

Explaining Disease: Correlations, Causes, and Mechanisms

PAUL THAGARD*

Philosophy Department, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada (e-mail: pthagard@watarts.uwaterloo.ca)

Abstract. Why do people get sick? I argue that a disease explanation is best thought of as causal network instantiation, where a causal network describes the interrelations among multiple factors, and instantiation consists of observational or hypothetical assignment of factors to the patient whose disease is being explained. This paper first discusses inference from correlation to causation, integrating recent psychological discussions of causal reasoning with epidemiological approaches to understanding disease causation, particularly concerning ulcers and lung cancer. It then shows how causal mechanisms represented by causal networks can contribute to reasoning involving correlation and causation. The understanding of causation and causal mechanisms provides the basis for a presentation of the causal network instantiation model of medical explanation.

Key words: explanation, disease, medicine, causality, mechanism, causal network

1. Introduction

Why do people get sick? Consider the case of Julia, a 50 year old lawyer who goes to her doctor complaining of stomach pains. After ordering some tests, the doctor tells her that she has a gastric ulcer. If this were the 1950s, the doctor would probably tell her that she needs to take it easy and drink more milk. If this were the 1970s or 1980s, Julia would probably be told that she suffered from excessive acidity and be prescribed Zantac or similar antacid drug. But since this is the 1990s, her well-informed doctor tells her that she probably has been infected by a newly discovered bacterium called *Helicobacter pylori* and that she needs to take a combination of antibiotics that will eradicate the bacteria and cure the ulcer.

The aim of this paper is to develop a characterization of disease explanations, such as the explanation that Julia got her ulcer because of a bacterial infection.¹ Medical explanation is very complex, because most diseases involve the interplay of multiple factors. Many people with *H. pylori* infection do not get ulcers, and some people have ulcers without having an infection. I will offer a proposal that a disease explanation is best thought of as a *causal network* instantiation, where a causal network describes the interrelations among multiple factors, and instantiation consists of observational or hypothetical assignment of factors to the patient whose disease is being explained. Explanation of why members of a particular class of people (women, lawyers, and so on) tend to get a particular disease is also causal network instantiation, but at a more abstract level.

Section 2 discusses the inference from correlation to causation, integrating recent psychological discussions of casual reasoning with epidemiological approaches to understanding disease causation. I will primarily use two cases to illustrate disease explanation: the development since 1983 of the bacterial theory of ulcers, and the evolution over the past several decades of ideas about the causes of cancer, particularly lung cancer. Both of these developments involved progression from observed correlations to accepted causal hypotheses (bacteria cause ulcers, smoking causes cancer), followed by increased understanding of the mechanisms by which the causes produce the diseases. Section 3 shows how causal mechanisms represented by causal networks can contribute to reasoning involving correlation and causation. The understanding of causation and causal mechanisms provides the basis in Section 4 for a presentation of the causal network instantiation model of medical explanation.

2. Correlation and Causes

Explanation of why people get a particular disease usually begins by the noticing of associations between the disease and possible causal factors. For example, the bacterial theory of ulcer originated in 1982 when two Australian physicians, Barry Marshall and J. Robin Warren, noticed an association between duodenal ulcer and infection with *Helicobacter pylori*, a previously unknown bacterium that Warren had microscopically discovered in biopsy specimens in 1979 (Marshall 1989, Thagard forthcoming). Marshall and Warren were aware that their study, which looked for relations between presence of the bacteria and various stomach elements in 100 patients who had had endoscopic examinations, did not establish a cause-and-effect relation between bacteria and ulcers (Marshall and Warren 1984, p. 1314). But they took it as evidence that the bacteria were etiologically related to the ulcers and undertook studies to determine whether eradicating the ulcers would cure the bacteria. These studies were successful, and by 1994 enough additional studies had been done by researchers in various countries that the U.S. National Institutes of Health Consensus Development Panel concluded that bacterial infection is causally related to ulcers and recommended antibiotic treatment (Marshall et al. 1988, National Institutes of Health Consensus Development Panel 1994)

A similar progression from correlation to causation has taken place with various kinds of cancer. Over two thousand years ago, Hippocrates described cancers of the skin, stomach, breast, and other body location, and held that cancer is caused, like all diseases, by an imbalance of bodily humors, particularly an excess of black bile. In the eighteenth century, rough correlations were noticed between cancers and various practices: using snuff and nose cancer, pipe smoking and lip cancer, chimney sweeping and scrotum cancer, and being a nun and breast cancer (Proctor 1995, pp. 27–28). The perils of causal reasoning are shown by the inferences of the Italian physician Bernardino Ramazzini who concluded in 1713 that the increased incidence of breast cancer in nuns was caused by their sexual abstinence, rather

than by their not having children. Early in the twentieth century it was shown that cancers can be induced in laboratory animals by radiation and coal tar.

Lung cancer rates increased significantly in Great Britain and the United States during the first half of the twentieth century, correlating with increase in smoking, but carefully controlled studies only began to appear in the 1950s (Hennekens and Buring 1987, p. 44). In one classic study conducted in England, 649 male and 60 female patients with lung cancer were matched to an equal number of control patients of the same age and sex. For both men and women, there was a strong correlation between lung cancer and smoking, particularly heavy smoking. By 1964, when the U.S Surgeon General's Report asserted a causal link between lung cancer and smoking, there had been 29 controlled studies performed in numerous countries that showed a high statistical association between lung cancer and smoking. Although the exact mechanism by which smoking causes cancer was not known, over 200 different compounds had been identified in cigarette smoke that were known carcinogens.

