INFECTIOUS CAUSATION OF DISEASE:
AN EVOLUTIONARY PERSPECTIVE

GREGORY M. COCHRAN, PAUL W. EWALD, and KYLE D. COCHRAN*

Introduction

Over the past two centuries, diseases have been separated into three categories: infectious diseases, genetic diseases, and diseases caused by too much or too little of some noninfectious environmental constituent. At the end of the 19th century, the most rapid development was in the first of these categories; within three decades after the first cause-effect linkage of a bacterium to a disease, most of the bacterial causes of common acute infectious diseases had been identified. This rapid progress can be attributed in large part to Koch's postulates, a rigorous systematic approach to identification of microbes as causes of disease. Koch's postulates were useful because they could generate conclusive evidence of infectious causation, particularly when (1) the causative organisms could be isolated and experimentally transmitted, and (2) symptoms occurred soon after the onset of infection in a high proportion of infected individuals. While guiding researchers down one path, however, the postulates directed them away from alternative paths: researchers attempting to document infectious causation were guided away from diseases that had little chance of fulfilling the postulates, even though they might have been infectious.

During the first half of the 20th century, when the study of infectious agents was shifting from bacteria to viruses, Mendel's genetics was being integrated into the study of disease. Some diseases could not be ascribed to infectious causes using Koch's postulates but could be shown to have

genetic bases, particularly if they were inherited according to Mendelian ratios. Mendelís genetics and Kochís postulates thus helped create a conceptual division of diseases into genetic and infectious categories, a division that persists today.

The third category—diseases resulting from noninfectious environmental causes—has a longer history. The known associations of poisons with illness provided a basis for understanding physical agents as causes of disease. The apparent "contagiousness" of some chemical agents, such as the irritant of poison ivy, led experts to consider that diseases could be contagious without being infectious. Even after the discovery of causative microbes during the last quarter of the 19th century, many infectious diseases were considered contagious through the action of poisons, but not necessarily infectious [1].

The difficulty of distinguishing noninfectious chemical insults from infection is often attributable to the similarities between the spread of diseases caused by a common source of toxins and the spread of infectious disease. Milk sickness, for example, resulted from the consumption of white snakeroot by milk cows and was widespread in the midwestern United States during the 19th century [2]. The spread of illness among those who drank contaminated milk could resemble the spread of infectious diseases, especially food-borne infectious diseases, and during the first quarter of the 20th century, there was disagreement over whether milk sickness was caused by infection or an environmental toxin [3]. Distinguishing microbes from environmental toxins as etiologic agents is still a source of controversy. The lytico-bodig of Guam, for example, which can resemble parkinsonism or amyotrophic lateral sclerosis, is attributed by some researchers to infection and by others to a toxin present in foods derived from cycads [4].

Diseases attributable to a paucity of an environmental constituent include scurvy (insufficient vitamin C), rickets (vitamin D), beriberi (vitamin B1), and goiter (iodine). After establishment of the germ theory, dietary deficiency diseases were sometimes mistakenly attributed to infection, but these uncertainties were generally resolved fairly rapidly through nutrient supplementation. During the first half of this century, for example, expert opinion was divided between infectious and dietary causation of pellagra, but by the middle of this century it was generally ascribed to niacin deficiency. Although the primary period of discovery of dietary deficiency diseases occurred prior to the middle of the 20th century, the recent association of folic acid deficiency with spina bifida demonstrates that some diseases may still be ascribed to dietary deficiency.

These three categories offer a conceptual framework for understanding diseases, but they pose a danger of canalizing thinking. They have, for example, contributed to the rejection of infectious causation when evidence in favor of noninfectious causes has been acquired but evidence

against infectious causation is lacking. This tendency to dismiss infectious causation has occurred in spite of the recognition that (1) infectious diseases are typically influenced by both host genetic and noninfectious environmental factors, and (2) some chronic diseases, such as tuberculosis and syphilis, have long been recognized as being caused by infection.

In this essay we analyze the present conceptions of disease etiology from an historical perspective and within the framework offered by evolutionary biology. We begin by analyzing the degree to which infectious causation has been accepted for different categories of disease over the past two centuries with an emphasis on (1) characteristics that make the infectious causes of different diseases conspicuous or cryptic, and (2) the need to detect ever more cryptic infectious causes as a legacy of the more rapid recognition of the conspicuous infectious causes. We then consider principles and approaches that could facilitate recognition of infectious diseases and other phenomena that are not normally considered to be of infectious origin.

Crypticity of Infectious Causation

HISTORICAL PATTERNS OF RECOGNITION

Rather than developing suddenly as a result of the work of the early bacteriologists, the acceptance of infectious causation has increased progressively over the past two centuries, albeit with sporadic acceleration and deceleration. An overview of this history reveals that the gradualness of this acceptance can be attributed to an overarching trend which continues in full force today: the recognition of ever more cryptic links between infection and disease. The degree of crypticity, in turn, can be attributed to a handful of characteristics, the importance of which depends on the intellectual traditions and technologies of the times.

One of the most obvious characteristics is size. During the last half of the 19th and first half of the 20th century, the infectious nature of diseases was recognized for progressively smaller parasites, with the discovery of multicellular parasites leading to the discovery of unicellular eukaryotes and large bacteria, and then smaller bacteria and viruses. The ability to recognize ever smaller parasites, however, only partly explains the trend toward an acceptance of infectious causation for an increasing number of diseases. Size of pathogens influences particularly the identification of pathogens, and though such identification had an important effect on the rigor with which infectiousness could be confirmed, the general acceptance of infectiousness for particular diseases depended more on other factors that contribute to the conspicuousness of chains of transmission.

The most conspicuous transmission chains occur when disease manifestations are externally apparent in a high proportion of infected individu-

als, when they occur soon after the onset of infection, and when contact between infected and susceptible individuals is easy to observe. Under these circumstances chains of transmission are apparent through everyday experience. Accordingly, acute respiratory diseases with conspicuous symptoms (such as coughing, sneezing, altered appearance of exposed skin, and severe illness) in a high proportion of infected individuals were generally accepted as infectious even before their causative organisms were identified. In such cases infectiousness was distinguished from contagiousness by the presumption of some entity in the diseased tissue that could grow and generate the same disease manifestations when transferred to susceptible individuals. ("Contagiousness" referred to the characteristic of being spread from diseased individuals to susceptible individuals, and could be attributed to poisons or infectious agents.)

The infectiousness of acute pulmonary diseases was confirmed and generally accepted long before the agents were identified. Measles offers an example. In 1846 Pierre Ludwig Panum was sent by the Danish government to investigate a measles epidemic on the Faeroe Islands, which lie midway between Norway and Iceland. The previous epidemic occurred in 1771, and the infectious spread and appropriate control measures were so apparent to some of the elderly inhabitants that they persuaded others to stay at home for a period that exceeded the duration of infectiousness. In so doing they probably saved nearly 20 percent of the Faroe Islands' inhabitants from infection [2]. Panum's report together with experimental transmission of measles in 1758 and 1842, led to a formal recognition of the infectiousness of measles, a century before the isolation of the measles virus [5].

Chains of transmission were generally more cryptic for sexually transmitted diseases than for acute respiratory tract diseases because early manifestations of sexually transmitted diseases tended to be silent and hidden by clothing, and because infectious sexual contact tends to be less observable. Those manifestations of sexually transmitted diseases that are externally apparent (e.g., tertiary syphilis or infertility) are difficult to link to infection because of the long and variable delay between the transmission event and their onset.

Evidence for this difference in recognition of infectiousness can be found in the medical texts of the late 19th century. The *Dictionary of Medicine,* for example, classified as infectious most of the common respiratory tract diseases, but none of the sexually transmitted diseases, even though *Neisseria gonorrheae* had been discovered five years earlier [1]. Syphilis was referred to as "A specific contagious, noninfectious disease; communicable by contact of the poison with a breach of surface, or by hereditary transmission" [6]. William Osler's *Principles and Practice of Medicine* referred to syphilis as infectious in the 1892 edition, but the infectious nature of gonorrhea was presented only in the later editions. The lag between pathogen identification and acceptance of infectious causation

was even longer for *Trichomonas vaginalis*, which was isolated and described in 1836, but was not generally accepted as a cause of vaginitis until more than a century later [7].

The potential for long-distance transmission was emphasized in the early 16th century by Girolamo Fracastoro, as a way of explaining how diseases like typhoid fever and plague sometimes seemed to leap from one population to the other. In the case of diarrheal diseases, however, this long-range transmission generated uncertainty about infectiousness even late during the 19th century, as evidenced by the argument over cholera transmission between supporters of Max von Pettenkopfer, who proposed that cholera welled up from the ground into air and then into people, and supporters of John Snow, who proposed waterborne transmission. The controversy continued for decades after the publication of Snow's conclusive evidence in 1854. Snow showed that cholera incidence was much greater among residents receiving water piped from a contaminated than from a relatively pure source, but the evidence was statistical rather than from a wealth of common experience. It was apparently this difference rather than a lack of scientific evidence that led to the prolonged disagreement. Even after the transmission of cholera was documented and the bacterial agents of typhoid fever and cholera were identified, experts walked the fence. "Asiatic cholera" was considered infectious, but transmission rather than infection was emphasized, and "simple cholera" was considered not to be transmissible [8, 9]. The presence of the typhoid bacterium in typhoid patients was mentioned, but the disease was attributed to transmissible "typhoid poison" [10]; and although unwholesome drinking water was cited as a "proximate and exciting cause of dysenteric disease," the pathogenic mechanism was attributed to a poison in the blood [11].

Recognition of the infectiousness of diarrheal diseases probably also was retarded because the high asymptomatic-to-symptomatic infection ratio of diarrheal pathogens would have made chains of person-to-person infection more cryptic, and because personal privacy makes mild diarrhea less conspicuous to others in the group than the sneezes, coughs, or dermal signs of acute respiratory infections [12]. Accordingly, infectiousness was accepted for diarrheal diseases later than for acute respiratory tract diseases and at about the same time as for sexually transmitted diseases.

The other long-distance transmission mode, vectorborne transmission, had a more circuitous route through the controversies over infectious causation. Nicholas Chervin, studying yellow fever in Barcelona in 1822, found no evidence for direct transmission. Though he did not consider other possibilities for long-distance transmission, his work helped sway the general opinion against contagiousness not only for yellow fever but for other diseases as well [13]. By continuing to advocate infectious causes of disease during the 1840s, Jacob Henle was considered a "gallant defender of an old fashioned error" [14]. This anti-contagion mindset of the medical main-

stream may help explain the ironic situation of the second quarter of the 19th century, during which medical experts were arguing that diseases such as cholera were not contagious, whereas the general public, who often witnessed direct spread in and among families, were convinced of its contagiousness [15]. Unlike diarrheal diseases, most vector-borne diseases cannot be transmitted directly from person to person; chains of transmission are therefore more difficult to observe. Accordingly, vector-borne diseases tended to be accepted as infectious slightly later than diarrheal diseases. The exceptions were based on real or apparent person-to-person transmission. Plague was accepted on the basis of its actual respiratory transmission. Dengue was accepted by some experts on the basis of spread among close contacts, which was erroneously interpreted as spread through direct contact [16]. At the turn of the century, the studies of Walter Reed and his colleagues, which built upon Carlos Finlay's hypothesis of mosquito transmission, resolved the matter for yellow fever, as did similar studies of vector-borne transmission for sleeping sickness and malaria.

Most of the common acute infectious diseases had been identified as infectious soon after the turn of the century. Although visual confirmation of most viruses would not occur until they were photographed with the electron microscope decades later, transmission by filtrates confirmed their infectiousness. In these transmission experiments, as in the observations of acute respiratory illnesses a century earlier, infectiousness was ascribed to unseen entities on the basis of chains of infection rather than identification of the infectious agent. The filtrate experiments were more rigorous than the earlier observational evidence because they were based on Koch's postulates; yet they stretched the postulates because induction of infections using filtrates was not as rigorous as induction of infection using isolated, identified organisms.

The availability of animal models was obviously important in making infectious processes more conspicuous, but even without animal models, some diseases were categorized as infectious on the same kinds of evidence that had been used throughout the 19th century to ascribe infectiousness to acute respiratory tract pathogens. Infectious mononucleosis, for example, was widely accepted as infectious from observational evidence decades before the Epstein-Barr virus was isolated in 1964.

During the first half of the 20th century, researchers began to confront another major barrier of crypticity: long delays between the onset of infection and the onset of disease. Long delays make cause-effect linkages cryptic because other events that occur during the intervening time can form the basis of alternative causal explanations. As the delay in onset of symptoms increases, the number of such events and, hence, the number of alternative hypotheses of causation increases. The alternative hypotheses may focus on specific environmental insults, or may interpret delayed, persistent symptoms as natural wear and tear, particularly if infections are ubiquitous.

Infectious causation of diseases with long delays also tend to be cryptic because such may occur in a minority of those infected or are variable in expression when they do occur. Finally, if a disease requires years or decades to produce characteristic symptoms, fulfilling Koch's postulates by experimental transmission studies will require comparably long amounts of time. However, led by the long-term studies of infectious diseases such as syphilis and tuberculosis, which have both acute and chronic stages, researchers began to recognize that chronic diseases could be delayed consequences of infections with very different acute-phase manifestations; chickenpox and shingles, for example, were recognized as two phases of a single infectious process [17, 18].

