining an auto-

\(^{102}\) The clonal selection hypothesis (Burnet, 1959).

\(^{103}\) Remick (1995). The cells that produce cytokines include immunologically active T-cells, bone marrow stromal cells, monocytes, other leukocytes, platelets, and fibroblasts. See Remick (1995).
regulatory balance in the defence mechanisms that the challenge of an invader often stimulates. Each cytokine acts on physically identifiable and chemically triggered target cell receptors. Factors that an organism recognizes as adverse can thereby activate the cytokine system to produce a hierarchy of chemical messengers that travel to local or distant sites, causing changes there that manifest either directly or indirectly. These interactions thereby make the cytokine system integral to the mechanism of disease generation and manifestation. Their existence and functions require recognition by any explication and definition of disease if this is to command respect at the present time.

The beginning of the 20th Century witnessed a further development that arose out of improvements in chemical analyses. People had recognized the physical existence of such structures as the thyroid gland, pancreas, adrenals, pituitary, testes, ovaries, and parathyroids for many hundreds of years, but failed to understand their functional implications. Albrecht von Haller probably coined the phrase *glands without ducts* in 1766 to describe these organs and Claude Bernard (Paris, 1813-1878) lectured about *internal secretion* in 1855, but it was not until after 1904—when the French histologist, Adolphe Limon, introduced the term *endocrine* and Ernest Henry Starling (London, 1866-1927), invented the word *hormone*—that a realistic understanding of these organs developed. Their functional

104 Examples are *tumour necrosis factor, granulocyte-colony stimulating*
significance became apparent as improved chemical techniques introduced progressively throughout the 20\textsuperscript{th} Century enabled clinicians to assess the effects of overactivity and underactivity of each of them. A different clinico-pathological syndrome correlated in each case with identifiable perturbation of circulating hormone levels. These challenges also required accommodation in any general understanding of disease.

Recently acquired information suggests, however, that the production of hormones is not merely a biochemical process. Their release requires certain stimuli to act on the cells that secrete them. Those stimuli include factors in autoregulatory mechanisms and also components of the cytokine system. Genetic information is additionally necessary for the transcription and encoding of each of the two main types of hormones, the steroids and the polypeptides. This then draws attention to yet another important component of disease processes—the role of genetic information. Observers had long recognized that genetic issues do play significant roles in regard to disease. Many ancient writers had commented upon the propensity of humans, animals, and plants to breed true with regard to both desirable and undesirable traits. The Hippocratic corpus even outlined a theory to explain inherited characteristics, suggesting that each part of the body transmitted its secretion via the

sperm into the offspring. Animal husbandmen for millennia bred from stock with desirable disease characteristics to avoid feeble progeny. Pierre de Maupertuis (Paris, 1698-1759), indeed, provided the first systematic description of an inherited human condition in 1744 when he wrote about several generations of a family whose members had polydactyly (increased numbers of fingers and toes). He concluded that they must inherit this, as familial occurrence was too unusual to occur merely by chance. Information about such diseases accumulated slowly until Joseph Adams (London, 1756-1818) distinguished between acquired and hereditary diseases, and between congenital occurrence of a disease and hereditary predisposition to it. He pointed out that some hereditary diseases do not manifest themselves until later in life and that those that manifest themselves severely in childhood tend to die out quickly. He commented both on the role of inbreeding and the likelihood of de novo occurrence of hereditary diseases by mutation.

Heredity played a crucial role in Charles Darwin’s thinking as he formulated his *Origin of Species*, published first in 1859, in which he commented not only on inherited characteristics in humans and animals, but also in plants including potatoes, hyacinths, and dahlias. He stated

---

105 Perhaps a ‘humour’, although the writer did not use that word. See On *Airs, Waters and Places*, paragraph 14 (Adams, 1939, page 33). Aristotle also examined this view (585b30-586a15; 721b6-721b29-722a16; 769a1-25).
106 De Maupertuis (1744).
explicitly, however, that no one then understood the laws of biological inheritance or why some peculiarities were inherited and others not.\textsuperscript{108} Gregor Mendel (Brno, 1822-1884), about five years later, developed an explanation for the mathematical precision by which diseased peas in his garden bred true, although his report attracted little interest until its rediscovery in 1900.\textsuperscript{109} Chromosomes that Walther Flemming (Kiel, 1843-1905) had discovered in cell nuclei in 1882\textsuperscript{110} explained the intracellular location of the mediators of these inherited characteristics, and by 1909 Archibald Garrod (Oxford, 1857-1936) was identifying the chemicals responsible for damaging tissues in inherited diseases that he labelled as \textit{inborn errors of metabolism}.\textsuperscript{111} This step greatly assisted in developing an understanding of disease processes in general by proving a biochemical link between the genetic programme and the clinical manifestations.

\textsuperscript{107} Adams (1814).
\textsuperscript{108} Darwin (1860), pages 12-13.
\textsuperscript{109} Mendel (1933). Hugo de Vries (Amsterdam, 1848-1935), Carl Correns (Tübingen, 1864-1933), and Erich von Tschernak (Vienna, 1871-1962) rediscovered it almost simultaneously. Translations of their papers and that of Mendel appeared as a supplement to \textit{Genetics}, 1950, volume 35. Mendel, a self-effacing monk, who lived in the quiet seclusion of an Augustinian monastery in Brünn in Austrian Moravia, systematically bred a plot of garden peas, identified why some had smooth surfaces and others wrinkled, and took the crucial step towards illuminating the ‘dimly seen laws’ (to quote Darwin) that controlled this phenomenon. He noted that his peas reproduced in certain mathematically distinct patterns over generations. He formulated explanations for this and reported his views to the local Natural Science Association in 1865.
\textsuperscript{110} Flemming (1882).
\textsuperscript{111} Garrod (1909). The diseases were alkaptonuria, albinism, cystinuria, and pentosuria.
Increasing knowledge about many diseases with an inherited basis accumulated throughout the early 20\textsuperscript{th} Century and long guided understanding of animal and human genetics. Plant geneticists—whose manipulations had profound economic implications—also made many theoretical advances as they sought to breed disease-resistant strains. Entomologists, such as Hermann Muller (New York, 1890-1967) who observed in 1927 that X-irradiation increased the rate of mutation in the \textit{Drosophila} fruit fly, similarly contributed by demonstrating that environmental injuries could cause genetic damage.\textsuperscript{112} These insights, like each of the other technical developments previously mentioned, required accommodation in the explication, and in any successful definition, of disease as a whole. That exercise was assisted as the genetic origin of many clinico-pathological conditions became apparent during the 20\textsuperscript{th} Century, for which chemical investigations then elucidated responsible mechanisms. The subsequent mid-century observation that desoxyribose nucleic acid (DNA) is a transforming factor in bacteria, the one gene-one enzyme hypothesis, and the identification of a functional model of the DNA molecule, were important consequences as was the recently completed Human Genome Project that identified the whole genetic code of human beings.\textsuperscript{113}

\textsuperscript{112} Muller (1927).
The psychiatric and social aspects of disease were another area—initially apparently quite unrelated, but now increasingly accepted as closely related to those of chemistry and genetics—in which major insights developed during the 20th Century. Knowledge about various clinico-pathological syndromes affecting the brain had accumulated in earlier years, but these left much unexplained at the turn of the century. Several factors had contributed to the separate development of ideas about emotional diseases from those of the clearly organic diseases of the brain. These included the long-standing enigma of the interaction between the mind and the body; the frequent need for urgent physical control and sedation of those who suffered from acute mental disturbances; the unavailability of animal models of disease upon which to experiment; the huge relative power that physicians tended to exercise over insane patients; complex relationships that often existed between communities and their psychiatrists; and the fierce opposition that existed between many protagonists of psychiatric disease as a pure thought disturbance and proponents of it as primarily a structural or chemical abnormality.¹¹⁴

The idea of madness was hardly new. Moses had indicated to the Jews of the Old Testament that insanity was a potential punishment that could

¹¹³ Geneticists working on the Human Genome Project in the 1990s sought to identify the locations of all 70 000 or so genes in each human, and to determine the chemical sequences of their three billion or so nucleotides.
¹¹⁴ Shorter (1997).
result from breaking God’s laws.\textsuperscript{115} Physical explanations held greater appeal in the Greco-Roman world than in the Old Testament biblical one\textsuperscript{116}, although moralistic views loomed large again in the medieval and Renaissance periods. The reasons for this perhaps included a dominant religious climate and the absence of a persuasive alternative to replace the declining theory of the humours.\textsuperscript{117} Some people considered that those ‘possessed by the devil’ (in effect, those with psychotic illnesses), as well as others who suffered from chronic physical diseases, had succumbed to the malign influence of witches.\textsuperscript{118} Epidemics of these conditions sometimes swept through closed communities such as convents. Acceptance of such explanations declined throughout the 17\textsuperscript{th} Century\textsuperscript{119}, aided in part by the publication in of a personal account of depression by the Oxford don, Robert Burton (1577-1640) in 1621.\textsuperscript{120} He listed many possible causes of diseases of the mind, ranging from the

\begin{itemize}
\item \textsuperscript{115} See, for example, Deuteronomy, chapter 28, verses 15, 27, and 28.
\item \textsuperscript{116} One example was the Hippocratic treatise entitled \textit{The Sacred Disease} that dealt with epilepsy (Adams, 1939, page 347).
\item \textsuperscript{117} The witchcraft cult developed in most European countries after the 13\textsuperscript{th} Century, two Dominican inquisitors who pursued witches in Northern Germany having vilified it in about 1490 in \textit{Malleus Maleficarum}; ‘The Witches’ Hammer’). Students of the phenomenon seem unable to explain its rise, although Macfarlane (1970) suggests that it might occur in a community during periods of social stress.
\item \textsuperscript{118} Robbins (1959), pages 392-398.
\item \textsuperscript{119} The beliefs of King James VI of Scotland (1566-1625) provide a good example of this decline: he published a book in 1597 in which he railed against the dangers of witchcraft and sorcery, but after becoming King James I of the United Kingdom in 1603 took the opportunity of personally interrogating some alleged witches and realized that malevolent people often used accusations of witchcraft to accuse innocent people falsely, whereupon he became quite sceptical of witchcraft.
\end{itemize}
stars to idleness, from anger to ambition, from gambling to excessive study, and from the loss of liberty to the loss of friends. Little advance thereafter occurred in the understanding of mental illness until the early 19th Century when physicians such as John Haslam (London, 1764-1844) and Philippe Pinel (Paris, 1745-1826) started to distinguish individual clinical syndromes such as that which gained the name of schizophrenia. Physicians later in that century developed an ability to identify people with structural and chemical brain damage associated with neurosyphilis, acute and chronic intoxications, endocrine disorders, and other physical conditions that presented clinically as behavioural disturbance, but this left the major diseases of schizophrenia and manic-depressive psychosis, as well as many minor psychological states, unexplained.

The psychoanalytical proposals of Sigmund Freud (Vienna, 1856-1939) provided one approach to the understanding of these conditions. Finding purely physical explanations of mental diseases unpersuasive, he moved progressively towards emotional ones. He postulated a form of memory that allowed ideas, especially of a sexual nature that an individual acquired in childhood and initially suppressed, later to emerge in disguised form during adult life and manifest as neuroses.\(^{121}\) He subsequently extended his theory by claiming the existence of an id/ego/superego psychological system that contrasted to his previous emphasis on id

\(^{120}\) Burton (1989).
forces alone. He then defined the ego as ‘the coherent organization of mental processes’ that represented reason; the id as the other part of the mind into which that ego extended and which represented passion; and the superego as ‘the higher nature…this ego-ideal…the representative of or relation to our parents’. Instinctual forces, often of a sexual nature, in the subconscious mind under the dominance of the id would drive a person to perform various acts. These modifications moved his theory of mental disease from one of a causal nature (traumatic childhood experiences causing manifestations in the form of fantasies and neuroses) to one of an imbalance nature in which the ego (the conscious state of the person’s mind) resulted from imbalance between the person’s id (their narcissistic impulses) and their superego (demands of their conscience). This manifested as mental health when balance existed, but psychiatric disease when imbalance prevailed. These insights that Freud provided into disease processes posed yet another facet that a successful definition of disease had to accommodate.

Psychiatrists in the first half of the 20\textsuperscript{th} Century increasingly sought to explain the major psychoses, in addition to the minor neuroses, in terms of Freud’s formulation. Further extensions of it gradually obliterated the distinction in the minds of many psychiatrists between mental health and mental disease. Some even came to perceive all humans as mentally

\textsuperscript{121} Freud (1895).
diseased. One eminent American psychiatrist, Karl Menninger, thus wrote in 1956\textsuperscript{122}:

Gone forever is the notion that the mentally ill person is an exception. It is now accepted that most people have some degree of mental illness at some time, and many of them have a degree of mental illness most of the time. This really should not surprise anyone, for do not most of us have some physical illness some of the time, and some of us much of the time?

Many people, however and perhaps unremarkably, seemed disinclined to admit that they had mental illness even for some of the time.

Political considerations also influenced ideas about mental disease in the period between 1930 and 1950. The views of two theoreticians, Karl Binding and Alfred Hoche, had a strong influence in the German-speaking world in the period between the two World Wars.\textsuperscript{123} They abhorred mental disease, suggesting that people often feigned it, and labelling many psychiatric patients (with whom they also lumped the feeble minded, the deformed, and retarded children) as human ballast and empty shells of humanity who were unworthy of life.\textsuperscript{124} Some governments used their ideas after 1933 to justify nefarious political and police actions, the very

\begin{enumerate}
\item[122] Menninger (1977), page 33.
\item[123] Binding was a retired Professor of Law from Leipzig, and former humanitarian, who had suffered profoundly after the death of his own son during World War I. Hoche was Professor of Psychiatry in Freiberg. They combined to write Die Freigabe der Vernichtung Lebensunwerten Lebens (‘The Permission to Destroy Life Unworthy of Life’).
\item[124] Lifton (1986), pages 46-47.
\end{enumerate}
memory of which continues to revolt the vast majority of humans. The flight of Freud in 1938, like that of so many other intellectuals from Central Europe, left the field of psychiatry in the Germanic world to researchers who were seeking clinico-pathological correlations for mental disease. Psychoanalysis lost credibility there just as these same émigré psychiatrists were arriving in the English-speaking world and pushing its accepted explanations in a psychic direction.

The theory of disease upon which psychoanalysis relied appealed to many people. Treatment overshadowed theory, as had so often previously occurred in psychiatry. The assumption prevailed that mental disease existed, but its nature proved too difficult to explain or define in terms that achieved consensus. Those in the West who considered themselves sick flocked to the couches of the psychoanalysts. Not all psychiatrists, however, found the psychoanalytical model satisfying. Others tried physical methods of treatment including pharmacological agents, electroconvulsive treatment, and surgical methods (such as lobotomy). Each of these relieved the symptoms of some seriously affected people, but none relieved the symptoms of all. They suggested, nevertheless, that since they often influenced diseases by physico-chemical methods the

\[126\] These included such drugs as sedatives, alcohol, opium, bromide chloral hydrate, hyoscine, barbiturates, insulin (to produce coma), penicillin (for neurosyphilis), lithium (for manic-depressive psychosis), chlorpromazine and its derivatives, and imipramine and its derivatives.
diseases might have physical, and not merely psychic, mechanisms. Psychiatrists who followed this clue sought further evidence for a physical basis for mental disease. Sir Thomas Clouston (Edinburgh, 1841-1915) had remarked in 1901 upon the tendency of schizophrenia to run in families, in contrast to other mental diseases, although he did not identify that this familial tendency had a genetic basis. An unsubstantiated, but much publicized, claim in 1913 that a factor circulating in the blood mediated schizophrenia influenced thinking towards a physical mechanism, whilst Hans Luxenberger in Munich answered the question about a genetic mechanism when he carried out comprehensive twin studies that strongly suggested an inherited tendency for schizophrenia.¹²⁷ These ultimately triggered a wealth of presently continuing experimentation to identify subtle structural features in the brains of psychotics and to identify specific genes associated with psychoses. This has encouraged a materialist interpretation to gain increasing support in the English-speaking world in the late 20th Century: one that perceives genetic factors as predisposing to structural or chemical mechanisms for psychiatric disease as a prelude to creating the typical clinical manifestations.¹²⁸ The credibility of psychoanalytical explanations started to decline after about 1960 for several reasons. Psychiatrists found that pharmacological and physical methods of treatment offered distinct benefits over psychoanalytical ones to severely troubled people. Evidence

¹²⁷ Luxenberger (1928).
accumulated about a genetic or other physical basis for various diseases previously labelled as psychoses. Internecine disputes occurred between the various psychoanalytical schools. People resented the suggestion that everyone is a little mad and countered by suggesting that perhaps the psychoanalysts should look more closely at themselves. They meanwhile were bringing their own methods into disrepute by entering into political disputes about such conditions as post-traumatic stress disorder and homosexuality. ‘Is homosexuality a disease?’ they asked. They had formerly classified it as such, but changed their stance after a public relations campaign targeted them.129 These factors all contributed to their decline such that by 1990 they had lost credibility for explaining anything other than minor neuroses. The task of explicating psychiatric disease, whether of physical or emotional provenance, nevertheless remained. Some people sought to counteract the declining impact of purely psychological explanations of disease by emphasizing its sociological consequences as its fundamental characteristic. One of these, Talcott Parsons described illness as ‘a socially institutionalised role-type’ in which an individual lost the ability to perform normally.130 Effective as were such approaches to explain some implications of disease, they failed to provide persuasive explanations of why disease occurred or the mechanisms whereby it damaged organisms. They also disregarded the whole issue of

129 Kirby (2000).
asymptomatic diseases. Such weaknesses meant that they failed to gain credence as any more than secondary explanations for the nature of disease.

§1.5 CONCLUSIONS

This brief overview of some of the more important theories of disease that have emerged over the centuries, and of the more recent scientific discoveries that have produced fundamental information that any acceptable theory of disease must encompass, demonstrates the enormous complexity of the topic. It leads one to seek some degree of organization of the theories to reduce them to an intelligible account.

The simplest of the theories are those based upon manifestations, as espoused by the Greco-Roman Empiricists and the early Christians. Disease in these terms is just that which people see it to be; it is nothing more than what those who suffer from it and those who observe it can immediately distinguish. This view has always attracted a strong following, but it ultimately fails to provide much insight into the origin of the manifestations. Later followers of manifestation views included Sylvius and, most importantly, many of the 19th Century clinico-pathologists. Inadequate as were their theories in the overall context, they did demonstrate the great value of careful observation to provide a basis for further investigation.
Many have, over the centuries, found themselves enquiring about the origins of the manifestations that they have observed, with the result that a plethora of causation theories have arisen. These have ranged from a focus upon obvious external attacks (such as the consequences of physical injuries, snake or insect bites) through the Greco-Roman pneumas and Paracelsus's entia to the consequences of infection and genetic variations. The various religions have historically espoused another subclass of causal theories by focussing upon the results of divine anger (as in Mesopotamia, Egypt, and Old Testament Judaism) or inappropriate human desire (as in Buddhism).

A much more complex group of theories are those that have postulated internal phenomena within an organism as the sine qua non of disease. One might describe these as internal mediation (or pathogenetic) theories. They have included such hypotheses and observations as the Egyptian whdw, imbalance of force theories (the classical Chinese yin-yang concept, the Indian Ayurvedic theory, the Greco-Roman Dogmatist theory of the four humours, and the Greco-Roman Methodist theory). Iatrochemical theories (such as that of Paracelsus, van Helmont’s archei, and Willis’s theory of fermentation) as well as iatromechanical theories (such as that of Hoffmann) fall within this category. Other mediation theories were those of the animists (for example, Stahl’s idea of the soul),
vitalists (von Haller’s irritability; as well as Glisson’s, de Bordeu’s, Bichat’s, and Brown’s ideas), and Rokitansky’s dyscrasias. More recently, the enormous volume of information relating to biochemical, endocrine, immune-mediated and molecular biological factors has fallen within this general category of internal mediation ideas.

The baffling complexity of information available about disease, and the apparently irreconcilable inconsistencies between many of the theories developed to explicate it, gave rise at a very early time to the formulation of a sceptical viewpoint that despaired of ever understanding the basis of disease. This school of thought has always attracted a band of followers, as has its antithesis, the eclectic theory, whose many devotees have seen merit in all the theories. Galen was their most influential proponent, followed by Boerhaave. Much contemporary medicine and biology seems instinctively to adopt this approach, shying away from commitment to any particular explicatory formulation. The travails that psychiatrists, in particular, have sustained during the last century typify this situation.

This overview, then, suggests that the strong historical support for many competing theories will require that any acceptable explication and definition of disease must account for the phenomenon in all living organisms, and not merely in humans; that it must deal with aspects of manifestation, of mechanism, and of causation; that it must address the
perceptions of people about disease as well as the reality of the phenomenon; and that it must deal with the implications of disease for society as a whole. Many scientists have struggled with these issues without success, suggesting that a deeper insight is necessary than a purely scientific analysis can provide. It is not therefore surprising that development of an understanding of disease has come in recent times to attract the attention of philosophers of science. This has really only represented a continuation of the efforts of several of the 17th and 18th Century thinkers mentioned above who were themselves as much (or more) philosophers as scientists. The purpose of §2 will therefore be to examine some of the more recent philosophical views that have emerged on the subject.

* * * * * * * * *
2. 20th CENTURY PHILOSOPHICAL IDEAS
ABOUT DISEASE

The physicians and scientists who debated the nature of disease had, during the 19th Century, largely focussed upon clinico-pathological aspects, with a tendency towards the end of that period to examine some causal and pathogenetic aspects. People occasionally, however, alluded to more philosophical issues: thus Sir Clifford Allbutt wrote in 1896 that ‘disease is a state of a living organism...the disease itself contains no elements essentially different from those of health, but elements presented in a different and less useful order’.\(^1\) The possibility of a philosophical analysis started to attract serious attention from thinkers in the mid 20th Century as a wider interest developed in the history and philosophy of science, and especially as people sought to understand the nature of psychiatric disease and to contrast it with somatic disease.

§2.1 SOME PROPOSALS
One of the 20th Century pioneers who thought about philosophical aspects was a physician, A.P. Cawadias. He perceived that an important contrast existed between what he called realist, conceptualist, and nominalist ideas of disease.\(^2\) He suggested that realists perceived diseases as separate

\(^{1}\) Quoted by King (1954).
\(^{2}\) Cawadias (1937).
things in themselves, more local than general, usually having a single cause, and this often having a microbial nature. **Conceptualists**, in contrast, perceived diseases as ideas that existed in the minds of observers, and closely resembled phenomena that the observers recognized. **Nominalists**, again in contrast, perceived diseases as names that identified artificial categories of morbid phenomena. His own position was basically a nominalist one in that he favoured what he called a **Neohippocratic** view of disease. He meant by this that disease is a word used to describe a functional species of combinations of symptoms that occur with some regularity as a result of the mind-body complex facing noxious external agents. He therefore used the word disease to represent the struggle of an organism against external evils. He wrote about the body’s constitution (or diathesis, or pre-disposition), by which he meant an hereditary or acquired transmissible or non-transmissible variation in the structure or function of the tissues that led them to react in particular ways to certain external stimuli. His ideas had definite vitalist overtones in that he postulated that every body is animated by a purposive striving to live, to develop, and to immortalize itself for the attainment of its ends and by procreation, ‘as if’ directed by an inner force that he likened to Hippocrates’ physis and to the vital force of some later physicians. His views were novel and represented a determined effort to engage in a fundamental way with the idea of disease. They proclaimed an unalloyed belief that all disease has its genesis in the impact of disadvantageous
external agents upon an organism. Their major weaknesses were, however, that they expounded a vitalism that was largely discredited, and they drained diseases of their individuality by devaluing the huge scientific advances made during the previous century or more.