To grasp how disease explanations work, we need to understand what correlations are, what causes are, and how correlations can provide evidence for causes.² Patricia Cheng's (1997) 'power PC' theory of how people infer causal powers from probabilistic information provides a useful starting point. She proposes that when scientists and ordinary people infer the causes of events, they use an intuitive notion of causal power to explain observed correlations. She characterizes correlation (covariation) in terms of probabilistic contrasts: how much more probable is an effect given a cause than without the cause. The association between an effect e and a possible cause c can be measured by: $\Delta P_c = P(e/c) - P(e|\sim c)$, i.e. the probability of e given c minus the probability of e given not- c . However, in contrast to a purely probabilistic account of causality, she introduces an additional notion of the *power* of a cause c to produce an effect e , p_c , which is the probability with which c produces e when c is present.³ Whereas $P(e/c)$ is an observable frequency, p_c is a theoretical entity that is hypothesized to explain frequencies, just as theoretical entities like electrons and molecules are hypothesized to explain observations in physics. On Cheng's account, causal powers are used to provide theoretical explanations of correlations, just as theories such as the kinetic theory of gases are used to explain laws such as ones linking observed properties of gases (pressure, volume, temperature).

According to Cheng, a causal power p_c is a probability, but what kind of probability? Philosophers have debated about whether probabilities are frequencies, logical relations, or subjective states, but the interpretation of probability that seems to fit best with Cheng's view is that a probability is a propensity, i.e. a dispositional property of part of the world to produce a frequency of events in the long run. The causal power p_c cannot be immediately inferred from the observed frequency $P(e/c)$ or the contrast ΔP_c , because the effect e may be due to alternative causes. Celibate nuns get breast cancer more than non-nuns, but it is non-pregnancy rather than celibacy that is causally related to breast cancer. To estimate the causal power

of c to produce e , we need to take into account alternative possible causes of e , designated collectively as a . If there are no alternative causes of e besides c , then $P(e/c) = p_c$, but they will normally not be equal if a is present and produces e in the presence of c , i.e. if $P(a/c) * p_a > 0$, where p_a is the causal power of a to produce c . In the simple case where a occurs independently of c , Cheng shows that p_c can be estimated using the equation:

$$p_c = \Delta P_c / (1 - P(a) * P_a).$$

The causal relation between e and c can thus be assessed by considering positively the correlation between e and c and negatively the operation of other causes a . When these alternative causes do not occur independently of c , then ΔP_c may not reflect the causal status of c .

Cheng's characterization of the relation between correlations and causal powers fits well with epidemiologists' discussions of the problem of determining the causes of diseases.⁴ According to Hennekens and Buring (1987, p. 30), a causal association is one in which a 'change in the frequency or quality of an exposure or characteristic results in a corresponding change in the frequency of the disease or outcome of interest.' Elwood (1988, p. 6) says that 'a factor is a cause of an event if its operation increases the frequency of the event.' These statements incorporate both ΔP_c , captured by the change in frequency, and the idea that the change in frequency is the result of the operation of the cause, i.e. a causal power. Further, epidemiologists stress that assessing whether the results of a study reveal a causal relation requires considering alternative explanations of the observed association, such as chance, bias in the design of the study, and confounding alternative causes (see also Evans 1993, Susser 1973). Thus the inference from correlation to cause must consider possible alternative causes, p_a .⁵

Hennekens and Buring summarize their extensive discussion of epidemiologic studies in the framework reproduced in table 1. Questions A1–A3 reflect the need to rule out alternative causes, while questions B1 and B3 reflect the desirability of high correlations ΔP_c . Cheng's account of causal reasoning captures five of the eight questions relevant to assessing causal power, but the remaining three questions are beyond the scope of her model, which is restricted to induction from observable input. Hennekens and Buring state (p. 40) that 'the belief in the existence of a cause and effect relationship is enhanced if there is a known or postulated biologic mechanism by which the exposure might reasonably alter the risk of the disease.' Moreover (p. 42), 'for a judgment of causality to be reasonable, it should be clear that the exposure of interest preceded the outcome by a period of time consistent with the proposed biological mechanism.' Thus according to Hennekens and Buring, epidemiologists do and should ask mechanism-related questions about biologic credibility and time sequence; this issue is discussed in the next section. Finally, Hennekens and Buring's last question concerns the existence of a dose-response relationship, that is, the observation of a gradient of risk associated with the degree of exposure. This relation is not just ΔP_c , the

Table I. Framework for the interpretation of an epidemiologic study.
From Hennekens and Buring 1987, p. 45.

-
- A. Is there a valid statistical association?
1. Is the association likely to be due to chance?
 2. Is the association likely to be due to bias?
 3. Is the association likely to be due to confounding?
- B. Can this valid statistical association be judged as cause and effect?
1. Is there a strong association?
 2. Is there biologic credibility to the hypothesis?
 3. Is there consistency with other studies
 4. Is the time sequence compatible?
 5. Is there evidence of a dose-response relationship?
-

increased probability of having the disease given the cause, but rather the relation that being subjected to more of the cause produces more of the disease, for example when heavy smokers get lung cancer more than light smokers.

Hennekens and Buring show how answers to the questions in table 1 provide a strong case for a causal connection between smoking and lung cancer. Many studies have shown a strong association between smoking and cancer, with a 9 to 10-fold increase in lung cancer among smokers, (B1, B3), and the high statistical significance of the results makes it unlikely that the association is due to chance (A1). The conduct of the studies ruled out various sources of observation bias (A2), and researchers controlled for four potential confounding factors: age, sex, social class, and place of residence (A3). By 1959, cigarette smoke was known to contain over 200 different compounds that were known carcinogens, providing possible mechanisms that establish the biologic credibility of hypothesis that smoking causes cancer (B2). Moreover, there was evidence of a temporal relationship between smoking and cancer, because people obviously get lung cancer after they have been smoking for a long time, and people who stop smoking dramatically drop their chances of getting cancer (B4). Finally, there is a significant dose-response relationship between smoking and lung cancer, in that the risk of developing lung cancer increases substantially with the number of cigarettes smoked per day and the duration of the habit.