Rheumatic fever illustrates how these crypticity factors played out during this period. The lag between the early symptoms of acute streptococcal infection (e.g., sore throats) and rheumatic fever typically range from a few weeks to a few months [19]. Toward the end of the 19th century, epidemiological links were noted between rheumatic fever and both sore throats and scarlet fever; infectious causation was proposed in 1895 by Arthur Newsholme, who argued against the generally accepted view that it was a hereditary disease [19]. During the first two decades of the 20th century, researchers actively investigated infectious causation of rheumatic fever, but could not find bacteria at the sites of damage. From the 1930s through the mid-1950s, research first associated and then causally linked rheumatic fever with *Streptococcus pyogenes*. The experimental control of rheumatic fever through the use of antibiotics from 1939 through 1955 provided the last batch of evidence necessary for general acceptance of infectious causation [19]. The process from suspicion of infectious causation to general acceptance therefore spanned a half century of scientific effort. Infectious causation of rheumatic fever was cryptic because of the lag between the onset of infection and the onset of disease, differences between early- and late-phase disease, the low frequency of disease relative to infection, and the role of autoimmunity, which can dissociate infectious organisms from the damage that they indirectly induce through the immune system.

THE CURRENT ROLE OF CRYPTICITY

Health science is still grappling with crypticity of infectious causation. The linking of infectious processes to cancer illustrates the difficulties. At the beginning of the 20th century, Peyton Rous discovered that cancer in chickens could be caused by a transmissible agent, later named Rous's sarcoma virus [20]. Infectious causation was much more apparent for this cancer than for human cancers, however, because Rous's sarcomas typically occurred about two weeks after the onset of infection. Human cancers sometimes occur several decades after the initial infection, as is the case, for example, with leukemia caused by HTLV-1 (known as the human

T-cell lymphotropic virus type one, or, alternatively, the human T-cell lymphoma/leukemia virus type one). Mother-to-offspring transmission of HTLV-1 makes its role in cancer causation even more cryptic because such transmission resembles transmission of genetic predispositions. When confronted with such sources of crypticity, experts have tended to attribute the cancers solely to genetic defects, senescence, and environmental insults. This predisposition may have been fostered by the discoveries during the first half of the century that environmental insults such as radiation and mutagenic chemicals could cause cancer relatively soon after exposure.

The discovery during the 1960s that Epstein-Barr virus and *Plasmodium falciparum* jointly caused Burkittís lymphoma renewed interest in infectious causation of human cancer, and made apparent another source of crypticity: a single disease may be caused by the joint action of more than one kind of pathogen. Substantial research support was directed to virologists during the early 1970s as part of Nixon's "War on Cancer." But the infectious connections were still too cryptic to allow quick establishment of infectious causation. The empirical linking of pathogens with cancer began its steady growth only around 1980 when HTLV-1 was linked to blood cell cancers [21]. About 15 percent of human cancers in the early 1990s were attributable to the infectious agents then known to cause cancer [22, 23], and since then the list of accepted infectious agents of cancer has grown to include human herpes virus 8, hepatitis C, and *Helicobacter pylori* [24–28].

There is no reason to believe that these discoveries have completed the list of infectious causes of human cancer; the demonstration of infectious causation has continued steadily over the past two decades, dismissal of infectious causation typically has been made on the basis of little if any evidence, and many cancers of unknown etiology show signs of infectious causation [29]. In breast tumors, for example, infectious causation is suspected, and proteins and nucleic acids that are homologous with those of oncogenic retroviruses have been observed more frequently in breast tumors than in normal tissue [30–32]. A recent study has also documented an association between Epstein-Barr virus and breast tumors [33].

The past two decades have revealed infectious causation in other chronic diseases as well. During the 1970s and 1980s, medical texts typically attributed peptic ulcers to gastric acidity, stress, smoking, alcohol consumption, and genetic predispostions (e.g., [34]). Infectious causes were not mentioned, even though evidence of infectious causation had been accumulating from the late 19th century through the mid-20th century: a spiral bacterium was associated with gastric ulcers at the end of the 19th century, ulcers had been experimentally transmitted in lab animals during the second decade of the 20th century, and peptic ulcers had been successfully treated with antibiotics in New York City hospitals during the late 1940s [35–37]. Infectious causation of peptic ulcers is now generally accepted by medical authorities, though this acceptance has grown only gradually since

the mid-1980s. In spite of the accumulated evidence, several attributes of ulcers made infectious causation cryptic: the loose correlation between infection and ulcers, the internal site of infection, and variable delays between the onset of infection and the onset of overt disease. The net effect is a chain of transmission that is so cryptic that the transmission mode of *H. pylori* is still unclear today.

The acceptance of infectious causation of atherosclerosis appears to be following a similar script. An infectious origin for atherosclerosis was proposed over a century ago, and incrimination of inflammation as an initiation step dates back to the early 19th century [38]. Consideration of infectious causation waned in the 20th century, in spite of a great amount of supportive evidence from both animal models and humans and no evidence that could justify rejection of infectious causation [38]. Particularly during the third quarter of the 20th century, medical texts restricted attention to cholesterol, high fat diets, stress, smoking, and genetic predispositions as primary causes of atherosclerosis. In the late 1970s, infectious causation was reconsidered, and in 1988, Saikku, Leinonen, and colleagues published serological evidence that implicated *Chlamydia pneumoniae* [39]. Since then a large body of evidence has supported a primary role for infection, with *C. pneumoniae* as the leading suspect [40–43]. The results have been mixed, however, with some studies failing to confirm these associations. Such contradictions are to be expected as ever more cryptic cases of infectious causation are studied: the tools that are useful in identifying infectious causes of acute diseases cannot be expected to function uniformly well for chronic diseases. When the tools function well, infectious causation is accepted. Left in the wake are those chronic diseases that may be caused by infection but may be difficult to detect with the conventional tools and approaches at hand. In the case of *C. pneumoniae*, serological positivity to acute pulmonary infections make detection of serological positivity to chronic infections particularly difficult, and may therefore contribute to discrepancies among studies. As with peptic ulcers, the site of infection is internal and the chains of infection are cryptic. As was the case with both peptic ulcers and gonorrhea, the presence of bacteria in lesions has been dismissed by critics as harmless bacteria colonizing damaged tissue. *C. pneumoniae* is also a leading suspect in Alzheimerís disease, but in this case the long process from the first critical studies to a resolution of the matter has just begun [44].

Because genetics can alter the course of infection, we expect to find that genetic determinants of disease may sometimes be best explained as genetic influences on infection. This situation is illustrated by recent work on the chronic diseases for which *C. pneumoniae* is implicated. The epsilon 4 allele of the human apolipoprotein E gene has been identified as a genetic risk factor for atherosclerosis, stroke, and Alzheimer's disease. It also appears to increase susceptibility to *C. pneumoniae* infection [45]. The

genetic risk imposed by the epsilon-4 allele may therefore result from a genetic vulnerability to infection rather than a direct influence of epsilon-4 on disease progression. The high allelic variability of genes that are involved in resistance to pathogens (e.g., HLA variability) suggests that this situation may be common.

Similarly, exposures to noninfectious agents may exacerbate presently unknown effects of infection. Atomic bomb survivors infected with HTLV-1, for example, had increased rates of diseases for which HTLV-1 has not yet been identified as a cause [46]. Lack of exposure to trace nutrients may also generate disease by increasing a person's vulnerability to infection. Both epidemiological associations and experiments with mice indicate, for example, that coxsackie viruses are more likely to cause the myocardial damage characteristic of Keshanís disease under conditions of selenium deficiency [47].

Such examples are generating a growing sense among some researchers that many more chronic diseases will prove to be caused by pathogens that may be (1) familiar causes of acute infection, (2) identified but not yet associated with disease, or (3) not yet identified [26, 27, 48, 49]. Understanding the sources of crypticity should facilitate recognition. The key problem is how to facilitate recognition of infectious causation among these diseases. One step toward resolution of this problem involves increased awareness of the sources of crypticity that we are likely to encounter in ascribing infectious causation. One source of crypticity is the increasing difficulty in obtaining suitable animal models. Few mammals live as long as humans. It is therefore difficult to find experimental animals that can be infected by an organism thought to cause long-delayed chronic disease and that then survive long enough to demonstrate the same chronic disease found in humans. Even if possible, these procedures may be prohibitively expensive. We can therefore expect experimental documentation of disease transmission to be less feasible in the future than it has been in the past. This problem has been encountered with AIDS research, even though the time between infection and AIDS is only moderately long. HIV causes immunodeficiency disease in other animal species, but the details of these disease syndromes differ from AIDS. Although it is generally accepted that HIV "causes" AIDS, the difficulty in finding animal models is one of the problems that has led to the questioning of this causal link.

Little if any acute pathology near the onset of infection is another source of crypticity that may inhibit recognition of infectious causation among chronic diseases. Pathogens are often classified as relatively harmless or even commensal without sufficient long-term study to warrant such a classification. The historical record illustrates the consequences of this error. Epstein-Barr viruses and human papillomaviruses were once thought of as relatively harmless on the basis of their linkage to relatively benign diseases that occur soon after infection (infectious mononucleosis and warts

respectively). But each virus can cause lethal cancers. Bacteroides was once thought to be a harmless commensal, but recent evidence indicates that it may be linked to ulcerative colitis [50].

Co-infectious causation may be another source of crypticity among those diseases that are yet to be linked to infectious causation. The joint action of more than one pathogen has also been disproportionately represented in recent discoveries of infectious causation. For example, replication of the hepatitis D virus, which causes an unusually severe hepatitis, is dependent on hepatitis B infection [24]. Kaposi's sarcoma and other AIDS-associated illnesses often require the predisposing effects of HIV infection. The body of evidence for atherosclerosis, for example, indicates that more than one pathogen may often be involved [38].

When the discoveries of infectious causation during the last quarter of the 20th century are viewed in the broader context of the 19th and 20th centuries, the inescapable conclusion is that we are still in the midst of a long process of recognizing infectious causation. In other words, we are still in the midst of establishing the scope of the germ theory of disease. The process has been prolonged because the crypticity of infectious causation is highly variable among diseases. The challenge is to identify those diseases that are caused by infection so cryptically that the tools and approaches of the past will no longer definitively establish infectious causation. To meet this challenge we must first identify those diseases that are likely candidates for infectious causation.

Markers of Infectious Etiology

The Relevance of Evolutionary Fitness

One useful tool for diagnosing infectious causation can be derived from the central principle of evolutionary biology: evolutionary fitness. Estimates of the fitness costs that are attributable to a particular disease (averaged over the entire population) can be used as an indicator for assessing whether the disease could reasonably be ascribed to genetic as opposed to infectious causation. For a genetic disease to be maintained at equilibrium in a population, the loss of the allele for the disease must equal the rate at which the allele is reintroduced. If the allele does not provide a fitness benefit, the loss due to the fitness costs of the disease would need to equal the rate at which the allele is generated through mutation. This reasoning leads to the conclusion that estimates of fitness costs can provide a sense of whether the disease is attributable to something other than simple genetic causation. Any human disease with a frequency that is too high to be maintained by the mutation rate is implicated as being caused by something other than just human genes. If the disease is inherited in Mendelian ratios and is too widespread to be accounted for by founder effects or

genetic drift, and if the time has been sufficient for natural selection to drive the frequencies of deleterious alleles to low levels, the allele must have conferred some compensating fitness benefit. The only such compensating fitness benefit that has been documented for major human genetic diseases is resistance to infection (see below).

How much time is necessary for a close approach to mutational equilibrium for deleterious alleles? Recorded history spans about 5,000 years, and the frequencies of deleterious alleles can change substantially in a small fraction of that time. Consider the sickle-cell allele, which protects heterozygotes from malaria and causes a lethal anemia in homozygotes. If a population with a sickle-cell allele frequency of 20 percent were transferred to an environment without malaria, the negative effects of sickle-cell anemia should cause the frequency of the sickle-cell allele to decrease by a factor of about three within about 10 generations. This estimate accords with geographic differences. In the United States, the sickle-cell allele frequency is about half of what would be expected from African source populations after accounting for admixture [51]. A geographic comparison within the Caribbean between malaria-free and malaria-endemic areas is also consistent: the frequency of the sickle-cell allele is low in Curaçao, where malaria has not been endemic, but not in Surinam, where malaria has been endemic [51].

Some diseases that are too severe to have been maintained over evolutionary time can be attributable to new environments [52]. If a new environment does not cause disease in all genetic variants, then its net effect is to make some alleles disadvantaged. If the resulting disease is even moderately severe, the selective differentials so generated will reduce the frequencies of the causative alleles to low levels within a few hundred generations. We therefore do not expect to see genetic diseases that are both severe and common (e.g., a frequency of 1 percent for homozygous recessive diseases like sickle-cell diseases) to persist solely as a legacy of a pre-agricultural environment if people without the disease syndrome have existed in the population. The restrictions are more stringent on genetic diseases due to dominant alleles than on those due to recessives. The restrictions on either category of diseases would be still more stringent if there had never been compensating advantages, because then one would have to explain how the frequency of the disadvantaged allele could ever have increased above the mutational equilibrium. Diseases caused by environmental changes therefore are expected to decrease with time as the new environment becomes an old environment, whenever the new environment causes a disease in only a portion of exposed individuals. In this context the distinction between diseases due to deleterious alleles and diseases due to new environments vanishes, because the negative effects of new environments diminish as the alleles that are poorly suited to the new environments are lost.

When the necessary variation is present in a population of vertebrates, evolutionary responses can be very rapid, with substantial change occur-

ring in 10 to 20 generations [53, 54]. If environmental changes such as the agricultural diet of the past few thousand years were the primary cause of a disease that substantially affects fitness, then we would expect natural selection to have reduced the negative effects, especially when the existing variation in disease among people exposed to the same diet indicates that the variation necessary for natural selection exists (e.g., variation in vulnerability to atherosclerosis, stroke, and Alzheimer's conferred by variation at the apolipoprotein E locus).