Lester S. King was another physician who entered the philosophical discussion in the mid-1950s by asking the question ‘What is disease?’ He considered that defining health as freedom from disease, and disease as an aberration from health, merely produced an unproductive circular argument. Diseases instead represent patterns or relationships, as proposed by Allbutt and Cawadias, rather than existing as things like rocks and trees and rivers. He also followed Allbutt in considering that, because disease interferes with usefulness and usefulness carries a connotation of value-judgement, disease implies something bad, in contrast to health that is good and desirable. He nevertheless believed that the existence of asymptomatic disease meant that whereas a sense of wellbeing correlates with health, lack of it does not necessarily correlate with disease. Abnormality also is not a close correlate of disease, as states of biological abnormality exist that are the antithesis of disease—for example having an intelligence quotient of 180, or being able to run 100 yards in 9.4 seconds. The association, nevertheless, of statistical abnormality with malaise or pain or death does form part of the idea, and ‘deviation from

\[\text{King (1954).}\]
the ideal’ may come closer still. This would account for such facts as that, whilst having 32 intact and well-aligned teeth may be statistically abnormal for an adult, it remains the ideal if the ideal is described as an observed characteristic that is a matter of general desire in a particular population. He also distinguished between disease and the names of diseases. He therefore suggested that ‘Disease is the aggregate of those conditions which, judged by the prevailing culture, are deemed painful, or disabling, and which, at the same time, deviate from either the statistical norm or from some idealized status’. Names of diseases, he then suggested, refer to complex patterns of factors that hang together, that recur, and that people create in their minds to organize their experience and to approximate to reality. King’s ideas so influenced discussion of this topic that The American Psychiatric Association relied upon them as recently as 1987 when it redefined mental disorders. Its statement, paraphrased, identified a mental disorder as a mental condition that causes distress or disability, and is not merely an expectable response to a particular event.\footnote{Wakefield (1992).} The outstanding weakness, however, of King’s definition and its derivatives was their focus upon the manifestations of disease without providing any insight into the origins of these. Disease, on this account, could as much arise by spontaneous generation as by any other means: the definition described the outcome without giving any attention to the causes and mechanics that had led to that outcome.
Sir Henry Cohen (later Lord Cohen of Birkenhead) was a contemporary of King who became interested in the historical evolution of the concept of disease. He suggested that two major traditions of disease concepts have existed—one ontological (perceiving disease as an entity) and the other normative (perceiving disease as an abnormality). Cohen discounted the former tradition on the grounds that it was weak and ‘mind-numbing’. He also criticised the latter on the grounds that normal is a range that varies with such factors as age and sex. He nevertheless preferred the latter, and described deviations from normal as the symptoms and signs; commonly recurring consistent patterns of these as creating symptom complexes (or syndromes); and syndromes dealing with one or more of three aspects of disease (its site; its associated functional disturbances; and causative factors in terms of morbid anatomy, physiology, psychology, and aetiology). Cohen thereby went far beyond King by adding issues of causation and mechanism to his analysis, and furthermore tried to remove any evaluative element from it by predicing it entirely upon deviation from statistical normality. The latter theme had an immediate appeal, the implications of which were to reverberate through successive definitions and philosophical debate for several decades whilst people sought to assess all their advantages and disadvantages.

\[5\] Cohen (1955).
When Owsei Temkin, some ten years after Cohen, examined the division of theories of disease into *ontological* (by which he implied *anatomical*) and *physiological* types he saw weaknesses with both, a need for aspects of both, and some merit on balance in the latter.\(^6\) He thought that the weakness with the ontological theory was that it ended up as a causal agent classification in which huge numbers of diseases reflected huge numbers of causal agents where these were apparent, whilst overlooking other important diseases where they were not. The challenge to the physiological theory was where to draw the line between normal and abnormal. This led him to decide that the question ‘Does disease exist or are there only sick persons?’ is an abstract one to which there is no meaningful answer. Robert Hudson and others later supported his ideas while emphasizing that different people (such as a patient, a physician, an anatomical pathologist, a bacteriologist, and a public health official) often have different concepts of the same disease process, and that concepts of disease have changed markedly over centuries and over cultures.\(^7\)

Defining ‘normal’ causes major problems, especially in the case of mental disease, an aspect that also later troubled Ruth Macklin because of

\(^6\) Temkin (1963).

\(^7\) See Hudson (1966). Hudson re-emphasized this point in a later paper (Hudson, 1977) entitled ‘How Diseases Birth and Die’, in which he stated that ‘The concept of disease is a dynamic, human creation which is changing now just as it always has’. The consequence is that diseases are forever being born and dying. He gave as examples various environmental diseases, iatrogenic diseases, pseudo-diseases, cultural diseases, the English sweating sickness of the 16\(^{th}\) Century, and Legionnaire’s disease of the 20\(^{th}\) Century.
differences in racial and social classes, appropriate selection of reference populations, quantitative assessment of mental characteristics, subjectivity of clinical assessment, and conflicts between multiple and mutually inconsistent psychological theories.\(^8\)

Deviation from normal nevertheless had many supporters with J.G. Scadding prominent among them. He had started examining the idea of disease in the late 1950s as he grappled with the problem of how to define several conditions that affect the respiratory tract.\(^9\) He suggested that the term disease referred to those abnormal phenomena that are common to a group of living organisms with disturbed structure or function, the group being defined in a stated way. The final clause of this definition, and the emphasis on disturbance, sought to provide an escape route from difficulties over abnormality of the type mentioned by King. Scadding returned repeatedly to this topic over the next few years, clarifying some aspects and refining his definition. He suggested that the definition must apply to all living organisms: plants, animals and humans. He tried to define normal on a statistical scale based on standards set up by studying many individuals who represented a fair sample of the relevant population, but free from evident structural or functional defect or exposure to noxious agents.\(^10\) He recognized, however, that this idea of normal would create

---

\(^8\) Macklin (1972).
practical difficulties, and therefore suggested that a definition of any particular disease should be a brief statement of observations that in principle (but not always in practice) identified with certainty that a given case belonged to a particular group. He also defined a description as a summarized statement of all available knowledge about the disease; and diagnostic criteria as the best discriminatory features derived from the description that produced a contrast with other diseases. He later refined his definition to state that ‘A disease is the sum of the abnormal phenomena displayed by a group of living organisms in association with a specified common characteristic or set of characteristics by which they differ from the norm for their species in such a way as to place them at a biological disadvantage’. The defining characteristics could be symptoms, signs, or syndromes; or anatomical, biochemical, functional, genetic, or causal factors.

Even this redefinition, however, later failed to satisfy Scadding when he examined the topic further. His colleagues and he therefore again reworded it to limit it to a medical application and to make it specifically normativist. It then stated: ‘In medical discourse, the name of a disease refers to the sum of the abnormal phenomena displayed by a group of living organisms in association with a specified common characteristic or set of characteristics by which they differ from the norm of their species in

11 Scadding (1967).
such a way as to place them at a biological disadvantage’. A commentary on this proposal identified that it pitched an essentialist or realist view of disease (disease as an entity with a shadowy metaphysical existence of its own that is responsible for the symptoms of the people afflicted by it) against a nominalist view (disease as an arbitrary concept in which a convenient name is given to a specified group of phenomena, but liable at any time to be adjusted or discarded). These approaches, however, failed to satisfy everyone. F. Kräupl Taylor, for example, objected that they did not identify the nature of people classified as patients. He proposed that the distinguishing attributes of patients are that they are not only abnormal (judged in his approach by the standards of the population and/or the norms of the individual), but also are the subject of therapeutic concern (whether experienced by themselves, by others around them, or by medical personnel). He openly admitted that in his hands these were subjective criteria and thereby indicated a belief that the purely objective criteria sought by Scadding had a fundamental inadequacy.

14 Taylor (1980). This criticism seemed to discourage Scadding as he replied to it (Scadding, 1980) in rather defeatist tones, commenting that “The search for a definition of ‘disease-in-general’ or for a ‘unified concept of disease’ seems to me to be unlikely to be profitable”, and that ‘It is a waste of time to look for a unified concept of disease. Indeed, I am tempted to compare the struggles of those searching for one with the grappling of theologians with the Problem of Evil’.
Scadding's final definition of the word *disease* as used in medical practice contrasted with the ideas of Christopher Boorse who published his first paper on the subject in 1975. Boorse tried to restrict the term *disease* to a purely theoretical use. He claimed that disease and illness are different. Disease was to him a theoretical concept analysable in biological terms and applicable to all species, whereas illness was a subclass of disease with narrative features applicable only to humans. A disease was an illness only if it was serious enough to be incapacitating and therefore was undesirable for its bearer, providing a title to special treatment and a valid excuse for otherwise criticisable behaviour. He doubted that the term mental illness should be used at all. He considered that health was ‘functional normality’ and therefore had a statistical component, whilst normal was that which functions in accordance with its design. He argued elsewhere that a function was a contribution to a living organism's goals such as survival and reproduction. He also concluded that, apart from a theory of the structure and functions of the human mind, virtually all assertions about mental health were either misuses of language or flatly conjectural. He made, however, the general observation that ‘An organism is healthy at any moment in proportion as it is not diseased; and a disease is a type of internal state of the organism which: (i) interferes with the performance of some natural function—i.e., some species-typical

---

15 Boorse (1975).
16 Boorse, ‘Wright on functions’ (1976).
17 Boorse, ‘What a Theory of Mental Health should be’ (1976).
contribution to survival and reproduction—characteristic of the organism’s age; and (ii) is not simply in the nature of the species, i.e. is either atypical of the species or, if typical, mainly due to environmental causes’.

Boorse later expanded his ideas on this topic when he moved from the restricted area of mental disease to the general area of disease as a whole.\(^{18}\) He then argued that ‘the medical conception of health as absence of disease is a value-free theoretical notion’. Health, he suggested, was normal functioning where normality was statistical and the functions were biological (he specifically excluded, however, ‘universal diseases’). Diseases then were ‘internal states that depress a functional ability below species-typical levels’ (apart from ‘universal environmental injuries’). He specifically included ‘all kinds of injuries’ as types of disease, but he objected to several alternative theories of disease on various grounds. A weakness of a disvalue theory (lack of physical well-being) was that shortness and ugliness were not diseases. A weakness of a ‘things that doctors treat’ theory was that there were diseases that they could not treat on the one hand, and non-diseases (for example, circumcisions, abortions, and provision of contraception) that they did treat on the other. A weakness of focusing on symptoms was that asymptomatic diseases existed. A weakness of focusing on disability was that pregnancy, or an inability of a baby to walk, or an inability of an adult

\(^{18}\) Boorse (1977).
to swim, were not diseases; whereas eczema (which gave no disability) was a disease. A weakness of focusing on lack of adaptation to the environment was that failure of a school student in a mathematics examination was not a disease. A weakness of defining disease in terms of disruption of homeostasis was that locomotion, growth, and reproduction all upset homeostasis, but were not diseases. Boorse based his idea of species design on the observation that the human species and others are highly uniform in structure and function. A reference class as a fraction of a species was however necessary since intra-species variations occur, but are not diseases (involving such items as blood groups, eye colour, skin pigmentation, gender, age, and race). He later characterized the reference class as ‘an age group of a sex of a species’.¹⁹

One might reasonably paraphrase Boorse’s theory as: health is normal functioning; disease is failure to function normally; and normal is measured statistically for that organism. Boorse defended his proposals energetically, despite himself identifying several weaknesses with his theory and seeking (perhaps not entirely persuasively) to explain these. They included ‘structural diseases’ (sic) that lacked functional components, universal diseases (he claimed that body parts have to function in their ‘typical’ way, but claimed his theory to be ‘value-free’ whilst failing to explain how judgement of what is typical can be anything

other than value-laden), and the effects of age. A further weakness that he himself failed to identify was his assumption that the contemporary edition of the American Medical Association’s *Standard Nomenclature of Diseases* incontrovertibly and permanently identified all diseases by which to measure the success of his formulation, without acknowledging that the contents of this list varied from edition to edition depending on the value judgements of the physicians who compiled it.

Boorse labelled his ideas as the *Biostatistical Theory of Disease*. Many authors have subsequently objected to this theory, including on the grounds that it was circular, vague\textsuperscript{20}, covertly normativist\textsuperscript{21}, relied upon bad biology and upon bad medicine, that it failed to fit medical disease classification, and simply that it was not overtly evaluative (normativist)—which ideas of disease generally must be.\textsuperscript{22} Boorse attempted to refute each of these criticisms in a later review.\textsuperscript{23} The success of his arguments must remain open to judgement; however, the extent of the opposition to his theories suggested that they had failed to achieve broad acceptance.

Boorse, like several of his predecessors, relied heavily on the concept of normality in his writings, but—as Jirí Vácha pointed out later—normality is

\textsuperscript{20} See, for example, Hare (1986).
\textsuperscript{21} See, for example, Bunzl (1980), Agich (1983) and Fulford (2001).
\textsuperscript{22} Hare (1986).
\textsuperscript{23} Boorse (1997).
itself a highly problematic concept.\textsuperscript{24} Scientific chaos surrounds this word, causing it to lack precise definition. It can imply commonness (usualness), averageness (highest frequency), typicalness, attaining adequate performance, attaining optimal performance, attaining ideal performance, or a naïve (everyday) complex of several of these connotations. The immense range of variability in humans precludes the existence of an average individual. Fitness also is a quantitative phenomenon, not an alternative one; extremeness does not necessarily mean disease; and structural and functional normality have only a loose relationship to each other. The limits of a statistically constructed norm are that those who construct it can avoid evaluating, even in selecting a reference sample. Absolute normality is probably an impossible aspiration, whereas a more realistic one is to describe intuitively the common image of normality as ‘that which corresponds to the pattern’.

Boorse had also made the assumption that a pathologist’s diagnosis of a disease was entirely objective—that it was entirely value-free. William Stempsey pointed out, however, that pathology is far from value-free: that it is laden with conceptual values, to the extent that not just disease, but even such words as nerve and organ convey value-laden concepts.\textsuperscript{25} He justified this claim by pointing out that we use conceptually based conventions when we describe something as a nerve or an organ, since

\textsuperscript{24} Vácha (1978).

\textsuperscript{25}
(for example) what we call the trigeminal nerve actually has two roots, a motor one and a sensory one, so it becomes a matter of value judgement whether a pathologist should call this one nerve or two nerves. Similarly, the pancreas has an exocrine component and an endocrine component so it is a matter of value judgement that a pathologist calls it one organ: he or she could as well refer to it as two organs. Pathological descriptions, Stempsey claimed, are riddled with value judgements such as these, yet Boorse implicitly denied their existence. Scott De Vito was another author who argued that ‘values infect our concepts of health and disease at all levels’.26 He pointed out that scientific concepts upon which writers like Boorse have relied are themselves value laden, that goods such as life and reproduction are value laden, that the choice of a reference class for normality is value laden, and that value-eliminated theories can lead to the conclusion that conditions such as blindness or deafness that do not reduce life-span are not diseases, and that the person who has an average life span but fewer than an average number of children is diseased. These are clearly absurd propositions. The underlying reason for this is presumably that some observations and ideas only make sense against a background of conceptual structuring; whereas others rely upon additional theoretical interpretations, some of which may be well accepted, others contentious; whereas others again are underpinned by social and

26 De Vito (2000).
even moral beliefs. All of these ambiguities raise issues of value judgements in the area of pathological classifications.

The inconsistencies evident in definitions and explicatons based upon normality led several other authors to adopt quite different approaches to those of Scadding and Boorse. Bernard Rollin, for example, explored the relationship between symptoms and illness, pointing out that some people with all the symptoms of illness function normally and that symptoms only become illnesses by virtue of their role and designation in the public domain. Socio-cultural considerations therefore determine the points at which given sets of symptoms become illnesses.27

Stephen Kellert, on the other hand, looked at health and disease from a rather different viewpoint to the traditional medical one.28 He contended that—whereas the medical view is to speak of symptoms, signs, pathological conditions, diagnoses, and treatment—a socio-cultural view is for a social group to make judgements according to its accepted standards. Hence an individual’s decision to consult a physician depends largely on how well the individual can cope and adapt to the exigencies of life in the face of the condition from which the person suffers. This novel perspective was one that must resonate well with some, if not many, non-professional people.

An economic analysis of health and disease was yet another unconventional approach that emphasized the social implications and thereby gave some insights into the nature of these conditions. The argument goes that both physical and mental health affect labour force participation and productivity; and they are influenced by genetics, by the socio-economic environment, and by personal behaviour. This suggests that an individual has some degree of control over his or her own health. The importance of these issues loomed large also in the comments that James Birch made when he pointed out that, as disease is a social construct with a history that antedates modern technology, it is a matter of convention that can vary from age to age and from culture to culture.

A range of authors interested in the sociological aspects of disease also commented on its definition. Writers such as William F. Goosens perhaps expounded the most extreme positions when they argued the normativist case that diseases are any conditions that pose ‘some threat to well-being’. Joseph Margolis introduced more moderate social considerations by suggesting that ‘disease is whatever is judged to disorder or to cause

---

28 Kellert (1976).
29 Fuchs (1976).
30 Birch (1979).
31 Goosens (1980). He failed to make clear whether he would exclude such threats to well being as a fighter aeroplane, an abusive wife, or a low-paid job from his categories of disease; yet few would perceive these as diseases.
to disorder, in the relevant way, the minimal integrity of body and mind relative to prudential functions’. The functions that he considered as prudential were avoidance of death, prolongation of life, restriction of pain, gratification of desires, and assurance of security for a person, their property, and their associates. This definition made the assessment of dysfunction rely explicitly upon value-laden criteria, although it accounted poorly for disease in plants and animals as they, unlike humans, can hardly exercise prudential judgement. Robert Sade, recognizing this, made the sole ultimate goal of every living being the preservation of life. Problems that Sade did not resolve, however, were that having a goal seemed to require some conscious effort, yet insensate living organisms that lack conscious ability can certainly suffer from disease, and that some sensate organisms consciously undertake life-endangering activities. Harold Mersky’s definition was somewhat similar and suffered similarly from its anthropocentricity and professional autocracy: ‘Disease is a state of malfunction of body or mind that is a matter of concern to the patient, his doctors, and other relevant persons, subject to the qualifications that the malfunction has to be defined from case to case and that the consequences of the disease for the patient’s obligations to others

32 Margolis (1976).
34 An interesting twist on this aspect that he could, perhaps, have adduced if he had desired was that proposed by Stanley Falkow (1998) that, from the point-of-view of a microbe, being pathogenic for another organism that it parasitises is merely a successful strategy for survival; however even
(and theirs to him) will be determined by the patient and his doctors with the consent of other relevant persons'.\textsuperscript{35} Yet another definition that had some elements in common with these was that of Edmund Emson: ‘disease is a state of the human organism that actually or potentially disadvantages a person for survival, reproduction, or full enjoyment of life (characteristic for age) other than by sole reason of social circumstance or by temporary and reversible environmental change’.\textsuperscript{36} The obvious problems with this, however, were that it applied only to humans, that it did not provide for non-life threatening non-disabling diseases, or for asymptomatic ones; and it would have difficulty with socially-generated diseases and those (such as heat stroke) that result from reversible environmental change.

Yet another approach was that of Robert D’Amico who emphasized the contrast between conventionalist and naturalist ideas about disease. He criticized Lawrie Reznek’s assertion that disease is not a natural kind term.\textsuperscript{37} He analysed ideation about disease into a conflict that follows the line: ‘Are classifications of disease conventions of labelling or the result of discovering some fact of the matter in nature?’ He perceived authors such as Boorse as typical naturalists who believed that knowledge about

\textsuperscript{35} Mersky (1986).
\textsuperscript{36} Emson (1987).
\textsuperscript{37} D’Amico (1995).
disease depended upon objective scientific research devoid of value judgements, in contrast to conventionalist authors such as Margolis who believed that disease is a social concept with a thoroughly ideological basis. D'Amico's own bias favoured the former position, although Lawrie Reznek then responded energetically to argue against the possibility of any value-free account of the concept of disease.\textsuperscript{38} He reasoned that a natural kind was a class of objects that shared some deep or theoretically interesting underlying nature to explain the cluster of properties that each member of the class shared. He considered that diseases have too diverse a collection of explanatory natures to enable us to find a single explanatory nature, although he admitted that his failure to find an explanation did not prove that none exists. Peter Zachar later entered this argument to suggest that psychiatric disorders are not natural kinds on the grounds that natural kinds are exhaustively defined by reference to inherent properties, so since they do not have clear boundaries one cannot classify them as natural kinds; rather, he said, they are 'practical kinds' that have stable patterns identifiable with varying levels of reliability and validity.\textsuperscript{39} Juha Räikkä adopted a slightly different stance on these issues when he subsequently sought to draw a distinction between the social concept of disease (that he labelled as value-laden) and the medical

\textsuperscript{38} Reznek (1995).
\textsuperscript{39} Zachar (2000).
concept of disease (the value-ladenness of which he considered still to be a matter of legitimate debate). 40

Linguistics also have an important place in understanding disease. Suzanne Fleischman has recently pointed out that disease to a linguist is a social construct and that people use words such as disease and illness with many nuances. 41 This harked back to an earlier series of arguments that originated with H. Tristram Engelhardt, Jr., in 1976 who suggested that the words disease, illness, and sickness have borders when used in everyday language. 42 Engelhardt had tried to draw distinctions between illness, a disease state, and disease, although he himself admitted that the distinctions he drew were his own idiosyncratic invention rather than a reflection of everyday usage. He proposed to define an illness as something that had unpleasant subjective manifestations, in contrast to a disease state that corresponded to a medical syndrome (a set of symptoms and physical signs that made a recognizable collection and that suggested some level of explanation of their existence), in further contrast to disease that identified ‘those pathophysiological or psychopathological generalizations used to correlate the elements of a disease state in order to allow (1) an explanation of its course and character, (2) prediction or prognosis concerning its outcome, and (3) therapy through manipulating

40 Räikkä (1996).
41 Fleischman (1999).
42 Engelhardt (1976).
variables important to the course of the illness'. This led him to suggest that disease states were the explananda (phenomena to be explained) whereas diseases were the explanantia (explanations of those phenomena). Two later authors who also sought to distinguish between disease and illness were Jeremiah A. Barondess (to whom disease was in the final analysis objective and illness subjective\(^{43}\)) and David Jennings (to whom disease conveyed a sense of pathological change, and illness a sense of suffering\(^{44}\)). K.W.M. Fulford adopted a different approach when he suggested that the idea of illness offered a bridge between biological and social definitions of disease\(^{45}\), although his view of illness (a negative value concept materializing as action failure where someone was unable to do something that the person would normally do in the absence of external impediment) in turn attracted criticism from Christopher McKnight who pointed out that Fulford’s theory required an intention to cause the action, but that this element would be absent if the action would have occurred anyway.\(^{46}\)

F.C. Redlich criticized the approaches of both Margolis and Englehardt in that they failed to follow popular usage when they offered their idiosyncratic and quite personal definitions of disease. He suggested that, although they had a perfect right to do as they had done, they might

\(^{43}\) Barondess (1979).
\(^{44}\) Jennings (1986).
\(^{45}\) Fulford (1993).
preferably have used symbols or some other method rather than words that were in everyday usage with different nuances. No human being, nevertheless, was completely free from disease throughout a lifetime, so no perfect and enduring state of health could occur, even at a minimal level. The roles played by patients and physicians were important to Redlich’s own thinking, with the result that he considered ideology to influence strongly the concepts of health and disease. The idea of normality was important, but a distinction existed between normality as functioning according to design and normality as a statistical concept. Whilst a neat definition of normality may not be possible, he believed that one might well identify in this subjective area an abnormal person as one who was in need of certain helpful interventions. The role of the physician must then loom large in this explication of disease: having treatment would determine whether a person would remain in the sick role so disease should come to be defined in terms of medical positivism—disease would become ‘what is treatable’. Gerald L. Klerman took those ideas even further in his comments about mental illness when he proposed a concept of disease that came close to being anything that inhibited the pursuit of happiness.\textsuperscript{47} Jozsef Kovács adopted an equally broad description when he characterized disease as ‘that which has to be changed’.\textsuperscript{48} This set of ideas inevitably provoke several serious objections, outstanding among

\textsuperscript{46} McKnight (1998).
\textsuperscript{47} Klerman (1977).
\textsuperscript{48} Kovács (1989).
which are their applicability only to human disease on the one hand and
their breadth of inclusion on the other that, in the case of Kovács' proposals, could even include unhappy marriages, or potholes in the road outside a person’s front gate, as examples of disease.

Another approach to the nature of disease that relied upon use of an alternative word was that proposed by Clouser, Culver, and Gert when they focussed upon malady. They suggested using this word to encompass the ideas conveyed by disease, illness, sickness, and injury. ‘A person has a malady’, they suggested, ‘if and only if he or she has a condition, other than a rational belief or desire, such that he or she is suffering or at increased risk of suffering, an evil (death, pain, disability, loss of freedom or opportunity, or loss of pleasure) in the absence of a distinct sustaining cause’. This definition, however, had several immediately obvious weaknesses that, curiously, they seemed not to recognize and counter. It would exclude from having a disease people who rationally believed that they had that disease. It would also exclude people whose disease had an obvious and sustaining cause. It would, on the other hand, include as diseased everyone incarcerated in a jail or living under a dictatorial regime (they had lost their freedom) or who disliked a job or a spouse (loss of pleasure).

---

A much more physiological view of disease appealed to Caroline Whitbeck.\textsuperscript{50} She argued that diseases are complex processes about which we have only limited information, and that we identify them because of their clinical manifestations or pathological findings. She argued against viewing them as entities or as things lodged within the host, against viewing them as having single proximate causes, and against accepting them as causes of their characteristic symptoms and signs. She preferred to view them as states of affairs that existed within the organism that encompass multiple causes, multiple processes, and multiple manifestations. Commentators such as W.M. Brown criticized her for reasons such as the exclusively human (in contrast to animal and plant) application of her definition as originally expressed\textsuperscript{51}, however her view resonated with another view increasingly espoused by anatomical pathologists that disease has four sequential aspects: cause (aetiology), mechanism of development (pathogenesis), structural alterations (morphological changes), and functional consequences (clinical manifestations).\textsuperscript{52}

The plethora of philosophical proposals about disease was, by 1986, becoming confusing and overwhelming. This led H. Tristram Englehardt, Jr. to review the various proposals on offer by that time. He listed them

\textsuperscript{50} Whitbeck (1977).
\textsuperscript{51} Brown (1985).
\textsuperscript{52} Robbins \textit{et al.} (1984).
under five major categories: evaluative (having something wrong), atypicality (for example, Boorse’s Biostatistical Theory), causing suffering, as a lesion, and social (placing the patient in a sick role).  

