11

A Model of Inexact Reasoning in Medicine

Edward H. Shortliffe and Bruce G. Buchanan

Inexact reasoning is common in the sciences. It is characterized by such phrases as "the art of good guessing," the "softer aspects of physics" (or chemistry, or any other science), and "good scientific judgment." By definition, inexact reasoning defies analysis as applications of sets of inference rules that are expressed in the predicate logic. Yet it need not defy all analysis. In this chapter we examine a model of inexact reasoning applied to a subdomain of medicine. Helmer and Rescher (1960) assert that the traditional concept of "exact" versus "inexact" science, with the social sciences accounting for the second class, has relied on a false distinction usually reflecting the presence or absence of mathematical notation. They point out that only a small portion of natural science can be termed exact—areas such as pure mathematics and subfields of physics in which some of the exactness "has even been put to the ultimate test of formal axiomatization." In several areas of applied natural science, on the other hand, decisions, predictions, and explanations are made only after exact procedures are mingled with unformalized expertise. The general awareness regarding these observations is reflected in the common references to the "artistic" components in the "science of medicine."

During the years since computers were first introduced into the medical arena, researchers have sought to develop techniques for modeling clinical decision making. Such efforts have had a dual motivation. Not only has their potential clinical significance been apparent, but the design of such programs has required an analytical approach to medical reasoning, which has in turn led to distillation of decision criteria that in some cases

This chapter is a shortened and edited version of a paper appearing in *Mathematical Biosciences* 23: 351–379 (1975).

had never been explicitly stated. It is both fascinating and educational for experts to reflect on the inference rules that they use when providing clinical consultations.

Several programs have successfully modeled the diagnostic process. Many of these have relied on statistical decision theory as reflected in the use of Bayes' Theorem for manipulation of conditional probabilities. Use of the theorem, however, requires either large amounts of valid background data or numerous approximations and assumptions. The success of Gorry and Barnett's early work (Gorry and Barnett, 1968) and of a similar study by Warner and coworkers using the same data (Warner et al., 1964) depended to a large extent on the availability of good data regarding several hundred individuals with congenital heart disease.

Although conditional probability provides useful results in areas of medical decision making such as those we have mentioned, vast portions of medical experience suffer from having so few data and so much imperfect knowledge that a rigorous probabilistic analysis, the ideal standard by which to judge the rationality of a physician's decisions, is not possible. It is nevertheless instructive to examine models for the less formal aspects of decision making. Physicians seem to use an ill-defined mechanism for reaching decisions despite a lack of formal knowledge regarding the interrelationships of all the variables that they are considering. This mechanism is often adequate, in well-trained or experienced individuals, to lead to sound conclusions on the basis of a limited set of observations.[1]

The purpose of this chapter is to examine the nature of such nonprobabilistic and unformalized reasoning processes and to propose a model by means of which such incomplete "artistic" knowledge might be quantified. We have developed this model in response to the needs of a computer program that will permit the opinions of experts to become more generally available to nonexperts. The model is, in effect, an approximation to conditional probability. Although conceived with medical decision making in mind, it is potentially applicable to any problem area in which real-world knowledge must be combined with expertise before an informed opinion can be obtained to explain observations or to suggest a course of action.

We begin with a brief discussion of Bayes' Theorem as it has been utilized by other workers in this field. The theorem will serve as a focus for discussion of the clinical problems that we would like to solve by using computer models. The potential applicability of the proposed decision model is then introduced in the context of the MYCIN system. Once the problem has been defined in this fashion, the criteria and numerical characteristics of a quantification scheme will be proposed. We conclude with a discussion of how the model is used by MYCIN when it offers opinions to physicians regarding antimicrobial therapy selection.

---

[1]Intuition may also lead to unsound conclusions, as noted by Schwartz et al. (1973).

11.1 Formulation of the Problem

The medical diagnostic problem can be viewed as the assignment of probabilities to specific diagnoses after analyzing all relevant data. If the sum of the relevant data (or evidence) is represented by e, and d_i is the ith diagnosis (or "disease") under consideration, then $P(d_i|e)$ is the conditional probability that the patient has disease i in light of the evidence e. Diagnostic programs have traditionally sought to find a set of evidence that allows $P(d_i|e)$ to exceed some threshold, say 0.95, for one of the possible diagnoses. Under these circumstances the second-ranked diagnosis is sufficiently less likely (<0.05) that the user is content to accept disease i as the diagnosis requiring therapeutic attention.[2]

Bayes' Theorem is useful in these applications because it allows $P(d_i|e)$ to be calculated from the component conditional probabilities:

$$P(d_i|e) = \frac{P(d_i)\ P(e|d_i)}{\Sigma\ P(d_j)\ P(e|d_j)}$$

In this representation of the theorem, d_i is one of n disjoint diagnoses, $P(d_i)$ is simply the *a priori* probability that the patient has disease i before any evidence has been gathered, and $P(e|d_i)$ is the probability that a patient will have the complex of symptoms and signs represented by e, given that he or she has disease d_i.