The development of the bacterial theory of ulcers can also be interpreted in terms of Cheng's theory of causality and Hennekens and Buring's framework for epidemiologic investigation. In 1983, when Marshall and Warren first proposed that peptic ulcers are caused by bacteria, most gastroenterologists were highly skeptical. They attributed the presence of bacteria in Warren's gastric biopsies to contamination, and they discounted the correlation between ulcers and bacterial infection as likely the result of chance or incorrect study design. Moreover, an

alternative explanation that ulcers are caused by excess acidity was widely accepted because of the success of antacids in alleviating ulcer symptoms. But attitudes toward the ulcer hypothesis changed dramatically when numerous other researchers observed the bacteria in stomach samples, and especially when other research teams replicated Marshall and Warren's finding that eradicating *Helicobacter pylori* usually cures ulcers.

The key question is whether bacteria cause ulcers, which requires attributing to *H. pylori* the causal power to increase the occurrence of ulcers. Initial evidence for this attribution was the finding that people with the bacteria more frequently have ulcers than those without, $P(\text{ulcers/bacteria}) > P(\text{ulcers/no bacteria})$, but the early studies could not establish causality because they did not address the question of possible alternative causes for the ulcers. Whereas lung cancer investigators had to use case-control methods to rule out alternative causes by pairing up patients with lung cancers with similar patients without the disease, ulcer investigators could use the fact that *H. pylori* can be eradicated by antibiotics to perform a highly controlled experiment with one set of patients, comparing them before eradication and after. The results are striking: the frequency of ulcers drops substantially in patients whose bacteria have been eliminated, and long-term recurrence rates are also much lower. These experiments thus show a very high value for ΔP , $P(\text{ulcers/bacteria}) - P(\text{ulcers/no bacteria})$, under circumstances in which no alternative causal factors such as stress, diet, and stomach acidity were varied.

Dose-response relationship has not been a factor in the conclusion that ulcers cause bacteria, since it is not easy to quantify how many bacteria inhabit a given patient's stomach. Time sequence is not much of an issue, since the common presence of the bacteria in children implies that people get the bacteria well before they get ulcers.⁶ But biologic credibility, concerning the mechanism by which bacterial infection might produce ulcers, has been the subject of much investigation, as I will discuss in the next section.

In sum, much of the practice of physicians and epidemiologists in identifying the causes of diseases can be understood in terms of Cheng's theory that causal powers are theoretical entities that are inferred on the basis of finding correlations and eliminating alternative causes. But mechanism considerations are also often relevant to assessing medical causality.

3. Causes and Mechanisms

What are mechanisms and how does reasoning about them affect the inference of causes from correlations? A mechanism is a system of parts that operate or interact like those of a machine, transmitting forces, motion, and energy to one another. For millennia humans have used simple machines such as levers, pulleys, inclined planes, screws, and wheels. More complicated machines can be built out of these simple ones, all of which transmit motion from one part to another by direct contact. In the sixteenth and seventeenth centuries, natural philosophers came more and

Table II. Sketch of some important mechanisms in science.

science	parts	changes	interactions
physics	objects such as sun and planets	motion	forces such as gravitation
chemistry	elements, molecules	mass, energy	reactions
evolutionary biology	organisms	new species	natural selection
genetics	genes	genetic transmission and alteration	heredity, mutation, recombination
geology	geological formations such as mountains	creation and elimination of formations	volcanic eruptions, erosion, etc.
plate tectonics	continents	motion such as continental drift	floating, collision
neuroscience	neurons	activation, synaptic connections	electrochemical transmissions
cell biology	cells	growth	cell division
cognitive science	mental representations	creation and alteration of representations	computational procedures

more to understand the world in terms of mechanisms, culminating with Newton's unified explanation of the motion of earthly and heavenly bodies. His concept of force, however, went beyond the operation of simple machines by direct contact to include the gravitational interaction of objects at a distance from each other. In the history of science, progress has been made in many sciences by the discovery of new mechanisms, each with interacting parts affecting each other's motion and other properties. Table 2 displays some of the most important of such mechanisms. The sciences employ different kinds of mechanisms in their explanations, but each involves a system of parts that change as the result of interactions among them that transmit force, motion, and energy. Mechanical systems are organized hierarchically, in that mechanisms at lower levels (e.g. molecules) produce changes that take place at higher level (e.g. cells).

Medical researchers similarly are highly concerned with finding mechanisms that explain the occurrence of diseases, for therapeutic as well as theoretical purposes: understanding the mechanism that produces a disease can lead to new ideas about how the disease can be treated. In cancer research, for example, major advances were made in the 1970's and 1980's in understanding the complex of causes that lead to cancer (Weinberg 1996). There are over a hundred different kinds of cancer, but all are now thought to result from uncontrolled cell growth arising from a series of genetic mutations, first in genes for promoting growth (oncogenes) and then in genes for suppressing the tumors that are produced by uncontrolled cell growth. The mechanism of cancer production then consists of

parts at two levels - cells and the genes they contain, along with changes in cell growth produced by a series of genetic mutations. Mutations in an individual can occur for a number of causes, including heredity, viruses, and behavioral and environmental factors such as smoking, diet, and exposure to chemicals. Figure 1 sums up the current understanding of the mechanisms underlying cancer. This understanding is currently generating new experimental treatments based on genetic manipulations such as restoring the function of tumor suppresser genes (Bishop and Weinberg 1996).

Ulcer researchers have also been very concerned with the mechanism by which *Helicobacter pylori* infection produces ulcers. Figure 2 displays a mechanism similar to one proposed by Graham (1989) that shows some of the interactions of heredity, environment, infection, and ulceration. Research is underway to fill in the gaps about these processes (e.g. Olbe et al. 1996).