The flexibility of defense and repair systems may also contribute to a lack of disease caused by new noninfectious threats. Cancers, for example, are commonly believed to be consequences of newly introduced chemicals in the diet. Yet in contrast to infectious agents, little evidence implicates typical doses of dietary chemicals as primary causes of human cancer, probably because humans have evolved effective flexible enzymatic systems for degrading potentially carcinogenic chemicals [55]. Even aflatoxins, which are one of the most carcinogenic of dietary constituents, may exert their negative effects largely in conjunction with viral infection [56].

When diseases are not inherited in Mendelian ratios, geneticists have argued that more complex genetic effects could be occurring (e.g., multilocus effects or segregation distortion). But if a disease is a "genetic disease" caused by damaging allelic instructions rather than a disease that depends on genetic predispositions to threats from infection or the physical environment, the diseases should show very distinct patterns among twins. Because monozygotic twins are virtually identical in genetic makeup, the monozygotic twin concordance for such genetic diseases should be virtually 100 percent. (Somatic mutations and mitochondrial differences could in theory cause some discordance, but it is difficult to imagine how such effects could account for more than a trivial deviation below 100 percent concordance for damaging diseases.) Even moderately high monozygotic twin concordances together with a reduced dizygotic twin concordance is insufficient evidence to exclude infection as a primary cause. Leprosy, for example, has about 60 to 80 percent monozygotic concordance and about a 20 percent dizygotic concordance; the analogous figures for tuberculosis are about 50 percent and 20 percent [51, 57]. The difference between monozygotic and dizygotic twin concordance is as one might expect from Mendelean inheritance. Susceptibility to leprosy and tuberculosis may have a substantial genetic basis, but we do not categorize them genetic diseases. A disease is categorized as infectious if elimination of the infectious agent(s) would eliminate the disease.

EVOLUTIONARY MAINTENANCE OF SEVERE INFECTIOUS DISEASES

According to current evolutionary theory, highly damaging infectious diseases (e.g., those causing a fitness loss of over 1 percent) can persist

indefinitely in human populations [58]. The underlying logic recognizes that natural selection favors pathogen variants that compete most effectively for the resources in their hosts. In some circumstances (e.g., when healthy hosts are important for transmission), the most successful competitors will be benign. In other circumstances (e.g., when severely ill hosts are efficient sources of transmission), the most successful competitors may be those that severely exploit their hosts and thus have a strongly negative effect on host fitness. From the hosts' perspective, the less damage the better. But because pathogens have short generation times and high population growth rates, they have a great potential for continually leading the coevolution away from benignity. If a host population has a particularly effective defense against a pathogen, the high rate of pathogen evolution may allow the pathogens to evolve around the defense before all of the hosts are resistant.

This argument does not specify the details by which selection on pathogens maintains the disease over time. The characteristics of chronic disease, for example, could result from within host evolution that generates increased virulence over time, as with the causation of AIDS [59]; or, chronic manifestations could result from compromises to host defenses, the aging process, reactivation of infections from a latent state, and/or coinfection with other pathogens. The important point is that damaging diseases can be maintained indefinitely over time when infectious agents are the cause.

This modern perspective on the long-term stability of relatively high virulence is now directly supported by archaeological evidence. Molecular evidence from Peruvian and Egyptian mummies demonstrates the persistence for thousands of years of tuberculosis and Chagas disease in South America, and smallpox, polio, tuberculosis, and falciparum malaria in the Old World [14, 60–63].

Fitness Load as an Indicator of Infectious Causation

The preceding considerations suggest that when diseases have been common in human populations for many generations and still have a substantial negative impact on fitness, they are likely to have infectious causes. These considerations suggest that the most important of the human diseases that are not now thought to be caused by infection will eventually be shown to have infectious causes. These infectious causes may be causes in the proximate sense (e.g., *Mycobacterium tuberculosis* causes tuberculosis) or in an evolutionary sense (*Plasmodium falciparum* is an evolutionary cause of sickle-cell anemia). According to these considerations, the major exceptions will occur when diseases are caused by an environmental change too recent to allow for evolutionary adaptation by the host [52]. If effects on fitness are so slight that they can be maintained solely by mutation rate, the

disease will not fall in the "most important" category because its effects are slight on afflicted individuals or very rare in the population.

The practical application of this insight depends on estimations of the evolutionary pressure against the disease syndrome or, more specifically, on quantifying the negative effect of a disease on the fitness of an average individual in the population under study. This negative effect can be estimated by multiplying the disease prevalence by the fractional decrease in fitness attributable to the disease among those who have the disease. This product, which we call "fitness load," allows evaluation of whether maintenance of the disease in the population is likely to depend on an infectious agent.

Fitness in this analysis refers to "inclusive fitness." As defined by Hamilton, inclusive fitness combines the direct effects of a characteristic on the individualís own survival and reproduction together with the indirect effects of the characteristic on other individuals, such as siblings, who carry the alleles that code for the characteristic [64]. These indirect effects are weighted by the probability that the other individuals carry the allele. The approach considers age-dependent fitness [65], and then estimates the loss in this age-dependent fitness that would result from the age-dependent manifestations of disease, such as death and infertility. Fitness load is similar to genetic load and estimates of selective differences between genotypes [66–68]. If fitness load were calculated for a known dominant genetic disease with known frequencies of disease-causing alleles, it would be essentially the same as these genetically based quantities. Fitness load, however, is calculated as the loss in fitness due to disease (weighted according to standard inclusive fitness ratios) rather than losses due to the selective inferiority of alternative genotypes. If the disease is infectious, the losses due to disease may result partly or entirely from characteristics of disease organism rather than from host alleles.

We assume that noninfectious environmental and genetic influences will be present even when the primary cause is found to be infectious. We use fitness load to identify those diseases that appear to be too damaging for too long to be maintained without infectious causation. We estimate the fitness load that would occur without antibiotics, surgery, and modern forms of birth control, because we are assessing the feasibility of different causal hypotheses for the long-term maintenance of the disease in the population. The most relevant data are from populations with the highest incidences of the disease, so long as these high incidences have been present over many generations, and hence can be considered relatively stable. Considering the potential effects of many variables that have not yet been accurately quantified (such as age-specific inclusive fitness effects associated with decreased stamina or menopause), we do not expect the accuracy of our estimates to be much better than an order of magnitude. However, variation in fitness load among entities recognized as diseases spans several orders of magnitude. Estimates accurate to within an order of magnitude

are therefore useful for illustrative purposes and for directing attention to diseases that are unlikely to be maintained without infectious causation.

We first consider the estimation of fitness load for three categories of diseases. The first category, illustrated by juvenile diabetes, may be directly triggered by infection. The second, illustrated by cystic fibrosis, is caused by a deleterious allele, which increases vulnerability to damaging secondary infections but appears to be maintained in the population at least partly as a result of protection against infection among heterozygotes. The third, illustrated by falciparum malaria, is known to be infectious.

Juvenile diabetes (also called type I diabetes or insulin-dependent diabetes) is caused by cessation of insulin production, usually during adolescence or childhood. Without treatment, death usually occurs within a few years, generating a fitness reduction among cases of at least 50 percent. As juvenile diabetes typically occurs in 0.2 to 0.3 percent of a population [51], the fitness load is calculated to be slightly above 0.001.

Cystic fibrosis is a recessive hereditary disease that results in chronic obstructive pulmonary disease and pancreatic insufficiency. It is most frequent among people of European origin (about 1 out of 2,400 births). Most males with cystic fibrosis are infertile, and females show reduced fertility. The immediate cause of death is typically secondary infection associated with poor clearance of viscous secretions. Before the development of antibiotics, the fitness of homozygotes was close to zero (few patients lived to 20). Our estimate of fitness load due to cystic fibrosis is therefore 0.0004.

For a few diseases, such as falciparum malaria, an indirect but more inclusive method of estimating fitness load is feasible. This indirect method involves calculation of the fitness load associated with a genetic defense. For malaria, the genetic defense is sickle-cell anemia. Before modern medical treatment, homozygotes almost always died during childhood, and therefore had near-zero fitness [51]. If the sickle-cell allele is in equilibrium, if its gene frequency is 20 percent, and if sickle-cell homozygotes have zero fitness, then the fitness load attributable to malaria would be approximately 0.16. This indirect measurement of the fitness load is probably more accurate than estimates based on life history markers, because it subsumes hard-to-measure effects such as reductions in fertility, fetal viability, and investment by parents and grandparents. Of course, this method for calculating the fitness load for known infectious diseases, such as falciparum malaria, is not directly useful for calculating the fitness load for diseases of unknown etiology, because it requires recognition of infectious causation; however, because the method obviates the necessity of making questionable assumptions that enter into calculation of fitness load from life history data, it provides a benchmark value of fitness load for one of the most damaging of human infectious diseases.

If a disease is caused by a dominant mutant allele, generates no other benefit, and is at equilibrium, the fitness load per generation should

approximately equal the per-generation mutation rate. That is, the loss of deleterious alleles through selection should equal the introduction of deleterious alleles through mutation. Neurofibromatosis is the most common damaging disease that is due to a dominant mutation and is maintainable by strictly by new mutations, which are generated at a rate of about 10^{-4} (i.e., one new mutation per 10^4 persons per generation). This extremely frequent generation of neurofibromatosis through mutation is possible because the neurofibromatosis gene is extremely long. Accordingly, the fitness load of neurofibromatosis is also about 10^{-4}.

The most damaging infectious diseases have fitness loads much larger than could be maintained by mutational rate (Table 1), and most seem to have existed for too long to have been caused by a new environment. By looking only at fitness load, one could therefore have predicted that smallpox, tuberculosis, malaria, typhoid, bubonic plague, yellow fever, and cholera all were of infectious origin, even if we had known nothing of their epidemiology. We would have to know only that they were not rare, significantly decreased fitness, and had existed in human populations over many generations.

Several common disorders reduce reproduction sufficiently to generate high fitness loads. Schizophrenia has an fitness load that is at least 0.005, much larger than that of any autosomal dominant genetic disease (Table 2). Many diseases ascribed to autoimmunity, such as Crohn's disease, ulcerative colitis, rheumatoid arthritis, multiple sclerosis, juvenile diabetes, and lupus, have fitness loads above or only slightly below 0.001 (Table 1), high enough to implicate infectious origins.

All of the diseases with high fitness loads (>0.01) and known causes are either directly caused by infection (Table 1) or are genetic diseases that are maintained at high frequency by infection (e.g., sickle-cell anemia; see Table 2). Although the lists in Tables 1 and 2 are not comprehensive, we know of no exceptions to this generalization.

Diseases with fitness loads less than 0.01 but above 0.001 tend to fall into these two categories or are chromosomal abnormalities (which may be generated by infection, as discussed in the next section). A greater proportion of diseases in this range of fitness loads cannot yet be classified in terms of causation. The evolutionary perspective may prove especially useful for this category by emphasizing the need to investigate hypotheses of infectious causation. It thus emphasizes the need to investigate infectious causation for diseases such as schizophrenia and breast cancer.

Any of the three categories of disease causation could stably maintain diseases with low fitness loads (i.e., <0.001). Infectious causation is therefore neither implicated or counterindicated for such diseases.

The value of a new perspective depends largely on the new predictions it generates. Without invoking the logic of evolutionary biology, some researchers have concluded or strongly suspected that diseases such as mul-

tiple sclerosis, schizophrenia, breast cancer, and atherosclerosis are caused by infection. What is an example of a disease that we predict to be infectious on the basis of evolutionary principles even though infectious causation is not currently being pursued as an hypotheses? We offer polycystic ovary disease. It is a cause of infertility among women second only to the tubal scarring caused by *Chlamydia trachomatis* and *Neisseria gonorrhea*. Polycystic ovary disease typically occurs in about 5 to 10 percent of women and is characterized by waxy white ovaries, diabetes-like disturbance of insulin levels, and a variety of hormonal abnormalities, which contribute to high miscarriage, reduced conception, and absence of menstruation. It has a fitness load above 0.01, primarily as a result of its negative effects on fertility. This fitness cost is too high for it to be maintained as a genetic disease by mutation pressure, and the disease is not inherited in simple Mendelian ratios.

With fitness loads of 0.01 weeding out strictly genetic diseases, assessments of post-menopausal contributions to inclusive fitness become theoretically important, because post-menopausal contributions are probably of this magnitude. If so, inclusive fitness contributions of post-menopausal women may weed out genetic causes of severe diseases that occur after 50 years of age. It is difficult to estimate the fitness impact of diseases that strike women in late middle age. Reproduction after menopause is zero, and the inclusive fitness effects from aid to children, grandchildren, and other related individuals is difficult to measure. If female inclusive fitness after menopause is comparable to male inclusive fitness at the same ages, as evolutionary theory suggests, the expected inclusive fitness of women who reach age 60 would be about 2 percent (i.e., because male fitness due to reproduction is that high). If people in these older age groups are making such fitness contributions, diseases that kill at relatively high frequencies in older age groups can generate fitness loads that are considerably higher than that maintainable by mutation pressure; atherosclerosis, for example, would have a fitness load that is probably slightly above 0.01 as a consequence of severe effects, such as myocardial infarction, stroke, and impotence, which tend to occur after reproduction but while increments in inclusive fitness can still be generated.