His problem was that he found weaknesses in all of these and could not accept any of them alone as persuasive, so he failed to achieve consensus except perhaps for acceptance of an ill-defined eclectic view. Robert Sade has more recently attempted a similar exercise by proposing a different classification into reductionist theories (purposefully value-free, for example Boorse’s theory) and relativist theories (purposefully value-based, for example Clouser, Culver and Gert’s theory). He viewed the ultimate goal of all living organisms as the maintenance of life, and he considered that this was sufficiently universally accepted as to be objective.  

József Kovács, in a further more recent review, proposed a classification of naturalist theories (value-free theories of disease that rely upon empirically discernable notions, for example Boorse’s theory) in contrast to normativist (value-based) theories that he then subdivided into normativist and objectivist theories (those that perceive values as moral goods that are objective and that people desire since they are good; with disease being undesired because it is in this context bad, for example Sade’s theory), and normativist and subjectivist theories (those that perceive values as subjective desires and as people’s personal vital goals, for example Clouser, Culver and Gert’s theory). His own analysis rejected

---

53 Engelhardt (1986).
the naturalist and normativist-objectivist theories in favour of a normativist-subjectivist approach that led him to a definition of disease as the opposite of health that he then defined as: ‘The healthier a physical or mental characteristic, process, reaction is, the more it makes it possible for the individual to adapt to reasonable social norms without pain and suffering, and the longer, and happier a life it will be able to ensure him in that society’.\textsuperscript{55} His immediate problem was of course that of anthropocentricity in his definition.

A sense of disquiet about all these efforts to encapsulate the idea of disease in a simple definition surfaced from time to time in the philosophical literature of the late 20\textsuperscript{th} Century. R.M. Hare expressed this clearly when he commented\textsuperscript{56}:

\begin{quote}
At this point it may strike us that one is perhaps being over-ambitious if one thinks that one will be able to capture our understanding of words like ‘health’ and ‘disease’ in cut-and-dried definitions. Wittgenstein has made us familiar with the idea that a word may have a spread of meanings; there are a whole lot of conditions for its use, and perhaps none of them is necessary or sufficient. On a particular occasion the word will be understood although one of these conditions is absent. So, although, for example, a word like ‘disease’ is used of men and of other animals and of plants in the same sense, in a way, yet in another way it is being used in subtly different senses. Understanding its use consists, not in being able to propound a hard-and-fast definition which will work for all cases, but in having learnt to recognise all these conditions, and when they are present or absent in a
\end{quote}

\textsuperscript{54} Sade (1995).
\textsuperscript{55} Kovács (1998).
\textsuperscript{56} Hare (1986), page 178.
particular case. Doctors should not need reminding of this, because they will often agree that a patient has, say, dengue, even though one of the common symptoms of that disease is absent, provided that he has the rest.

Germund Hesslow then expressed a much more radical view in 1993 when he argued that the role of the disease concept is illusory, and that a distinction of it from health represents a conceptual straightjacket. ‘Sophisticated and mature clinical decision making’, he claimed, ‘requires that we free ourselves from the concept of disease’.57 His reasons for arguing thus were that the grounds that might justify a definition of disease are all irrelevant as (he claimed) a definition is unnecessary when disease is used as grounds for medical treatment, for medical insurance, for special rights, or for reduced responsibilities. He based this argument on the observation that some conditions described as diseases (for example, benign tumours and birth marks) are not treated, whilst treatment is sometimes given for reasons that are not diseases (for example, cosmetic surgery and sex change operations); some conditions that are diseases fail to attract insurance benefits ‘in most countries’ (for example, provision of spectacles for refractive errors) whereas the health insurance system (in his country, Sweden) pays for treatments for some conditions that are not diseases (for example, cosmetic surgery); it is not the disease per se that gains special benefits for a patient, but the pain and discomfort that working and other tasks create for the diseased person that achieve this;

57 Hesslow (1993).
and exempting mentally diseased people from moral responsibility is not because of the disease *per se*, but because the punishments that they would otherwise incur would be unlikely to be successful in reforming them. The weakness with all these arguments, one might argue, is that they rely upon a particular definition of disease that is far from universally accepted, and after doing so then disregard the common situations in order to focus on occasional, extraordinary, and rare situations without affording them the special analyses that they deserve. Hesslow’s claims provoked significant disagreement, with Lennart Nordenfelt in particular arguing against them.  

They received support, however, a few years later from John Worrall who wrote provocatively that he found no advantage in seeking a definition of disease on the grounds that ‘Conceptual analysis is nowadays, surely correctly, regarded as of limited interest even amongst erstwhile linguistic philosophers’ and that the correct view ‘is not that disease-in-general is a value-laden term, but rather that there is no reason to suppose that there is any such thing—there is no reason to think that it names a ‘natural kind’, as opposed to “lung cancer”, “diabetes mellitus” etc, which surely and unfortunately do’.  

He seemed nevertheless to disregard the fact that people do speak of disease, so they must surely believe that they have some object of their discussions.

---

Kazem Sadegh-Zadeh labelled Hesslow’s position as one of philistinism when he sought to classify the various theories of disease. He summarized it as a theory that held the notion of disease as being superfluous to medicine and irrelevant to clinicians. He contrasted it with normativist theories (such as those of Engelhardt and Margolis) that were based on value judgements of what ought not to be; descriptivist or naturalist theories (such as that of Boorse) based on biological facts free of value judgements; naïve normalist theories (as found in many medical textbooks) that health is normal, disease is abnormal, and normal is unexplained; fictionalist theories that there are no diseases but only sick people; and metaphorical theories (for example, Szasz) that there are no psychiatric diseases and that mental illness is a myth or a metaphor. His own view was that of fuzzy disease, based on the perception that a person can be healthy and not healthy, ill and not ill, diseased and not diseased at the same time: that health and disease are not dual and are not mutually exclusive. He viewed a fuzzy set as a collection of objects with grades of membership that lack sharp borders between members and non-members. An obvious problem, however, with his theory was that it failed either to define or to explicate the ideas of health or disease, so in reality it made little advance.

---

60 Sadegh-Zadeh (2000).
§2.2 SOME CONCLUSIONS

How can one summarize this huge morass of philosophical analysis? The discussion of the nature and definition of disease in which philosophers have engaged during the past half-century has clearly ranged over a very broad expanse. Many possibilities of definition have emerged yet none of them has achieved universal (or indeed even a high level of) acceptance. Virtually every time, however, that someone has commented upon the difficulty of defining disease some other commentator has emerged and pointed out how important such a definition is for the peaceful functioning of society. Raanan Gillon expressed this clearly in 1986 when, after reviewing the then extant definitions and pronouncing them all inadequate, concluded that—even though no-one had by then succeeded—further effort was essential since so much social decision-making in workplaces and in law depends on having some satisfactory resolution of the problem. 61

An attempt to analyse the proposed definitions suggests, indeed, that—far from any consensus arising for any particular definition—philosophers who

61 Gillon (1986). The theories that Gillon examined and dismissed as inadequate were the realist theory (disease as a thing) on the grounds that it did not consider the whole person; the nominalist theory (as expounded by Scadding) on the grounds that it would include anything that doctors want it to include, but exclude other items such as celibacy or poverty; the value-free concept (for example, Boorse’s Biostatistical Theory) on the grounds that it would make a champion high-jumper or a courageous soldier diseased; and Culver and Gert’s theory on the grounds that its authors themselves posed difficulties that they could not resolve.
have addressed the problem have failed even to reach any satisfactory agreement about the nature of the problem. One reason for this may be that many of them have perceived this as a single problem, whereas their discussions suggest that it is a series of problems. An overview of their discussions, indeed, suggests that they were dealing with a broad topic containing several subsidiary topics, each layering upon and interacting with others (see diagram).
Adoption of that proposal would lead one to conclude that the broad topic that philosophers have sought to address is the idea of disease. Some of them have then subdivided this broad topic into the narrower ones of disease as a word that they have sought to define, in contrast to disease as a real thing (a ‘natural kind’) that they have sought to understand. Others again have seen their task as that of dealing with a combination of both of these aspects. This subdivision, interestingly, closely reflects the early proposals made by Cawadias—relatively slight, however, as has been the attention paid to them.

Those who have dealt with the linguistic aspects of disease have tended to contrast its definition in a purely medical context with its definition in a broader social context. A few of the latter have viewed the word as expressing a unitary idea whereas many of them have then further subdivided the implications of the word disease by attributing various connotations—often of a rather idiosyncratic nature—to it and to other words that deal with the same subject, such as illness, sickness, and
Those philosophers who have striven to understand disease as a natural kind have tended to subdivide into those who have perceived it in the narrower sense of being a structure (for example, Cawadias’s realists, Cohen’s ontologists and the 19th Century clinico-pathologists) and those who have perceived it as being a process (for example, Boorse and Margolis), although again a few have straddled both camps (for example, Scadding in his first definition). Natural kind philosophers, regardless of whether they favour structural or functional explanations, have then tended to subdivide into even narrower groups depending upon the ways in which they have considered disease to impinge upon its host. There have thus been those who have focussed on disease as that which causes an organism to fail to function according to its design (for example, Christopher Boorse). Others have focussed on its interference with the ability of an organism to survive and/or reproduce (for example, Scadding’s later definitions, Margolis and Emson). Some others have focussed on disease as a cause of financial cost (for example, Fuchs) or of discomfort (for example, Emson and Klerman). Again, some philosophers have accepted that disease can be a natural kind that can impinge on more than one of these aspects of life.
A further narrowing of the natural kind topic has related to the way in which judgement is obtained of impingement by disease on the structure or functions of an organism. This has come about as a result of the method proposed by which to judge the effects of the disease. Some (for example, Boorse) have proposed that this must be entirely objectively, such as by comparison with normal, usually defined statistically. Others (for example, Stempsey, de Vito, and Reznek) have considered that a subjective element always pertains when labelling a phenomenon as a disease, thereby inducing an element of value judgement that the condition is bad.

Finally, occasional philosophers have advanced so many views without clearly favouring any of them that one must label them as eclectics (for example, H. Tristram Engelhardt, Jr.), whilst others have found the level of disagreement over the whole topic so disheartening that they have ended up adopting sceptical views (for example, Heslow and Worrall).

The philosophical analyses of the topic therefore appear not to have succeeded up to the present time any more effectively than have their medical and scientific precursors in resolving the issues of how to explicate and then to define disease. The present proposals represent a further attempt to resolve this most unsatisfactory impasse.
3. THE ROLES OF EXPLICATION AND DEFINITION

Previous philosophical analyses of the idea of disease have often mixed attempts at defining the word with attempts at explicating the reality. A primary intention of the present dissertation is to provide a clear demarcation between these two exercises. Achievement of that demarcation requires discussion of the nature of each of these terms, explication and definition, and the differences between them.

§3.1 THE NATURE OF EXPLICATION

Explication means, literally, the unfolding of something.\(^1\) A secondary meaning is the unravelling of something that is tangled. A major problem with disease relates to the tangles of scientific and philosophical ideas that have confounded understanding of the subject. The idea of unravelling these, of revealing the nature of disease by unfolding its tightly wrapped package of ideas, is both appealing and accurately describes one focus of this thesis.

This use of the term explication conforms to the usage of it by several previous philosophers. Rudolf Carnap wrote that ‘By the explication of a familiar but vague subject we mean its replacement by a new exact

\(^1\) Derived from the Latin: ex- in the sense of ‘removal of’; and plicare, ‘to fold’.
concept’.² W.V. Quine later described *explication* as ‘Philosophical analysis’.³ A.R. Lacey discussed the topic in the following terms⁴:

*Explication*, when not simply a synonym for ‘explanation’, is the process whereby a hitherto imprecise notion is given a formal definition, and so made suitable for use in formal work. The definition does not claim to be synonymous with the original notion, since it is avowedly making it more precise.

The advantage of using the word *explication* rather than *explanation* in the present context is that the implications of the latter have attracted extensive discussion in philosophical circles over the past century. People interested in the topic of explanation have developed several conflicting theories, none of which have great relevance in the present context and allusion to some of which would merely distract from it. *Explication*, on the other hand, has not attracted that involvement, has connotations that fit closely with the present aims, and so appears more appropriate for the present purpose.

§3.2 THE NATURE OF DEFINITION

The word *definition* has, over centuries, also had several contrasting implications.⁵ Philosophers who have discussed the subject have

---

² Carnap (1956), page 7.
³ Quine (1960), page 259.
⁴ Lacey (1996), pages 105-106.
⁵ For a more detailed discussion of this, see Robinson (1972), pages 12-13. Richard Robinson undertook a comprehensive analysis of the subject of definition, providing an account that has subsequently received
traditionally tended to fall into three groups—often described respectively as realists, nominalists, and conceptualists—depending upon whether they perceived definitions as dealing with things, or words, or concepts.

Adoption of this classification would place Socrates and Plato as realists because they perceived definitions as defining things. Plato indeed saw a definition as marking the end of the process of getting to know the ultimate reality about something, as exemplified by his theory of forms. Aristotle, Milton, and Spinoza followed him in this approach. J.S. Mill, in contrast, was a nominalist in that he perceived a definition as giving an account of a word or words: ‘The simplest and most correct notion of a Definition is, a proposition declaratory of a meaning of a word; namely either the meaning which it bears in common acceptation, or that which the speaker or writer, for the particular purpose of his discourse, intends to annex to it.’ Kant exemplified the third possibility, that of the conceptualist, because he wrote that ‘To define, as the word itself indicates, really only means to present the complete, original concept of a thing within the limits of its concept’.

considerable support and relatively little criticism. His views have provided a basis for the approach that I shall adopt in this thesis.

6 Mill (1891), page 86. Robinson also classified John Locke as a nominalist, although Locke’s comment ‘definition being nothing but making another understand by words what idea the term defined stands for’ (Locke, 1959, volume 2, page 20) had conceptualist overtones.

7 Kant (1986), A727.
The word *definition* has itself previously attracted many definitions. Suggested possibilities have ranged from Plato’s ‘Revealing one’s thoughts by means of speech’, to Wittgenstein’s ‘rules for the translation of one language into another’, to Carnap’s ‘rule for mutual transformation of words in the same language’.\(^8\) Plato, indeed, considered three alternatives and gave as his favoured suggestion the view that definition is ‘being able to name some mark by which the thing one is asked about differs from everything else’.\(^9\) He considered that this was the meaning that most people would give, but suggested that a definition must also announce the content of the thing defined.

Robinson considered that the word *definition* refers to ‘A certain human activity, intellectual in character’. The great strengths of that starting point are that all definitions (as we presently understand them) are verbal expressions that would fail to exist if humans failed to exist, and that the utterance of meaningful words results from intellectual activity. It does, however, only identify the venue in which definition occurs—without itself defining *definition*—since many other human activities also exist that are of an intellectual nature. Robinson went on to clarify that the purpose of definition is to deal with associations, and that those associations can be of words and/or of things. He suggested that, as a result of this, three

---

\(^8\) See Robinson (1972), pages 2-3, for a summary of a dozen of these alternatives.

\(^9\) Plato (1961), *Theaetetus* 208c.
possibilities exist: word-word, word-thing, and thing-thing definitions. The first two constitute the realm of what commentators have traditionally called nominal definitions, whilst the last is what they have called real definitions. He argued against persisting with the idea of real definitions because some of the entities so described are not indeed definitions, and the others are actually word-thing definitions. He considered, furthermore, that word-word definitions are of relatively little importance in contexts such as the present since they merely identify two words that have identical meanings—as, for example, the translation of a word from one language to another without elucidating the meaning of the words in either language.

This then leaves word-thing definitions as the crucial variety, of which there are two principal types: lexical definitions and stipulative (or legislative) definitions. The former describe the ways in which actual persons have historically used actual words, recording an already existing association of a name with an object. The latter explicitly and self-consciously create new words, or meanings for words, by assigning a novel name to an object. Dictionaries primarily deal with lexical definitions,

———

10 They included such issues as searching for essences, searching for a cause, seeking a key to explain a mass of facts, adopting an ideal, abstracting, analysing to realize that a certain form is a certain complex of forms, synthesizing to realize that a certain form is a certain part of a certain complex form, or improving one’s concepts (see Robinson, 1972, pages 189-190).
although they can sometimes stipulate in usually minor ways.¹¹ Lexical definitions involve three types of people: the definer, the person who reads the definition, and the person (who may have existed several centuries earlier) who used the word with the meaning that the definer has ascribed to it. Each of these parties may place subtle subjective interpretations on the word described. Lexical definitions (at least to the extent, but only to the extent, that they accurately record usage and avoid subjective interpretations) have truth-value, whereas stipulative definitions (at the moment when they are stipulated) have none. Robinson furthermore claimed that words are signs and that ‘every sign is lexically definable by somebody to somebody at some time and place’.¹² Living language, however, is always in flux and the meanings of words can change. Stipulative definitions, in contrast to lexical ones, have several purposes including the removal of error from a commonly used word, the improvement of concepts, the removal of arbitrariness from the language, the abbreviation of language, and the naming of a newly recognized object. People create them to serve their intercurrent purposes and they can use various methods to achieve this purpose—of which Robinson identified seven different possibilities, to each of which he assigned a distinctive name.¹³

¹² Robinson (1972), page 42.
¹³ See Robinson (1972), pages 93-148. The synonymous method of stipulative definition states another word identical with the one being
These many approaches to stipulative definition each have advantages and disadvantages. They are not mutually exclusive in that each of them is necessary on some occasions to achieve the creation of new ideas that a fluent use of language requires. Some of them are more effective than others to deal with the definition of nouns, so a stipulative approach using one or other of the methods available may become necessary in dealing with a topic such as disease where one can argue that existing lexical definitions fail to provide adequate insight. The multiplicity of approaches to stipulative definitions should not, however, suggest that inventing a stipulative definition is some sort of trivial game. Whereas a lexical definition tries to identify the understanding that people have of its subject and will stand or fall depending on its success in achieving this, a

---

defined—as occurs in two-language or brief one-language dictionaries (for example, automobile means motor-car; chien means dog). The analytic method consists of breaking the word down into its component parts that explain its meaning (for example, octagon means a polygon with eight sides). The synthetic method consists of indicating the relation of the thing denominated by the word to other things (for example, red means the colour produced by light of wavelength 7000-6500 Å). The implicative method suggests the meaning of a word by using it in a sentence without saying exactly what it is (for example, diagonal is elucidated by saying a square has two diagonals, each of which divides it into a right-angled isosceles triangle). The denotative method involves giving examples of the thing (for example, birds are things such as swans and robins and larks and geese, but not such things as bats or butterflies or aeroplanes). The ostensive method involves naming the object to be defined whilst physically indicating it in some way (for example, uttering the word car whilst pointing to a motor-car). Finally, the rule-giving method involves always using a word according to some pre-existing conventions, as occurs in the case of many conjunctions and prepositions (for example,
stipulative definition will try to rectify the inadequacies of existing and unsatisfactory lexical definitions of the subject. People will then assess its value by its success in clarifying a previously obscure subject. The definition and explication of disease is an incontrovertibly contentious subject still shrouded for many in obscurity. The importance of the distinction between the two types of definition to this subject will therefore become apparent in later chapters of this thesis. They will demonstrate that the present explication raises issues that no previously proposed definition adequately addresses. The definition proposed will inevitably include a stipulative element in an attempt to resolve these. Judgements about its value will ride upon its success in improving understanding of the subject.

§3.3 ANTICIPATIONS FROM THE EXPLICATION AND DEFINITION OF ‘DISEASE’

What, then, might one expect of a new explication and definition of disease? The outstanding anticipation will be for a simple, but comprehensive, account that will help people working in the practical and non-theoretical world to achieve the socially responsible ends at which they aim. But, desirable as is this intention, it is not in itself enough. A persuasive explication and definition must also contain elements that will withstand critical theoretical analysis—a status not achieved by any

always using the word and to link together two related phrases, or always
previously proposed explication or definition. The enterprise must, therefore, aim at achieving a persuasive theoretical account that would withstand robust usage in the practical world.

What are the basic criteria that an account must meet if it is to achieve these ends? Several features will be required if it is to have any chance of meeting these goals.

It must, firstly, hold true across the whole spectrum of biology since discussion of disease has a fundamental similarity whether dealing with humans, with animals, or with plants. An account that provides a satisfying explication and definition for only one or a limited number of species is likely to contain flaws of sufficient magnitude that it will implode when they eventually become apparent.

It must, secondly, deal with the enormous wealth of scientific data that have accumulated about disease and diseases over the past three millennia, but do so in a succinct way such that it does not retreat into a morass of unmanageable complexity. It will, to achieve this, at least have to recognize the contributions made by manifestations, causes, and mechanisms that occur in organisms when these suffer from disease. It will have to provide explanations for each of these aspects,
comprehending many previous discussions, accommodating each of them, and binding them together in an intelligible way. It will have to resolve the tension that exists between structural and functional descriptions of disease and superimpose a social interpretation upon this to accommodate new information about disease as it emerges.

It will also have to deal with the tension that exists within the idea of disease in particular cases, in groups of cases that have characteristics in common, and as a universal idea. This will oblige accounting for the way in which disease lends itself to diagnosis (its epistemology) as well as to classification (its nosology). It will have to address explicitly the differences between realist and nominalist views, ideally by quarantining its explication from its definition. It must in doing so, however, link these two aspects as complementary viewpoints to each other in ways that minimize internal inconsistencies.

A successful account will, in short, have to deal with the many scientific and philosophical analyses undertaken in recent years, with the difficulties that they have exposed in explicating the subject, and with their most obvious implications. It will, furthermore, have to achieve this in a manner that is both persuasive and succinct.

* * * * * * * * * *
4. A NEW DEFINITION OF DISEASE

The new definition proposed for disease in this thesis is:

*Disease* is a word used by observers to describe a process that occurs when one or more external factors interact with a living organism to produce physical and/or mental changes within the organism that the observers consider disadvantage the organism as compared with its former state.

Some implications of this definition are that:

- Changes cannot occur without causes;
- Causation requires both the organism and the external factors;
- The interaction is the central pathogenetic mechanism of disease;
- The causes and the changes may be overt or covert, physical or mental;
- Observers judge whether disadvantage occurs;
- Each observer's perceptions create a personal diagnosis;
- Consensus among observers strengthens diagnoses;
- Collective and universal ideas about disease develop from individual ones.
This definition lends itself to diagrammatic expression, as shown in the adjacent Figure, to which the following symbols apply:

CM: The material cause—the existence of the living organism
CE: The efficient cause—the external factor that interacts
P1: The first pathogenetic event in which energy is exchanged
M1: The first manifestation
P2: The second pathogenetic event in which energy is exchanged
M2: The second manifestation
MF: The final manifestation
O1: One observer
D1: The diagnosis in the mind of the first observer
O2: A second observer
D2: The diagnosis in the mind of the second observer
O3: A third observer
D3: The diagnosis in the mind of the third observer
DF: The final diagnosis, reached by consensus in society

An external factor (CE) interacts here with a living organism (CM). The interaction is a pathogenetic event that involves an exchange of energy (P1). That results in a change in the organism, that takes the form of one or more manifestations (M1). These manifestations may be the final manifestations (MF), or they may interact again with the organism (CM)
with a further exchange of energy ($P_2$) to produce one or more further manifestations ($M_2$). They, then, may constitute the final manifestations ($MF$). One or more further series of energy-exchanging interactions ($P_3...P_n$) may alternatively occur between the organism and those intermediary manifestations before the final manifestations ($MF$) develop. Various observers ($O_1, O_2, O_3...O_n$) may discern the situation that has occurred. The most obvious method whereby they can do this is by becoming aware of the final manifestation(s) ($MF$), but they may also notice aspects of any antecedent manifestations ($M_1...M_n$) and pathogenetic events ($P_1...P_n$). They each develop diagnoses ($D_1, D_2, D_3...D_n$) as they analyse their observations in their minds. They may well, in the process of doing this, repeatedly re-examine their observations. The various observers then interpret their diagnoses ($D_1...D_n$) within the cultural context in which they all live to reach a consensus (or final) diagnosis ($DF$) that they perceive as making sense within the limits of the knowledge available to them in their community. If they perceive this to represent a physically or mentally disadvantageous condition they speak of the living organism ($CM$) as having disease.
This definition starts by assuming that disease exists in the real world, that people are often aware of it, and that it is susceptible of definition. It takes these as self-evident premises that (at least for the present purposes) require no additional justification. It then makes several assertions, the validity of which is potentially less secure and which may therefore require justifying argument. The level of necessary argument varies depending upon the particular assertion.

Its first assertion, that *disease* is a word, is in itself hardly controversial, but does serve the important purpose of identifying the act of defining specifically as a nominalist exercise. It provides a contrast with the subsequent act of explicating that deals with the real entity that exists when disease occurs. This approach—of using a definition to deal with the nominal aspects and then separately using an explication to deal with the real aspects—achieves from the outset a clear separation of nominalist from realist considerations, whilst recognizing both to have relevance in any discussion of the topic of disease.

The second assertion made by the definition is that the word *disease* describes a process and changes that result from it. This relates to the question of whether disease is primarily a function or primarily a structure. The approach adopted is to regard it as a process (a function) that may
(but does not necessarily) have structural consequences. The consequences, or manifestations, cannot be present before the phenomenon of disease starts. They appear later. Something must occur to make them appear, so action must be involved. A phenomenon that involves action is a process, not a structure. The structural consequences, nevertheless, are an integral part of the overall disease idea so a satisfactory definition must include them together with the process that created them. The definition therefore speaks of ‘changes’ with the sense of ‘a process and its consequences’.