Recent psychological research by Woo-kyoung Ahn and her colleagues has found that when ordinary people are asked to provide causes for events, they seek out information about underlying causal mechanisms as well as using information about correlations (Ahn et al. (1995); Ahn and Bailenson 1996). For example, if people are asked to state the cause of John's car accident, they will not survey a range of possible factors that correlate with accidents, but will rather focus on the process underlying the relationship between cause and effect, such as John's being drunk leading to erratic driving, leading to the accident. Whereas causal attribution based on correlation (covariation) alone would ignore mechanisms connecting cause and effects, ordinary people are like medical researchers in seeking mechanisms that connect cause and effect.

As Cheng (1997) points out, however, the emphasis on mechanism does not by itself provide an answer to the question of how people infer cause from correlation: knowledge of mechanism is itself knowledge of causally related events which must have somehow been previously acquired. Medical researchers inferred that bacteria cause ulcers and that smoking causes cancer at times when little was known about the relevant causal mechanisms. Reasoning about mechanisms can contribute to causal inference, but is not necessary for it. In domains where causal knowledge is rich, there is a kind of feedback loop in which more knowledge about causes leads to more knowledge about mechanisms which leads to more knowledge about causes. But in less well understood domains, correlations and consideration of alternative causes can get causal knowledge started in the absence of much comprehension of mechanisms.

To understand how reasoning about mechanisms affects reasoning about causes, we need to consider four different situations that arise in science and ordinary life when we are considering whether a factor c is a cause of an event e :

1. There is a known mechanism by which c produces e .
2. There is a plausible mechanism by which c produces e .
3. There is no known mechanism by which c produces e .
4. There is no plausible mechanism by which c produces e .

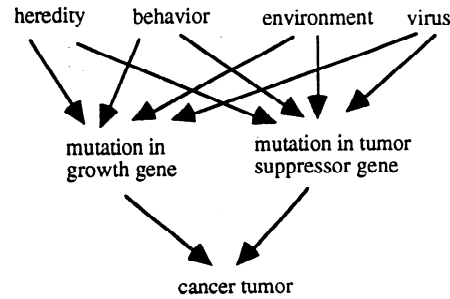


Figure 1. Mechanism of cancer production.

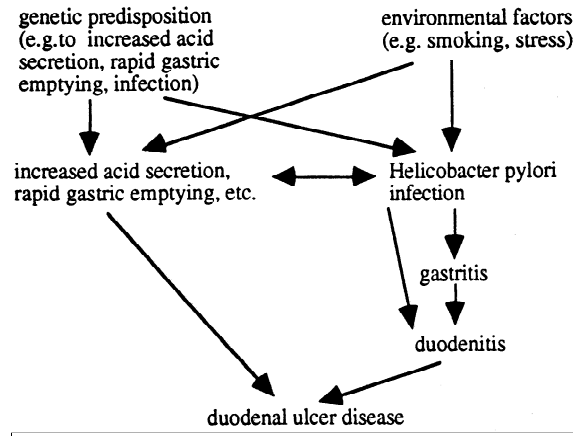


Figure 2. Possible mechanism of duodenal ulcer production. Modified from Graham 1989, p. 51. Gastric ulcer causation is similar.

For there to be a known mechanism by which c produces e , c must be a component of or occurrence in a system of parts that is known to interact to produce e . Only very recently has a precise mechanism by which smoking causes cancer become known through the identification of a component of cigarette smoke (Benzo[a]pyrene) that produces mutations in the tumor suppresser gene p53 (Denis-senko et al., 1996). As we saw above, however, there has long been a *plausible* mechanism by which smoking causes lung cancer.

When there is a known mechanism connecting c and e , the inference that c causes e is strongly encouraged, although careful causal inference will still need to take into account information about correlations and alternative causes, since a different mechanism may have produced e by an alternative cause a . For example, drunk driving often produces erratic driving that produces accidents, but even if John was drunk his accident might have been caused by a mechanical malfunction rather than his drunkenness. Similarly, even though there is now a plausible mechanism connecting *H. pylori* infection and ulcers, we should not immediately conclude

that Julia has the infection, since approximately 20% of ulcers are caused by use of non-steroidal antiinflammatory drugs (NSAIDs) such as aspirin. But awareness of known and plausible mechanisms connecting c and e clearly facilitates inference that c causes e , in a manner that will be more fully spelled out below. Another way in which the plausibility of a mechanism can be judged is by analogy: if a cause and effect are similar to another cause and effect that are connected by a known mechanism, then it is plausible that a similar mechanism may operate in the original case. There was a plausible mechanism by which *H. pylori* caused stomach ulcers, since other bacteria were known to produce other sores.

Sometimes causal inference from correlation can be blocked when there is no plausible mechanism connecting the event and its cause, that is when possible mechanisms are incompatible with what is known. When Marshall and Warren first proposed that bacteria cause ulcers, the stomach was widely believed to be too acidic for bacteria to survive for long, so that there was no plausible mechanism by which bacteria could produce ulcers. Later it was found that *H. pylori* produce ammonia which neutralizes stomach acid allowing them to survive, removing the implausibility of the bacteria-ulcer mechanism. Similarly, when Alfred Wegener proposed continental drift early in this century, his theory was rejected in part because the mechanisms he proposed for continental motion were incompatible with contemporary geophysics. Only when plate tectonics was developed in the 1960s was it understood how continents can be in motion.

The two cases just mentioned are ones in which the implausibility of mechanisms was overcome, but there are many cases where rejection of causal relations remains appropriate. Even though there are some empirical studies providing correlational evidence for ESP, it is difficult to believe that people have such powers as telepathy and telekinesis, which have properties such as being unaffected by spatial and temporal relations that conflict with known physical mechanisms. Similarly, homeopathic medicine using minute doses of drugs violates established views concerning the amounts of substances needed to be chemically effective. An even more extreme case is the theory of Velikovsky that the planet Venus once swung close to Earth causing many historical events such as the parting of the Red Sea for Moses. Such planetary motion is totally incompatible with Newtonian mechanics, so there is no plausible mechanism by which Venus' motion could have the claimed effect.