We have so far ignored the complex problem of identifying the "relevant" data that should be gathered in order to diagnose the patient's disease. Evidence is actually acquired piece by piece, the necessary additional data being identified on the basis of the likely diagnoses at any given time. Diagnostic programs that mimic the process of analyzing evidence incrementally often use a modified version of Bayes' Theorem that is appropriate for sequential diagnosis (Gorry and Barnett, 1968):

> Let e_1 be the set of all observations to date, and s_1 be some new piece of data. Furthermore, let e be the new set of observations once s_1 has been added to e_1. Then:

$$P(d_i|e) = \frac{P(s_1|d_i\ \&\ e_1)\ P(d_i|e_1)}{\Sigma\ P(s_1|d_j\ \&\ e_1)\ P(d_j|e_1)}$$

The successful programs that use Bayes' Theorem in this form require huge amounts of statistical data, not only $P(s_k|d_j)$ for each of the pieces of

[2]Several programs have also included utility considerations in their analyses. For example, an unlikely but lethal disease that responds well to treatment may merit therapeutic attention because $P(d_i|e)$ is nonzero (although very small).

data, s_k, in e, but also the interrelationships of the s_k within each disease d_j.[3] The congenital heart disease programs (Gorry and Barnett, 1968; Warner et al., 1964) were able to acquire all the necessary conditional probabilities from a survey of several hundred patients with confirmed diagnoses and thus had nonjudgmental data on which to base their Bayesian analyses.

Edwards (1972, pp. 139–140) has summarized the kinds of problems that can arise when an attempt is made to gather the kinds of data needed for rigorous analysis:

> My friends who are expert about medical records tell me that to attempt to dig out from even the most sophisticated hospital's records the frequency of association between any particular symptom and any particular diagnosis is next to impossible—and when I raise the question of complexes of symptoms, they stop speaking to me. For another thing, doctors keep telling me that diseases change, that this year's flu is different from last year's flu, so that symptom-disease records extending far back in time are of very limited usefulness. Moreover, the observation of symptoms is well-supplied with error, and the diagnosis of diseases is even more so; both kinds of errors will ordinarily be frozen permanently into symptom-disease statistics. Finally, even if diseases didn't change, doctors would. The usefulness of disease categories is so much a function of available treatments that these categories themselves change as treatments change—a fact hard to incorporate into symptom-disease statistics.
>
> All these arguments against symptom-disease statistics are perhaps somewhat overstated. Where such statistics can be obtained and believed, obviously they should be used. But I argue that usually they cannot be obtained, and even in those instances where they have been obtained, they may not deserve belief.

An alternative to exhaustive data collection is to use the knowledge that an expert has about the disease—partly based on experience and partly on general principles—to reason about diagnoses. In the case of this judgmental knowledge acquired from experts, the conditional probabilities and their complex interrelationships cannot be acquired in an exhaustive manner. Opinions can be sought and attempts made to quantify them, but the extent to which the resulting numbers can be manipulated as probabilities is not clear. We shall explain this last point more fully as we proceed. First, let us examine some of the reasons that it might be desirable to construct a model that allows us to avoid the inherent problems of explicitly relating the conditional probabilities to one another.

A conditional probability statement is, in effect, a statement of a decision criterion or rule. For example, the expression $P(d_i|s_k)=x$ can be read as a statement that there is a $100x\%$ chance that a patient observed to have symptom s_k has disease d_i. Stated in rule form, it would be

---

[3]For example, although s_1 and s_2 are independent over all diseases, it may be true that s_1 and s_2 are closely linked for patients with disease d_i. Thus relationships must be known within *each* of the d_j; overall relationships are not sufficient.

IF: The patient has sign or symptom s_k
THEN: Conclude that he has disease d_i with probability x

We shall often refer to statements of conditional probability as decision rules or decision criteria in the diagnostic context. The value of x for such rules may not be obvious (e.g., "y strongly suggests that z is true" is difficult to quantify), but an expert may be able to offer an estimate of this number based on clinical experience and general knowledge, even when such numbers are not readily available otherwise.