Another implication of this assertion is that the state of a diseased organism is always changed from its pre-morbid state. It does not require that the pre-morbid state was ‘normal’ or some idealized state for an organism of that species: it merely requires that the state of the organism whilst sustaining the disease is different from the state beforehand.

The third assertion made by the definition is that diseases occur within living organisms. One does not find evidence of people using the word disease to describe changes that occur in non-biological objects. We say, for example, that a machine is malfunctioning when it does not work according to design, rather than that it is diseased. We say that a building
is damaged after a storm lashes it, not that it is diseased.¹ Our convention is to restrict use of the term disease to biological objects, and this definition reflects that convention.

The fourth assertion made by the definition is that disease occurs when one or more external factors interact with an organism. It avoids stating that disease necessarily occurs when external factors interact with the organism by subsequently requiring that disadvantage must ensue. It thereby insulates itself from calling beneficial interactions (such as the everyday consumption of food) the harbingers of disease. Many cases of disease, nevertheless, are clearly the consequence of action against an organism by one or more external factors. Although people often call the less important items predisposing factors when more than one external factor exists without which the disease would not occur, these nevertheless remain as external factors, they influence the whole process, so they have some causal implications. Such considerations are hardly contentious. Some people may, however, question two other implications of this assertion: firstly, the implication that disease cannot occur unless one or more external factors act upon the organism; and, secondly, the implication that injuries, the consequences of bites, and certain other disadvantageous situations that occur to organisms are examples of

¹ The term virus is often nowadays used with regard to damaging intruders inserted into computer programmes, but I would suggest that this reflects
disease. I believe that powerful arguments exist to support both these contentions, but they require more detailed discussion to which I shall return in §5.1 when examining the causation of disease.

The fifth assertion, made by implication, is that an interaction occurs within the organism. This is an important aspect of the definition that requires detailed explication in order to accommodate many scientific issues raised by previous theories of disease and by recent discoveries. The development of a unified description of its method of occurrence must enhance understanding of the whole process. I believe that such a description is possible, and shall address this subject in §5.1 and §5.2 when addressing causation and pathogenesis.

The sixth assertion made by the definition is that manifestations (in the form of physical and/or mental changes) occur, and a corollary emphasizes that these may be overt or covert. People commonly recognize that they have disease—or that animals, plants, or other people within their cognisance have disease—when they perceive manifestations appearing of the disease. These manifestations usually take the form of symptoms, physical signs, and the results of special laboratory and other investigations. Although this aspect also is hardly contentious, the

merely a convenient shorthand way to describe by analogy a particular situation, rather than a primary use of the word.
possibility of asymptomatic disease warrants detailed exploration, so I shall return to it in §5.3 (that addresses manifestations) and §5.5.

The seventh assertion is that the manifestations can be of a physical nature, of a mental nature, or of both. Physical manifestations are hardly a matter of contention as they are the traditional way in which many diseases manifest. Mental manifestations, however, are more complex. One does not, for example, observe these in plants or lower animals. The definition, furthermore, may not on first examination appear to accommodate them easily (because of its causal implications), despite most people accepting that they do occur in humans and perhaps also in animals that have higher levels of cerebral development. Certain considerations, however, that I shall outline in §5.4 demonstrate that this definition does indeed provide a reasonable basis to explain these.

The eighth assertion is the implication that when observers use the word disease (and subsidiary ones to describe varieties of it) they are diagnosing the disease. Can, however, undiagnosed disease exist? Can there be disease in the absence of perception by observers? This relates to the issue of asymptomatic disease mentioned in the sixth assertion above that I shall discuss further in §5.5.
The ninth assertion is the implication that the observers may be either single or multiple. This seems hardly contentious, but raises the question of the criteria necessary for a person to qualify for identification as an observer. This will provide the topic for examination in §5.8.

The tenth assertion is that the changes must disadvantage the organism. This claim flies in the face of the definitions proposed by several previous commentators on disease, so it inevitably requires considerable exploration and justification. That will form the subject of §5.6.

A further assertion is the implication that changes cannot occur without causes. This indicates that the present use of the word *cause* is as a means of identifying some thing or things that are responsible for bringing about something else (the effects, which here are the manifestations). It thereby denies the possibility of spontaneous changes, a topic that will receive further notice in §5.1.

The next assertion is the implication that the organism and the external factors are all causes. Few would challenge that external factors are often causes of disease, although some might wonder about labelling the organism itself as a cause, and some might question whether all cases of disease must have external causes. This is an important component of the definition that introduces stipulative elements, so it warrants the
presentation of appropriate supportive arguments. I shall discuss it in detail in §5.1 whilst dealing with causation.

The thirteenth assertion is the labelling of the interaction as a pathogenic event. This statement is little more than an exercise in naming, or subsidiary defining, in order to clarify the discussion and simplify the use of terminology. It will, however, receive some further discussion in §5.2.

The fourteenth assertion is that each observer forms a diagnosis. This obliges an assessment of the nature of the phenomenon described by the word diagnosis, a word that people often use without defining. Many seem instinctively to believe that only one diagnosis (‘the right diagnosis’) is possible or at least legitimate, but I propose to challenge this view by suggesting that there are as many diagnoses as there are observers of each case of disease. That, being a stipulative aspect of the definition, may well stimulate some controversy, so I shall discuss it in more detail in §5.7.

The fifteenth assertion is that a consensus of perceptions can occur (indeed, I shall suggest, does often occur) among observers. I shall discuss this and the sixteenth assertion, that this consensus strengthens the diagnosis, in some detail in §5.8.
The final assertion is that collective and universal ideas of disease derive from particular ones. This alludes to varying usages of the word disease, sometimes in the sense of disease-as-a-particular, sometimes of disease-as-a-collective, and sometimes of disease-as-a-universal. This is a large topic that includes several important facets that illuminate the phenomenon. It warrants a detailed analysis that I shall give in §5.9.

* * * * * * * * * *
5. AN EXPLICATION OF DISEASE

The presently proposed definition has many implications and several important strengths. They each warrant notice. Perhaps it also has some weaknesses; and, if so, each of these requires careful analysis to determine whether or not it is justified. It is, of course, possible that reasonable explanations exist for apparent weaknesses and that a fresh examination of them will elucidate some previously obscure or contentious issues, permitting clarification of ideas about the disease process as a whole, and so strengthening the case for the new definition rather than damaging it.

Some topics that warrant particular examination in this regard include the causation of disease, its pathogenesis and manifestations, mental disease and genetic disease, asymptomatic disease, the disadvantage of disease, its diagnosis and the people who observe it, the contrast between disease and diseases, some borderline conditions (such as menstruation, menopause, pregnancy, and obesity), and the implications of various words that people use as synonyms or correlates of disease. The role of the present chapter is to examine these issues in some depth.
§5.1 CAUSATION

Every organism that develops disease displays the outstanding characteristic that it differs in some way whilst it has its disease from how it was before it developed the disease. The differences may affect the structure of the organisms, or the function, or they may affect both of these aspects. Something has, in every case, changed in each of these organisms in such a way that the resultant effects thereafter mark the organism as diseased for as long as they persist. These differences are changes and changes generally have causes. Four aspects of causation (or aetiology, as it is often called when used in medicine and the other biological sciences regarding disease) warrant particular consideration in relation to the new definition. These are the implications that the definition conveys that causes are necessary, that they are necessarily multiple, that they may take many forms, and that they necessarily lead to an expenditure of energy.

*The implication that causes are necessary*

The idea that causation plays an important role in ideas about disease is hardly new, either at the practical or at the theoretical level. The present definition emphasizes this role rather more forcefully than have some (although not all) of its predecessors and does so on the grounds that gaining a comprehensive understanding of each case of disease, and of the idea of disease in general, requires an understanding of how that case
has come about, and of how diseases in general come about—of its and their causes. Most medical, veterinary, and horticultural texts demonstrate the practical application of this idea when they discuss aetiology as an important component of each description of each particular type of disease. Following this line of argument then prompts questions about whether every disease must have a cause: of whether ‘uncaused disease’ can occur.

It is conceivable that uncaused disease could occur although this seems unlikely on at least two separate grounds. The first of these is the general one that uncaused changes seem not to occur throughout any other realm of nature: that whenever one examines in sufficient detail any active development that occurs in nature one can discover some cause or causes that underlie it. The development of a case of disease is an active development—it involves change in the status of the individual—so one would expect that it must have some cause. The second is that in the area of each category of disease that one examines one finds either that causes are either clearly identified or that knowledgeable people have proposed reasonable hypotheses to explain them. The universal occurrence of change when an organism develops disease makes it reasonable to include change as a necessary component of disease. The great unlikelihood of change occurring without causes to stimulate it then suggests that it is similarly reasonable to include causation as another
important component. This line of thinking at a theoretical level has influenced the present definition, whilst practical considerations have also had an important influence.

An examination of some practical examples of disease will illustrate these points. One observes that most conditions that people describe as diseases have some identifiable or postulated external causal factor or factors. Many of these are hardly controversial. No one now doubts that inhaled dust has played a causal role in the lung disease of a coal miner who develops pneumoconiosis. A child develops measles after contacting another child who suffers from measles: we all accept that a causal infectious agent or agents (in fact, a virus) must have passed from one to the other. A pregnant woman takes thalidomide and her baby is born with certain deformities: people nowadays recognise thalidomide as the obvious toxic chemical culprit. A motorcyclist smashes his machine and fractures his leg, the latter heals with a deformity, he thereafter walks with a limp, and goes on to develop arthritis in his knee: the trauma of the smash clearly produced the deformity, and the trauma of walking on the deformed leg presumably produced the arthritis. A fat man with a high cholesterol level suffers a heart attack: everyone blames the cholesterol-rich fast food that he systematically consumed; whilst beer catches the blame when his drinking companion develops cirrhosis of the liver. The
causal factors in these and many similar examples of disease are relatively uncontroversial.

One furthermore fails to achieve even a superficial understanding of each of these individual cases of disease until one manages to identify some persuasive external causal factor that has interacted with the organism and instigated the changes that occurred in the latter that we eventually recognize as disease. This underlines the fundamental importance that causation plays to the idea of disease as a whole. A problem does nevertheless arise over a relatively small number of diseases for which we have difficulty in recognizing obvious causes. A person who attempts to define disease and explicate it then has three alternative methods by which to deal with these. The first is to adopt the position that change is universal in disease, that causation is universal in other varieties of change, that empirical knowledge about disease remains incomplete, and that requirement of a role for causation is inappropriate in the absence of total empirical knowledge. The second is to adopt the position that change is universal in disease, that causation is universal in other varieties of change, that empirical knowledge about disease remains incomplete, that understanding causation is crucial to fully understanding disease, but that causation may presently remain covert and the definition should acknowledge this whilst not thereby denying it a crucial role in the idea of disease. The third is to adopt the position that change is universal in
disease, that causation is universal in other varieties of change, that empirical knowledge about change in disease remains incomplete, so empirical knowledge about causation in some types of disease is incomplete, but that this should not prevent the definition from stating without qualification that causation is a necessary component of disease.

The present definition adopts the middle course on this matter. It precludes labelling a state as a disease if that state has no causes (so it implies a belief that all cases of disease have causes), although it does not go so far as necessitating the conclusive identification of the causes before a state can be labelled as a disease. It thus argues against the idea of spontaneous generation of disease—of disease that comes about as a result of the intrinsic nature of an organism independently of the presence of any external action upon that organism\(^1\), whilst it accepts that the external cause(s) may presently remain obscure. Disease is a process, the present definition suggests, and processes by their nature require some trigger: something must have initiated the process, and that something is what the present definition identifies as (part, at least, of) the causes.

\(^1\) This gives the present definition something in common with a philosophical viewpoint that has dated at least from the time of Aristotle who commented that ‘Everything that is in motion must be moved by something’ (Physics, Book VII, 1, 241b 34; this and all other references to Aristotle refer to the edition edited by Jonathan Barnes [1984]), although it does not align it so closely with Aristotle’s analysis of causation as to
The implication that causes are necessary for the generation of all cases of disease throws out an immediate challenge to sceptics to try to identify examples of causeless conditions to which people usually apply the word disease—since the discovery of incontrovertible examples would pose serious problems for the new definition. The obvious candidates for ‘causeless diseases’ are, on the one hand, the genetic diseases, and, on the other, a series of conditions whose titles include such adjectives as essential, idiopathic, and autoimmune. Genetic diseases are a unique group of conditions with special attributes that require detailed analysis. I shall examine them in §5.10, where the conclusion that I shall reach is that they too are each likely to have identified or identifiable causes. This leaves the essential, idiopathic and autoimmune diseases to be analysed, some typical examples of which are essential hypertension and idiopathic thrombocytopenic purpura (known also as autoimmune thrombocytopenia). A discussion of them is equally applicable to other examples of such conditions.

People could reasonably assume that use of the term essential in these conditions indicates that the disease is ‘by the very nature of the person’; that idiopathic indicates ‘disease of one’s own origin’; and that autoimmune indicates ‘immune mediated damage directed at oneself’.
Three possibilities do however exist to explain the use of words such as *essential*, *idiopathic*, and *autoimmune* in the names of diseases. The first is that the conditions to which these words refer are in fact causeless, in which case they do pose genuine difficulties for the present definition. The second is that the diseases to which they refer are in fact caused, but the people who use these words create for themselves a semantic problem by adopting a misguided or imprecise terminology. The third possibility is that the diseases are caused, but the users of the words are limited by the present state of empirical knowledge, to the extent that contemporary technology does not permit them to understand the diseases well enough to propose persuasive explanations of their causes. I would suggest that the second and third explanations are, in combination, far more persuasive than the first. The reason for suggesting this is that a detailed examination shows that people with a particular interest in and knowledge about these diseases do in fact propose theories to explain their causation. These knowledgeable observers thereby imply that they expect to develop persuasive causal explanations as soon as they can gain adequate information that gives them the necessary insight. They find themselves limited, meanwhile, to proposing hypotheses, and to using adjectives that indicate their uncertainty about the causal nature of the diseases that they are discussing.
One can support this contention with many examples. A typical textbook definition of *essential hypertension* states: ‘Essential hypertension may therefore be defined as sustained high blood pressure not attributable to a single cause but reflecting the interaction of multiple genetic and environmental influences’.\(^2\) This definition suggests, not that the expert who wrote it believes that essential hypertension is uncaused, but that the expert believes it to have causes that presently remain obscure and difficult for even an expert to describe. It does indicate, nevertheless, a belief that some of the probable causes come from within the organism and some from without.

Another example is a typical textbook description of *autoimmune thrombocytopenia*. This indicates that circulating antibodies bind to and destroy the platelets of people who suffer from that disease under circumstances in which a preceding viral infection may have triggered the development of the antibodies in some victims, but in which the trigger still remains obscure in others.\(^3\) The implication again here is that we need more information, rather than that such information cannot exist. One finds, indeed, that practically every case of what some might call ‘causeless diseases’ conforms to this same template of postulated, but as yet unconfirmed, causal factors.

An approach, in cases such as these, of admitting contemporary uncertainty and of striving to gain appropriate information in the future, would appear inherently more productive than to deny by definition the possibility of ever discovering information. The wisdom of that approach is emphasized when one examines the sequential accumulation of knowledge about certain diseases of which the causes remained obscure until quite recently, but have now gained acceptance. Many of the items discussed in §1.3 above provide typical examples of these. Cigarette smoking and asbestos exposure, as causal factors in the development of various types of lung cancer, similarly demonstrate this point. Many people believed, fifty years ago, that cancer was a disease that was either causeless or at least of unknown cause. Epidemiological data subsequently became available that identified cigarette smoking as a powerful predisposing factor for one particular type of lung cancer, and asbestos exposure as a predisposing cause for another. No one denied that these were both types of disease, and most who previously claimed that their causes were unknown later came to accept that causal factors existed as persuasive information accumulated.

These considerations inevitably lead one to criticize as misleading the traditional choice of adjectives such as essential, idiopathic, and autoimmune to describe certain types of disease. Phrases such as

---

3 Ledingham and Warrell (2000), page 301.
hypertension of less than fully identified cause would then seem more appropriate that such shorthand terms as essential hypertension. The problem, of course, is that the wordiness of the more accurate phrases makes them unattractive. One nevertheless hardly finds persuasive a claim that ‘causeless disease’ exists merely because people have taken the short cut of using succinct but imprecise terminology in preference to voluminous precision.

The present definition is therefore stipulative to the extent of suggesting that similar circumstances pertain for all diseases currently labelled with such adjectives as essential, idiopathic, and autoimmune. It makes this conceptual claim on the bases both that changes occur in disease and that the occurrence of changes requires causes; and that empirical observation fails to identify persuasive cases of disease in which one cannot either identify causes or realistically postulate causes that the development of more refined methods of assessment would enable one to confirm. The clear implication, in summary, is that ‘causeless disease’ does not occur, although the causes of some types of disease still remain obscure, that a definition should ideally acknowledge this, and the terminology presently used to describe some diseases is less than ideal.
The implication that causes are necessarily multiple

An organism cannot develop a disease unless the organism exists, so the existence of the organism is a pre-requisite for the occurrence of disease, and a contributing factor to causing the disease. Something more than the mere existence of the organism must, however, occur for the development of disease since (as demonstrated above) diseases are unlikely to arise spontaneously, and since observation reveals that every organism does not eventually develop every possible variety of disease (which would presumably occur if the mere existence of an organism was sufficient for the development of disease). Some additional factor or factors, external to an organism, must therefore act upon it to cause it to develop disease. People often refer to the most obvious external factor in each case of disease as *the cause*, but in doing so they overlook the contributions made both by the organism itself through its own very existence and also by additional less obvious external factors.

This idea of multiple necessary factors contributing to causation has a long history in general philosophical discussion, despite a modern tendency to focus upon the most important external causal factor. It dates indeed at least from the time of Aristotle whose extensive analysis of the idea of causation in living creatures, as well as in the inanimate world and in logic, identified as causal factors more items than just this most important or immediate or obvious instrument responsible for an event. A
brief review of his ideas is useful in the present context. Aristotle postulated four distinct components of each case of causation, designating these as the matter (or material cause), the form (or formal cause), the mover (or efficient cause), and that for the sake of which (or final cause). He described the matter as ‘that from which (as immanent material) a thing comes into being, eg. the bronze of the statue’; the form as the pattern in which matter exists; the mover as ‘the source of the change’ or ‘what initiated the change’; and that for the sake of which as ‘the aim’. He perceived events as occurring in series, with the final cause of one segment often becoming the efficient cause of the next, but denied that such series could go on to infinity. His overall idea therefore seemed to be that the material cause is the object upon which an efficient cause acts; that the efficient cause is an agent external to the object upon which it acts; and that the final cause is the purpose for which that external agent acts.

---

5 Metaphysics, Book V, 2, 1013a 24-34. See also Physics, Book II, 3, 194b 16 – 195b 30.
6 Metaphysics, Book I, 3, 983a 24-32.
7 Posterior Analytics, Book II, 11, 94a 20-23.
8 Posterior Analytics, Book II, 11, 94a 20-23.
9 He thereby concluded that ‘there must be some first mover’—and so created a type of fifth, or prime, cause.
10 The explanation of the formal cause is the most contentious of these proposals. Most commentators follow the reasoning of Bunge (1959, page 32) rather than that of Taylor (1967, page 56) who perceived it as ‘that into which something is changed’.
The immediate relevance of these concepts to the present definition of disease is their creation of a template in which an impinging external factor or factors act upon an organism, at least two of which are necessary to cause the development of disease. Causes, on analyses such as this, are necessarily multiple. A person who perceives over-riding importance in the role of the efficient cause could, however, suggest that a claim such as this is either a non-sequitur or is merely trivially true on the grounds that it contains a suppressed premise, namely that a disease is always a disease of an organism. It nevertheless only becomes a non-sequitur if one emphasizes the role of efficient causes to such an extent that they annihilate the other causes to which Aristotle referred, and in this I believe one is safer to follow Aristotle’s lead if only to emphasize the fact that suppression of the role of the organism potentially leads to undesirable practical consequences. These include, on theoretical grounds, an inappropriate tendency to de-emphasize the initial and continuing role of the organism in each of the sequential interactions (P₁…n) that occur between the external cause (CE) and the sequential manifestations (M₁…n) that ultimately lead to the development of the final manifestation (MF) of disease (as shown in the diagram in §4); and, on practical grounds, the development of a de-humanized atmosphere within hospitals where the latter cease to be places to treat sick people and become merely scientific laboratories in which to study disease.
The implication that causes may take many forms

The assertion that causes may take many forms also deserves comment as it contains stipulative elements. These forms can relate, on the one hand, to the visibility of the causes; and, on the other, to their character.

Regarding visibility: external causes can be overt or covert. The information provided in §1.1 above suggests that people have always regarded some causes of disease as overt and uncontroversial. They have, for example, long recognised the disease-causing potential of such external insults as snake and spider bites, and they have rarely questioned the potentially dire consequences of these. Much of the enigmatic aura that has surrounded disease—producing such controversy about its nature—has, however, arisen from the speculation about the many diseases whose causes remained obscure for so long. The accumulation of objective information about the latter through systematic research has nevertheless progressively opened the curtain of obscurity and enabled people also to identify their causal aspects, to the extent that (as demonstrated in the previous section) very few, if any, now remain devoid of proven or reasonably hypothesized explanations. The numbers of covert external causes have thereby progressively declined with the accumulation of knowledge, converting them to overt causal explanations. Innumerable examples demonstrate this. Consider certain types of lung disease. People suspected for several hundred years that inhaled dust
could cause permanent scarring of the lungs of miners, but they had much more difficulty identifying external causes that led to pulmonary scarring in certain other cases. The inhalation of acid-fast bacteria as the mechanism of acquiring tuberculosis puzzled them for millennia—in part because its course was so protracted; in part because investigators lacked microscopes that would display the bacteria; and in part because, even with microscopy, they had to devise special stains to demonstrate them. Acid-fast bacilli therefore remained as the covert external cause of tuberculosis until such time as microbiologists developed accurate technical methods, whereupon the external cause of this disease became overt. Limitless speculation changed almost overnight to objective understanding. This then is a typical demonstration of the way in which causes of disease can have either covert or overt appearances and that the transition from the former to the latter has often mirrored a progressive accumulation of knowledge.

Regarding character: external causes, whether they are overt or covert, also fall into several distinct groups based upon their physical, chemical, microbiological, radiological, psychological, and other characteristics. Some of these are non-controversial, but others warrant some examination. The propensity for bacteria, viruses, and parasites to produce disease by invading other living organisms is nowadays well accepted. The potential of ionising radiation to cause disease in organisms
is also beyond debate. The emotional disturbance that can result from psychological trauma derived from threatening external situations has received much attention during the past century. These characteristic causes have now become entirely acceptable.

The issues of wounds due to injuries and of ill effects resulting from poisoning do, perhaps, require closer examination in that people often perceive them as examples of vicissitudes that befall organisms, that differ somehow from disease in that special names apply to them. Injurious blows and the consumption of toxic chemicals are, nevertheless, external causes of physical manifestations that have disadvantageous consequences and that not infrequently lead to states that everyone would agree are those of disease. Attempts to draw important distinctions between them and disease become somewhat tenuous, and indeed progressively less persuasive, the more deeply that one examines the situation. Certainly, the harm that injuries and poisonings cause to the organisms that they affect, and the ways in which society deals with their victims, differ little from the corresponding circumstances related to other types of disease: injured humans and poisoned humans receive medical care and social security benefits in just the same way as do other diseased humans, implying that society draws little practical distinction between the causal methods of these three types of vicissitude.

11 Section §5.4 below examines this aspect in greater detail.
A detailed examination of some of the implications of physical injuries provides additional support for this viewpoint. Consider the situation of a blow that produces a wound in an organism: perhaps the experience of a sportsman who injures his knee. The knee swells, is painful and immobile for several days, but then resolves. People who assess the situation a month after its occurrence find no evidence of sustained disadvantage, even though the blow was an external factor that had acted upon the sportsman and produced temporarily disadvantageous manifestations. They call it an *injury*. Should one classify sporting injuries such as this as members of some category of their own fundamentally different from disease, or are they really examples of a specialized type of disease dignified with its own name? An important determinant of the answer to that question is that physical wounds often produce late sequelae that everyone would instinctively describe as *diseases*. The same sportsman may have sustained the same injury to his knee; his joint may have swollen, become painful, and lacked mobility; and he may then have recovered promptly. The sportsman may have considered his condition cured a month after the injury occurred, but discovered a few years later that osteoarthritis was affecting his knee joint, the precursor of which was that very sporting injury, and the consequences of which were persistent disability. Everyone accepts that osteoarthritis is a disease. Its necessary external cause was the blow to the joint, so the blow was a cause of the
disease, and the resulting injury was an intermediate step in its development. The injury was, then, a brief manifestation of the disease of which the osteoarthritis was a more delayed manifestation. If one classifies the blow that caused the immediate reaction as a non-producer of disease, but that same blow that eventually caused the osteoarthritis as a producer of disease, the main distinguishing difference between them was the duration of observation before the judgement was made. The terminology nevertheless seems illogical when used in this way. The simplest method of resolving such situations is instead to accept that the word disease covers a broad range of circumstances that include specialized circumstances such as injuries. This would provide a reasonable explanation of how some trauma can eventually result in disease but other trauma not do so.