How can medical researchers and ordinary people combine information about mechanisms with information about correlations and alternative causes to reach conclusions about cause and effect? Recall Cheng's view that causes are theoretical entities to be inferred on the basis of correlations and alternative causes. Elsewhere I have argued that the justification of scientific theories including their postulation of theoretical entities is a matter of explanatory coherence, in which a theory is accepted because it provides a better explanation of the evidence (Thagard 1992). Explanatory coherence of a hypothesis is a matter both of the evidence it explains and of its being explained by higher level hypotheses. For example, Darwin justified the hypothesis of evolution both in terms of the biological evidence it explained

and in terms of evolution being explained by the mechanism of natural selection. Moreover, he explicitly compared the explanatory power of his theory of evolution by natural selection with the explanatory limitations of the dominant creation theory of the origin of species. These three factors - explaining evidence, being explained by mechanisms, and consideration of alternative hypotheses, are precisely the same considerations that go into evaluation of a causal hypothesis.

Figure 3 shows how the inference that c causes a disease d can be understood in terms of explanatory coherence.⁷ When medical researchers collect data that find a correlation between c and d , i.e. a high value for $P(d/c) - P(d/\sim c)$, there are several possible explanations for these data. That there really is a correlation in the relevant population between d and c is one possible explanation for the data, but experimenters must rule out explanations such as that the correlation in the data arose from chance or from experimental bias.⁸ Careful experimental designs involving such techniques as randomization and double blinding help to rule out bias, and appropriate techniques of statistical inference tend to rule out chance, leading to the acceptance of the hypothesis that there is a real correlation between c and d . However, before researchers can conclude that c causes d , they must have reason to believe that this hypothesis is a better explanation of the correlation than other confounding causes that might have been responsible for it. Again careful experimental design that manipulates only c or that otherwise controls for other potential causes is the key to concluding that c causes d is the best explanation of the correlation. In addition, the existence of a known or plausible mechanism for how c can produce d increases the explanatory coherence of the causal hypothesis. On the other hand, if all mechanisms that might connect c with d are incompatible with other scientific knowledge, then the hypothesis that c causes d becomes incoherent with the total body of knowledge. Evans (1993, p. 174) offers as one of his criteria for causation in medicine that “the whole thing should make biologic and epidemiologic sense.” As Hennekens and Buring (1987) suggest, a major determinant of whether a causal hypothesis makes sense is whether it comes with a plausible underlying mechanism.

Figure 3 points to a synthesis of Cheng’s ideas about causal powers, probabilities, and alternative causes with considerations of mechanism. Mechanisms are not a necessary condition for causal inference, but when they are known or plausible they can enhance the explanatory coherence of a causal hypothesis. Moreover, causal hypotheses incompatible with known mechanisms are greatly reduced in explanatory coherence. Inference to causes, like inference to theoretical entities in general, depends on explanatory coherence as determined by evidence, alternative hypotheses, and higher level hypotheses.

Inference to medical causes is similar to legal inference concerning responsibility for crimes. In a murder case, for example, the acceptability of the hypothesis that someone is the murderer depends on how well that hypothesis explains the evidence, on the availability of other hypotheses to explain the evidence, and on the presence of a motive that would provide a higher level explanation of why the

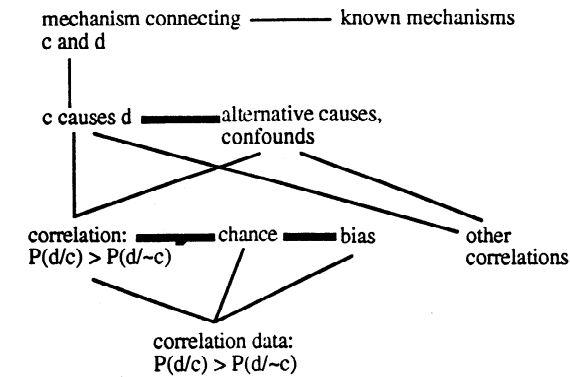


Figure 3. Inferring a cause c from correlation data about a disease d . That there is a correlation between d and c must be a better explanation of the observed correlation than chance or bias (or fraud). That c causes d must be a better explanation of the correlation and other correlations than alternative confounding causes. The existence of a mechanism connecting c and d provides an explanation of why c causes d . In the Figure, thin lines are explanatory relations, while the thick lines indicate incompatibility.

accused committed the murder. Motives in murder trials are like mechanisms in medical reasoning, providing non-essential but coherence-enhancing explanation of a hypothesis.

This section has discussed how knowledge of mechanisms can affect inferences about causality, but it has passed over the question of how such knowledge is obtained. There are three possibilities. First, some knowledge about basic physical mechanisms may be innate, providing an infant with a head start for figuring out the world. For example, it is possible that infants are innately equipped to infer a causal relation when one moving object bang into another object that then starts moving. Second, some of the links in the causal chains that constitute a mechanism may be learned by induction from observed correlations as described in Cheng's Power PC model. For example, we can observe the relations among pressure, temperature, and volume changes in gases and infer that they are causally connected. Third, sometimes mechanisms are abduced, that is posited as a package of hypothetical links used to explain something observed. For example, in cognitive science we posit computational mechanisms with various representations and processes to explain intelligent behavior. Darwin abduced the following causal chain:

variation + competition \rightarrow natural selection \rightarrow evolution of species.