A large set of such rules obtained from textbooks and experts would clearly contain a large amount of medical knowledge. It is conceivable that a computer program could be designed to consider all such general rules and to generate a final probability of each d_i based on data regarding a specific patient. Bayes' Theorem would only be appropriate for such a program, however, if values for $P(s_1|d_i)$ and $P(s_1|d_i \,\&\, s_2)$ could be obtained. As has been noted, these requirements become unworkable, even if the subjective probabilities of experts are used, in cases where a large number of diagnoses (hypotheses) must be considered. The first requires acquiring the inverse of every rule, and the second requires obtaining explicit statements regarding the interrelationships of all rules in the system.

In short, we would like to devise an approximate method that allows us to compute a value for $P(d_i|e)$ solely in terms of $P(d_i|s_k)$, where e is the composite of all the observed s_k. Such a technique will not be exact, but since the conditional probabilities reflect judgmental (and thus highly subjective) knowledge, a rigorous application of Bayes' Theorem will not necessarily produce accurate cumulative probabilities either. Instead, we look for ways to handle decision rules as discrete packets of knowledge and for a quantification scheme that permits accumulation of evidence in a manner that adequately reflects the reasoning process of an expert using the same or similar rules.

11.2 MYCIN's Rule-Based Approach

As has been discussed, MYCIN's principal task is to determine the likely identity of pathogens in patients with infections and to assist in the selection of a therapeutic regimen appropriate for treating the organisms under consideration. We have explained how MYCIN models the consultation process, utilizing judgmental knowledge acquired from experts in conjunction with certain statistical data that are available from the clinical microbiology laboratory and from patient records.

It is useful to consider the advantages provided by a rule-based system for computer use of judgmental knowledge. It should be emphasized that we see these advantages as being sufficiently strong in certain environments that we have devised an alternative and approximate approach that par-

allels the results available using Bayes' Theorem. We do not argue against the use of Bayes' Theorem in those medical environments in which sufficient data are available to permit its adequate use.

The advantages of rule-based systems for diagnostic consultations include:

1. the use of general knowledge (from textbooks or experts) for consideration of a specific patient (even well-indexed books may be difficult for a nonexpert to use when considering a patient whose problem is not quite the same as those of patients discussed in the text);
2. the use of judgmental knowledge for consideration of very small classes of patients with rare diseases about which good statistical data are not available;
3. ease of modification (since the rules are not explicitly related to one another and there need be no prestructured decision tree for such a system, rule modifications and the addition of new rules need not require complex considerations regarding interactions with the remainder of the system's knowledge);
4. facilitated search for potential inconsistencies and contradictions in the knowledge base (criteria stored explicitly in packets such as rules can be searched and compared without major difficulty);
5. straightforward mechanisms for explaining decisions to a user by identifying and communicating the relevant rules;
6. an augmented instructional capability (a system user may be educated regarding system knowledge in a selective fashion; i.e., only those portions of the decision process that are puzzling need be examined).

We shall use the following rule for illustrative purposes throughout this chapter:

IF: 1) The stain of the organism is gram positive, and
2) The morphology of the organism is coccus, and
3) The growth conformation of the organism is chains
THEN: There is suggestive evidence (.7) that the identity of the organism is streptococcus

This rule reflects our collaborating expert's belief that gram-positive cocci growing in chains are apt to be streptococci. When asked to weight his belief in this conclusion,[4] he indicated a 70% belief that the conclusion was valid. Translated to the notation of conditional probability, this rule seems

[4]In the English-language version of the rules, the program uses phrases such as "suggestive evidence," as in the above example. However, the numbers following these terms, indicating degrees of certainty, are all that is used in the model. The English phrases are not given by the expert and then quantified; they are, in effect, "canned-phrases" used only for translating rules into English representations. The prompt used for acquiring the certainty measure from the expert is as follows: "On a scale of 1 to 10, how much certainty do you affix to this conclusion?"

to say $P(h_1|s_1 \,\&\, s_2 \,\&\, s_3) = 0.7$ where h_1 is the hypothesis that the organism is a *Streptococcus*, s_1 is the observation that the organism is gram-positive, s_2 that it is a coccus, and s_3 that it grows in chains. Questioning of the expert gradually reveals, however, that despite the apparent similarity to a statement regarding a conditional probability, the number 0.7 differs significantly from a probability. The expert may well agree that $P(h_1|s_1 \,\&\, s_2 \,\&\, s_3) = 0.7$, but he becomes uneasy when he attempts to follow the logical conclusion that therefore $P(\neg h_1|s_1 \,\&\, s_2 \,\&\, s_3) = 0.3$. He claims that the three observations are evidence (to degree 0.7) *in favor* of the conclusion that the organism is a *Streptococcus* and should not be construed as evidence (to degree 0.3) *against Streptococcus*. We shall refer to this problem as Paradox 1 and return to it later in the exposition, after the interpretation of the 0.7 in the rule above has been introduced.