Atherosclerosis is an interesting test case because manifestations of arteriosclerosis can be observed in mummies of ancient Egypt [14]. Atherosclerosis has been attributed to dietary, physiological and genetic causes, but all known risk factors are insufficient to explain the total amount of atherosclerosis disease. Infectious causation of atherosclerosis is therefore implicated by the entirety of evidence: its moderately high fitness load, its lack of strong monozygotic twin concordance, its long history, the insufficiency of other risk factors to explain its magnitude, and the unmasking of its associations with infectious agents, particularly *C. pneumoniae.*

Some cancers similarly occur sufficiently early and often to generate fitness loads that are sufficiently large to implicate infection and/or new envi-

TABLE 1

ESTIMATES OF FITNESS LOAD FOR DISEASES THAT ARE CAUSED PRIMARILY BY INFECTION
OR NONINFECTIOUS ENVIRONMENTAL AGENTS, OR FOR WHICH CAUSES ARE UNKNOWN

Fitness Load*	Disease or Syndrome	Causation†	Comments‡
<1.0	Malaria, falciparum	i	West Central Africa
	Kuru	i	Fore tribe, New Guinea
<0.1	AIDS	i	East and Central Africa
	Atherosclerosis	i?	U.S.; includes impotency
	Bubonic/pneumonic plague	i	14th-century Europe
	Chagas disease	i	Venezuela, Brazil
	Cholera	i	mid-19th-century London
	Diphtheria	i	general
	Endometriosis	u	general
	Gonorrhea	i	general
	Hookworm	i	southeastern U.S., Puerto Rico
	Leishmaniasis, visceral	i	India, caused by *Leishmania*
	Leprosy	i	*donovani*
	Malaria, vivax	i	general
	Measles	i	18th-, 19th-century Europe
	Onchocerciasis	i	general
	Pelvic inflammatory disease	i	river valleys in tropical Africa
	Pertussis	i	general
	Pneumococcal pneumonia	i	general
	Polycystic ovary disease	u	general
	Rotavirus	i	U.S., Western Europe
	Scarlet fever	i	general
	Schistosomiasis	i	general
	Shigellosis	i	Egypt, China
	Sleeping sickness	i	South Asia
	Smallpox	i	sub-Saharan Africa
	Syphilis	i	general
	Tuberculosis	i	general
	Typhoid	i	general
	Yellow fever	i	general
			sub-Saharan Africa, Neotropics
<0.01	Ascariasis	i	Central Africa
	Breast cancer	u	U.S., Western Europe
	Cerebral palsy	u	U.S.
	Cervical cancer	i	U.S.
	Cysticercosis	i	Central America
	Dengue fever	i	Southeast Asia, Neotropics
	Endomyocardial fibrosis	u	tropical Africa
	Hepatitis B	i	West Africa, Southeast Asia, Brazil
	Hepatitis C	i	Egypt, Bolivia
	Influenza	i	U.S., Western Europe
	Juvenile (type I) diabetes	u	U.S., Western Europe

Fitness Load*	Disease or Syndrome	Causation†	Comments‡
	Kaposi's sarcoma	i	Central Africa
	Lung cancer	e, i?	U.S., Western Europe
	Clinical depression	i	U.S., Western Europe
	Meningitis	i	sub-Saharan West Africa
	Peptic ulcer	i	U.S., Western Europe
	Poliomyelitis	i	general
	Pre-eclampsia and eclampsia	u	U.S., Europe
	Rheumatic heart fever	i	general
	Rheumatoid arthritis	u	general
	Rubella	i	general
	Schizophrenia	u	U.S., Europe
	Tetanus	i	general
	Trichomoniasis	i	general
	Typhus	i	general, 19th-century Europe
	Yaws	i	tropical Africa
<0.001	Adult-onset diabetes	u	Amerindians, Aboriginal Autralians, Polynesians
	Anorexia nervosa	u	U.S.
	Burkitt's lymphoma	i	tropical Africa, New Guinea
	Crohn's disease	u	U.S., Western Europe
	Histoplasmosis	i	Central U.S.
	Hodgkin's disease	u	general
	Multiple sclerosis	u	U.S., Western Europe
	Nasopharyngeal cancer	i	South China
	Osteoporosis	u	Europeans
	Paget's disease	u	Great Britain
	Prostate cancer	u	U.S., Europe
	Rabies	i	general
	Sarcoidosis	u	U.S., Western Europe
	Ulcerative colitis	u	U.S., Western Europe
<0.0001	Anthrax	i	general
	Parkinson's disease	u	U.S., Europe
<0.00001	Botulism	i, e	general
	Lyme disease	i	Europe, North America
	Rocky Mountain Spotted fever	i	Western U.S.

*Each category in the fitness load column corresponds to estimated reductions in inclusive fitness that are below the value given but not below the next lowest category.

†i=infectious cause; e=noninfectious environmental cause; u=unknown cause; ?=cause suggestive but not certain

‡Geographic areas are given when calculations were based on a particular population; "general" refers to derivations from general statements about lethality, infertility, and/or prevalence without restriction of comments to a particular geographic area.

SOURCE. Most of the information on which these estimates were based was obtained from general texts such as [145, 146].

TABLE 2

FITNESS LOAD AND THE ROLE OF INFECTION IN GENETIC DISEASES

Fitness Load*	Disease or Syndrome	Frequency of Condition (%)	Relevance of Infection†	Comments‡
<0.1	Sickle-cell anemia	4	defense; infectious damage during crisis	West Central Africa; autosomal recessive; heterozygotes are resistant to *P. falciparum* malaria
	Glucose 6-phosphate dehydrogenase deficiency	25	defense	southern European males; X linked recessive; confers protection against malaria; 400 variant deficiency alleles; manifestations include favism, hemolytic anemia and neonatal jaundice
<0.01	Thalassemias	α : 20–40 β : 3–9	defense	Thailand; confers protection against malaria
	Klinefelter's syndrome (XXY)	0.1	genetic damage caused by infection?	supernumerary chromosomes
<0.001	Cystic Fibrosis	0.05	defense; tissue damage from infection	northern Europeans; confers protection against typhoid fever; pathology results from genetically based vulnerability to respiratory tract infections
	Turner's syndrome (XO)	0.04	genetic damage caused by infection?	among females; chromosomal deletion
	Hemochromatosis	5	defense?	northern Europeans; modification of membrane protein analogous to genetic defect of cystic fibrosis
	Duchenne's muscular dystrophy	0.02	none	X-linked; maintenance of relatively high fitness load by mutation may occur because the gene is very long; 2,300 Kbp
	Tay-Sachs	0.03	defense?	Ashkenazi Jews; different mutations with same phenotype indicate that ethnic and geographic distribution of disease is not simply a founder effect
	XXXY	0.04	genetic damage caused by infection?	among males; supernumerary chromosomes
<0.0001	XYY	0.1	genetic damage caused by infection?	among males; supernumerary chromosomes
	XXX	0.1 (in females)	genetic damage caused by infection?	supernumerary chromosome; sometimes causes mild mental retardation

	fitness load	tissue damage from infection?	
Polycystic kidney disease	0.1		dominant allele; late-onset; high mutation rate
Alpha-1-antitrypsin deficiency	0.07	defense?	Scandinavia; protease inhibitor; deficiency associated with emphysema and liver disease
Myotonic dystrophy	0.06	defense?	autosomal dominant; variable penetrance; trinucleotide repeat
Duffy negative	90–100	defense	West Africa; confers protection against vivax malaria
Gaucher's disease	0.04	defense?	European Jews
Neurofibromatosis	0.03	none	maintenance by mutation facilitated because gene is long (350 Kbp) and mutation rate very high
Fragile X	0.03	defense?	X-linked; trinucleotide-repeats
Color-blindness	8	no	West European males; X-linked recessive; fitness load is estimated from post-hunter-gatherer societies; high prevalence can occur because fitness costs of color-blindness are presumably very low in such societies
Phenylketonuria (PKU)	0.01	defense?	U.S. Caucasians; deficiency in phenylalanine hydroxylase; causes mental retardation; treatable by dietary means
Achondroplasia	0.02	none	frequency at birth; autosomal dominant; most frequent form of short-limb dwarfism
Hemophilia	0.01	none	among males; X-linked recessive
Prader-Willi syndrome	0.004	none	autosomal dominant; gene deletion
Tuberous sclerosis	0.004	none	autosomal dominant
Adenomatous polyposis of colon	0.008	none	autosomal dominant
<0.00001 Huntington's disease	0.006	defense?	autosomal dominant; trinucleotide repeats
Apert syndrome	0.001	none	autosomal dominant

*Each category in the fitness load column corresponds to estimated reductions in inclusive fitness that are below the value given but not below the next lowest category.

†"Defense" indicates genetic characteristic that confers protection against infection; "none" indicates that infection is neither known nor suspected of being relevant; "?" indicates that the role is unproven but that there are theoretical reasons for expecting the role to be occurring.

‡Geographic regions and subpopulations correspond to those for which frequencies were derived.

SOURCE. Most of the information on which these estimates were based was obtained from general sources, such as [51, 145–147].

ronmental changes. For most of these cancers, new noninfectious environmental causes seem insufficient to explain the high incidence, whereas cryptic infectious causes are feasible and often already identified or implicated by suggestive evidence. Breast cancer, ovarian cancer, and Hodgkin's disease are in the implicated category; Hodgkin's disease, for example, has a fitness load comparable to that of ovarian or breast cancer (Table 1) and is associated, perhaps causally, with Epstein-Barr virus [69, 70].

Models of infection-induced cancers suggest that compounds generated from infection may interact with noninfectious environmental factors in the oncogenic process (e.g., for associations of *H. pylori* and smoked foods with stomach cancer, see [71]). Lung cancer has a high fitness load and may be the strongest example of a chronic widespread disease associated with a high fitness load and caused primarily by a new environmental agent: cigarettes. Yet even in this case, a causative interaction between smoking, infection, and lung cancer is consistent with available evidence because smoking and lung cancer are correlated with infection, and some lung cancers are associated with viral infection independently of smoking [72–75]. The association between asbestos exposure and mesothelioma is similarly confounded with viral infection [76].

Trends in identification of causal mechanisms provide a sense of what we can expect in the near future. Over the past two decades the evidence implicating infectious causation of cancer has increased steadily, whereas the evidence implicating noninfectious environmental causation has not increased to the same extent despite much attention to the latter. The recent evidence for infectious causation therefore corresponds to the overall trend that has occurred over the past century and a half, namely a progressive, albeit somewhat erratic increase in the recognition of increasingly cryptic infectious causation.

EPIDEMIOLOGICAL FOOTPRINTS AND INTERVENTIONS

In some cases the emphasis on identification of an agent has probably hindered progress. Just as the acceptance of infectious causation of diseases such as yellow fever, cholera, and mononucleosis depended on evidence other than isolation of the agent, the infectiousness of currently questioned diseases can also be accepted on epidemiological evidence even before a pathogen is identified. Rejecting infectious causation on the basis of incomplete evidence may be more damaging than accepting infectious causation on the basis of incomplete evidence, especially if the crypticity of infectious causation may generate lags of many years or even decades between when infectious causation can be ascribed with reasonable certainty and when an infectious agent can be identified.

Multiple sclerosis (MS) offers an illustration. MS is characterized by an autoimmune reaction against the myelin sheaths surrounding neurons.

Epidemiological patterns implicate childhood infections that are particularly transmissible in northern latitudes. People who spent their childhood years in the tropics experience less MS, thus implicating winter respiratory infections [77]. Epidemic waves of MS also implicate infectious causation; for example, a surge of MS cases occurred in the Faeroe Islands from the early 1940s to the early 1970s [78]. Although various pathogens have been suggested, including a newly identified retrovirus, the matter remains unresolved [79–81]. Virtually all MS patients are infected with Epstein-Barr virus, raising the possibility that MS may result from coinfection of Epstein-Barr virus with one or more other pathogens [82].

Seasonal and geographic variations of mental illness similarly suggest infectious causation. Schizophrenia and severe depression disorders are about 10 percent more frequent among babies born during winter and spring, when respiratory infections tend to be more common [83, 84]. The geographic association of mental illness with degree of urban residency predates 20th-century society and coincides with high density living situations regardless of the specific details of the situation; the association between urban life and mental illness thus accords well with infectious causation [85, 86].

Long-continued rheumatoid arthritis causes distinctive changes to the joints that can be recognized in Amerindian skeletons from the Mississippi valley going back several thousand years, but not in Old World skeletons from before ad 1500. This epidemiological footprint implicates an infectious agent that was brought back to Europe from the New World by early explorers [87].

Cancers too may show changes over time and space that implicate infectious causation. Adult T-cell leukemia (by HTLV-1) and Kaposi's sarcoma (by human herpes virus 8) were identified as infectious largely as a result of spatial clusters. Similar spatial or temporal patterns occur in other cancers, which are not yet widely recognized as infectious. Esophageal cancer, for example, has increased eightfold over the past 20 years in Scandinavia [88].

Given the historical track record for the eventual discovery of infectious causation on the basis of such epidemiological footprints, hypotheses of transmission and control that are based on infectious causation warrant increased attention. When the probable routes of infection can be deduced from epidemiological data, programs for reduction of transmission can be introduced even before identification of the agents. On the basis of past work and perhaps a few key additional studies, transmission routes of the multiple sclerosis agent(s), for example, might be identified and transmission curbed. Just as we did not have to isolate the agents of yellow fever, measles, and cholera to enact control measures to reduce transmission, we do not necessarily have to identify the agents of more cryptic infectious diseases to take steps toward their control. Similarly, even if the agent is not known, studies of effects of anti-pathogen drugs can be

evaluated in controlled trials; if effects are positive, the drugs can be used prior to conclusive identification of the causative infectious agents. Such an approach might have provided decades of improved therapy to patients suffering from ulcers from the 1950s to the 1990s. Drugs with anti-pathogen effects also appear to be effective against tropical sprue and rheumatoid arthritis [89, 90].