The difference between abductive and inductive inference about mechanisms is that in inductive inference the parts and processes are observed, while in abductive inference they are hypothesized. Knowledge about mechanisms involving theoretical (nonobservable) entities must be gained abductively, by inferring that the existence of the mechanism is the best explanation of the results of observation

and experiment. Different domains vary in the extent to which knowledge about mechanisms is innate, induced from correlations, or abductive.

4. Disease Explanation as Causal Network Instantiation

The above description of the interrelations of correlations, causes, and mechanisms provides the basis for an account of the nature of medical explanation. First we can eliminate a number of defective alternative accounts of explanation, including that explanation is essentially deductive, statistical, or involves single causes.

1. *Explanation is not deductive.* The deductive-nomological model of Hempel (1965), according to which an explanation is a deduction of a fact to be explained from universal laws, clearly does not apply to the kinds of medical explanation discussed here. Deductive explanations can be found in other fields such as physics, in which mathematical laws entail observations. But there are no general laws about the origins of ulcers and cancer. As we saw, most people with *H. pylori* do not get ulcers, and many people without *H. pylori* do get ulcers because of NSAIDs. Similarly, most smokers do not get lung cancer and some non-smokers get lung cancer. The development of ulcers, like the development of cancer, is far too complex for general laws to provide deductive explanation.

2. *Explanation is not statistical.* Statistics are certainly relevant to developing medical explanations, as we saw in the contribution of $P(\text{ulcers/bacteria}) - P(\text{ulcers/no bacteria})$ to the conclusion that bacteria cause ulcers. But correlations themselves have no explanatory force, since they may be the result of confounding alternative causes. As we saw in Figure 3, the conclusion that there is a causal and hence an explanatory relation between a factor and a disease depends on numerous coherence considerations, including the full range of correlations explained, the applicability of alternative causes, and the availability of a mechanism by which the factor produces the disease. A medical explanation need not show that a disease was to be expected with high probability, since the probability of getting the disease given the main cause may well be less than .5, as is the case for both ulcers/bacteria and lung cancer/smoking.

3. *Explanation is not in terms of single causes.* Although it is legitimate to see bacteria as the major causal factor in most ulcers and smoking as the major causal factor in most lung cancers, it is simplistic to explain someone's ulcer only in terms of bacterial infection, or someone's lung cancer only in terms of smoking. As Figures 1 and 2 displayed, ulcer causation and cancer causation are complex processes involving multiple interacting factors. Medical researchers are increasingly stressing the multifactorial nature of disease explanations. Adult-onset diabetes, for example, is now understood as arising from a complex of factors including heredity, obesity, and inactivity, all contributing to glucose intolerance, possibly because of a mechanism that involves a protein that reduces glucose uptake.

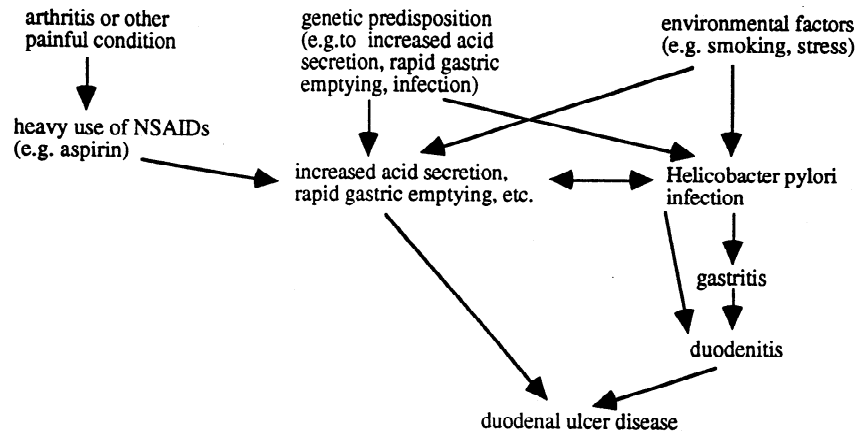


Figure 4. General causal network for duodenal ulcers, expanding figure 2.

I propose instead that medical explanation should be thought of as causal network instantiation (CNI).⁹ For each disease, epidemiological studies and biological research establish a system of causal factors involved in the production of a disease. The causal network for cancer is a more elaborate version of Figure 1, and the causal network for ulcers is a more elaborate version of Figure 4, which expands Figure 2. Crucially, the nodes in this network are not connected merely by conditional probabilities, $P(\text{effect}/\text{cause})$, but by causal relations inferred on the basis of multiple considerations, including correlations $P(\text{effect}/\text{cause}) - P(\text{effect}/\sim\text{cause})$, alternative causes, and mechanisms. We then explain why a given patient has a given disease by instantiating the network, that is by specifying which of the factors operate in that patient. To go back to the Julia case with which this paper began, the physician can start to instantiate the network in Figure 4 by determining whether Julia takes large quantities of NSAIDs, for example because she has arthritis. Different instantiation can take place on the basis of tests, for example endoscopy or a breath test to determine whether her stomach is infected with *H. pylori*. Some instantiation will be abductive, making hypotheses about the operation of factors that cannot be observed or tested for.¹⁰ The physician might make the abduction that Julia has a hereditary inclination to excess acidity, which would explain why she unlike most people with *H. pylori* has an ulcer; the hereditary abduction would be strengthened if her parents and other relatives had ulcers. Similarly, to explain patients' lung cancers, we instantiate a causal network with information about their smoking, their other behaviors, their heredity, and so on.

Instantiation of a causal network such as the one in Figure 4 produces a kind of narrative explanation of why a person gets sick. We can tell several possible stories about Julia, such as:

1. Julia became infected with *H. pylori* and because of a predisposition to excess acidity she got an ulcer.

2. Julia took a lot of aspirin for her arthritis which produced so much acidity in her stomach that she got ulcers.

But medical explanation is not just story telling, since a good medical explanation should point to all the interacting factors for which there is causal evidence and for which there is evidence of relevance to the case at hand. A narrative may be a useful device for communicating a causal network instantiation, but it is the ensemble of statistically-based causal relations that is more crucial to the explanation than the narration.