A detailed examination of the circumstances surrounding poisoning suggests that this also is a specialized variety of disease. Poisoning can occur under many different circumstances, including exposure of an insect to an insecticide, a person to an inadvertent overdose of a medication, or a person to a homicidal attempt upon his or her life. The time course of the consequences of the poisoning may be brief or it may be long: the organism may die within a few hours or it may survive for years and
display permanent deleterious effects.\textsuperscript{12} Few would dispute that those cases displaying chronic deleterious manifestations are examples of disease to which they could also comfortably apply the word \textit{poisoning}. Many people might, however, look upon the situation of organisms that die promptly as suffering from \textit{poisoning}, with the word \textit{disease} as a somewhat inappropriate description of their state. The situation in the latter, however, closely resembles the situation of the early consequences of physical trauma discussed previously in which a distinctive word (\textit{injury}) is commonly used to describe cases where the cause is obvious, immediate, and specialized; and the more general term (\textit{disease}) is commonly used to describe the chronic and often more obscure situation. It does, however, seem rather artificial to argue in favour of drawing major distinctions between cases in which a slightly different dose of the same poison results in prompt demise rather than in protracted ill health, in some people but not in others. The solution again here seems to be to adopt an encompassing approach of identifying all cases of deleterious consequences of poisoning as disease, whilst recognizing that in immediately obvious cases the more specific term is appropriate. The overarching consideration is that in each of these cases an organism is exposed to an external causal factor that interacts with it to produce

\textsuperscript{12} Some typical examples would be analgesic nephropathy resulting from the protracted consumption of compound analgesics; peptic ulceration resulting from the ingestion of aspirin; Cushingoid features resulting from corticosteroid therapy; or membranous glomerulonephritis resulting from gold treatment for arthritis.
disadvantageous physical and/or mental manifestations, which observers can diagnose.

The conclusion, then, that one would draw from this discussion is that causes of disease can take many forms. The organism itself is one. The external factors are others. Some external causes are overt, others covert. Some causes result in brief deleterious consequences, others in chronic disability. Some types of external factors, of which injuries and poisons are prime examples, have distinctive natures, to describe which people commonly use specialized terms. The consequences of the interaction between the external factor and the organism are nevertheless deleterious physical and/or mental manifestations that observers can diagnose, so they all support the present proposal as an appropriate definition of disease.

The implication that causes necessarily lead to an expenditure of energy

The fourth stipulative aspect of causation created by the present definition that needs some comment relates to the identification of the causal interaction as one that consumes energy. The mere co-existence of a living organism with a potentially injurious external agent does not, by itself, ensure the development of disease. They are both necessary prerequisites, but are not sufficient prerequisites, to result in disease since not only must they be present, but they must also interact.
A bee, to take an example, can sit on the skin of a human: the human (to use Aristotle’s terminology) is a potential material cause, the bee is a potential efficient cause, but no disease (whether a local red and painful lesion, or a generalized anaphylactic reaction) occurs unless the bee stings the human. The bee has, only then when it drives it sting into the human, performed an action. The conclusion must be that the action of creating a disease has required some factor(s), additional to the mere presence of the human and the bee in close apposition. The bee had to consume some energy in order to perform its action.

Two people, to take another example, can each consume large numbers of compound analgesic tablets; one of them develops analgesic nephropathy (a kidney disease that results from consuming many such tablets), but the other does not do so: something must therefore have occurred in the person who developed the disease differently from that which occurred in the person who did not. The organism and the external causes were (at least superficially) comparable in both, so some additional factor has differed. That factor was, upon analysis, the development of an active association in one person, but the failure of such a development in the other. The reason for the failure in this case was likely to be the presence of some subsidiary protective factor in the person who failed to develop the disease (perhaps having an habitually large fluid consumption
that diluted the concentration of the toxic chemical in the urine). This then suggests that disease only results when the circumstances of the multiple necessary causes allow and facilitate an interaction to occur between them in consequence of which their manifestations develop.

Awareness of the need for an intervening step that links the presence of causal factors with the manifestations of the causal process is far from new. Aristotle recognized this when he wrote of the meson (a word usually translated as the middle term) in the contexts both of logic and of causation. He used it in logic to describe a necessary item that links two statements of the type: ‘when A may be predicated of all B, and B may be predicated of all C, then A may be predicated of all C’, where he described B as the middle term, with A and C as extremes. The middle term was thus that through which the premise produced (or proved) the conclusion.\textsuperscript{13} He similarly described the role of the middle term in causation as\textsuperscript{14}:

Since we think we understand when we know the explanation, and there are four types of explanation (one, what it is to be a thing; one, that if certain things hold it is necessary that this do; another, what initiated the change; and, fourth, the aim), all these things are proved through the middle term.

The implication of the middle term here is that it describes that which links the antecedent components of the causal process to the succeeding

\textsuperscript{13} Grote, 1883, pages 148-149; Smith, 1995, page 36.
\textsuperscript{14} Posterior Analytics, Book II, 11, 94a 20-23.
components—similarly to its role in logic where it linked the premises with the conclusion. This idea is pertinent for the present theory of disease as here the middle term represents the pathogenetic mechanism that links the causes with the manifestations.\footnote{The German pathologist E.L. Wagner probably introduced the word \textit{pathogenesis} in his \textit{Manual of General Pathology}, the first use of it in English appearing in the 1876 translation of that work.}

What then is the essential nature of a pathogenetic mechanism? The concept is far wider than of a mechanism solely relating to disease. The work of C.J. Ducasse and his successors on general aspects of causation illustrated this.\footnote{Ducasse, 1924.} Ducasse critiqued many previous philosophical interpretations of causation as a prelude to developing his own formulation in which \textit{energy} was a key idea. He described energy as 'the power of any kind of thing to cause any kind of change or state in any other thing'. Energy was, fundamentally, power. Jerrold Aronson later developed this further by suggesting that, when two objects come into contact in the process of causation, a quantity (such as velocity, momentum, kinetic energy, or heat) transfers to one of them that thereby transforms into the effect-object, which has sustained an unnatural change.\footnote{Aronson, \textit{Synthese}, 1971; Aronson, \textit{Studies in History and Philosophy of Science}, 1971.} W.V. Quine expressed the same idea in the following terms\footnote{Quine, 1973.}:  

\begin{quote}
\textit{[\ldots]}\textit{energy was a key idea. He described energy as ‘the power of any kind of thing to cause any kind of change or state in any other thing’.
Energy was, fundamentally, power. Jerrold Aronson later developed this further by suggesting that, when two objects come into contact in the process of causation, a quantity (such as velocity, momentum, kinetic energy, or heat) transfers to one of them that thereby transforms into the effect-object, which has sustained an unnatural change. W.V. Quine expressed the same idea in the following terms:} \end{quote}
The imparting of energy still seems to be the central idea [to the causal idiom]. The transfer of momentum from one billiard ball to another is persistently cited as a paradigm case of causality. Thus we might see a simpleminded or root notion of causality in terms of the flow of energy. Cause and effect are events such that all energy in the effect flowed from the cause.

The exchange of energy has subsequently remained an attractive proposition to many philosophers for the explication of causation. It also provides a useful basis upon which to found a definition and explication of disease. An independent understanding of the nature of energy is however necessary if this idea is to fulfil its greatest potential since Ducasse’s explanation of causation suffers from the weakness of making it true by definition that cause involves energy due to his making a somewhat circular argument by defining energy in terms of cause. I do not however believe that this creates an impossible impasse as it does appear

---

19 Ducasse himself expanded upon it in a later work (1926). Aronson (1982) ably defended his thesis against several attacks that related particularly to the direction of causation and the direction of transfer of energy. David Fair (1979) argued in favour of ‘a physicalistic reduction of the causal relation to one of energy momentum transference in the technical sense of physics. Adrian Heathcote (1989) applied quantum field theory to explain the interactions that occur between the causal objects, and later tried to reunite the physicalist explanation of causation with laws of nature by using an a posteriori identification of repeated causal sequences (Heathcote and Armstrong, 1991). Phil Dowe (1992a and 1992b) supported the idea of transfer of energy when he proposed that ‘A causal interaction is an intersection of world-lines which involves exchange of a conserved quantity’ using Minkowski diagrams. Wesley Salmon (1998, pages 253-259) even went so far as to abandon his previously strongly argued views about causation and to accept Dowe’s views, albeit with modifications.
feasible to me to develop an independent understanding of energy, a task that I shall address in §5.2.

The conclusions, then, that one might reach about some potentially controversial issues of causation of disease are several. Firstly, causes are almost certainly necessary; albeit with the presumption that the idea of idiopathic, essential, autoimmune, and other comparable types of disease as being causeless is more likely to exemplify contemporary ignorance than to prove that causeless diseases occur. Secondly, causes of disease are necessarily multiple, in that both the organism that sustains the disease and some external factor(s) that act(s) upon that organism must exist before disease can occur. Thirdly, causes of disease may take various forms, both covert and overt, including (in the broadest sense of the use of the term disease) such specialized varieties as injuries and poisons. Fourthly, causes of disease necessarily lead to an expenditure of energy before the manifestations of disease can appear. Acceptance that a reasonable case supports these items removes most grounds for objection to a definition that stipulates them.

§5.2 PATHOGENESIS

I presented in §5.1 some reasons for claiming that the expenditure of energy plays a central role in the creation of disease. This role, however, whilst appearing superficially simple, becomes progressively more
complex on closer examination. Some reasons for this include the need to understand the nature of the thing called energy; the occurrence of sequential causal processes in a diseased organism; and the layered levels of understanding that people develop about those processes.

The formal analysis of the idea of energy is a relatively modern phenomenon despite the classical implications that the Greek derivation of the word implies. Most contemporary physicists define it in such terms as ‘The capacity for doing work’. The type of work that it originally described was that done by kinetic energy (‘the capacity for doing work that matter possesses because it is in motion’), but it soon extended to include potential energy (‘the capacity to do work that a body or system has by virtue of its position or configuration’) and internal energy (‘the sum of the kinetic energy of the molecules of a system and the potential energy of intermolecular attraction’). Varieties of internal energy include heat, chemical, electrical, nuclear, and mass energy. The various types of energy can transform between each other, although the way in which they ultimately achieve this at a subatomic level remains somewhat

---

20 The word is derived from the Greek *energeia*, that in turn came from *en*, ‘in’, and *ergon*, work’. The word is not itself of classical provenance, having first been used in a physical sense in 1807 by Thomas Young (1773-1829) to identify a relationship between the mass of a body and its velocity.

21 Roller and Nedeisky (1976). This and each of the following definitions come from this source.
The suggested existence of a finite quantity of energy led to proposals that energy is conserved, and that it can be neither created nor destroyed. Experimental physics have subsequently supported these principles, but have not excluded the possibility that future empirical observations may challenge their veracity. Such an objection, however, raises general issues beyond those relating merely to disease, and far beyond the scope of the present discussion, so it seems reasonable to accept for the present purposes that energy exists, is convertible, and is conserved.

The generation of disease involves various forms of energy, depending upon the type of disease. The simplest to understand is kinetic energy: if an external object travelling at high speed comes to a halt buried within an organism upon which it has impacted it transfers its kinetic energy to the organism and the resulting injury that the organism sustains is a manifestation of that transfer of energy. The implications of the transfer of some other forms of energy are similarly obvious. A nearby fire will transfer some of its heat to an organism and thereby create disease: this may take the form of redness of the skin if the heat is only modest and

---

22 Sir Benjamin Thompson (also known as Count Rumford, 1753-1814) first suggested this possibility in 1798 and the principal has subsequently received widespread support.

23 The First Law of Thermodynamics and The Theory of the Conservation of Energy: the physicist, James Joule (1818-1889) and the physicians, Julius von Mayer (1814-1878) and Hermann von Helmholtz (1821-1894)
localized, of a penetrating burn if the heat is intense but focussed, or of
generalized hyperthermia if the heat is protracted and generalized. Exposure to nuclear energy can cause local tissue damage if localized and modest, persisting local structural and functional alterations if more intense, genetic mutations if generalized and yet more intense, and death if even more severe. Many toxic substances harm organisms by transferring chemical energy, such as when invading microbial pathogens act on host organisms by releasing chemical exotoxins or endotoxins. These, then, illustrate a few of the many ways in which external factors can interact with organisms to generate disease by transmitting energy.

An aspect arising out of the transformation of energy is that of entropy, which also has implications for the present definition of disease. Certain experimental evidence demonstrates that although many forms of energy can inter-convert, some thermal energy cannot convert into mechanical work. A consequence of this is ‘that in any process in a completely isolated system, the entropy of the total system must increase’ where entropy represents the non-convertible thermal energy.\(^{24}\) This (for reasons that are beyond the scope of the present discussion to examine in detail) then provides an explanation for the direction that the causal process invariably takes in time, with causes always preceding effects, rather than

\(^{24}\) Undertook some of the key research into this topic in the 1840s (see Theobald, 1966, pages 55-59). Crawley (1975), pages 44-45.
vice versa. 25 An application of this idea to the present definition of disease explains the invariable direction in time in which disease develops, with the causes necessarily preceding rather than following the manifestations.

The development of the final manifestations of disease in an organism often involves sequential causal processes. An external factor or factors act(s) upon an organism to initiate the disease process, and some time (that may be anything between picoseconds and years) elapses before the final manifestations appear. The pathogenetic interaction occurs between this beginning and this end, but is often not merely a single interaction: indeed, it is only rarely a single interaction. The developments that occur in living organisms, rather, are usually a series of intervening pathogenetic interactions, the ultimate outcome of which is the development of the final manifestations. Each of these intervening pathogenetic interactions then consists of the organism in which the disease is occurring (corresponding to Aristotle’s material cause), an external cause (a manifestation of the preceding causal process that may indeed lie within the organism, but is foreign to its usual biological functions; and which corresponds to Aristotle’s efficient cause), an interaction between the two, and one or more manifestations (often covert, but nevertheless identifiable by observers who have at their disposal appropriately sensitive testing apparatus) that in turn become the

‘external’ (or efficient cause[s]) of a further pathogenetic interaction or interactions in the continuing chain that eventually leads to the final manifestation or manifestations. This sequence appears in the causal diagram (see §4) as $C_E$, the original external (or efficient) cause interacting with the organism ($C_M$) to produce the first pathogenetic event ($P_1$) and thereby the first manifestations ($M_1$). Each of these then becomes an efficient cause that interacts with the organism ($C_M$) in a second pathogenetic event ($P_2$) to produce new manifestations ($M_2$). That sequence may thereafter recur, and often does so on numerous occasions, until the final manifestations ($M_F$) eventuate.

One can examine any well-defined disease process to get some idea of how adequately this theoretical model reflects reality. Take, as an example, post-streptococcal glomerulonephritis. This disease results from the infection, usually of the throat or skin, of some (but not all) humans with a strain of streptococcal bacteria. Currently available information suggests that these organisms release a toxic chemical, that this acts as an antigen to stimulate an immune reaction in the host, that the host’s lymphocytes (a particular family of cells) produce antibodies, that the antibodies combine with antigen and complement molecules to form immune complexes, that the immune complexes travel in the circulation to the glomeruli of the host’s kidneys (the structures that filter
the urine from the blood) where they become deposited. Additional evidence suggests that the deposited immune complexes in the glomeruli stimulate the production of cytokines that attract other circulating cells (polymorphonuclear leukocytes) into the glomeruli and also damage existing cells and membranes there. These developments are visible on histological preparations of the glomeruli. The damaged glomeruli leak blood into the urine in most victims, protein into the urine in some victims, and fail to excrete other toxic chemicals in some victims. The bleeding can manifest as obvious blood in the urine; the leakage of protein can cause a decrease of the protein level in the plasma that then allows water to seep into the tissues and so can manifest as wide-spread swelling of the tissues; and the failure of clearance of toxic chemicals can cause these to accumulate in the body generally where they cause secondary damage that manifests in such ways as vomiting, itch, and coma. The clinically apparent disease of post-streptococcal glomerulonephritis usually takes between about ten and twenty days to develop from the time of infection to the time of development of clinically obvious symptoms.

The streptococcal bacteria and the host are the causes in this example. The fact that only some humans appear to be susceptible suggests that some genetic factor in humans may help identify susceptible people. The interaction between the susceptible host and the bacteria leads to the first

---

26 The technical information presented in this example is based upon the
pathogenetic process, the manifestation of which is the release into the host of antigens. The antigens then become the efficient cause for the second pathogenetic process when they interact with the host’s lymphocytes to stimulate the production of antibodies that then become the manifestations of the second pathogenetic process. The combination of the antigens with the host’s antibodies and the host’s complement become the third pathogenetic process, with the production of circulating immune complexes as the manifestation of this. The deposition of the circulating immune complexes in the host’s glomeruli becomes the fourth pathogenetic process, with the histologically demonstrable presence of these in the glomeruli (as identified by light microscopy, immunofluorescent microscopy, and electron microscopy) as the manifestation. The induction of release of cytokines by the host’s cells in response to these deposited complexes becomes the fifth pathogenetic process, the detection of which by appropriate methods is its manifestation. The action by the cytokines on the host to induce the infiltration of polymorphonuclear cells is the sixth process (the manifestation of which is also identified by histology); whilst the action by immune complexes to damage the glomeruli is the seventh process, the manifestations of which are the leakage of blood and protein and the failure to excrete other chemicals that the victim would usually excrete. The excessive excretion of protein in the urine of the host is the eighth description of this disease provided by Couser and Johnson (1997).
pathogenetic process, the identification of which by appropriate chemical methods is its manifestation. The consequent lowering of the plasma protein level in the host is the ninth pathogenetic process, the manifestation of which is the development of generalized swelling. This outline, then, suggests the practical way in which the present suggestion of sequential pathogenetic processes works—and central to all of them is the important element of expenditure of energy.

This description of the mechanisms by which a bacterial infection damages a host also draws attention to an important aspect relating to the pathogenesis of disease. External agents appear, at first sight, to damage organisms by widely differing mechanisms. Thermal injuries are superficially quite different from nutritional, and these are quite different from chemical. Poisons appear to damage organisms in different ways from infectious agents, and these differ again from the mechanisms operative in genetic diseases. Each of these injuries does, however, damage cells, and that damage is often identifiable at a microscopic level. Two or more levels of examination are therefore possible, and whilst vast macroscopic differences can exist between them, some unexpected cellular similarities can pertain. Cells nevertheless display diverse and

27 One must emphasize that, as additional information accumulates in the future with ongoing research into the processes involved in this disease, details of the intervening pathogenetic processes may alter, however the basic principles will remain unchanged. A similar proviso would apply to any particular disease that one might choose to examine.
distinctive patterns of reaction to injury. Each cell, furthermore, contains many subcellular organelles—including a nucleus, nucleolus, chromosomes, endoplasmic reticulum, and mitochondria—at the level of one or more of which damage may have occurred that left other subcellular organelles unimpaired. Similarities and differences may thereby appear only at the subcellular level.

The different levels of understanding facilitated by examination at levels of magnification involved in the pathogenetic mechanism suggests that successive layers of understanding may exist and become discernable depending upon the resolution of the investigative equipment available to the observer to examine the energy-consuming ways in which disease develops. The depth of examination certainly does not even stop at some superficial level as chemical mechanisms and methods of analysis play important roles. Appropriate apparatus will display molecular mechanisms that will in turn resolve into atomic mechanisms, and then into subatomic ones. This resolution of the mechanisms into ever more refined layers of causality presumably eventually reaches a level of analysis in which the pathogenesis involves little more than an exchange of energy between fundamental particles that in turn are themselves merely a form of energy. This raises the possibility that somewhat similar, if not identical, interactions underlie all pathogenetic mechanisms, with a commonality to
all forms of disease, but in which the distinctive manifestations are the consequence of the distinctive causes.

An analysis such as this suggests a possible ideation of disease that exists in layers of explanation, each layer legitimate and complete in its own right, but with each underpinned by an ever more fundamental layer. One method of description of this is that when a particular organism develops disease it does so because some part of it fails to function in its usual manner (*the functional level of disease*). This part is, in humans and animals, often an organ. When the organ fails to function it does so because a particular external factor has acted upon it, expending energy to cause a particular structural derangement (*the organ, or structurally macroscopic, level of disease*). That structural derangement produces, at a dissected macroscopic level, damage to particular tissues within the organ (*the tissue level of disease*). Those tissues are deranged because the cells that compose them have suffered damage (*the cellular level of disease*). That damage to the cells has occurred because of disorganisation of various electron-microscopically visible organelles that exist within the cells (*the organelle level of disease*). That disruption to organelles has occurred because of the action of certain complex molecules such as the cytokines that have acted upon them (*the molecular biological level of disease*). Those molecules consist of
biochemical components such as genes (the biochemical level of disease). Those biochemical components consist of individual atoms (the atomic, or chemical, level of disease). They again, in turn, come about by the conversion of energy (the transfer of energy level of disease). Each of these levels is explanatory in its own right, yet none can ultimately produce manifestations of disease without involving most or all of the others. An external (or efficient) cause acts upon the organism (or material cause) in each one of them to produce an effect (or manifestation), and the expenditure of energy occurs at every step along the way.

Another implication follows from that analysis. The historical overview provided in §1 above, of the development of ideas about disease, demonstrated quite emphatically the distinctions that have occurred between the methods of investigation used by people working in various scientific fields related to disease—fields as diverse as anatomical pathology, microbiology, biochemistry, immunology, molecular biology, endocrinology, genetics, psychiatry, and sociology. Each of these fields has developed its own techniques of examination of the information available. Each has tended to emphasize the centrality of its conceptualisations. Each, however, is legitimate in its own right, such that insights provided by each have penetrated yet another layer of obscurity that has enveloped the totality of the topic. The distinctive (and largely

---

Sterelny and Griffiths (1999, pages 112-148) discuss the idea of layering
non-interchangeable) methods of investigation used by those interested in each layer have reinforced the successive layers of understanding of disease. This layering has tended initially to promote confusion rather than clarity. A comprehensive understanding of the whole process, however, requires an ability to transcend the layers if one is to perceive the interrelationships created by the sequences of pathogenetic mechanisms that typify every example of disease that inevitably incorporates aspects of most, if not all, of the layers: and that understanding requires an awareness that every change that occurs within the organism involves an energy-converting event.

§5.3 MANIFESTIONS

Disease manifests in so many ways that these defy easy tabulation. They vary from species to species, and from organism to organism, reflecting the unique nature of each pathogenetic interaction peculiar to the individual in which disease occurs. Interactions often produce manifestations that initiate further items in a pathogenetic chain linking the earliest causes via intermediate manifestations to the final manifestations. Each such chain is unique to the organism in which it occurs. Manifestations do, however, have some general aspects that warrant notice. They include the issues of whether one might reasonably perceive disease merely as manifestations and as nothing else; whether disease

of explanation in contexts such as this.
can exist without manifestations; of the fundamental nature of manifestations; of their quantification; and of their potential duration.

Disease as more than mere manifestations

Can one reasonably consider disease as manifestations and as nothing else? Some people\textsuperscript{29} have portrayed disease as consisting of nothing more than a manifestation (or set of manifestations)—focussing sometimes on subjective symptoms, sometimes on structural abnormalities, and sometimes on malfunction. They have perceived the manifestations as the disease, suggesting that nothing else is relevant. Others, however, have branded this approach as over-simplistic and unrealistic since it immediately begs the question of the origin of the manifestations. If the disease is nothing more than the manifestations, then the manifestations presumably either have no origin, or their origin is irrelevant to an understanding of disease.

The idea of a spontaneous generation of manifestations (if one were to consider disease as nothing more than manifestations) appears intrinsically unpersuasive for reasons similar to those set out in §5.1 above. Here, as there, some preceding process must surely initiate each observable item that appears. An examination of various examples of disease in the everyday world supports the validity of this approach,

\textsuperscript{29} See §1.
although—as the reasoning that applies to them all varies little—a single example will suffice to illustrate the point.

Consider the circumstances of a densely black lump that suddenly develops on the skin of a person with a fair complexion. We all accept that this is a disease of some sort. Its physical substance is, as a manifestation, an obvious part of our idea of it as a disease. Those with unenquiring minds might perceive it as all there is to the disease. Why, though, has it suddenly appeared on that part of the skin of this particular individual? Unenquiring people presumably assume that nothing predisposed to its development and that it will have no adverse consequences beyond the mere cosmetic. Its unexpected development might, however, stimulate others to examine whether some identifiable factor was involved in its genesis. Do other fair-skinned people who suddenly develop black lumps have similar factors in common? Some additional pieces of information become available. Histological examination of the lump shows that it has the cellular characteristics to which pathologists apply the name *malignant melanoma*. The person who has developed it is of European ancestry and has worked in a sun-exposed occupation in the Australian bush throughout his life. The lump is on a sun-exposed part of his body. Accumulated data about people who have developed similar lumps with similar cellular characteristics suggest that they mostly have fair complexions, and had exposed the parts of their
skin in which their melanomas developed to sunlight over protracted periods of time. Progressive accumulation of such information has led to the development of a consensus of opinion that excessive exposure of fair skin to ultraviolet irradiation present in sunlight is likely to cause the development of black lumps that go under the name of malignant melanomas. The disease would not have occurred without exposure to the irradiation, so the irradiation becomes a necessary part of the total disease process. The irradiation must have induced some invisible, and possibly quite complex, changes to occur in the cells of the skin that resulted eventually in their conversion to black colour. Longer observation, furthermore, of the person in whom the malignant melanoma has developed reveals that the distinctive cells that constitute it have an ability to disseminate to distant parts of the body and start growing there, eventually sapping all the strength of the patient and leading to his death. Such reasoning then makes the suggestion that the mere manifestation of a black lump as representing the totality of the disease seem inappropriately superficial.