The reliance on epidemiological footprints and anti-pathogen drugs as alternative indicators of infectious causation may become increasingly valuable as the linkage of particular infectious agents with particular diseases becomes increasingly difficult. The identification of the particular causative pathogen is obviously a worthwhile goal, but its value need not blind us to the value of identifying infectious causation and intervening on the basis of this knowledge even without identifying the particular pathogen. The history of medicine raises an irony in this regard. Identification of the causative pathogen has been deemed important partly because of its usefulness in developing control measures such as vaccines and antibiotics. If the causative pathogen could be isolated and studied, the potential for developing interventions against it are obviously increased. During the 20th century, however, the elements of this tradeoff changed. The difficulty of identifying the causative pathogens with a given investigative approach increased because we identified the infectious agents that were the easiest to identify. The difficulty of developing effective vaccines increased because we developed the effective vaccines that were easiest to develop. The utility of emphasizing the pathogen identification before development of effective intervention has therefore changed. We often may be more effective in controlling disease if we investigate and enact epidemiological and anti-pathogen interventions before we agree on the causative agent.

The Interplay of Genes, Environment, and Infection

SELF-DESTRUCTIVE DEFENSES

Genetic diseases with relatively high fitness loads are generally too damaging to be maintained by typical rates of mutation, and therefore may require infectious causation to explain their evolutionary persistence. Ironically, such self-destructive genetic diseases may be maintained evolutionarily precisely because they damage biological machinery. Alleles will be lost in proportion to the fitness costs that they impose, and will be generated only slowly by mutation; thus, only small fitness loads can be maintained solely by mutation. An allele that causes a severe genetic disease at a frequency that is too high to be maintained by mutation must generate fitness benefits to itself that compensate for the extra fitness costs it incurs. It is difficult to imagine how severe damage to biological machinery could

somehow improve some other aspect of the basic functioning of the machinery sufficiently to compensate for the damage. It is easy, however, to envision how genetic damage to biological machinery can guard against even greater damage from a parasite.

Imagine that a company produces jet engines with poor quality control and that a jet with one defective engine and one good engine can still fly safely, albeit with reduced power. If the only requirement were optimal power performance, the poor quality control would be a liability. But suppose that when fighter jets are under attack from heat-seeking missiles, jets having one good engine and one defective engine emit so much less heat that the heat-seeking mechanism cannot find them. Planes with two defective engines, however, almost always crash. As the intensity of attack increases relative to the need for reserve power, the frequency of poor engines in the fleet increases, even though they cause more crashes because more planes have two of them.

In the absence of such tracking pressures, one expects that natural selection will continue to improve the mechanical functioning of an organism; however, when a parasite uses a particular characteristic of the host to exploit it, damage to that characteristic can provide a net fitness benefit to the host. Much as poor quality control reduces rather than improves the function of jet engines, random mutations are more likely to reduce rather than improve the function of gene products. Because there are so many more ways of generating damage than improving mechanical function, mutations that generate such self-destructive defenses should be relatively common. When the damage is incurred primarily among homozygous recessives, it increases disproportionately as the frequency of the damaging allele increases (as q^2 increases as a function of q). When such a self-destructive defense is present, it therefore should not rise to anywhere near fixation, and hence should not select intensely for pathogen variants that circumvent the defense. So long as these pathogen variants are not as successful as the variants that attack the hosts who do not harbor the defensive allele, any selection for such circumventing pathogens in heterozygote hosts will be diluted by the selection for success in individuals homozygous for the normal allele.

The major genetic diseases as a rule occur at frequencies that are too high to be maintained by mutation without some compensating benefit. Several are defenses against malaria (Table 2, rows 1–3). The cystic fibrosis allele appears to be maintained in relatively high frequency because it provides protection against *Salmonella typhi*, the cause of typhoid fever. *S. typhi* uses the normal protein for entry into cells but is obstructed by the defective protein [91].

Which other genetic diseases are caused in an evolutionary sense by infection? We suspect hemochromatosis. Like cystic fibrosis, hemochromatosis is a genetic disease of northern Europeans. It is associated with

iron buildup throughout the body. Homozygotes may have liver disease, diabetes, infertility, heart failure, and increased susceptibility to pathogens. About one in 10 northern Europeans carry a single copy of the hemochromatosis allele. The normal protein is related to the MHC proteins and is likewise expressed on the outer surface of the cell membrane; it appears to down-regulate iron accumulation by associating with the transferrin receptor [92]. The hemochromatosis protein, however, accumulates in the cell rather than on the membrane. This is just the kind of biochemical change that could inhibit entrance of a pathogen that uses the normal protein for entrance into the cell; moreover, the hemochromatosis allele is too common to be maintained by mutation and seems too widespread and persistent in populations to be the result of a founder effect. Some researchers have argued that the hemochromatosis allele could be useful in heterozygote form through an increase in iron absorption, but the geographic distribution of the gene does not fit this explanation. The allele occurs in areas where the diet has contained iron-rich foods such as meats, but not in areas with iron-poor diets. Also, natural selection should not have to destroy a geneís function in order to absorb more iron. We therefore predict that hemochromatosis will, like cystic fibrosis, eventually be recognized as a defense against some damaging pathogen that has been common in northern Europe.

We also predict that alpha-1-antitrypsin deficiency will eventually be recognized as a self-destructive defense. Alpha-1-antitrypsin deficiency is common in northern Europeans, causing emphysema and liver disease in homozygotes. The normal protein controls the level of an enzyme that increases the potential for leukocyte movement by increasing permeability of elastic cartilage. The net result of the antitrypsin deficiency is higher levels of this enzyme. A consequence may be increased protection against pathogens, but at a cost: the increased tissue damage that eventually leads to emphysema and cirrhosis of the liver. If this argument is correct, we should see some positive effects of alpha-1-antitrypsin deficiency on infection. Although this prediction awaits definitive tests, a recent study reported that alpha-1-antitrypsin deficiency was associated with reduced rather than increased respiratory dysfunction in cystic fibrosis patients, "contrary to current thinking" [93]. The overactivity of alpha-1-antitrypsin could well improve respiratory function by helping to control the infections that cause problems in cystic fibrosis patients. This example illustrates a practical reason why we should try to understand whether genetic diseases are self-destructive defenses against pathogens: we may decide not to treat them. If we could treat a cystic fibrosis patient for concurrent alpha-1-antitrypsin deficiency, we might be exacerbating the situation. Over the longer term, understanding the defensive functions of such alleles may improve decisions about "corrective" gene therapy. If alpha-1-antitrypsin deficiency heterozygotes are protected against a still-threatening

disease such as *Mycobacterium tuberculosis*, then correcting these genes could have a negative overall effect on the person's health. The problem may be especially complex for alpha-1-antitrypsin deficiency, because it may generate a general escalation of immunological attack and may therefore protect against a variety of pathogens.

CHROMOSOMAL ABERRATIONS

Genetic diseases with fitness loads that are too high to be maintained solely by mutation (i.e., above 0.001) tend to fall into three categories: (1) they are known to be self-destructive defenses against infectious disease (e.g., sickle-cell anemia); (2) they have characteristics that implicate such a defensive function (e.g., hemochromatosis); or (3) they are caused by chromosomal aberrations (see Table 2). Unlike genetic diseases that are too severe to depend solely on underlying mutation rates for their persistence, genetic diseases caused by chromosomal aberrations can be maintained in the population solely by the frequency of new aberrations. An evolutionary perspective, however, raises a question: why is this frequency of new aberrations so high? The frequency of chromosomal aberrations and the frequency of miscarriages, which are largely attributable to chromosomal aberrations, is far higher in humans than in other mammals [94, 95]. We see no reason why human chromosomes should be inherently more fragile during the course of meiosis and mitosis, because vulnerability to chromosomal aberrations should have been selected against throughout our history, right up to the present. Infections, however, are known to contribute to chromosomal aberrations; hepatitis B, polyoma viruses, and papillomaviruses, for example, are associated with chromosomal damage [96–102], and human papillomaviruses occurred in the majority of material from first-trimester spontaneous abortions, being three times more common than in material from elective abortions [103]. These considerations suggest that spontaneous abortions and associated chromosomal damage, so prevalent in humans, may be caused by infection.

Two characteristics of human populations that would make us more likely to have infection-induced chromosomal damage are (1) the high frequency and diversity of infections that can be maintained in the large and ever-mixing human population, and (2) continual receptivity for sexual intercourse. We consider sexually transmitted pathogens to be the most likely culprits, because they are uniquely suited to prosper from continuous sexual receptivity and because the damage would have to occur early during the reproduction, probably by affecting the meiotic process of the egg cells prior to the first reduction division. Sexually transmitted pathogens probably have more ready access to the eggs in this stage. Also, sexually transmitted pathogens are selected to be long-lived within individual hosts (because they must remain infectious over time periods greater than

the intervals between sexual partner changes); they therefore must avoid the immune system. Intracellular infection may contribute to this long-term persistence, and prolonged, intracellular persistence seems especially likely to cause chromosomal damage. A useful beginning point for such study would be to look for pathogens in oocytes of women who have had a history of miscarriage. A virus that causes miscarriages and fetal malformations has been found in oocytes of cattle [104].

GENETIC DISEASE AND GENETIC VULNERABILITY

If one looks beyond genetic diseases that represent evolutionary responses to infection (such as sickle-cell anemia) and genetic aberrations, one is left with very little evidence for frequent, severe damage from strictly genetic diseases. Muscular dystrophy and hemophilia, for example, are damaging to the individuals involved, but do not generate a large average fitness reduction within the population (Table 2). The label "genetic disease" requires that concordances are essentially 100 percent for monozygotic twins and drop from this value among less closely related family members, according to Mendelian proportions. Genetic diseases that involve resistance to infection, such as sickle-cell anemia, occur in such a pattern, as do rare disorders, which can be maintained by mutational pressure. But such ratios are absent among the widespread and prevalent diseases that are not known to be defenses against infection. Researchers have attempted to explain this absence by invoking unspecified environmental factors or collections of deleterious alleles at multiple loci; but despite decades of effort, researchers have not identified alleles hypothesized by these complex models for any damaging and pervasive disease. These arguments may be applicable to diseases that impose very low fitness costs per person in the population [105], but such diseases, though they may be important to the afflicted individuals, are relatively minor concerns of the health sciences as a whole. In this context we consider pleiotropic effects and hitchhiking genes to be of short-term rather than long-term importance. If, for example, separation of linked genes would improve an organism's fitness by eliminating a disease, then we expect that over time this linkage disequilibrium would diminish through recombination.

Genetic associations with disease may result from alleles that increase protection from one pathogen but increase vulnerability to another [106]. The genetic variation responsible for these associations may persist indefinitely, because the success of a variant is negatively related to its frequency. As a pathogen variant increases, it favors increased frequency of a host variant that controls it; as this host variant becomes more prevalent, it increasingly disfavors the pathogen variant. "Dynamical polymorphisms" of resistance genes may thus persist by this process of nonprogressive seething [107, 108]. If underlying infectious causes are not recognized, the reduced

success of some host variants (due to the inherent costs of the particular genetic defense or to the increased vulnerability to the existing mix of pathogens) could be misconstrued as a genetic disease.

Variations in human leukocyte antigens (HLA) appear to be a particularly important example of this process, because different HLA alleles are associated with differences in recognition and presentation of foreign antigens, or with differences in vulnerability to autoimmune disease as a consequence of cross reactivity between pathogen and host antigens [109]. HLA variants are associated with vulnerability to infectious diseases such as malaria and tuberculosis, and infection has been causally linked to HLA-associated autoimmune diseases such as herpes stromal keratitis and Reiter's syndrome [51, 110–112]. The high degree of polymorphism of such alleles implicates a role for infection in many diseases that exhibit moderate heritibilities [113, 114].

HLA variants are associated with many chronic diseases, such as juvenile diabetes, rheumatoid arthritis, systemic lupus erythematosus, scleroderma, Crohn's disease, myasthenia gravis, Hodgkin's disease, multiple sclerosis, Grave's disease, narcolepsy, and pernicious anemia [51]. Individuals with HLA B-27, for example, comprise about 5 percent of the population and are about 80 times more likely to have ankylosing spondylitis, which is often manifested by severe arthritis-like conditions [51]. The monozygotic twin concordance for most of these HLA-associated diseases is less than 50 percent—something other than host genetics is therefore making a major contribution [115]. Infectious causation has not been investigated for some of these HLA associations; for others there is suggestive evidence. Juvenile diabetes, for example, is due to the gradual disappearance of insulin production. Its temporal association with infection, the similarity of viral antigens with autoimmune-associated HLA antigens, and regional and seasonal patterns, implicate infectious causation. Coxsackie virus is a leading suspect [115].

The reasons for the presence and maintenance of trinucleotide repeat diseases such as fragile X and myotonic dystrophy are largely unknown. These diseases occur after a small, "subclinical" length of trinucleotide repeats lengthens rapidly over several generations as a result of errors associated with a hairpin transcriptional structure. The lineages of these people then tends to dwindle to extinction. Individuals with subclinical repeat lengths are virtually absent in some regions and too widespread in other regions to be attributed solely to founder effects. The low frequency of new haplotypes among such individuals suggests that the low repeat numbers are generated too rarely by new mutations to be maintained solely by mutation. This set of characteristics is consistent with the subclinical repeats being a defense against infectious agents and are difficult to explain without invoking such a benefit; we therefore suggest that the possibility of a defensive function of subclinical repeat lengths warrants investigation.