Causal networks provide an explanatory schema or pattern, but they differ from the sorts of explanatory schemas and patterns proposed by others. Unlike the explanatory patterns of Kitcher (1981, 1993), causal networks are not deductive. Deductive patterns may well have applications in fields such as mathematical physics, but they are of no use in medicine where causal relationships are not well represented by universal laws. Unlike the explanation patterns of Schank (1986), causal networks are not simple schemas that are used to provide single causes for effects, but instead describe complex mechanisms of multiple interacting factors. My account of medical explanation as causal network instantiation is compatible with the emphasis on mechanistic explanations by Salmon (1984) and Humphreys (1989), but provides a fuller specification of how causal networks are constructed and applied. As already mentioned, my CNI account is not compatible with interpreting the relations between factors in a causal network purely in terms of conditional probabilities.

Like explanation of a disease in a particular patient, explanation of why a group of people is prone to a particular disease is also a matter of causal network instantiation. People in underdeveloped countries are more likely to have gastritis than North Americans, because poorer sanitation makes it more likely that they will acquire *H. pylori* infections that produce ulcers. Nuns are more likely to get breast cancer than other women, because women who do not have full-term pregnancies before the age of 30 are more likely to get breast cancers, probably because of some mechanism by which pregnancy affects breast cell division. When we want to explain why a group is more likely to get a disease, we invoke the causal network for the disease and instantiate the nodes based on observations and abductions about the disease factors possessed by members of the group. Thus CNI explanations of both individual and group disease occurrence are structurally identical.¹¹

5. Conclusion

This paper has shown how correlations, causes, and mechanisms all figure in the construction of causal networks that can be instantiated to provide medical explanations. The main criterion for assessing a model of disease explanation is whether it accounts for the explanatory reasoning of medical researchers and practitioners. We have seen that the causal network instantiation model of medical explanation fits well with methodological recommendations of epidemiologists such as Hen-

nekens and Buring, as well as with the practice of medical researchers working on diseases such as ulcers and lung cancer. Additional examples of development and application of causal networks could easily be generated for other diseases such as diabetes. My account of medical explanation as causal network instantiation gains further credibility from the fact that its assumptions about the relations of correlations, causes, and mechanisms are consistent with (and provide a synthesis of) Cheng's and Ahn's psychological models of human causal reasoning.

This paper makes no claims about application of the CNI model beyond medicine. For some fields such as physics, the existence of universal laws and mathematical precision often make possible explanations that are deductive. On the other hand, in fields such as economics the lack of causal knowledge interrelating various economic factors may restrict explanations to being based on statistical associations. I expect, however, that there are many fields such as evolutionary biology, ecology, genetics, psychology, and sociology in which explanatory practice fits the CNI model. For example, the possession of a feature or behavior by members of a particular species can be explained in terms of a causal network involving mechanisms of genetics and natural selection. Similarly, the possession of a trait or behavior by a human can be understood in terms of a causal network of hereditary, environmental, and psychological factors. In psychology as in medicine, explanation is complex and multifactorial in ways well characterized as causal network instantiation.

Notes

*For research support, I am grateful to the Natural Sciences and Engineering Research Council of Canada, and the Social Sciences and Humanities Research Council of Canada. Thanks to Patricia Cheng, Herbert Simon, and Rob Wilson for comments on earlier drafts.

¹This paper is not concerned with the diagnosis problem of finding diseases that explain given symptoms, but rather with finding causes of diseases that patients are known to have. On medical diagnosis, see for example Peng and Reggia (1990).

²Terminological note: I take "correlation" to be interchangeable with "covariation" and "statistical association." Correlations are not always measured by the statistical formula for coefficient of correlation, which applies only to linear relationships.

³Similarly, Peng and Reggia (1990, p. 101f.) use "probabilistic causal models" that rely, not on conditional probabilities of the form $P(\text{effect/disease})$, but on "conditional causal probabilities" of the form $P(\text{disease causes effect/disease})$. Both probabilistic and causal power ideas have a long history in philosophy. On probabilistic causality, see for example Suppes (1970), Eells (1991), and Shafer (1996). On causal powers, see for example Cartwright (1989) and Harré and Madden (1975).

⁴It also fits with the view of Chinn and Brewer (1996) that data interpretation is a matter of building mental models that include alternative explanations.

⁵Is the consideration of alternative explanations in causal reasoning descriptive or prescriptive? Both: I am offering a model of medical reasoning that is "biscrptive", i.e. that describes how people make inferences when they are in accord with the best practices compatible with their cognitive capacities (Thagard 1992, p.97).

⁶The correlation between ulcers and bacteria might be taken to suggest that ulcers cause bacterial infections rather than the other way around. But the presence of bacteria is too widespread for this to be plausible: $P(\text{bacteria/ulcers}) - P(\text{bacteria/no ulcers})$ is not high, since the bacteria are quite common, infecting up to 50% of the population. Moreover, *H. pylori* were not found to be prominent on gastric ulcer borders, suggesting that the ulcers were not responsible for bacterial growth.

⁷For the full theory of explanatory coherence and its implementation in the computational model ECHO, see Thagard (1989, 1992).

⁸Mayo (1996) provides a thorough discussion of the use of statistical tests to rule out errors deriving from chance and other factors. Another possible source of error is fraud, when the observed correlations are based on fabricated data.

⁹Recent work on causal networks includes: Glymour et al. (1987); Iwasaki and Simon (1994); Pearl (1988); Shafer (1996).

¹⁰Abductive inference is inference to explanatory hypotheses. See for example Thagard (1988) and Josephson and Josephson (1994).

¹¹Note that I have not attempted to define cause in terms of explanation or explanation in terms of cause. Causes, mechanisms, explanations, and explanatory coherence are intertwined notions.