It is tempting to conclude that the expert is irrational if he is unwilling to follow the implications of his probabilistic statements to their logical conclusions. Another interpretation, however, is that the numbers he has given should not be construed as probabilities at all, that they are judgmental measures that reflect a level of *belief*. The nature of such numbers and the very existence of such concepts have interested philosophers of science for the last half-century. We shall therefore digress temporarily to examine some of these theoretical issues. We then proceed to a detailed presentation of the quantitative model we propose. In the last section of this chapter, we shall show how the model has been implemented for ongoing use by the MYCIN program.

11.3 Philosophical Background

The familiar *P*-function[5] of traditional probability theory is a straightforward concept from elementary statistics. However, because of imperfect knowledge and the dependence of decisions on individual judgments, the *P*-function no longer seems entirely appropriate for modeling some of the decision processes in medical diagnosis. This problem with the *P*-function has been well recognized and has generated several philosophical treatises

[5]The *P*-function may be defined in a variety of ways. Emanuel Parzen (1960) suggests a set-theoretical definition: Given a random situation, which is described by a sample description space s, probability is a function P that to every event e assigns a nonnegative real number, denoted by $P(e)$ and called the probability of the event e. The probability function must satisfy three axioms:

Axiom 1: $P(e) \geq 0$ for every event e;
Axiom 2: $P(s) = 1$ for the certain element s;
Axiom 3: $P(e \cup f) = P(e) + P(f)$ if $ef = 0$ or, in words, the probability of the union of two mutually exclusive events is the sum of their probabilities.

during the last 30 years. One difficulty with these analyses is that they are, in general, more theoretical than practical in orientation. They have characterized the problem well but have offered few quantitative or theoretical techniques that lend themselves to computer simulation of related reasoning processes. It is useful to examine these writings, however, in order to avoid recognized pitfalls.

This section therefore summarizes some of the theory that should be considered when analyzing the decision problem that we have described. We discuss several interpretations of probability itself, the theory on which Bayes' Theorem relies. The difficulties met when trying to use the *P*-function during the modeling of medical decision making are reiterated. Then we discuss the theory of confirmation, an approach to the interpretation of evidence. Our discussion argues that confirmation provides a natural environment in which to model certain aspects of medical reasoning. We then briefly summarize some other approaches to the problem, each of which has arisen in response to the inadequacies of applied probability. Although each of these alternate approaches is potentially useful in the problem area that concerns us, we have chosen to develop a quantification scheme based on the concept of confirmation.

11.3.1 Probability

Swinburne (1973) provides a useful classification of the theories of probability proposed over the last 200 years. The first of these, the Classical Theory of Probability, asserts that if the probability of an event is said to be *P*, then "there are integers *m* and *n* such that $P = m/n$ . . . such that *n* exclusive and exhaustive alternatives must occur, *m* of which constitute the occurrence of *s*." This theory, like the second and third to be described, is called "statistical probability" by Swinburne. These interpretations are typified by statements of the form "the probability of an *A* being a *B* is *P*."

The second probability theory cited by Swinburne, the Propensity Theory, asserts that probability propositions "make claims" about a propensity or "would-be" or tendency in things. If an atom is said to have a probability of 0.9 of disintegrating within the next minute, a statement has been made about its propensity to do so.

The Frequency Theory is based on the familiar claim that propositions about probability are propositions about proportions or relative frequencies as observed in the past. This interpretation provides the basis for the statistical data collection used by most of the Bayesian diagnostic programs.

Harré (1970) observes that statistical probability seems to differ syntactically from the sense of probability used in inference problems such as medical diagnosis. He points out that the traditional concept of probability refers to what is likely to turn out to be true (in the future), whereas the other variety of probability examines what has already turned out to be true but cannot be determined directly. Although these two kinds of prob-

lems may be approached on the basis of identical observations, the occurrence or nonoccurrence of future events is subject to the probabilistic analysis of statistics, whereas the verification of a belief, hypothesis, or conjecture concerning a truth in the present requires a "process" of analysis commonly referred to as *confirmation*. This distinction on the basis of tense may seem somewhat artificial at first, but it does serve a useful purpose as we attempt to develop a framework for analysis of the diagnosis problem.