Psychological Phenomena

The fitness load approach also applies to psychological phenomena. Psychological and behavioral alterations among infected individuals may represent manipulations of hosts for the spread of pathogens, defenses against the pathogen, or side effects of infection that benefit neither host nor pathogen [116]. Rabies is a dramatic example of a manipulation: the pathological rage induced by the virus increases the chance of biting and hence transmission of the virus in the saliva. Syphilitic insanity appears to be an example of a side effect. These examples could be dismissed because they are so extreme, but it is the extreme examples that are the most conspicuous and that therefore will tend to be discovered without a concerted research effort. Just as it was inappropriate to reject infectious causation of human cancers on the basis of the conspicuous progression of Rous's sarcoma in chickens, it would be inappropriate to reject the idea that pathogens could cause subtle changes in human behavior on the basis of the conspicuous psychological effects of rabies and syphilis. Had *Treponema pallidum* been a less conspicuous organism, causing less conspicuous mental illness, and less conspicuous links between early symptomatic infection and mental illness we might not recognize even today that this spirochaete can cause mental illness.

The intricacies of neuronal circuitry suggest that behavioral changes may result from slight changes in brain structure; and such slight changes may be more difficult to link to infectious organisms than the gross pathological changes found in other less intricate tissues, such as the changes in the lungs caused by *M. tuberculosis*. Moreover, generating appropriate animal models for testing of infectious causation often may be very difficult or even impossible because of the intellectual complexity of human behavior, the subtleties of variations in human behavior, and the difficulties of understanding the thought processes of nonhuman subjects. Considering these problems, we cannot use the absence of conspicuous evidence of infection-induced behavioral alterations as evidence of the absence of such alterations.

In spite of these limitations, a combination of comparative study and animal experimentation are already proving useful for investigations of some psychiatric disorders. Borna disease virus, for example, has a tropism for the limbic system and is known to cause mood disorders in sheep, horses, and rats [117, 118]. It has been isolated from human brain tissue [119], and infections in humans are significantly associated with schizophrenia and clinical depression (i.e., bipolar disorder) [117, 120, 121]. Evidence also implicates infectious causation of some obsessive compulsive disorders by streptococcal infections [122, 123].

Most of the examples of infectious causation discussed thus far are considered to be infectious by at least some experts who are studying the phe-

nomena from a largely nonevolutionary perspective; and there has been over the past few years a growing recognition that infection may cause many of the diseases previously ascribed to noninfectious causes [26, 27, 29, 48, 49]. The evolutionary logic presented here provides support for this growing attention to hypotheses of infectious causation. But to get a sense of the broad scope of insights into infectious causation that might be provided by an evolutionary perspective, it is appropriate to consider phenomena that are not now considered by experts to have infectious causes, but are nevertheless identified by the evolutionary framework as candidates for infectious causation. As is the case for any new scientific perspective, the value of the evolutionary framework for the recognition of infectious causation depends on the extent to which it can help identify infectious causation of phenomena that we intuitively feel are not caused by infection. We shift to this category of phenomena, realizing that the line of logic may lead to hypotheses that we may find disturbing, distasteful, or socially disruptive, but we do so with the belief that it is better to know how nature works than to live in ignorance. We consider human homosexuality.

Phenomena that strongly reduce the evolutionary fitness of their bearers cannot be maintained by strictly genetic causation at frequencies far above the rates at which they could be generated by mutation. The fitness costs of male homosexuality place it in this category [124]. Perhaps more importantly, each of the hypotheses that have been put forward to explain male homosexuality have critical flaws that, if not sufficient to cause their outright rejection, are sufficient to severely weaken them. A full consideration of these issues is beyond the scope of this essay, but we briefly summarize the most salient points below.

Human homosexuality can be traced back at least several thousand years. A substantially genetic cause could not be maintained over this time because of the great fitness costs that homosexuality imposes, unless there is some compensating benefit. But no evidence for a compensating benefit exists. Inclusive fitness benefits, for example, seem insufficient to overcome the reduction in reproduction, because male homosexuals do not channel their resources into the well-being of kin at an increased level [125]. One could try to rescue the genetic causation hypothesis by arguing that reproductive effects of homosexuality have changed substantially in recent generations, with homosexual men having had just as many offspring as heterosexual men in previous centuries. This possibility seems unlikely from what is known about modern homosexual behavior, but it could be investigated empirically from historical records. A primarily genetic basis, however, is negated by the low monozygotic twin concordance, which is about 20 percent [126]. Accordingly, the allele that purportedly conferred homosexual orientation has not stood up to independent testing [127].

Another hypothesis is that male homosexuality results from novel sociocultural influences, but the attraction to reproductive partners of a repro-

ductively feasible sex seems so critical to evolutionary fitness that one would expect these attractions to be strongly buffered against social effects that could generate a preference for exclusive sexual relations with members of the same sex. The occurrence of exclusive male/male sexual preferences in sheep shows that cultural "powers of suggestion" are not necessary to generate the phenomenon [127–130]. (Sheep do not watch television, read newspapers, or discuss alternatives lifestyles.) A brief consideration of ungulate and human infections should be sufficient to counter the natural inclination to dismiss the phenomena in sheep as irrelevant to humans; many important human infectious agents, such as HIV, herpes viruses, the measles virus, *Mycobacterium tuberculosis*, borna disease virus, and prions, have related pathogens in sheep or other ungulates that have similar effects.

In contrast with difficulties of noninfectious explanations of homosexuality, the hypothesis of infectious causation does not incorporate critical logical flaws or contradictions of fundamental biological principles. Indeed, anecdotal reports indicate that changes in human sexual orientation have occurred following changes in the limbic area due to trauma or infection [131, 132]. One possible route would be sexual, whereby homosexual behavior could facilitate spread because of the larger numbers of partners homosexual males may have on average, relative to heterosexual males. Alternatively, transmission could be partly or entirely by one or more nonsexual routes, and homosexual orientation be a side effect of the infection that is unrelated to transmission.

Although this hypothesis of infectious causation may generate a negative knee-jerk response, such responses are not reliable indicators of the validity of scientific hypotheses. The critical weaknesses of the alternative hypotheses draw attention to the need for rigorous testing of any hypothesis that has a sound theoretical basis, even if we find the hypothesis disturbing and disorienting. The presence of the phenomenon in sheep allows for experimental tests.

Focusing on Infectious Causation

Fulfillment of Koch's postulates confirms infectious causation, but doing so has become less feasible over the past century [133]; it is likely to become even less feasible in the future because many of the characteristics that make infectious causation cryptic also hinder fulfillment of the postulates. The long lags between the onset of infection and the onset of symptoms, for example, often make experimental transfer and documentation of disease impracticable.

With the clarity of hindsight we can assess whether the fulfilling of Koch's postulates has been necessary for recognition of infectious causation, or whether it has been an important but largely supplementary pro-

cess. Infectious causation, after being accepted on the basis of epidemiological or therapeutic evidence, could be questioned on the basis of a failure to fulfill Koch's postulates (as has been done by Duesberg in the case of HIV as a cause of AIDS [134]) and then disproved. If this sequence of events were common, we would have evidence for the importance of fulfilling Koch's postulates before acceptance of infectious causation. To evaluate this idea we can look at medical history spanning the period from the mid-19th to the mid-20th century. Over this period there have been several hundred disease entities for which infectious causation was considered very probable prior to the fulfilling of Koch's postulates. Infectious causation has been confirmed for the vast majority of these disease entities and is still a feasible hypothesis for almost all of the others.

In only a few cases has infectious causation been accepted by a substantial proportion of medical authorities and then shown to be incorrect. Infectious causation of milk sickness, for example, was accepted in some medical circles, particularly among those who had little if any firsthand experience. In contrast, those with firsthand experience often adamantly expressed their well-founded and correct opinion that the disease was caused by a milk-borne toxin that cows ingested when they fed on white snakeroot. The evidence for infectious causation was very weak, because it did not distinguish between spatial variation in exposure to infectious versus noninfectious agents. The history of milk sickness therefore illustrates the standards of evidence that are too low to represent trustworthy indicators of infectious causation.

The reliance on Koch's postulates has already diminished out of necessity. The postulates have not been fulfilled for most of the diseases for which infectious causation has been accepted over the past two decades, such as cervical cancer, liver cancers, adult T-cell leukemia, and AIDS. Koch himself was relatively cautious about the utility of his postulates; he proposed them as useful guidelines but did not suggest that they should be the only guidelines. The historical record, including Koch's own views, therefore supports a reduced emphasis on Koch's postulates for future efforts to recognize infectious causation. When evidence other than confirmation of Koch's postulates has led to the general acceptance of infectious causation, this acceptance has been trustworthy.

Although limitations of Koch's postulates have been addressed in medical science, the emphasis has been on ways in which causative organisms can be identified even without Koch's postulates (e.g., [133, 135]). Our focus is different. Considering the increasing difficulty of identifying etiologic agents as the causes of chronic diseases, we emphasize the need to focus on discovery of infectious causation as a goal that may sometimes be distinct from the identification of etiologic agents. In addition to allowing the development of effective interventions prior to the discovery of etiologic agents, this shift may actually facilitate the identification of causative

organisms, because recognition of infectious causation and transmission routes can guide researchers to cryptic causative organisms. Kuru, for example, was controlled without identifying an infectious agent by using the evidence for infectious transmission to break the transmission cycle. The acceptance of infectious causation of kuru and similar diseases helped to facilitate the discovery of prions, a novel class of infectious agents.

Interventions with antimicrobial drugs offer a similar application of this strategy. If particular drugs are effective, particular categories of etiologic agents will be incriminated. This approach serves as one of the most direct experimental tests of cause-and-effect association, though it has the drawback associated with effects of antimicrobials other than the negative effects on pathogens (e.g., anti-inflammatory effects) [135]. A major benefit of this approach is the generation of effective treatment even before the etiologic agent is identified. Had this approach been more aggressively and systematically pursued, ulcers, for example, could have been cured and a bacterial cause implicated decades earlier. "Non-gonorrheal urethritis" was controlled by this approach before its primary etiologic agent, *Chlamydia trachomatis*, was identified. This control helped guide researchers to a bacterial agent, but ectopic pregnancies and infertility due to tubal scarring could have been linked to a bacterial cause and controlled earlier as well.

The same missed opportunity will apply to atherosclerosis, if *C. pneumoniae* proves to be its primary cause and if it is successfully treatable with antibiotics. A logical starting point for such research is to make greater use of records of antibiotic usage to assess whether natural variation in antibiotic usage correlates with variation in the occurrence of diseases suspected of being caused by infection. This approach could have been invoked decades ago but has only been applied to heart disease over the past few years. It has provided additional evidence of infectious causation of first-time myocardial infarction in one study but not in another [136, 137]. Such discrepancies are not surprising for this kind of approach because many variables are uncontrolled: associations could be generated by spurious correlations, and lack of association can occur because real effects of antibiotics could be overwhelmed by other uncontrolled effects. Discrepancies therefore are best interpreted as evidence justifying further study, rather than evidence for the rejection of the hypothesis.

As therapeutic agents against viruses become available, the approach can be applied to implicate viral causation. Ribavirin, for example, has inhibitory effects on borna disease virus [138]. Tests of the efficacy of ribivirin on patients with schizophrenia or bipolar disorder could simultaneously provide evidence of infectious causation, the probable etiological agent, and effective therapy. A similar argument could be made for immunological therapy, which also appears to be effective against borna disease virus [139, 140]. Lithium treatment has antiviral effects and normalizes markers of increased immunological activity in bipolar patients

[141, 142]. The hypothesis that lithium treatment helps bipolar patients by inhibiting viruses (such as the borna disease virus) deserves testing.

Our arguments do not negate the value of identifying infectious agents. Rather they emphasize the danger of allowing a focus on the agent to direct attention away from recognizing infectious causation. Technological innovations will undoubtedly strengthen both processes. They can make cryptic infectious causes more conspicuous, as representational difference analysis has done for Kaposi's sarcoma; and they can reveal chains of infection and statistical associations between infection and disease even when infections rarely generate overt disease [133, 143, 144]. Nevertheless, we can expect that the overall process that has been occurring over the past two centuries will continue: each new technique and its associated guidelines for identifying infectious causation will generate an increase in the identification of etiologic agents that can be identified by the approach, but will leave in its wake cases of infectious causation that are cryptic to information generated by the approach. To maintain progress in identification of infectious causation and infectious agents, we can expect an ongoing need to change approaches.

The insights into infectious causation generated by consideration of fitness load direct attention to especially damaging diseases that are probably caused by infection; it thus directs attention toward those diseases for which investigations of transmission and control through alternative approaches are especially warranted. The identification of etiologic agents and the demonstration of infectious causation by Koch's postulates still serves as a gold standard of evidence for infectious causation. We need not, however, let the inability to meet the gold standard for understanding infectious causation lead us to be overly skeptical of assigning infectious causation based on silver or bronze standards, nor to use this assignment to enact effective interventions before etiologic agents have been identified. The history of infectious diseases indicates that the silver and bronze standards have been extremely valuable. We can expect them to be even more valuable in the future as we are left with those diseases for which attainment of the gold standard for identification of etiological agents will be increasingly difficult.

At a practical level we need to understand the causes of disease to understand how best to decrease the suffering from disease. At a more basic level, as Nesse and Williams have emphasized, we need to understand the causes of disease, because the mix of superb engineering and seemingly underbuilt components in the human body is a mystery [52]. Is a circulatory system prone to atherosclerosis, for example, really like a Mercedes Benz with the soda-straw fuel line? Or is a state-of-the art fuel line simply prone to microbial sabotage? The resolution of this question and the broader set to which it belongs is academic, but it also promises to reorient medical research and may improve fundamentally our states of health.