References

- Ahn, W. and Bailenson, J. (1996), 'Causal Attribution as a Search for Underlying Mechanism: An Explanation of the Conjunction Fallacy and the Discounting Principle'. *Cognitive Psychology*, 31, pp. 82–123.
- Ahn, W., Kalish, C. W., Medin, D. L. and Gelman, S. (1995), 'The Role of Covariation Versus Mechanism Information in Causal Attribution'. *Cognition*, 54, pp. 299–352.
- Bishop, J. M. and Weinber, R. A. (ed.). (1996), *Scientific American Molecular Oncology*. New York: Scientific American Books.
- Cartwright, N. (1989), *Nature's Capacities and Their Measurement*. Oxford: Clarendon Press.
- Cheng, P. W. (1997), 'From Covariation to Causation: A Causal Power Theory'. *Psychological Review*, 104, pp. 367–405.
- Chinn, C. A. and Brewer, F. (1996), 'Mental models in data interpretation'. *Philosophy of Science*, 63 (Proceedings supplement), S211–219.
- Denissenko, M. F., Pao, A., Tang, M. and Pfeifer, G. P. (1996), 'Preferential Formation of Benzo[a]pyrene Adducts at Lung Cancer Mutational Hotspots in p53'. *Science*, 274 (5286), pp. 430–432.
- Eells, E. (1991), *Probabilistic Causality*. Cambridge: Cambridge University Press.
- Elwood, J. M. (1988), *Causal Relationships in Medicine*. Oxford: Oxford University Press.
- Evans, A. S. (1993), *Causation and Disease: A Chronological Journey*. New York: Plenum.
- Glymour, C., Scheines, R., Spirtes, P. and Kelly, K. (1987), *Discovering Causal Structure*. Orlando: Academic Press.
- Graham, D. Y. (1989), 'Campylobacter Pylori and Peptic Ulcer Disease'. *Gastroenterology*, 6, pp. 615–625.
- Harré, R. and Madden, E. (1975), *Causal Powers*. Oxford: Blackwell.
- Hempel, C. G. (1965), *Aspects of Scientific Explanation*. New York: The Free Press.
- Hennekens, C. H. and Buring, J. E. (1987), *Epidemiology in Medicine*. Boston: Little, Brown.
- Humphreys, P. (1989), *The Chances of Explanation*. Princeton: Princeton University Press.
- Iwasaki, Y. and Simon, H. (1994), 'Causality and model abstraction'. *Artificial Intelligence*, 67, pp. 143–194.
- Josephson, J. R. and Josephson, S. G. (ed.). (1994), *Abductive Inference: Computation, Philosophy, Technology*. Cambridge: Cambridge University Press.
- Kitcher, P. (1981), 'Explanatory unification'. *Philosophy of Science*, 48, pp. 507–531.
- Kitcher, P. (1993), *The Advancement of Science*. Oxford: Oxford University Press.
- Marshall, B. J. (1989), 'History of the Discovery of C. Pylori', In M. J. Blaser (eds.), *Campylobacter Pylori in Gastritis and Peptic Ulcer Disease* (pp. 7–22). New York: Igaku-Shoin.
- Marshall, B. J., Goodwin C. S., Warren, J. R., Murray, R., Blincow, E. D., Blackburn, S. J., Phillips, M., Waters, T. E. and Sanderson, C. R. (1988), 'Prospective Doubleblind Trial of Duodenal Ulcer Relapse After Eradication of Campylobacter Pylori'. *Lancet*, 2(8626/8627), pp. 1437–1441.
- Marshall, B. J. and Warren, J. R. (1984), 'Unidentified Curved Bacilli in the Stomach of Patients With Gastritis and Peptic Ulceration'. *Lancet*, 1(8390), pp. 1311–1315.
- Mayo, D. (1996), *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

- National Institutes of Health Consensus Development Panel (1994), 'Helicobacter Pylori in Peptic Ulcer Disease'. *Journal of the American Medical Association*, 272, pp. 65–69.
- Olbe, L., Hamlet, A., Dalenbäck, J. and Fändriks, L. (1996), 'A Mechanism by Which Helicobacter Pylori Infection of the Antrum Contributes to the Development of Duodenal Ulcer'. *Gastroenterology*, 110, pp. 1386–1394.
- Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems*. San Mateo: Morgan Kaufman.
- Peng, Y. and Reggia, J. (1990), *Abductive Inference Models for Diagnostic Problem Solving*. New York: Springer-Verlag.
- Proctor, R. N. (1995) *Cancer Wars: How Politics Shapes what we Know and don't Know about Cancer*. New York: BasicBooks.
- Salmon, W. (1984), *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
- Schank, R. C. (1986), *Explanation Patterns: Understanding Mechanically and Creatively*. Hillsdale, NJ: Erlbaum.
- Shafer, G. (1996), *The Art of Causal Conjecture*. Cambridge, MA: MIT Press.
- Suppes, P. (1970), *Probabilistic Theory of Causality*. Atlantic Highlands, NJ: Humanities Press.
- Susser, M. (1973), *Causal Thinking in the Health Sciences*. New York: Oxford University Press.
- Thagard, P. (1988), *Computational Philosophy of Science*. Cambridge, MA: MIT Press/Bradford Books
- Thagard, P. (1989), 'Explanatory Coherence'. *Behavioral and Brain Sciences*, 12, pp. 435–467.
- Thagard, P. (1992), *Conceptual Revolutions*. Princeton: Princeton University Press.
- Thagard, P. (forthcoming). 'Ulcers and Bacteria I: Discovery and Acceptance'. *Studies in History and Philosophy of Science*.
- Weinberg, R. A. (1996), *Racing to the Beginning of the Road: The Search for the Origin of Cancer*. New York: Harmony Books.