The reasoning applied in this example of the case of a person who developed a black lump on his skin is typical of the reasoning that one can apply to every case in which a manifestation of disease occurs. The manifestation in each of these cases becomes merely the trigger for analysis of the processes that have led up to its appearance and also of
further processes that its development may initiate that will result in further manifestations.

Such circumstances then create significant difficulties for anyone who would suggest that any given manifestation of disease is the totality of the disease.

*Unmanifested disease*

Can disease exist without manifestations? This question is the converse of the previous one. The present definition claims that manifestations are signifiers of disease and that the manifestations are an integral part of the idea of disease. It does not, however, go so far as to claim that they must be obvious. It indeed accepts that a disease can occur with presently unidentified manifestations provided that these meet certain criteria. The most important of the latter are that the manifestations will be demonstrable at some future time, that observers will then judge those manifestations as disadvantageous to the organism, and that the manifestations are the eventual consequence of a pathophysiological interaction between an external factor and the organism. The definition is very broad in permitting observers to use any available form of technology to identify the manifestations.
Many manifestations are readily discernible, so for them such considerations are irrelevant: disease is present. The definition clearly disallows the possibility of disease in the total absence of manifestations, overt or covert, now or in the future: in such circumstances there is no disease. A more difficult situation exists in the case of conditions where no manifestations are presently overt, but where they will eventually develop and where they presently exist if observers use appropriate methods of examination. The observation of such covert manifestations often depends upon the availability of appropriate technology with which to observe them and an understanding of the potential implications of the observations made with that technology. The situation is therefore somewhat akin to that of causes (see §5.1 above) where the definition accepts that covert and overt examples can occur, but anticipates that the number of covert examples tends gradually to decrease and the number of overt examples gradually to increase as people apply increasingly sophisticated techniques of observation. An example will illustrate the application of the proposed definition to this situation.

Consider the example of a woman who is bitten by a mosquito whilst travelling in New Guinea. She feels quite well for the following twelve days then develops a fever with rigors. A blood test reveals the presence of malaria parasites. Everyone accepts that she has the disease of malaria. The mosquito that bit her was presumably infected with this and injected
some parasites into her blood. She was in what people describe as the *incubation period* during the intervening twelve days. Most people would furthermore accept that she had the disease of malaria during that period, but that it was not overtly manifested. If, however, someone had taken a blood sample from her during that period and examined it appropriately, they might have observed some malaria parasites in occasional of her red blood cells, in which case the previously covert manifestation (the unobserved parasitised red cells) would have become overt. Disease, then, existed throughout this period, but it remained unmanifested unless or until she developed symptoms that prompted someone to examine her blood film. This example thus demonstrates that unmanifested disease can occur.

*The nature of manifestations*

Another issue of interest, namely the fundamental nature of manifestations, also emphasizes the importance of the additional necessary criteria of disease. The definition is, on this issue, simultaneously broad in some aspects, but specific in others. Its breadth comes from its acceptance that a manifestation can be any physical or mental characteristic that develops in an organism. Its specificity comes from the additional criterion by which it focuses that acceptance, and particularly its requirement that the process that culminates in manifestations must disadvantage the organism. The judgement of
disadvantage is itself potentially quite broad (to the extent that it is within the individual assessment of anyone who beholds the organism). Here again, however, the definition and its corollaries create some specificity by obliging the interested observers to reach a consensus that disadvantage exists. The practical consequence of these provisos is to make manifestations objective items in that they are physical and/or mental characteristics, but social items in that their interpretation occurs in a social context.

**The quantification and duration of manifestations**

Are manifestations quantifiable; and, if so, is there a need to quantify them? Some manifestations of disease are clearly serious: the individual who lives in a permanently vegetative state after suffering a severe stroke provides an obvious example. Other manifestations are quite mild: the port-wine stain created on a person’s forehead by a haemangioma carries no greater disadvantage than the cosmetic consequence of appearing unsightly. Observers must judge how seriously disadvantageous the

---

30 For more details about the observers and their diagnoses, see §5.7 and §5.8 below.
31 The example of the port-wine stain illustrates the potential fluidity of the interface between disease and non-disease, since occasional people (for example, perhaps Mr Mikhail Gorbachev) have managed to use such superficial blemishes to their personal advantage by achieving improved public identification that has assisted them at least temporarily in their political and other social aspirations. One could well suggest that when circumstances make the manifestation advantageous they disqualify it from acceptance as disease, at least for the duration that it remains advantageous.
implications are of any particular manifestation before they can label a condition as a disease. They often find that, whilst never doubting the presence of disease, they also have to measure the amount of disadvantage that it creates. An obvious example of this occurs when financial compensation becomes necessary as a result of negligence on the part of some person who is accused of having contributed to the development of the disease. Law courts often face the task of having to quantify the level of disadvantage suffered by an individual (whether human or animal) to quantify the level of compensation that the instigator of the disease must make to the patient or animal’s owner.

The possible duration of manifestations is another topic of interest. Manifestations can be overt or covert, brief or sustained, or any combination of these qualities. Some examples will demonstrate this.

(i) A woman travelling on a railway train is stung by a bee that flies in through a window: she immediately starts to wheeze as she tries to breathe, but knowing that she has an allergy to bee stings gives herself an injection of adrenaline from a syringe that she always carries with her, and a few minutes later is entirely free of symptoms. The manifestations of her disease were both overt and brief.

(ii) A woman travelling on a railway train hits her head on a metal pillar when the train suddenly brakes: she immediately loses consciousness and develops weakness down one side of her body. The weakness persists
even though she regains consciousness. Investigations show that she has
developed a stroke due to a sub-dural haematoma, the manifestations of
which persist for the remainder of her life. The manifestations of her
disease are both overt and sustained.

(iii) A woman travelling on a railway train inhales some viral particles
coughed out by a fellow passenger: two days later she develops a sore
throat with fever, cough, and sputum. The symptoms persist for five days
before she recovers completely. She has suffered from influenza, the
manifestations of which were briefly covert, then briefly overt.

(iv) A previously well woman travelling on a railway train suddenly slumps
to one side and is found to have developed weakness that persists down
one side of her body throughout her remaining life despite intensive
treatment. Investigations carried out upon her arrival at hospital
demonstrate that her illness is due to a brain abscess that must have
developed asymptotically over the preceding few days. Her disease
was covertly brief, but became overtly sustained.

(v) An apparently previously well woman travelling on a railway train
suddenly lost consciousness and died before she arrived at a hospital. An
autopsy revealed that she had suffered sub-arachnoid bleeding from an
aneurysm (a weakness in the wall) of an artery in her brain.

Understanding of the natural history of such conditions suggests that the
aneurysm must have developed many years earlier without causing any
overt manifestations, but for some reason it suddenly ruptured and bled,
causing her to die immediately. She had covertly sustained, but overtly brief, disease.

(vi) A middle-aged woman had grown up in a home in which her father had suffered from tuberculosis. She first developed cough, sputum, breathlessness, and the coughing of blood as a teenager. Investigations showed that she suffered from tuberculosis. She noticed that, despite treatment, she still experienced breathlessness when she exercised many years later. Her disease, the manifestations of which were for long covert, subsequently remained overt for a prolonged period of time.

These several examples demonstrate that disease can have manifestations that are overt or covert, brief or sustained. None of these differences have great implications with regard to the general idea of disease, important as they are to the particular organism that suffers from disease. They nevertheless emphasize that disease requires manifestations, but that it is more than just manifestations; that the fundamental nature of the manifestations is objective, but that they must be interpreted in ways that satisfy a consensus of observers; that they are quantifiable; and that they can have highly variable durations.

§5.4 MENTAL DISEASE

The issue of mental illness has always posed an important challenge for anyone seeking to define and explicate disease. The brief outline
presented in §1.4 above of the historical development of ideas about mental disease demonstrates the huge diversity of opinion on the topic that any definition must encompass. Views have ranged from the claim that no such thing as mental disease exists, to the claim that everyone is a little mad; and from the claim that all mental disease is psychogenic, to the claim that it all has a physical basis. The chasms of understanding that divide these claims, and the potential legal implications that they pose as society grapples with the issues surrounding the mentally ill, make the crafting of a definition as important as it is difficult. I suggest, however, that the present definition does manage to identify an acceptable path through this intellectual maze, and certainly a more acceptable path than have many previous definitions.

The present definition does not legislate that mental disease exists. It does, however, indicate that if mental disease exists it must develop in a living organism by means of a process that has causes and that has consequential manifestations. Such a process must, as demonstrated by the previous discussion (see §5.2 above), take the form of a pathogenetic mechanism. The overwhelming majority of people believe that mental disease exists, and the overwhelming majority of those who do so accept that pathogenetic mechanisms and causes lead to the manifestations: where they differ is over the nature of those mechanisms. An outstanding advantage of the present definition is that it succeeds in avoiding their
arguments as it does not, and indeed does not need to, legislate on the
nature of the pathogenetic mechanisms: it merely necessitates some form
of pathogenetic mechanism in each case of mental disease. It equally
allows the mechanisms to be psychic or physico-chemical, favouring
neither, and drawing no fundamental distinction between either of these
alternatives.

The definition does nevertheless require that the pathogenetic
mechanisms have causes and have disadvantageous manifestations of a
physico-chemical and/or a psychological type. The causes of most
diseases generally classified as ‘mental’ are presently either securely
established or are postulated by theories that have varying degrees of
supportive evidence. The existence of the living human being (perhaps
also occasionally the higher animal) in whom mental disease occurs is a
necessary component in the causal process. External factors with which
the organism comes into contact, or of which it develops deficiencies, are
often also identifiable. Some examples are repeated minor head trauma in
professional boxers making them ‘punch drunk’; Treponema pallidum
bacteria in tertiary syphilis, manifesting as ‘general paralysis of the insane’
or as delusions of grandeur; alcohol consumption in Korsakoff’s
psychosis, or its sudden withdrawal in delirium tremens; narcotic use by
people with addictions; consumption of medications in various
hallucinatory and paranoid states; vitamin B12 deficiency in pernicious
anaemia due to failure of intestinal absorption, in turn the consequence of
damaged gastric mucosa; and schizophrenic and manic-depressive states
in Wilson’s disease, where a genetically-induced failure to excrete copper
results in cerebral accumulation of that mineral. The pathogenetic
mechanisms in certain other types of mental disease involve rather more
complex sequences of steps, many of the details of which presently
remain obscure, but in all of which one or more external factors ultimately
operate. An example of these occurs with various forms of senile
dementia and Alzheimer’s disease, where combinations of genetic and
acquired factors interact to impair cerebral blood flow and mediate a
variety of cellular changes. They often marshal a multitude of external
causal factors that include excessive cholesterol consumption causing
atherosclerosis, excessive salt consumption causing hypertension,
excessive sugar consumption destabilizing diabetes, and probably many
others.

The external causal factors in many cases of the major psychoses (such
as schizophrenia and manic-depressive psychosis) still remain
controversial. This does not, however, necessarily imply that they will
forever remain hidden. The twin studies mentioned in §1.4 above, the
identification of ever more classes of causative chemicals, and the
response of patients suffering from these diseases to chemical treatments,
all suggest that mechanisms and causal agents of a physical and/or chemical nature will eventually emerge for each of them.

The genetically mediated group of mental diseases form another important class. The ability of the definition to accommodate them, whilst at first perhaps seeming to defy explication through external causal factors, will stand or fall on the same grounds as it will over other genetic diseases (see §5.10 below).

The minor emotional states, such as anxiety and many cases of depression, might at first sight seem not to conform to the present general definition on the grounds that they lack obvious external causal factors. This, however, is only the case if one confines one’s search for such factors to those of a physical or chemical nature. The proposed definition does not, though, limit the external factors in this way. It clearly accepts that external psychic influences can act upon a human to initiate purely psychological interactions that manifest in ways (whether emotional or physical) which observers consider disadvantageous, and that these are what people usually include within their idea of mental disease. A child, for example, whom a parent repetitively lashes with unjustified verbal abuse may react by developing an anxious personality and subsequently lack the ability to cope with stresses that others reared in more caring environments will easily accommodate. That anxiety state may then
manifest, when the child-become-adult is exposed to emotional stress, as inappropriate utterances, as avoidance of threatening situations, or as physical symptoms such as tachycardia. Exogenous depression is a similar situation, where affected individuals react in a psychological and psychosomatic manner to the awareness of some threatening or unpleasant information for which they have no more adequate alternative defence mechanisms. The external pathogenetic agent in cases such as this is the unpleasant information directed from an external source that impinges upon an individual who then develops the disease of exogenous depression.

The proposed definition, whilst accommodating both physical and psychic external stimuli, does require that the manifestations of pathogenetic mental interactions are disadvantageous to the individual in which they occur before they qualify for the appellation of disease. This protects it from labelling various advantageous mental interactions as disease. An example that clarifies this situation is the difference between fear and anxiety. The traditional distinction between these conditions is that fear is a word used to describe an awareness of apprehension and dread that has an identifiable stimulus, whereas anxiety is a word used to describe an awareness of apprehension and dread that lacks a clearly identifiable stimulus.\(^\text{32}\) The emotional manifestations displayed by a passenger in a

---

\(^{32}\) Spraycar (1995), pages 108 and 634.
motor vehicle that swerves out of control may be identical to those displayed by another person every time that the latter travels in a motor vehicle. The reaction in the first instance is one that most observers would label as justified fear, would judge as advantageous in that it might stimulate the passenger to take life-protective actions, and would not describe as a form of mental disease. The reaction in the second instance, in contrast, is one that most observers would label as anxiety, would consider as inappropriate, would judge as disadvantageous in that it might inhibit the passenger from making beneficial journeys, and would speak of as a form of mild disease. The proposed definition identifies that latter situation as disease, whilst not permitting this terminology to apply to the former situation. It thus facilitates important distinctions such as these between situations that most people would perceive as everyday psychological reactions and other situations that they would perceive as mental disease.

Many previous attempts to define and explicate mental disease have tended to deal with it separately from somatic disease. They have, by drawing that distinction, created a host of difficult issues, not least of which is where to draw a distinction between mental and physical disease. These have provoked telling criticisms of many of the previous definitions of disease outlined in §2 above. A peculiar advantage of the present proposal is that it does not need to draw any such distinction and so
navigates through these dangerous waters by stipulating a formal template that is as applicable to mental as it is to physical disease, with limits that are sufficiently broad as to accommodate conflicting theories of the mechanisms of mental disease without simultaneously favouring any particular one of these. An outstanding mark of this is that it can accommodate Freudian theory about the genesis of mental disease as well as it can accommodate purely chemical and physical explanations. Its strength comes through requiring that some mechanism must exist to explain the linkage of the cause with the manifestations of the disease while taking an entirely neutral position as to the nature of the mechanism or mechanisms that intervene so long as they occur. The adoption of this position protects it from the challenges that many previous definitions of mental disease in particular, and also of disease in general, have encountered.

§5.5 ASYMPTOMATIC DISEASE

People who examine the presently proposed definition of disease may question its pertinence in view of the frequent use in medical discourse of the term asymptomatic disease.\(^3\) The implication of this term is that some cases of disease have no symptoms. One might then reasonably ask how observers can diagnose something that has no symptoms, and how it is

\(^3\) This topic relates closely to that of unmanifested disease that was discussed in §5.3 above, but raises some additional points that are worthy of discussion.
that something without symptoms can disadvantage a person? Could one not suggest that something that fails to produce symptoms thereby fails to fulfil the criteria of disease nominated by the present definition, and so that the presently proposed definition fails?

An examination of this issue requires assessment of whether asymptomatic disease can indeed occur. Is the problem merely that the commonly used terminology misleads by suggesting that asymptomatic disease occurs when in fact no reasonable examples exist of such a phenomenon? The answer to this question is clear: the proposition that symptomless disease does occur seems established beyond doubt. Take, for example, lung cancer. No reasonable person would deny that lung cancer is a disease. A man may have a cancer present in his lung for many months before it causes any symptoms. He is not aware of it and nor is anyone else. People can then reasonably describe him as having had asymptomatic disease during the period that passed between the time when the lump first developed and when he first became aware of it. This example is, furthermore, far from unique, as many other comparable examples of asymptomatic disease present themselves. The available evidence therefore suggests that asymptomatic disease does indeed occur.
The presently proposed definition, however, avoids difficulties in this regard by using the word *manifestations* rather than *symptoms*. Symptoms are merely one form of manifestation, so the questions next arise of whether other types of manifestations can occur apart from symptoms, and whether asymptomatic disease may fall within the definition by displaying these other manifestations rather than symptoms. Here again the answers are clear: other forms of manifestations apart from symptoms do occur, and their existence brings at least some asymptomatic disease within the present definition. The evidence for this is that contemporary clinicians use the word *symptom* to contrast with the word *sign* (often sub-specified in terms such as *physical sign* or *radiological sign*), following a convention established in the early 19th Century by the French physician, René Laennec (1781-1826), who tried to transform medicine into an objective science. He considered that the views of patients often obscured, rather than elucidated, an understanding of the processes affecting them. He therefore promoted the idea of physical signs identified by qualified observers as objective and reliable, in contrast to the symptoms that patients experienced, that he dismissed as subjective and unreliable.34 The result of this is that people nowadays generally take a *symptom* to mean a subjective physical or mental change that a someone perceives when suffering from disease; in contrast to a *sign*, that they take to mean a physical or mental change that an external observer identifies in

---

34 The term *sign* had prognostic implications before the time of Laennec,
a victim of disease independently of the victim’s own perceptions. The man with lung cancer previously mentioned as an example of asymptomatic disease may have particular abnormalities that observers can discern by physical examination or by radiological techniques. Signs then become the manifestations of disease in the absence of (or, indeed, often while awaiting the development of) symptoms.

But is one justified in saying that disease is present before anyone applies the necessary technology to an asymptomatic patient to elucidate the presence of signs? One would instinctively assume that disease is indeed present in each case throughout the period that has elapsed from the time of the pathogenetic interaction of the causal factors until the time when some observer first identifies some manifestation. Does this invalidate the definition? The definition deals with this possibility by not requiring observers to diagnose disease at its earliest time of onset. It merely indicates that disease is a word that they use to describe a disadvantageous process within the organism when they become aware of it, without standing in the way of them applying it retrospectively. The point needs making, indeed, that the proposed definition does not even make it obligatory for observers to diagnose the disease in any detail: it requires them to consider it disadvantageous to the organism, and it

but thereafter came to apply largely to diagnostic aspects.
permits them to speculate about its nature, but it does not specify the extent to which each observer must form perceptions about the disease.

A further significant consideration is of three important asymptomatic states that exist, that everyone would accept as examples of disease and that the present definition also comfortably accommodates. These are the situations of certain disadvantageous conditions occurring in unconscious people, in sleeping people, and in insensate organisms. The background to each of these circumstances is that symptoms require awareness, and awareness requires the existence of a well-functioning brain. Not all living organisms possess this: unconscious humans lack it as do plants and lower animals, whilst sleeping humans at least partially lack it. All disease is therefore asymptomatic for them. The causes, pathogenetic mechanisms, manifestations in the form of signs, and observers’ judgement of disadvantage all nevertheless exist in these people. The present definition supports observers in using the word disease to describe their observations in these situations on the same grounds as those outlined above relating to the presence of manifestations other than signs, and the flexibility with regard to retrospective description. This then suggests, on the one hand, that asymptomatic disease can exist in cases such as this; and, on the other, that the formulation of the present definition creates no difficulties in these cases.
The conclusions that analysis of the issue of asymptomatic disease leads one to reach are, firstly, that symptoms are but one variety of manifestations; secondly, that signs of various types are another variety of manifestation that allow one to accept the existence of much asymptomatic disease within the bounds of the present definition; and, thirdly, that an ability to provide retrospective descriptions (a procedure permitted by the proposed definition) mirrors a common application of the word *disease*. This analysis furthermore, by challenging the implications of the definition in these circumstances and finding them comfortably accommodated, not only provides some useful insights into the whole idea of disease, but also strengthens support for the definition.

§5.6 DISADVANTAGE

The proposed definition requires a condition to produce disadvantageous physical and/or mental manifestations if it is to be considered as a disease. This provision distinguishes it categorically from some other recent definitions, especially those associated with Christopher Boorse and his predecessors—whose Biostatistical Theory of Disease sought objectivity by replacing ‘disadvantage’ with ‘deviation from normal’ as a central notion (see §2.1 above). The principal reason for taking the present stance is the very close, and virtually inevitable, connection that everyday observation reveals between disease and disadvantage.
Diseases are hardly conditions that people desire to acquire. One has great difficulty in naming any disease that, on balance, is better acquired than not acquired. Diseases cause pain and suffering, they impede the enjoyment of fruitful lives by the organisms that develop them. They often reduce longevity. People dread the unexpected appearance of diseases, they take innumerable precautions to avoid them, and they expend vast sums of money to treat them. Diseases, whether they occur in humans or in animals or in plants, destroy economic potential. People regard the appearance of disease with displeasure rather than with equanimity. They feel sorry for others who suffer from diseases, offering assistance and special dispensations to make life more bearable in circumstances that often approach the intolerable. Diseases, then, are archetypically undesirable and disadvantageous things.35

It is therefore hardly surprising that disadvantage closely mirrors the derivational meaning of the word disease in the English language and of cognate words for it in many foreign languages. The linguistic association of dis (‘deprivation of’) with privation and disadvantage is undeniable. If

Evidence of the level of disadvantage that society instinctively associates with the idea of disease is the fact that occasional individuals feign to suffer from diseases that they do not have, or magnify the extent of diseases that they do indeed have, in order to obtain by false pretences the benefits that society, out of pity, often grants to those who do suffer from disease. Occasional people even develop a disease (hypochondriasis) in which they display an unfounded belief that they are ill when indeed they are not, with the subconscious motivation often of avoiding responsibilities that they otherwise might have to shoulder.
disease involved no privation then the *dis* prefix would become redundant and one would merely say that an organism was *at ease* when one wanted to indicate that something was physically and/or mentally wrong with it. Such a terminology would obviously be absurd.³⁶

Analysing the understanding that people have of a particular word is always difficult in the absence of information derived from extensive surveys—and these are unavailable in the case of *disease*. Some guide to community views is, however, possible by examining the words that people use to express ideas opposite to those conveyed by the particular word, and by examining the actions that people take when faced by circumstances that the word conveys. The antonym of disease, *health*, carries suggestions of goodness, wholeness, helpfulness, healing, and even holiness. These are all positive and advantageous states that

³⁶ Few words can have more disadvantageous connotations than do the cognates of the English word *disease* in other European languages, suggesting that the intrinsic ideas associated with them run true in most societies. Many words for disease in the Romance languages include the prefix *mal*, ‘bad’ (French *malaise*, Italian *malattie*, Spanish *mal*, and Portuguese *mal* and *malestia*), whilst the Dutch *kwaal* relates to *kwaad*, ‘evil’, and the Russian *bolzin* derives from the Old Norse *bol*, also ‘evil’. The German word *krankheit* literally means ‘the state of being weak’, as does the Greek *asthenia* (‘lack of strength’). Two other Greek words *nosos* (cf. the English *noxious*) and *pathos* (cf. the English *pathetic*), and the Latin *morbus* (derived from *mors*, ‘death’; and cf. the English *morbid* and *mortuary*) have equally ominous overtones; while most Scandinavian words (Norwegian *sykdom*, Danish *sygdom*, Swedish *syukdom*, Icelandic *sjukur*, and cf. the Dutch *ziek* and Old German *siech*) have the same unattractive connotations related both to disease in general and to
contrast quite strongly with the negativity implied by disease. They reinforce the intrinsic advantage that the idea of health conveys, and the intrinsic disadvantage that the idea of disease conveys. People furthermore demonstrate by their actions that disease causes them difficulties, rather than enhancing their lives. They tend, when faced by disease, to discontinue their usual activities rather than embarking upon new ones. They often seek help from others rather than offering to relieve others of burdens. Their incomes tend to fall rather than to rise. They make statements such as ‘It is good to be wealthy, but health is more important than wealth’. People, even when they are healthy, demonstrate their fear of potential disease by taking insurance to provide for their treatment if they fall ill, and by demanding that their governments provide effective health services and social security benefits for them when they suffer disease. People, by their actions, demonstrate that they fear disease and believe that it disadvantages them.

One might expect that, in the face of such fundamental evidence derived from the intrinsic meaning of the word disease and from the actions of people affected by it, any definition of it would unequivocally state that disease is disadvantageous. Yet this is not the case. Not all of the definitions previously proposed for the word disease indicate that it must convey a notion of disadvantage, and indeed some of them have vomiting as does the English sick. None of them suggests an attractive
specifically denied such a notion. Another way, therefore, of examining whether one should embed the idea of disadvantage in a new definition is to examine the durability of those approaches that have excluded it.