REFERENCES

1. Quainn, R. *A Dictionary of Medicine*. New York: Appleton, 1884.
2. Roueché, B. *Curiosities in Medicine*. New York: Berkley, 1958.
3. Cone, T. E., Jr. Milk sickness or tremetol poisoning. In *Cambridge World History of Human Disease*, edited by K. Kiple. Cambridge: Cambridge Univ. Press, 1993. 880–83.
4. Sacks, O. W. *The Island of the Colorblind and Cycad Island*. New York: Knopf, 1996.
5. McKhann, C. F. Measles. In *Textbook of Medicine*, 6th ed., edited by R. L. K. F. Cecil. Philadelphia: W. B. Saunders, 1943. 21–29.
6. Cooper, A. Syphilis. In *A Dictionary of Medicine*, edited by R. Quainn. New York: Appleton, 1884. 1575–85.
7. Lossick, J. G. Epidemiology of urogenital trichomoniasis. In *Trichomonads Parasitic in Humans*, edited by B. M. Honigberg. New York: Springer Verlag, 1990. 311–23.
8. MacNamara, C. Asiatic cholera. In *A Dictionary of Medicine*, edited by R. Quainn. New York: Appleton, 1884. 239–43.
9. MacNamara, C. Choleraic diarrhea. In *A Dictionary of Medicine*, edited by R. Quainn. New York: Appleton, 1884. 243–45.
10. Broadbent, W. H. Typhoid fever. In *A Dictionary of Medicine*, edited by R. Quainn. New York: Appleton, 1884. 1679–90.
11. Ewart, J. Dysentery. In *A Dictionary of Medicine*, edited by R. Quainn. New York: Appleton, 1884. 408–16.
12. Ewald, P. W. Waterborne transmission and the evolution of virulence among gastrointestinal bacteria. *Epidemiol. Infect.* 106:83–119, 1991.
13. McNeill, W. H. *Plagues and Peoples*. Garden City, NJ: Anchor, 1976.
14. Ackerknecht, E. H. *A Short History of Medicine*. New York: Ronald Press, 1955.
15. Rosenberg, C. E. *The Cholera Years*. Chicago: Univ. of Chicago Press, 1962.
16. Fayrer, J. Dengue. In *A Dictionary of Medicine*, edited by R. Quainn. New York: Appleton, 1884. 339–41.
17. Von Bókay, J. Uber den aetiologischen zusammenhang der varizellen mit gewissen fallen von herpes zoster. *Wien. Klin. Wschr.* 22:1323, 1909.
18. Seiler, H. E. A study of herpes zoster particularly in its relationship to chickenpox. *J. Hygiene* 47:253–62, 1949.
19. Stollerman, G. H. *Rheumatic Fever and Streptococcal Infection*. New York: Grune & Stratton, 1975.
20. Rous, P. A. Sarcoma of the fowl transmissible by an agent separable from the tumor cells. *J. Exp. Med.* 13:397–411, 1911.
21. Poiesz, B., F. W. Ruscetti, A. F. Gazdar, et al. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc. Natl. Acad. Sci. USA* 77:7415–19, 1980.
22. Zur Hausen, H. Viruses in human cancers. *Science* 254:1167–73, 1991.
23. Pisani, P., D. M. Parkin, N. Muñoz, and J. Ferlay. Cancer and infection: Estimates of the attributable fraction in 1990. *Cancer Epidemiol. Biomarkers Prev.* 6:387–400, 1997.
24. White, D., and F. Fenner. *Medical Virology*, 4th ed. San Diego: Academic, 1994.
25. Parsonnet, J., S. Hansen, L. Rodriguez, et al. *Helicobacter pylori* infection and gastric lymphoma. *N. Engl. J. Med.* 330:126–771, 1994.
26. Cassell, G. H. Infectious causes of chronic inflammatory diseases and cancer. *Emerg. Infect. Dis.* 4:475–87, 1998.
27. Lorber, B. Are all diseases infectious? *Ann. Int. Med.* 125:844–51. 1996.
28. Neipel, F., J. C. Albrecht, and B. Fleckenstein. Human herpesvirus 8: The first human Rhadinovirus. *J. Natl. Cancer Inst. Mono.* 23:73–77, 1998.

29. Parsonnet, J. Introduction. In *Microbes and Malignancy: Infections as a Cause of Human Cancer*, edited by J. Parsonnet. New York: Oxford Univ. Press, 1999.

30. Labat, M. L. Possible retroviral etiology of human breast cancer. *Biomed. Pharmacother.* 52:6–12, 1998.

31. Pogo, B. G. T., and J. F. Holland. Possibilities of a viral etiology for human breast cancer: A review. *Biol. Trace Elem. Res.* 56:131–42, 1997.

32. Rakowicz-Szulczynska, E. M., B. Jackson, and W. Snyder. Prostate, breast and gynecological cancer markers RAK with homology to HIV-1. *Cancer Lett.* 124:213–23, 1998.

33. Bonnet, M., J. M. Guinebretiere, E. Kremmer, et al. Detection of Epstein-Barr virus in invasive breast cancers. *J. Natl. Cancer Inst.* 91:1376–81, 1999.

34. Grossman, M. I. Peptic ulcer: Pathogenesis and pathophysiology. In *Cecil Textbook of Medicine*, 15th ed., edited by P. B. Beeson, W. Mcdermott, and J. B. Wyngaarden. Philadelphia: Saunders, 1979. 1502–7.

35. Doenges, J. L. Spirochetes in gastric glands of *Macacus rhesus* and humans without definite history of related disease. *Proc. Soc. Exp. Biol. Med.* 38:536–38, 1938.

36. Kidd, M., and I. M. Modlin. A century of Helicobacter pylori: Paradigms lost, Paradigms regained. *Digestion* 59:1–15, 1998.

37. Fremont-Smith, P. Letter to the editor. *Atlantic Monthly* 283(5):12, 1999.

38. Nieto, F. J. Infections and atherosclerosis: New clues from an old hypothesis? *Am. J. Epidemiol.* 148:937–48, 1998.

39. Saikku, P., M. Leinonen, K. Mattila, et al. Serological evidence of an association of a novel *Chlamydia*, TWAR, with chronic coronary heart disease and acute myocardial infarction. *Lancet* 2:983–86, 1988.

40. Gupta, S., and A. J. Camm. Chlamydia pneumoniae and coronary heart disease. *BMJ* 314:1778–79, 1997.

41. Saikku, P. Chlamydia pneumoniae and atherosclerosis: An update. *Scand. J. Infect. Dis.* 104(suppl.):53–56, 1997.

42. Campbell, L. A., C. C. Kuo, and J. T. Grayston. *Chlamydia pneumoniae* and cardiovascular disease. *Emerg. Infect. Dis.* 4: 1998.

43. Moazed, T. C., L. A. Campbell, M. E. Rosenfeld, et al. *Chlamydia pneumoniae* infection accelerates the progression of atherosclerosis in apolipoprotein E-deficient mice. *J. Infect. Dis.* 180:238–41, 1999.

44. Balin, B. J., H. C. Gérard, E. J. Arking, et al. Identification and localization of *Chlamydia pneumoniae* in the Alzheimer's brain. *Med. Microbiol. Immunol. (Berlin)* 187:23–42, 1998.

45. Gérard, H. C., G. F. Wang, B. J. Balin, et al. Frequency of apolipoprotein E (APOE) allele types in patients with *Chlamydia*-associated arthritis and other arthritides. *Microb. Pathog.* 26:35–43, 1998.

46. Arisawa, K., M. Soda, M. Akahoshi, et al. Human T-lymphotropic virus type-I infection, antibody titers and cause-specific mortality among atomic-bomb survivors. *Jpn. J. Cancer Res.* 89:797–805, 1998.

47. Levander, O. A., and M. A. Beck. Interacting nutritional and infectious etiologies of Keshan disease: Insights from coxsackie virus Binduced myocarditis in mice deficient in selenium or vitamin E. *Biol. Trace Elem. Res.* 56:5–21, 1997.

48. Friedman, H., and M. Bendinelli. Introduction. In *DNA Tumor Viruses: Oncogenic Mechanisms*, edited by G. Barbanti-Brodano, H. Friedman, and M. Bendinelli. New York: Plenum, 1995.

49. Fredricks, D. N., and D. Relman. Infectious agents and the etiology of chronic idiopathic diseases. *Curr. Clin. Top. Infect. Dis.* 18:180–200, 1998.

50. Rath, H. C., H. H. Herfarth, J. S. Ikeda, et al. Normal luminal bacteria, especially Bacteroides species, mediate chronic colitis, gastritis, and arthritis in

HLAB27/human beta2 microglobulin transgenic rats. *J. Clin. Invest.* 98:945–53, 1996.

51. Vogel, F., and A. G. Motulsky. *Human Genetics*, 3rd ed. Berlin: Springer Verlag, 1997.
52. Nesse, R. M., and G. C. Williams. *Why We Get Sick: The New Science of Darwinian Medicine.* New York: Times Books, 1994.
53. Williams, G. C. *Natural Selection: Domains, Levels, and Challenges.* New York: Oxford Univ. Press, 1992.
54. Reznick, D., F. Shaw, F. Rodd, and R. Shaw. Evaluation of the rate of evolution in natural populations of guppies *Poecilia reticulata*. Science 275:1934–37, 1997.
55. Ames, B. N., and L. S. Gold. The causes and prevention of cancer: gaining perspective. *Environ. Health Persp.* 105:865–73, 1997.
56. Lunn, R. M., Y. J. Zhang, L. Y. Wang, et al. p53 mutations, chronic hepatitis B virus infection, and aflatoxin exposure in hepatocellular carcinoma in Taiwan. *Cancer Res.* 57:3471–77, 1997.
57. Strickberger, M. W. *Genetics*, 2nd ed. New York: Macmillan, 1976.
58. Ewald, P. W. *Evolution of Infectious Disease.* New York: Oxford Univ. Press, 1994.
59. Bell, G. Pathogen evolution within host individuals as a primary cause of senescence. *Genetica* 91:21–34, 1993.
60. Salo, W. L., A. C. Aufderheide, J. Buikstra, and T. A. Holcomb. Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *Proc. Natl. Acad. Sci. USA* 91:2091–94, 1994.
61. Arriaza, B. T., W. Salo, A. C. Aufderheide, and T. A. Holcomb. Pre-Columbian tuberculosis in northern Chile: Molecular and skeletal evidence. *Am. J. Phys. Anthropol.* 98:37–45, 1995.
62. Guhl, F., C. Jaramillo, R. Yockteng, et al. *Trypanosoma cruzi* DNA in human mummies. *Lancet* 349:1370, 1997.
63. Miller, R. L., S. Ikram, G. J. Armelagos, et al. Diagnosis of *Plasmodium falciparum* infections in mummies using the rapid manual *Para*Sight™-F test. *Trans. R. Soc. Trop. Med. Hyg.* 88:31–32, 1994.
64. Hamilton, W. D. The genetical evolution of social behavior. *J. Theor. Biol.* 7:1–52, 1964.
65. Hamilton, W. D. On the moulding of senescence by natural selection. *J. Theor. Biol.* 12:12–45, 1966.
66. Muller, H. J. Our load of mutations. *Am. J. Hum. Genet.* 2:111–76, 1950.
67. Crow, J., and M. Kimura. *An Introduction to Population Genetics Theory.* New York: Harper & Row, 1970.
68. Cavalli-Sforza, L. L., and W. F. Bodmer. *The Genetics of Human Populations.* San Francisco: Freeman, 1971.
69. Dolcetti, R., and M. Boiocchi. Epstein-Barr virus in the pathogenesis of Hodgkin's disease. *Biomed. Pharmacother.* 52:13–25, 1998.
70. Stiller, C. A. What causes Hodgkin's disease in children? *Eur. J. Cancer* 34:523–28, 1998.
71. Blanchard, T. G., and S. J. Czinn. Immunological determinants that may affect the *Helicobacter pylori* cancer risk. *Alim. Pharm. Ther.* 12:83–90, 1998.
72. Kurtti, P., R. Isoaho, L. Vonhertzen, et al. Influence of age, gender and smoking on Streptococcus pneumoniae, Haemophilus influenzae and Moraxella (Branhamella) catarrhalis antibody titres in an elderly population. *Scand. J. Infect. Dis.* 29:485–89, 1997.
73. Laurila, A. L., T. Anttila, E. Laara, et al. Serological evidence of an association between Chlamydia pneumoniae infection and lung cancer. *Int. J. Cancer* 74:31–34, 1997.

74. Von Hertzen, L., H. M. Surcel, J. Kaprio, et al. Immune responses to *Chlamydia pneumoniae* in twins in relation to gender and smoking. *J. Med. Microbiol.* 47:441–46, 1998.

75. Chen, F. F., J. J. Yan, W. W. Lai, et al. Epstein-Barr virus-associated nonsmall cell lung carcinoma: Undifferentiated "lymphoepithelioma-like" carcinoma as a distinct entity with better prognosis. *Cancer* 82:2334–42, 1998.

76. Carbone, M., P. Rizzo, P. M. Grimley, et al. Simian virus-40 large-T antigen binds p53 in human mesotheliomas. *Nature Med.* 3:908–12, 1997.

77. Gale, C. R., and C. N. Martyn. Migrant studies in multiple sclerosis. *Prog. Neurobiol.* 47:425–48, 1995.

78. MacDonald W. I. Multiple sclerosis. In *Cambridge World History of Human Disease*, edited by K. Kiple. Cambridge: Cambridge Univ. Press, 1993. 883–87.

79. Perron, H., J. A. Garson, F. Bedin, et al. Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis: The Collaborative Research Group on Multiple Sclerosis. *Proc. Natl. Acad. Sci. USA* 94:7583–88, 1997.

80. Bachmann, S., and J. Kesselring. Multiple sclerosis and infectious childhood diseases. *Neuroepidemiol.* 17:154–60, 1998.

81. Jacobson, S. Association of human herpesvirus-6 and multiple scerosis: here we go again? *J. Neurovirol.* 4:471–73, 1998.

82. Myhr, K. M., T. Riise, and E. Barrett Connor, et al. Altered antibody pattern to Epstein-Barr virus but not to other herpesviruses in multiple sclerosis: A population based case-control study from western Norway. *J. Neurol. Neurosurg. Psychiatry* 64:539–42, 1998.

83. Yolken, R. H., and E. F. Torrey. Viruses, schizophrenia, and bipolar disorder. *Clin. Microbiol. Rev.* 8:131–45, 1995.

84. Castrogiovanni, P., S. Iapichino, C. Pacchierotti, and F. Pieraccini. Season of birth in psychiatry: A review. *Neuropsychobiol.* 37:175–81, 1998.

85. Torrey, E. F., A. E. Bowler, and K. Clark. Urban birth and residence as risk factors for psychoses: An analysis of 1880 data. *Schizophr. Res.* 25:169–76, 1997.

86. Torrey, E. F., and R. H. Yolken. At issue: Is household crowding a risk factor for schizophrenia and bipolar disorder? *Schizophr. Bull.* 24:321–24, 1998.

87. Rothschild, B. M., and R. J. Woods. Does rheumatoid polyarthritis come from the New World? *Rev. Rhum. Mal. Osteoartic.* 57:271–74, 1990.

88. Bytzer, P., P. B. Christensen, P. Damkier, et al. Adenocarcinoma of the esophagus and Barrett's esophagus: A population based study. *Am. J. Gastroenterol.* 94:86–91, 1999.

89. Haghighi, P., and P. L. Wolf. Tropical sprue and subclinical enteropathyy: A vision for the nineties. *Crit. Rev. Clin. Lab. Sci.* 34:313–41, 1997.

90. Tilley, B. C., G. S. Alarcón, S. P. Heyse, et al. Minocycline in rheumatoid arthritis: A 48-week, double-blind, placebo-controlled trial. *Ann. Int. Med.* 122:81–89, 1995.

91. Pier, G. B., M. Grout, T. Azidi, et al. *Salmonella typhi* uses CFTR to enter intestinal cells. *Nature* 393:79–82, 1998.

92. Gross, C. N., A. Irrinki, J. N. Feder, and C. A. Enns. Co-trafficking of HFE, a nonclassical major histocompatibility complex class I protein, with the transferrin receptor implies a role in intracellular iron regulation. *J. Biol. Chem.* 273:22068–74, 1998.

93. Mahadeva, R., R. C. Westerbeek, D. J. Perry, et al. Alpha-antitrypsin deficiency alleles and the Taq-I G→A allele in cystic fibrosis lung disease. *Eur. Respir. J.* 11: 873–79, 1998.

94. Chandley, A. C. The origin of chromosomal aberrations in man and their

potential for survival and reproduction in the adult human population. *Ann. Genet.* 24:5–11, 1981.

95. Roberts R. M. Embryonic loss and conceptus interference production. In *Uterine and Embryonic Factors in Early Pregancy,* edited by J. R. I. Strauss and C. R. Lyttle. New York: Plenum, 1991. 21–31.

96. Hildt, E., P. H. Hofschneider, and S. Urban. The role of hepatitis B virus (HBV) in the development of hepatocellular carcinoma. *Sem. Virol.* 7:333–47, 1996.

97. Livezey, K. W., and D. Simon. Accumulation of genetic alterations in a human hepatoma cell line transfected with hepatitis B virus. *Mut. Res.* 377:187–98, 1997.

98. Neel, J. V. An association, in adult Japanese, between the occurrence of rogue cells among cultured lypmphocytes (JC virus activity) and the frequency of "simple" chromosomal damage among the lymphocytes of persons exhibiting these rogue cells. *Am. J. Hum. Genet.* 63:489–97, 1998.

99. Southern, S. A., M. F. Evans, and C. S. Herrington. Basal cell tetrasomy in low-grade cervical squamous intraepithelial lesions infected with high-risk human papillomaviruses. *Cancer Res.* 57:4210–13, 1997.

100. Southern, S. A., and C. S. Herrington. Interphase karyotypic analysis of chromosomes 11, 17 and X in invasive squamous-cell carcinoma of the cervix: Morphological correlation with HPV infection. *Int. J. Cancer* 70:502–7, 1997.

101. Southern, S. A., and C. S. Herrington. Molecular events in uterine cervical cancer. *Sex. Trans. Infect.* 74:101–9, 1998.

102. Steenbergen, R. D. M., M. A. J. A. Hermsen, J. M. M. Walboomers, et al. Non-random allelic losses at 3p, 11p and 13q during HPV-mediated immortalization and concomitant loss of terminal differentiation of human keratinocytes. *Int. J. Cancer* 76:412–17, 1998.

103. Hermonat, P. L., L. Han, P. J. Wendel, et al. Human papillomavirus is more prevalent in first trimester spontaneously aborted products of conception compared to elective specimens. *Virus Genes* 14:13–17, 1997.

104. Fray, M. D., H. Prentice, M. C. Clarke, and B. Charleston. Immunohisto-chemical evidence for the localization of bovine viral diarrhea virus, a single-stranded RNA virus, in ovarian oocytes in the cow. *Vet. Pathol.* 35:25–39, 1998.

105. Roff, D. A. Evolution of threshold traits: The balance between selectional selection, drift and mutation. *Heredity* 80:25–32, 1998.

106. Hamilton, W. D. Sex versus non-sex versus parasite. *Oikos* 35:282–90, 1980.

107. Hamilton, W. D. Seething genetics of health and the evolution of sex. In *Evolution of Life, Fossils, Molecules, and Culture,* edited by S. Osawa and T. Honjo. Berlin: Springer, 1991. 229–51.

108. Hamilton, W. D. Haploid dynamical polymorphism in a host with matching parasites: Effects of mutation/subdivision, linkage, and patterns of selection. *J. Hered.* 84:328–38, 1993.

109. Oldstone, M. B. A. Molecular mimicry and immune-mediated diseases. *FASEB J.* 12:1255–65, 1998.

110. Hill, A., C. E. Allsopp, D. Kwiatkowski, et al. Common west African HLA antigens are associated with protection from severe malaria. *Nature* 352:595–600, 1991.

111. Goldfeld, A. E., J. C. Delgado, S. Thim, et al. Association of an HLA-DQ allele with clinical tuberculosis. *JAMA* 279:226–28, 1998.

112. Zhao, Z., F. Granucci, P. A. Schaffer, and H. Cantor. Molecular mimicry by herpes simplex virus type 1: Autoimmune disease after viral infection. *Science* 279: 1344–47, 1998.

113. Hamilton, W. D. Pathogens as causes of genetic diversity in their host populations. In *Population Biology of Infectious Diseases: Report of the Dahlem Workshop*

on Population Biology of Infectious Disease Agents, Berlin 1982, March 14–19, edited by R. M. Anderson and R. M. May. Berlin: Springer Verlag, 1982. 269–96.

114. Hamilton, W. D., R. Axelrod, and R. Tanese. Sex as an adaptation to resist parasites. *Proc. Natl. Acad. Sci. USA* 87:3566–73, 1990.

115. Weiss, K. M. *Genetic Variation and Human Disease: Principles and Evolutionary Approaches.* Cambridge: Cambridge Univ. Press, 1995.

116. Ewald, P. W. Evolutionary biology and the treatment of signs and symptoms of infectious disease. *J. Theor. Biol.* 86:169–76, 1980.

117. Dietrich, D. E., M. Schedlowski, L. Bode, et al. A viro-psycho-immunological disease-model of a subtype affective disorder. *Pharmacopsychiat.* 31:77–82, 1998.

118. Pletnikov, M. V., S. A. Rubin, G. J. Schwartz, et al. Persistent neonatal Borna disease virus (BDV) infection of the brain causes chronic emotional abnormalities in adult rats. *Physiol. Behav.* 66:823–31, 1999.

119. Haga, S., M. Yoshimura, Y. Motoi, et al. Detection of borna disease virus genome in normal human brain tissue. *Brain Res.* 770:307–9, 1997.

120. Iwahashi, K., M. Watanabe, K. Nakamura, et al. Clinical investigation of the relationship between Borna disease virus (BDV) infection and schizophrenia in 67 patients in Japan. *Acta. Psychiatr. Scand.* 96:412–15, 1997.

121. Iwahashi, K., M. Watanabe, K. Nakamura, et al. Positive and negative syndromes, and Borna disease virus infection in schizophrenia. *Neuropsychobiol.* 37:59–64, 1998.

122. Rapoport, J. L. The biology of obsessions and compulsions. *Sci. Amer.* 260:82–89, 1989.

123. Rapoport, J. L. The new biology of obsessive-compulsive disease: implications for evolutionary psychology. *Persp. Biol. Med.* 41:159–75, 1998.

124. Bell, A. P., and M. S. Weinberg. *Homosexualities: A Study of Diversity among Men and Women.* New York: Simon and Schuster, 1978.

125. Bailey, J. M. Empirical tests of two evolutionary hypotheses of male homosexuality. Paper presented at the Annual Meeting of the Human Behavior and Evolution Society, 5 June 1999.

126. Bailey, J. M., M. P. Dunne, and N. G. Martin. Genetic and environmental influences on sexual orientation and its correlates in an Australian twin sample. *J. Pers. Soc. Psychol.* (in press 2000).

127. Rice, G., C. Anderson, N. Risch, and G. Ebers. Male homosexuality: Absence of linkage to microsatellite markers at Xq28. *Science* 284:665–67, 1999.

128. Perkins, A., and J. A. Fitzgerald. Luteinizing hormone, testosterone, and behavioral response of male-oriented rams to estrous ewes and rams. *J. Anim. Sci.* 70:1787–94, 1992.

129. Perkins, A., J. A. Fitzgerald, and G. E. Moss. A comparison of LH secretion and brain estradiol receptors in heterosexual and homosexual rams and female sheep. *Horm. Behav.* 29:31–34, 1995.

130. Resko, J. A., A. Perkins, C. E. Roselli, et al. Endocrine correlates of partner preferences behavior in rams. *Biol. Reprod.* 55:120–26, 1996.

131. Miller, B. L., J. L. Cummings, H. Mcintyre, et al. Hypersexuality or altered sexual preference following brain injury. *Neurol. Neurosurg. Psychiatry* 49:867–73, 1986.

132. Gorman, M. R. Male homosexual desire: neurological investigations and scientific bias. *Persp. Biol. Med.* 38:61–81, 1994.

133. Fredricks, D. N., and D. A. Relman. Sequence-based identification of microbial pathogens: A reconsideration of Koch's postulates. *Clin. Microbiol. Rev.* 9:18–33, 1996.

134. Duesberg, P. H. AIDS acquired by drug consumption and other noncontagious risk factors. *Pharmacol. Ther.* 55:201–77, 1992.

135. Sutter, M. C. Assigning causation in disease: Beyond Koch's postulates. *Persp. Biol. Med.* 39:591–92, 1996.
136. Meier, C. R., L. E. Derby, S. S. Jick, et al. Antibiotics and risk of subsequent first-time acute myocardial infarction. *JAMA* 281:427–31, 1999.
137. Jackson, L. A., N. L. Smith, S. R. Heckbert, et al. Lack of association between first myocardial infarction and past use of erythromycin, tetracycline, or doxycycline. *Emerg. Infect. Dis.* 5:281–84, 1999.
138. Jordan, I., T. Briese, D. R. Averett, and W. I. Lipkin. Inhibition of Borna disease virus replication by ribavirin. *J. Virol.* 73:7903–6, 1999.
139. Rubin, S. A., T. A. Yednock, and K. M. Carbone. In vivo treatment with anti-alpha(4) integrin suppresses clinical and pathological evidence of Borna disease virus infection. *J. Neuroimmunol.* 84:158–63, 1998.
140. Hallensleben, W., and P. Staeheli. Inhibition of Borna disease virus multiplication by interferon: Cell line differences in susceptibility. *Arch. Virol.* 144: 1209–16, 1999.
141. Rapaport, M. H., L. Guylai, and P. Whybrow. Immune parameters in rapid cycling bipolar patients before and after lithium treatment. *J. Psychiatr. Res.* 33: 335–40, 1999.
142. Rybakowski, J. K. The effect of lithium on the immune system. *Human Psychopharmacol. Clin. Exp.* 14:345–53, 1999.
143. Moore, P. S., and Y. Chang. Kaposi's sarcoma (KS), KS-associated herpesvirus, and the criteria for causality in the age of molecular biology. *Am. J. Epidemiol.* 147:217–21, 1998.
144. Relman, D. A. The search for unrecognized pathogens. *Science* 284:1308–10, 1999.
145. Kiple, K. *Cambridge World History of Human Disease*. Cambridge: Cambridge Univ. Press, 1993.
146. Manson, P., and G. C. Cook. *Manson's Tropical Diseases*, 20th ed. London: Saunders, 1996.
147. Online Mendelian Inheritance in Man, OMIM (TM). Center for Medical Genetics, Johns Hopkins Univ. (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD). http://www3.ncbi.nlm.nih.gov/omim/.