Boorse’s Biostatistical Theory is the outstanding example of the alternative approaches, in that it replaces the notion of disadvantage with the notion of deviation from normal. The detailed analysis of his theory given previously (see §2.1 above) demonstrates that it contains serious weaknesses that have attracted trenchant criticisms. Boorse has himself made valiant attempts to answer these, yet an independent observer must surely conclude that those attempts remain ultimately unconvincing for the reasons set out previously (§2.1). This lack of persuasion is compounded by their reliance on the concept of normality and their failure to address the inherent confusion associated with the idea of normality that authors such as Jirí Vácha have described.37 What is the norm when one is considering a biological species? Take, for example, the case of humans: is the norm a man or a woman? Is it a person with black hair or blond hair or a redhead? Is it a person with European facial features or African or East Asian? Is it a tall person or a short person, a strong person or a weak person, a young person or an old person? No one would consider that any of these characteristics would necessarily make a person diseased, yet they all make people different from each other, and one or other of them state.
must be nominated as the ideal for the species ('the normal') if one is to create a background against which to measure the possibility of disease on the grounds of deviation from the norm.

This points up a fundamental weakness of Boorse's whole approach. His idea of seeking some 'objective' definition of disease, which is clearly his primary aim, indicates a positivistic desire to rid the enterprise of all value judgements. It makes the assumption that the mere mention of 'disadvantage' indicates the presence of value judgement. People can, however, assess disadvantage dispassionately and indeed often measure it. The aim of being 'scientific' thus seems to lead him to overlook the fact that, on the one hand, the mere adoption of a value-free terminology will not necessarily enable him to achieve his goal; and, on the other, that disadvantage is a potentially measurable parameter. The ultimate consequence is to put him at risk of entirely missing the point of why anyone might wish to develop a notion of disease.

These difficulties encountered by Boorse and others who have tried to predicate a definition of disease upon some hypothetical norm for a species have been exacerbated by the conceptual contortions that they have had to make in their attempts to resolve their problems. Their case has appeared weaker with every one of these contortions. This therefore

---

37 Vácha (1978).
becomes an important consideration for the present discussion. If one regards Boorse’s work as the most substantial attempt to replace the value-laden idea of disadvantage in disease with a value-free statistical one based on norms, and if he fails, then subsequent attempts at providing a definition and explication of disease have little alternative but to stipulate the need for a disadvantageous physical or mental state developing against a background of a previously non-disadvantageous state, rather than seeking yet again and ever more unsuccessfully to eliminate the whole idea of disadvantage.

The present proposal suggests that the question of whether disease must always be predicated upon deviation from the norm is a meaningless consideration. It arises instead from a base that assumes no such thing as a norm to exist for living organisms, at least in the context of deciding whether disease has developed. It maintains, rather that disease involves change to a disadvantaged physical and/or mental state from a previously non-disadvantaged state, thereby entirely avoiding having to impose dubious criteria of normality.

It is, however, important to recognize that although the present definition stipulates that disease must disadvantage the victim, it does not specify when that disadvantage must occur. This enables one to consider as diseased (or potentially diseased) an asymptomatic organism that has
evidence of some presently non-disadvantageous condition, but who has a high probability of developing disadvantage in the future due to that condition. An individual faced with such circumstances may well recognize their dire future implications and so decide upon actions that take account of that likelihood. Prospective disadvantage then becomes as pertinent a form of disadvantage as does present disadvantage for the purposes of the proposed definition. Observers (including the organism suffering from the disease if this is a human) are the people who must decide about these matters. They may well then label a condition as a disease before it has caused any measurable disadvantage if they have good grounds to predict that future disadvantage will accrue on the basis of information about the natural history of similar conditions in other organisms or the progressive nature of the present condition in the present organism.

The considerations therefore that have influenced the inclusion of disadvantage as a necessary component of the present definition are the common understanding of the connotations of the word; the actions of people who suffer from, or anticipate suffering from, disease; and the failure of alternative non-disadvantage theories to survive critical analysis. One is left with little alternative, then, but to accept that the idea of disadvantage is deeply embedded in the idea of disease.

§5.7 OBSERVERS
Consideration of the observers of disease involves consideration of three important issues: their necessity, their nature, and their functions.

To address first the necessity of the observers, the discussion in §2.1 above pointed out that some previous writers (such as A.P. Cawadias) drew distinctions between realist views of disease, conceptualist views, and nominalist views. Realists, they suggested, perceive each occurrence of disease as a separate thing in itself. Conceptualists perceive disease (and perhaps each occurrence thereof) as an idea in the mind of an observer. Nominalists perceive a disease as the name that observers use to identify an artificial category of biological phenomena. This classification has the advantage of emphasizing that, on a realist assessment, the causes, pathogenesis and manifestations of disease are independent of any observer. They exist even if no one ever perceives them. The practical weakness of the realist understanding of disease is, however, that no one will determine that an organism has suffered disadvantage unless some observer observes the disadvantaged organism. No one, furthermore, will apply the word disease to it, no one will study its causes and mechanisms, and no one will distinguish between the different types of disease that occur unless someone observes what is happening. Each of these latter actions is intrinsic to our contemporary understanding of disease; and, because each requires someone to observe the diseased organism, each of them enhances the role of observers in the
contemporary idea of disease. The actions of observers therefore constitute an integral part of what we presently perceive as disease. They make a purely realist understanding of disease inadequate to explicate the contemporary idea. Any successful definition and explication of disease must accept this. The proposals now under consideration have the great advantage of recognizing this by building an obligatory conceptualist layer upon a realist base and making them both part of the unified edifice.

The diagram in §4 clarifies the position of these aspects to the present proposal. The causes of disease ($C_E$ and $C_M$) interact as a pathogenetic reaction ($P_1$) to produce a manifestation or manifestations ($M_1$) that may or may not become apparent to observers. These initial manifestations not infrequently interact again ($P_2$) with the organism ($C_M$) to produce a further manifestation or manifestations ($M_2$), and this sequence can recur on many occasions until a final manifestation or manifestations ($M_F$) result. Disease has then occurred within the organism, but the word disease is not used to describe it until an observer ($O_1$) or observers ($O_1\ldots O_n$) perceive it to have occurred. The identity and functions of potential observers clearly have important implications for the understanding of disease and they warrant further discussion.

To address next the nature of observers, one would suggest that anyone who becomes aware of the manifestations of disease in an organism is an
observer of that organism’s disease. People require no particular qualifications to recognize manifestations, provided that they possess the sensory apparatus necessary to perceive the manifestations, and the intellectual apparatus necessary to process those perceptions. Those requirements do, however, make humans the only observers of disease who can perform this role in an intelligent way. Humans, furthermore, occupy a unique category among organisms in this regard since they are capable of perceiving that they themselves are sick, whereas members of other species lack that capacity, either completely or to a very large degree. This means that diseased people themselves form an important group amongst those who are observers of disease. Unaffected lay people form a second group. A third group consists of professional people who specialize in studying disease—medical practitioners, veterinarians, plant pathologists, entomologists, and the like.

This unique role for humans, resulting from their personal perceptions of their own diseases, has created a particular convention that does not extend to other species. The discussion of manifestations in §5.3 indicated that for the past two centuries clinicians dealing with human disease have distinguished between symptoms (the manifestations observed by a patient) and signs (the manifestations observed by others, especially by medical practitioners). Some investigators have then denigrated symptoms as subjective, and therefore inherently unreliable,
interpretations of manifestations. They have in contrast dignified signs as objective, and less value-laden than symptoms, since someone (whose impartiality is assumed) beyond the patient has observed them. This distinction between symptoms and signs only applies in the case of human disease as other organisms have at best an unformulated awareness of symptoms.

Symptoms and signs do nevertheless tend to merge with each other. Some symptoms (for example, pain) are only perceptible to those who suffer from disease, whilst some others (such as hearing and visual impairment) closely approach the same situation. An astute patient may well have a keener appreciation of disease manifested in these ways than does any professional observer. Certain conditions, on the other hand, have subtleties that only a person with specialized knowledge can recognize. Symptoms and signs in yet other conditions overlap to the extent that any effort to distinguish between them, and so between the observers of them, becomes relatively unimportant. Take a simple example, mentioned previously, to illustrate this. Consider a lump that grows on a person’s skin. Those who would draw an important distinction between a symptom and a sign would claim that the lump that the patient notices is a symptom, but the lump that the patient’s doctor notices is a sign. The lump is, nevertheless, the same lump; its causes, its pathogenetic mechanism, its natural history, and its outcome will all
remain the same regardless of the word used to describe it, and of who observes it. This might suggest that the case for distinguishing between symptoms and signs, and so between the observations of those who suffer from disease and those who stand outside it, is often a relatively weak one—at least in so far as they are capable of making mere observations. The distinction between them does, however, often gain significance in the subsequent process of interpreting the observations that forms the remainder of the process commonly known as diagnosis (see §5.8 below).

This brings the discussion to the third important consideration concerning observers of disease, namely the functions that they undertake. Their primary function must inevitably be to discern the manifestations of disease, in the absence of which none of their subsidiary actions is possible. They must however then perform several other functions relating to diagnosis that occupy pivotal places in the contemporary understanding of disease. Successful accomplishment of those roles requires observers to possess certain crucial attributes. These include possession of the intellectual apparatus necessary to conceptualise the observations that they make, possession of an ability to judge whether disadvantage results from the process that is occurring in the organism (since disease only occurs if the organism is disadvantaged and if someone, namely the observer or observers, determines that issue), and possession of an ability
to interact with other observers. This last attribute is particularly important since the closer the interpretive consensus of the observers about each case of disease, and about the generalities that similar cases raise, the more durable their conclusions become. The functions of observers, therefore, are both physical and intellectual: physical because they must use their senses to receive the information about the manifestations and the genesis of these; and intellectual because they must process that information with ingenuity, judgement, balance, and diplomacy.

One can therefore conclude that the observers of disease are necessary and are its conceptualists. Anyone who is aware of the manifestations of disease is an observer, so this includes everyone capable of sensing the manifestations in the diseased organism—the patient himself or herself in the case of human disease, the lay person who stands nearby, and the professional who specializes in understanding disease. The functions of observers then go far beyond merely discerning the manifestations: they have both to judge whether the process that produces the observable characteristics is a disadvantageous one for the organism (without which no disease would occur), and to seek a consensus about their observations with other observers in order to obtain a persuasive explanation of the diseased organism's condition. They are the key performers in the social interpretation of disease without which no use of the term disease would be possible.
§5.8 Diagnosis

The observers are, as indicated in §5.7, an integral part of the proposed definition of disease. Observers who examine an organism that they consider might suffer from disease seek to gain knowledge about it, undertaking an exercise commonly known as diagnosis. This is, in effect, a specialized form of epistemology and needs to be examined against the background of that discipline. The diagnostic exercise has two aspects, the first involving investigation of the disease process that is affecting a particular organism, and the second involving categorization of the knowledge thereby gained within a previously ordained classification of disease processes. The latter will form the topic of §5.9 below (‘Amalgamating diseases’) whereas the current discussion will address the former. It raises several significant issues that include the identities of the people entitled to participate in the diagnostic process, the methods that they use, the levels of discernment that they achieve, the certainty with which they can hold their knowledge, and the ways in which they deal with their knowledge.

To address firstly the issue of who can participate in the diagnosis of disease: diseases affect all types of living organisms, but diagnosis of them is an exercise effectively restricted to humans. Any human, indeed, who observes disease in an organism, whether in him/herself or in some
other organism, is at least starting on a diagnostic exercise since that human is gaining some knowledge about the diseased organism. This makes diagnosis the prerogative of all humans; although, as in any area of human activity in which there is a long history of practice and discussion, those who have studied the topic in greatest depth are those most likely to display the greatest skill in their endeavours. Professionals thereby achieve a crucial role in the diagnostic process that inevitably tends to overshadow, but never entirely to obliterate, the role of amateurs including the victims themselves of disease.

The methods that diagnosticians use to investigate a disease process in a particular organism are primarily those of applying the senses of the observer to the affected organism: listening to his/her story if it is another human, looking at it, feeling it, and using technical devices that produce information of various kinds that the observer can sense. Those devices take many forms; they involve a variety of physical and chemical methods; they can examine the diseased organism as a whole and in parts; they can examine its exterior and its internal components, sometimes imaged and sometimes extracted for the purpose of the particular examinations undertaken. Observers thus use many technical methods in the process of diagnosis, but the common feature that underpins all of them is that, regardless of the wide variety of procedures involved, they all provide information that impinges upon the sensory awareness of the observers.
What level of discernment must an observer achieve to determine that disease exists? A huge range is possible. Most observers can identify gross manifestations. The visibility of some manifestations may, however, be very small. The most astute observer is often one who notices that a minor variation is occurring in the organism at a time when all other observers notice no change from the usual situation. Less observant observers, however, will have to await the development of more profound manifestations before they realize that changes are occurring, and so that disease is developing. The disease becomes clear to every observer when these changes culminate in grossly disadvantageous manifestations. The changes in the organism thus progress independently of the acuity of the observers, but what can vary is that degree of observational acuity. This then emphasizes on the one hand the importance of the realist aspect of disease processes, whilst reiterating on the other the relevance of the conceptualist and the nominalist aspects to the whole.

How certain, then, must an observer be that disease has occurred before he or she can reasonably claim that this is the case? As wide a range of variation occurs in this aspect as in the other aspects of diagnosis. Several reasons contribute to this. The manifestations, like the causes and the degree of disadvantage, may be overt or covert. Observers vary in their ability to perceive any of these components, with what appears overt
to one observer being covert to another. The levels of experience acquired by the various observers of a diseased organism then come to play an important role in their ability to diagnose disease. Much diagnosis, indeed, results from pattern recognition that in turn comes from repeated previous practical experiences with similar examples of disease, or from systematic theoretical study of the characteristics of various types of disease. Continuous recapitulation of a particular pattern is, in such circumstances, likely to facilitate certainty. The closer a diagnostic consensus that multiple observers can achieve about the processes occurring in a particular organism the greater that certainty is likely to become. The more closely, furthermore, the knowledge thereby gained from the diagnostic process lends itself to categorization within accepted classifications, the more confident the diagnosticians can become in their conclusions. This thereby creates a linkage between the diagnostic processes as they affect a particular organism to the process that creates formal classifications of diseases (see §5.9, Amalgamating diseases, below).

The ways in which the observers deal with their diagnoses are important. Each observer forms his or her own diagnosis (D1…Dn in the diagram in §4) on the basis of the available information. Each of these observers, however, often has the subsequent opportunity of comparing his/her observations with those of other observers who have examined the same diseased organism. They may or may not all concur in their assessments,
but often find themselves in a position of debating these assessments until some reasonable agreement emerges (although it is certainly not guaranteed to do so) that takes the form of a consensus (or final) diagnosis. The finality even of this nevertheless remains limited, as each of the observers is limited by his or her cultural milieu and also by the technological imperatives of the modalities of investigation available to them. The form taken by the final diagnosis is often that of a word or phrase to categorize the process in terms of the accepted classifications of disease extant at the time that the observers make their diagnoses and formulate these into a final diagnosis.

The conclusion that one reaches about diagnosis is then that it constitutes a necessary part of the understanding of disease. Aspects of it relate to understanding the processes that occur in each individual case, and to the development of broad categories that share common features and that lead to classifications of types of disease. All humans who observe diseased organisms can form some level of diagnosis of the processes occurring: they include the individuals who themselves suffer from disease, but the greater the depth of professional study applied by an observer the greater the understanding that he or she is likely to acquire. Observers use many types of sensory modalities—some of them simple, but others highly technical—to acquire the information that leads them to their diagnoses. This produces very broad ranges of discernment and of
certainty in the efforts of the observers, but consensus enhances their certainty, as does an ability to recognize common patterns among disease processes that lead in turn to the development of systems of classification. When several observers discuss their diagnoses they are likely to formulate a consensus that incorporates the features common to each of their observations; and this final diagnosis, influenced often by the cultural and classificatory environments in which it occurs, tends to reflect the deepest understanding of the disease process available to the particular observers.

§5.9 AMALGAMATING DISEASES

A major weakness of previously proposed definitions and explications of disease is that, without exception, they have failed to address a crucial issue that revolves around the use of the word *disease*. People use this single word in three quite different contexts—contexts certainly that are related to each other, but nevertheless contexts that, if left undifferentiated from each other, are liable to result in a morass of confusion. These contexts, then, require clear distinction if one desires to introduce some clarity into an analysis of the topic, so the purpose of the present section is to clarify this area that has previously suffered from remarkable neglect.

The discussion so far presented about the proposed definition of disease has focussed upon a particular type of phenomenon that affects a
particular organism in a particular way for a finite time. This, however, is only one of three different uses that people apply to the word *disease*. They also use it to indicate the existence of several episodes of similar phenomena that have characteristics in common and that affect multiple organisms at various times. A third use is to indicate the generic existence of many episodes of similar phenomena that variously affect every living organism. I would argue that, whilst the first of these uses is the primary one, a persuasive definition of disease must accommodate all three usages, and that the present definition succeeds in accomplishing this.

Some typical examples may clarify the differences between these three usages of the word. Examples of the first (or single episode) use are statements such as ‘The disease in John Smith's lungs will soon kill him', or ‘Lack of trace elements in the soil has caused the disease in this tree’. Examples of the second (or group) use would include: ‘Polycystic kidney disease is an inherited condition that kills many people', or ‘American brood disease has now spread to Australia where it is infecting ever more hives of bees'. Typical examples of the third (or generic) use are: ‘Disease can affect plants as well as humans', or ‘The prevalence of disease increases during wartime'.

The use of the same word to express each of these ideas can easily create confusion. Analysis of the topic would be much simpler if
languages, including the English language, possessed three separate
words with which to convey the different ideas, but unfortunately most do
not. I shall, nevertheless, try to rectify this deficiency in the present
discussion by adopting three (unfortunately rather cumbersome) terms to
convey these three respective ideas: namely disease-as-a-particular,
disease-as-a-collective, and disease-as-a-universal.

Disease-as-a-particular

The definition proposed by this thesis intentionally states that disease is a
word used to describe a process that occurs in a living organism. It does
not, in so doing, preclude other uses of the word disease, although it
implies that this is the prime (or at least a particularly important) use. The
definition emphasises singularity on the grounds that if no individual
organism were to suffer disease, no collection of organisms could suffer it;
and if no collection suffered it, then it could hardly become a universal
experience. Empirical observation, furthermore, supports the use of the
singular as an appropriate starting point, in that when someone notices
something unusual about any living organism the initial focus of attention
is upon that particular case. An observer, for example, deciding that a
shrub in a garden is diseased would not notice anything unusual about the
index shrub if it resembled every other shrub of the same variety. Some
unusual characteristic draws the observer’s attention to that particular
shrub. The manifestations in individual organisms form the building blocks
that observers assemble when gathering evidence to create ideas about diseases-as-collectives and disease-as-a-universal. Single affected organisms thereby become the most appropriate index items to examine when one seeks to understand disease: and the manifestations in them are the hallmarks of disease-as-a-particular.

Since the several previous chapters of this discussion have all focussed upon the idea of disease-as-a-particular this aspect warrants little further attention now, in contrast to the other uses of the word.

Disease-as-a-collective

The present theory suggests that all organisms are susceptible to disease when they interact with certain external factors. One might then expect that if two organisms of the same species were to encounter similar external factors they would interact in similar ways. If many organisms did this, their pathogenetic mechanisms and manifestations would probably also resemble each other. Similarities would then pervade the group, despite each organism developing its own individual disease. This tendency is what makes the phrase disease-as-a-collective useful to describe categories of disease that have certain common features. The characteristics of the diseased organisms in any particular category then resemble each other closely, despite each member of the category having
its unique experience of disease and maintaining its own identity. Their similarities especially outweigh their dissimilarities in each of the important areas of causation, pathogenetic mechanisms and manifestations.

No two living organisms are, however, in all ways identical (even genetically identical twins differ in their mental processes) and no two interacting factors are identical. The variations in identity of the organisms and of the interacting factors that do occur will, if sufficiently great, cause variations of response to occur. These will, if sufficiently great, then encourage people to perceive them as different diseases-as-a-collective. The degree to which each disease-as-a-collective differs from each other disease-as-a-collective will depend upon how closely the involved organisms resemble each other and how closely the interacting factors resemble each other. An example will illustrate this: assume that the host organisms are all girls of similar age and ethnicity who study in a particular classroom in a particular school. Some have encountered chicken pox virus, whereas others have encountered measles virus at about the same time. All have interacted (become infected with) their respective viruses. Each of these infections causes a distinctive skin rash so, although all of them develop evidence of infections at about the same time, most people would suggest that members of the school class suffer from two different

---

38 Organisms that are identical twins and that are subject to the simultaneous action of identical external factors offer an interesting
diseases-as-collectives. Those diseases both have viral aetiologies in common, yet they have less in common than would be the case if all the girls had succumbed to an epidemic of chicken pox alone. The victims of chicken pox and measles do nevertheless still have more in common with each other than they have with a third group of girls who studied in the same classroom and who simultaneously developed Streptococcal impetigo, since the latter is a bacterial rather than a viral infection. The girls affected by the impetigo infection, nevertheless, have more in common with those affected by chicken pox and measles infections than with a fourth group of girls who suffered hockey-stick injuries to their legs at about the same time; and the hockey-stick-injured girls have more in common with the chicken pox-infected and measles-infected and impetigo-infected ones than with the diseased eucalypt trees growing in the school garden in which a nutritional deficiency had caused wrinkled leaves. All those organisms (the virally-infected girls, the bacterially-infected girls, the hockey-stick-injured girls, and the wrinkle-leaved eucalypt trees) do however have disease, and each of them resembles some others in whom the same external agent triggered their disease. Varying degrees of similarity thus exist among various diseases-as-a-collective, examination of the inter-relationships of which requires analysis of their causes, their pathogenetic mechanisms, their manifestations, and the perceptions that observers have of them. The present definition of demonstration in this regard since the diseases that they develop are
disease provides a useful template upon which to base an examination of that situation.

People find it convenient to distinguish the many diseases-as-collectives from each other for purposes of discussion. They do this by naming them. That undertaking has both advantages and disadvantages. People (both professionals and lay people) use names, on the one hand, to create a series of index items and so to reduce tedious and repetitive descriptions of each diseased organism that they discuss. Collective names thereby assist in the assembly, storage and publication of information. They thus facilitate efforts to address such practical issues as prognostication and treatment. The bestowal of names, on the other hand, fosters the idea that each disease-as-a-collective is an entity identified by its name. The present definition denies the suggestion that such entities exist except as intellectual constructs in the minds of observers. It emphasises, instead, the particular and material nature of each diseased organism, and the importance of perceiving disease as a process that occurs in the organism over time that observers must primarily interpret for that individual organism. The definition does not preclude the development of the category of disease-as-a-collective, indeed it encourages it; however it discourages people who create such categories from endowing them with overtones of physical existence. The challenge for those who group similar

identical, except that these occupy different positions in space.
cases of diseased organisms collectively then becomes one of understanding the limitations of the exercise in which they are engaging.

Western physicians and biologists have sought to collectivise diseases since classical Greco-Roman times, and have proposed systematic nosologies for over three hundred years. Their earliest efforts created several general categories such as fevers, deformities, and injuries; but they later subdivided these broad groups into ever-narrower ones. A confusing array of names proposed by individual practitioners each relied upon some distinctive feature or features, gaining credence principally by common usage. No individual or international body has ever achieved jurisdiction over the nomenclature, with authority to allocate a standardised title to each disease-as-a-collective, in clear contrast to the situation that exists for anatomical and botanical nomenclature. These titles of diseases-as-collectives have then tended to change over time, reflecting both the accumulation of new information and trends in fashion.\(^{39}\) Authors of textbooks and other publications have often unilaterally created precedents of nomenclature in their works, with indices such as *The International Classification of Diseases* reflecting, rather than to guiding, the allocation of names. Such indices have themselves usually employed multiple approaches by which to assign names to diseases-as-collectives.

\(^{39}\) Comparison of a medical dictionary published a century ago with those published recently reveals that many terms for diseases-as-a-collective
collectives by variously focussing upon anatomical variations, functional features, pathological processes, manifestations (including symptoms, signs, and syndromes), eponyms, geographical occurrence, and other idiosyncrasies. The great strength of these classifications has nevertheless been that, based upon observations made upon previous organisms that have fulfilled similar criteria for allocation to such categories, they facilitate prediction of the future course of disease in a particular organism and effective methods of therapy. The resultant hope of relief of the particular organism from the disadvantage inherent in all disease becomes an irresistible force impelling the allocation of each disease-as-a-particular into a disease-as-a-collective category.

These observations therefore suggest that, despite a continuing flux in systematic nosology, the allocation of individual cases of disease into collective groups does have significant practical advantages. This then makes the process of placing a given disease-as-a-particular into a disease-as-a-collective category an important component of medical and biological discourse. People require technical knowledge to participate meaningfully in that discourse, and also knowledge about the diseased organism. This specialised aspect of diagnosis (commonly referred to as differential diagnosis) enables observers to allocate each disease-as-
particular into some category of disease-as-a-collective. An effective explication of disease must accommodate this aspect, a task undertaken by the present formulation when it states that collective ideas about disease develop from individual ones, when it suggests that each of the observers (O₁...n in §4) reaches a personal diagnosis (D₁...n) that includes consideration of the process of differential diagnosis, and when it allows for appropriate discussion with other observers to reach a consensus by which the process affecting the particular organism is allocated a final diagnosis (D_FINISH) that corresponds to an accepted category of disease-as-a-collective.

This examination of disease-as-a-collective suggests, then, that the idea is a useful one, albeit one that must always remain subsidiary to the understanding of the disease process as focussing upon disease-as-a-particular. The great risk of this process is that emphasis upon the idea of diseases-as-collectives can easily create an illusion that such categories enjoy a material existence that is, in fact, non-existent. Their advantages, despite that, far outweigh their disadvantages and it is from them that more generalized ideas of disease develop.

_Disease-as-a-universal_

40 Standard medical textbooks are replete with examples of this. See, for example, Wilson _et alia_ (1991), or Weatherall _et alia_ (1996).
The reasons outlined above indicate why people who wish to discuss diseases often find it convenient to consider them in groups as diseases-as-collectives. They then often find further advantage in using some term that encompasses all cases of disease-as-a-particular and all groups of diseases-as-collectives to refer to the overall idea of disease. This is the third application of the word *disease*, for which I have adopted the term *disease-as-a-universal* to provide distinction from the particular and the collective uses. That term then refers to all cases of disease that have ever existed or will ever exist. Just as the present definition of disease suggests that one can identify the circumstances that surround disease-as-a-particular rather more easily than one can identify those surrounding disease-as-a-collective, so disease-as-a-universal poses an even more difficult idea to describe. A huge array of external forces can here interact upon a huge array of organisms. Here again, however, they have their physical existence only in the individual cases, with the result that the idea of disease-as-a-universal has as little physical existence as does disease-as-a-collective. It is an idea created in the minds of humans to refer to all the causes, processes, manifestations, and thoughts that observers generate about all past, present, and future cases of disease-as-a-particular. People have, over the course of history, proposed many theories to explain their observations of the resulting phenomena. I would indeed suggest that the further that these have tended to diverge from the idea of disease-as-a-particular the less satisfactory they have become in
describing the idea of disease-as-a-universal. This perhaps reiterates John Locke’s succinct comment upon the long-standing philosophical discussion about universals: ‘since all things that exist are only particulars how come we by general terms...?’.\footnote{Locke, 1959 (Essay concerning Human Understanding, Bk 3, Ch 3, §6),} This is, however, not the place to review the protracted debate that has surrounded ‘the problem of universals’. The fact is, of course, that we do come by general terms that reflect general ideas, that most people in most cultures accept that such ideas exist, and that no one can reasonably deny that people often use the word disease in the sense of disease-as-a-universal. Their problem becomes one, not of accepting this, but of explaining it. The present definition circumvents this whole issue, not by attempting to explain it, but rather by accepting the existence of an idea of universals, and by suggesting that an understanding of universal ideas is most effectively achieved by moving towards them from particular ideas through (in this case) the medium of collective ideas.

The present definition, in conclusion, focuses upon disease-as-a-particular as the occurrence of an individual episode of disease that affects an individual organism. It accepts, however, that the word disease has additional connotations beyond this that correspond to the ideas of disease-as-a-collective and disease-as-a-universal. It encompasses these as subsidiary aspects, but avoids any theoretical explanation of them.
beyond the assumption that they must derive from ideas of disease-as-a-particular.

§5.10 GENETIC DISEASES

Genetic diseases pose, at first sight, an important challenge to a theory of disease that centres upon the creation of a pathogenetic event due to an external factor interacting with an organism. An organism usually develops a genetic disease when it inherits a particular genetic code from one or both of its parents: hence these conditions apparently arise without the impact of any malign external influence upon the organisms that develop them. That would appear to contravene an important component of the presently proposed definition, with the implication either that these are not diseases or that the definition contains a serious flaw.

Are, then, these genetic conditions really diseases? Such inherited conditions as haemophilia and polycystic kidney disease certainly produce physical effects in organisms. Those effects are often disadvantageous. Observers, in turn, invariably assess the effects as disadvantageous. People, furthermore, generally accept that these conditions really are examples of disease. So can one seriously suggest that these are anything other than diseases? I believe not. Yet they apparently develop without any obvious external factors interacting with the organisms that
harbour them and this is the aspect of the genetic diseases that poses the greatest difficulty for the present definition. These diseases are, in addition, common; and they affect all types of organisms. One cannot, therefore, take the alternative course of blandly dismissing them as minor exceptions of little practical importance. A reasonable explanation for them must exist if the presently proposed definition is to survive.

A close examination of the nature of these conditions and their causation, taken with a broad interpretation of the proposed definition, does however suggest that a reasonable explanation exists that brings them within the compass of the proposed formulation. The argument that addresses this issue is a technical one that deals with the nature of the gametes—the ovum and the sperm—that interact to form an organism, with their relationship to each other, and with the origin of non-sexually-induced heritable changes that occasionally occur within them.

Each gamete is itself a living organism that experiences a brief lifespan. The continued existence of an ovum requires that a sperm fertilize it to enable it to develop into a mature organism capable of independent survival. The sperm that performs the act of fertilization is then an external factor that interacts with it to produce physical (and consequential, but delayed, mental) changes within it. The result usually advantages the ovum by enabling it to maintain its existence, but if the result
disadvantages the ovum then the proposed definition would label this interaction as creating an organism that suffers from disease. One way in which the result can disadvantage the ovum is if the sperm carries a genetic code that, upon interaction with the ovum, results in a disadvantageous reformulation of the latter’s genes. The ovum has then been advantaged in being enabled to maintain its existence, but disadvantaged in being impregnated with a genetic code that will disadvantage it during that maintained existence in contrast to the situation in which it would have found itself if an unimpaired sperm had fertilized it. Observers then perceive the consequent organism as diseased in a way that meets the requirements of the proposed definition.

A similar situation can, of course, also apply in reverse: the ovum is a factor external to another briefly living organism, the sperm. The ovum interacts with the sperm to change the sperm and create a new organism capable of independent and advantageous existence. If, however, the ovum carries a particular genetic arrangement that results, upon meeting the sperm, in a disadvantageous reformulation of the latter’s genes, then the consequent organism fulfils the criteria for suffering from disease. Analysis in this way of the mechanisms whereby genetic diseases are transmitted from generation to generation demonstrates that they come within the presently proposed criteria for the definition of disease.
The interesting question still remains unaddressed, however, of the mechanism whereby a gamete that carries potentially disadvantageous genes acquires those genes when most other members of its species escape them. A completely satisfying analysis should presumably also propose some explanation for that phenomenon of a type that is also consistent with the presently proposed definition. Development of such an analysis starts from the observation that some patterns of genetic arrangement clearly run in some families (or in some strains of organisms) and produce generation after generation of individuals who suffer from particular genetic diseases. A distinctive genetic pattern will survive if those who acquire it experience such mild effects that they can reproduce before they become fatally disabled. But how did the first members of the family who displayed the pattern acquire it? The generally accepted mechanism is by mutation: by a change in the genetic pattern from a former variety that did not result in phenotypes that display disease to a new variety that results in phenotypes that do display disease. The very use of the word mutation (derived from the Latin, mutare, to change) links genetically mediated conditions with disease, both in the minds of observers and in a manner consistent with the implications of the presently proposed definition. Some mutations then result in phenotypes that are so profoundly affected that they ensure the failure of perpetuation of the disease, but other milder ones are not so profoundly affected with
the result that they create recapitulating generations of diseased individuals.

How, though, do mutations come about? Do they develop spontaneously within a gamete or are they the result of adverse external influences? The term *spontaneous mutation* is certainly often used to describe the origin of some mutations for which no obvious externally triggered mechanism exists. The explanation usually provided for such mutations is that genetic mechanisms within organisms always include a function that resembles a proof reading system, such that when a deleterious genetic code eventuates by mutation the genetic apparatus will ideally eliminate it and so prevent a diseased organism from resulting: it is then only when that proof reading system fails to function adequately that a diseased strain results. Accurate as this interpretation is likely to be, it still does not answer the question of what triggers the mutation in the first place: it deals rather with what self-protective mechanisms an organism might possess to guard against the consequences of mutations once they have occurred.

The study of mutagenesis—the origin of mutations—is the area in which the answers lie to the fundamental question raised by this whole issue. A tantalizing enigma arises because the study of mutagenesis remains as yet only poorly developed. The available information about it comes largely from chance observations and a quite limited amount of systematic
research. The thrust of the available information, however, does suggest that identifiable mutations are most likely to result from the action of external factors (such as ionising radiation or noxious chemicals) upon a genotype. The range of possible external factors and their possible interactions, nevertheless, still remain only scantily understood. These circumstances then resemble those in the causation of idiopathic non-genetic forms of disease discussed in §5.1 above, although there is significantly less empirical knowledge available about them. Evidence certainly does exist that particular physical and chemical factors in the environment (such as minor variations in ambient temperature) can markedly alter the rate at which ‘spontaneous’ mutations occur. The more evidence that accumulates about the effect of multiple chemical, radiation, and other environmental factors that cause or modify mutations, the less the idea of entirely spontaneous events appeals and the more attractive their nomination as mutations of presently unidentified cause would become. This does not exclude the theoretical possibility (as also exists in any discussion of causation of disease, for which see §5.1 above) that uncaused mutations could conceivably occur. That possibility is, however, a less persuasive proposition than is the possibility of externally caused mutations for exactly the same reasons as applied to causation.

One other special situation relating to genetic diseases warrants some additional comment. Whereas many mutations are physically
disadvantageous, some others can confer advantage and thereby create fitter phenotypes that possess an improved ability to survive. Darwinian theory requires this as an inherent component of evolution; and where observers reach a consensus that no disadvantage develops for the individuals affected, neither people generally, nor the presently proposed definition, deem disease to exist so no conflict exists in this regard.

Another important aspect of genetic disease that requires detailed attention—not because it deals with a numerically huge component of genetic disease, but because of its potential to confuse discussion and distract from an understanding about the major body of genetic diseases—relates to a small group of mutations that can have ambivalent effects, the consequences of which are usually disadvantageous but which can occasionally create advantage. Sickle cell disease is perhaps the classical example of these.

The sickle cell gene is common among the inhabitants of certain geographic areas in the world. Each human has two loci at which the sickle cell gene can express. An individual who inherits the sickle cell gene from both parents acquires a double dose of the gene (the homozygous situation) that will result in profound symptoms due to anaemia and manifestations in several organ systems of the body. An individual, however, who inherits the sickle cell gene from only one parent is said to
have the *heterozygous* condition and to be a carrier of the sickle cell trait. This person will have no clinical manifestations unless he or she develops low blood levels of oxygen. The child of a person with sickle cell trait whose other parent is also a carrier of the sickle cell trait will have a 25 per cent chance (determined statistically) of being homozygous and of developing severe manifestations, a 50 per cent chance of being heterozygous, and a 25 per cent chance of receiving no sickle cell genes. The situation becomes more complicated when a heterozygous person lives in an environment where malaria is endemic. The trait may then partially protect individuals from developing some disadvantageous manifestations of malaria if they catch that disease.\(^{42}\)

The implications of the presently proposed definition seem to conform to the assessments of most observers who live in areas in whose populations the sickle cell gene is common. These are to consider a homozygous person with disadvantageous symptomatic manifestations as having disease; a heterozygous person with no disadvantageous symptomatic manifestations as merely having a potential for disease; and a heterozygous person who develops a low blood oxygen tension and so develops symptomatic manifestations as having disease. The causal factors involved in this last situation are both the presence of the sickle cell trait and the reason for the low oxygen tension. The person whose

sickle cell trait provides protection from malaria is then perceived as being physically advantaged by their genetic inheritance, but this is different to being physically advantaged by having disease since the presence of the sickle cell trait alone does not constitute disease.

This analysis therefore suggests that one can quite reasonably perceive genetic diseases as fulfilling the requirements of the proposed definition. This occurs, in the first instance, because the interaction of a variant gamete with a non-variant gamete is an external attack upon the latter that results in a diseased organism. It occurs, additionally, because the available information about mutagenesis suggests that this results, where its origin is understood (and here the same provisos apply as applied in the case of causation of 'idiopathic' diseases), from an external physical or chemical factor interacting with the genetic material of an organism to produce a deleterious mutation that may remain covert in that organism but then pass to its offspring and manifest in each new phenotype that carries the altered genome. These two perspectives explain the genetic origins of many phenomena that people look upon as disease. Some particular genetic situations furthermore, when analysed in detail, demonstrate that the presently proposed definition conforms quite closely to common understandings of what is disease and what is not disease.

§5.11 BORDERLINE CONDITIONS
A number of conditions exist that relate to circumstances affecting living organisms, but which people do not generally perceive as being types of disease. These warrant examination in order to demonstrate the ways in which the presently proposed definition provides a means of differentiating disease from borderline conditions that differ from disease. Among the conditions worthy of examination are menstruation, menopause, pregnancy, and obesity.

**Menstruation**

A woman who menstruates is at physical disadvantage during the course of this physiological process. Should this then suggest that menstruation is a form of disease? People do not regard menstruation (so long as it remains within the usual range and pattern of bleeding) as a disease state, so the question arises as to whether the presently proposed definition would bring it within the ambit of disease and thereby render the definition discordant with everyday perceptions. The reason that observers do not usually regard menstruation as a form of disease is presumably because they perceive its occurrence as a necessary component of the maintenance of fertility in a female, and they perceive maintenance of fertility as a desirable and healthy occurrence during an important portion of a woman’s life. People therefore usually perceive menstruation, on

---

Many might argue that she suffers some social disadvantage, but at a physical level she is also disadvantaged on the grounds (perhaps amongst
balance, as an indication of advantage rather than of disadvantage—of health rather than of disease. Menstruation furthermore occurs without some external factor interacting with the organism to cause it, so here again the proposed definition conforms to community attitudes when it fails to identify it as a disease. Menstruation is, however, occasionally accompanied by other symptoms and these then suggest the presence of some disease. A typical example of such symptoms is the pain associated with menstruation that occurs in dysmenorrhoea, the several possible factors involved in the causation of which include mechanical obstruction of the female genital tract, inflammation within the genital tract, or emotional disturbances. Presence of these types of accompaniments inevitably indicates the presence of disease.

Menopause

If menstruation is an indication of health, what then is the position of the menopause? Is this an indicator of disease? People do not usually consider it as such, provided that it occurs at about the usual time of life. Does however the presently proposed definition bring it within the category of disease? I would again argue that it does not, here upon the grounds that it protects a woman from having further pregnancies at a time when her body is becoming less physically capable of withstanding others) of being rendered more likely to develop iron deficiency and anaemia.
the stresses involved in them.\textsuperscript{44} The consequence is that menopause becomes protective of health rather than disadvantageous to health. It furthermore occurs without any external factor acting upon the woman to initiate it.\textsuperscript{45} This discounts it as a form of disease.

\textit{Pregnancy}\textsuperscript{46}

If menstruation and menopause are not examples of disease, what then is the situation in pregnancy? People do not usually look upon pregnancy as a disease, although they do perceive complications that occur during pregnancy as manifestations of disease. Yet in pregnancy a woman appears to have interacted with an external factor (a sperm) that has produced physical changes within her that can disadvantage her if complications arise. The presently proposed definition might therefore appear on superficial examination to encompass pregnancy within the category of disease, in contrast to the way in which people usually perceive pregnancy. Are they wrong, or is the definition flawed?

\textsuperscript{44} Her pelvic floor muscles, amongst many considerations, will generally be weaker, with inevitable mechanical implications for complications to develop during and as a result of a pregnancy embarked upon at an older age. Statistically, furthermore, maternal and foetal outcomes are worse for elderly mothers.

\textsuperscript{45} This contrasts with circumstances where amenorrhoea occurs at times that are not appropriate for the menopause. People perceive this as a state of disease because they can then usually identify some process for which they can often find an external initiating factor that has acted upon the woman to cause it and thereby to disadvantage her.
A persuasive argument based on this definition does, however, exist to suggest that people are wise in not regarding pregnancy as a disease. It relies upon identification of the physical boundaries of an individual, examination of the anatomy of the female genital tract, and an understanding of the structural relations between a fetus and its mother. The situation differs somewhat from animal species to animal species, but the differences are of detail rather than of principal, so one can reasonably extrapolate to other species from the human situation.

Pregnancy is a unique event in which two individuals, the mother and the infant, exist temporarily in close proximity to each other. This raises the questions of whether a boundary lies between the two of them during the course of the pregnancy, if so where, and if so from when? Where, in more general terms, does the boundary lie between a living organism and its environment?

The obvious place to draw such a boundary is at the external surface of an individual’s skin. This, for much of the body, is hardly controversial: few can, for example, question where the outer surface of the skin lies on the palm of the hand. The situation regarding the orifices of the body cavities that open to the exterior—the mouth, nose, ears, lacrimal ducts, anus, urethra, vagina, and sweat glands—is perhaps more complex, although it

---

46 The technical aspects in this section reflect the commentary in Greenhill
too resolves satisfactorily upon close structural examination. The layer of epithelial cells that forms the skin and envelopes the whole body spreads continuously into each of these orifices to line its cavity in an unbroken manner, such that all that lies within the lumen of the cavity is technically outside the body, albeit in a highly protected position. An object lying within such a cavity would, though, have the potential to injure the adjacent epithelium and so to cause disease. One would thus regard a razor blade that one held in one’s clenched hand as being outside the body, but capable of injuring it and causing disease if one grasped the blade carelessly. One would similarly regard a razor blade that one held inside one’s mouth as being outside the body, but capable of damaging the tissues of the mouth. The child, in the case of a pregnancy, eventually becomes an external object to its mother, so when and how does this occur: how should one regard the pre-natal situation whilst the fetus still lies within the mother’s genital tract?

A pregnancy occurs when a sperm successfully fertilizes an ovum. Close structural examination reveals that this event technically occurs outside the mother’s body. This is because it occurs after the ovum has escaped into the lumen of the mother’s genital tract by rupturing the peritoneal epithelium (the layer of skin that is in unbroken continuity with the skin of the exterior and that overlies her ovary in which the ovum developed). The

fetus rapidly develops its own epithelium (its own layer of skin), with the result that the pregnancy then technically develops in a highly protected position outside the mother’s body, initially in one of her uterine tubes and later in her uterus. The outer layer of the fertilized ovum (the trophoblast) there fuses with her uterine epithelium to form a unique tissue that facilitates the diffusion of molecules between the quite separate maternal and fetal circulations. The developing fetus thus becomes parasitic upon its mother whilst effectively maintaining its individuality from her and (in most pregnancies) not harming her.

A mother and her fetus clearly interact to produce physical changes in each other, but the proposed definition only identifies disease as occurring if observers consider that those changes disadvantage one or other of the interacting individuals. The fetus is clearly not disadvantaged since it thrives and grows as a consequence of the interaction. The mother also is not disadvantaged unless complications arise—such as an ectopic pregnancy (implantation of the fertilized ovum at some site other than in the uterus, with secondary adverse consequences), hypertension, bleeding due to inappropriate placental separation, excessive vomiting, or emotional disturbances. Instinctive attitudes and the proposed definition agree in these circumstances that the mother is suffering disadvantage and so that disease is present. The conclusion that one therefore must reach is that an uncomplicated pregnancy is not a form of disease,
whereas when complications do arise from a pregnancy in a mother she does indeed suffer from disease.

**Obesity**

Is obesity a disease? Here, as in several of the borderline situations, the proposed definition would identify it as such if and only if it was causing disadvantage to the organism about which observers agreed. This also corresponds reasonably well with the implications of everyday speech in which people might describe someone as being fat without implying that this meant that he or she was diseased, but would perceive the individual as being diseased if the amount of fatness was such that it was causing identifiable physical or mental disadvantage.

The conclusion that one reaches from an examination of these several borderline situations is that each raises a possibility of disease, but upon detailed consideration fails to fall into that category either in general usage or according to the proposed definition.

**§5.12 DISEASE BY ANY OTHER NAME**

The English language has a number of words other than *disease* that describe much the same idea. These include *affliction, ailment, complaint, debility, disorder, distemper, disturbance, illness, ill health, indisposition, infirmity, malady, malaise, morbidity, sickness, suffering, unwell,* and
upset. Some previous commentators have sought to draw important distinctions between several of these, but their efforts have often seemed strained and artificial. The origins of these words are diverse, some reflecting usage in languages from which English is derived, others reflecting previous theories of disease (see §1.1, 1.2, and 1.3 above). The general approach adopted in the present analysis is to avoid trying to draw arbitrary distinctions between them, and rather to regard them all as variant ways of describing the intellectual area that is here called disease. They do not therefore seem to warrant undertaking any detailed analysis.

6. GENERAL CONCLUSION

Disease is a topic that has attracted an enormous volume of discussion over the centuries, with commentators proposing many different ideas to explicate it. The definition of the word disease has, furthermore, exercised many thinkers during the past half century. Unanimity has failed to emerge yet many practical considerations make it desirable that people who talk about disease have some clear definition about which they can agree. The purpose of this thesis is to propose and defend a new definition. This states that Disease is a word used by observers to describe a process that occurs when one or more external factors interact with a living organism to produce physical and/or mental changes within the organism that the observers consider disadvantage the organism as compared with its former state.
The definition carries several important implications that need stating in conjunction with it in order to clarify it. These are that:

- **Changes cannot occur without causes**;
- **Causation requires both the organism and the external factors**;
- **The interaction is the central pathogenetic mechanism of disease**;
- **The causes and the changes may be overt or covert, physical or mental**;
- **Observers judge whether disadvantage occurs**;
- **Each observer’s perceptions create a personal diagnosis**;
- **Consensus among observers strengthens diagnoses**;
- **Collective and universal ideas about disease develop from individual ones**.

Defence of this proposal has required an analysis of the ideas of causation, pathogenesis, manifestations, and disadvantage. The important roles played by observers are, furthermore, integral to everyday understanding of disease and are reflected in the proposed definition. A clear understanding has required one to draw a distinction between the ideas of *disease-as-a-particular*, *disease-as-a-collective*, and *disease-as-a-universal*. Analysis of each of these aspects has demonstrated the reasoning behind several of the provisions of the definition.
Several areas have threatened to cause difficulty for the definition. Prominent amongst these are mental disease, asymptomatic disease, and genetic disease. Each of these, however, has proven less of a challenge than it initially appeared. The fact that the proposed definition is sufficiently robust as to withstand attack from each of these quarters suggests that it may have a durability that surpasses the strengths of previous proposals that have emanated from other commentators. Several borderline conditions also exist—dealing with issues such as menstruation, menopause, pregnancy, and obesity—in each of which the implications of the proposed definition closely mirror everyday perceptions of the relationship of the conditions to disease, again supporting its veracity.

This definition of disease is quite novel. One would hope that it might provide a conclusive solution to the many previous attempts to address this somewhat vexed topic, whilst recognizing that it also inevitably draws attention to various issues that would benefit from further examination and research. The limits of the current presentation preclude one from discussing these further here, although an indication of some of them might help to identify some questions that remain incompletely resolved and that would provide priorities for future research.
The first is that readers will inevitably raise specific examples with which to challenge the present definition. Some readers have already questioned, for example, whether the definition could under certain circumstances mean that one would have to classify a tattoo, a bad haircut, or political dissent as forms of disease. Each of these and many other possibilities that people might suggest, warrant individual analysis and discussion.

Some other issues are that the definition and explication avowedly have prescriptive elements in addition to merely normative ones. These may well be some of its strongest attributes, but they also provoke the question of whether they are really justified. One obvious example of this would be the decision to include injuries and poisonings as specialized varieties of disease rather than in some quite different category of phenomena. I consider that they are justified, but others may wish to dispute this. Arguments that others might care to raise would need measured consideration and subsequent rebuttal or acceptance.

The present definition gives considerable power to observers. Is this excessive? Does it, furthermore, by equating the role of external observers with that of the patient in cases of human disease leave itself open to accusations of promoting an excessive level of paternalism? Does its anthropocentricity regarding observation, on the other hand, reduce its reliability regarding diseases of animals and plants?
Is the definition justified in dissolving distinctions between disease, sickness, illness, and malady, that some previous commentators have claimed exist and are important?

Evaluation of these, and quite possibly other, future lines of research and discussion that the present proposals may provoke can only serve to enrich the philosophical debate about the nature and limits of disease. Such enrichment is likely in turn to promote a clearer understanding, and that is then likely to move discussion closer to achieving a consensus of views on this topic that is of such considerable practical importance in so many aspects of everyday life.

* * * * * * * * * *


Behring, Dr., ‘II. Untersuchungen über das Zustandekommen der Diphtherie-Immunität bei Thieren’, *Deutsche Medicinische Wochenschrift*, 1890, volume 16, pages 1145-1148.

Behring, Dr., Dr Kitasato, ‘I. Ueber das Zustandekommen der Diphtherie-Immunität und der Tetanus-Immunität bei Thieren’, *Deutsche Medicinische Wochenschrift*, 1890, volume 16, pages 1113-1114.


Boyle, Robert. *The Sceptical Chymist: or Chymico-Physical Doubts & Paradoxes, Touching the Spagyrist’s Principles Commonly Call’d Hypostatical; As they Are Wont to be Propos’d and Defended by the Generalty of Alchymists. Whereunto is Prœmis’d Part of Another Difcourfe Relating to the Same Subject*. London: J. Cadwell, 1661.


Fleischman, Suzanne, ‘I am…I have…I suffer from…: A Linguist Reflects on the Language of Illness and Disease’, *Journal of Medical Humanities*, 1999, volume 20, pages 3-32.


Hesslow, Germund, ‘Do We Need a Concept of Disease?’, *Theoretical Medicine*, 1993, volume 14, pages 1-14.


Loeffler, F.A.J, and P. Frosch, 'Bericht der Kommission zur Erforschung der Maul-und Klauenseuche bei dem Institut für Infektionskrankheiten',