Swinburne also discusses two more theories of probability, each of which bears more direct relation to the problem at hand. One is the Subjective Theory originally put forward by Ramsey (1931) and developed in particular by Savage (1974) and de Finetti (1972). In their view, statements of probability regarding an event are propositions regarding people's actual belief in the occurrence (present or future) of the event in question. Although this approach fails as an explanation of statistical probability (where beliefs that may be irrational have no bearing on the calculated probability of, say, a six being rolled on the next toss of a die), it is alluring for our purposes because it attempts to recognize the dependence of decisions, in certain problem areas, on both the weight of evidence and its interpretation as based on the expertise (beliefs) of the individual making the decision. In fact, de Finetti (1972, p. 4) has stated part of our problem explicitly:

> On many occasions decision-makers make use of expert opinion. Such opinions cannot possibly take the form of advice bearing directly on the decision; . . . . Occasionally, [the expert] is required to state a probability, but it is not easy to find a convenient form in which he can express it.

Furthermore, the goals of the subjective probabilists seem very similar to those which we have also delineated (de Finetti, 1972, p. 144):

> We hold it to be chimerical for anyone to arrive at beliefs, opinions, or determinations without the intervention of his personal judgment. We strive to make such judgments as dispassionate, reflective, and wise as possible by a doctrine which shows where and how they intervene and lays bare possible inconsistencies among judgments.

One way to acquire the subjective probabilities of experts is suggested by Savage and described by a geological analyst as follows (Grayson, 1960, p. 256):

> The simplest [way] is to ask the geologist. . . . The geologist looks at the evidence, thinks, and then gives a figure such as 1 in 5 or 50-50. Admittedly this is difficult. . . . Thus, several ways have been proposed to help the geologist make his probability estimate explicit. . . . The leading proponent of personal [i.e., subjective] probabilities, Savage, proposes what seems to be the most workable method. One can, namely, ask the person not how he feels

but what he would do in such and such a situation. Accordingly, a geologist would be confronted with a choice-making situation.

There is one principal problem to be faced, however, in attempting to adopt the subjectivist model for our computer program—namely, the subjectivists' criticism of those who avoid a Bayesian approach. Subjectivists assert that the conditional and initial probabilities needed for use of Bayes' Theorem may simply be acquired by asking the opinion of an expert. We must reject this approach when the number of decision criteria becomes large, however, because it would require that experts be asked to quantify an unmanageably large number of interrelationships.[6]

A final point to be made regarding subjectivist theory is that the probabilities so obtained are meant to be utilized by the *P*-function of statistical probability so that inconsistencies among the judgments offered by the experts may be discovered. Despite apparently irrational beliefs that may be revealed in this way ("irrational" here means that the subjective probabilities are inconsistent with the axioms of the *P*-function), the expert opinions provide useful criteria, which may lead to sound decisions if it is accepted that the numbers offered are not necessarily probabilities in the traditional sense. It is our assertion that a new quantitative system should therefore be devised in order to utilize the experts' criteria effectively.

Let us return now to the fifth and final category in Swinburne's list of probability theories (Swinburne, 1973). This is the Logical Theory, which gained its classical exposition in J. M. Keynes' *A Treatise on Probability* (1962). Since that time, its most notable proponent has been Rudolf Carnap. In the Logical Theory, probability is said to be a logical relation between statements of evidence and hypotheses. Carnap describes this and the frequency interpretation of probability as follows (Carnap, 1950, p. 19):

> (i) Probability$_1$ is the degree of confirmation of a hypothesis h with respect to an evidence statement e; e.g., an observational report. This is a logical semantical concept. A sentence about this concept is based, not on observation of facts, but on logical analysis. . . .
>
> (ii) Probability$_2$ is the relative frequency (in the long run) of one property of events or things with respect to another. A sentence about this concept is factual, empirical.

In order to avoid confusion regarding which concept of probability is being discussed, the term *probability* will hereafter be reserved for probability$_2$, i.e., the *P*-function of statistical probability. Probability$_1$, or epistemic probability as Swinburne (1973) describes it, will be called *degree of confirmation* in keeping with Carnap's terminology.

---

[6]It would also complicate the addition of new decision criteria since they would no longer be modular and would thus require itemization of all possible interactions with preexisting criteria.

11.3.2 Confirmation

Carnap's interpretation of confirmation rests upon strict logical entailment. Several authors, however, have viewed the subject in a broader context, such as our application requires. For example, just as the observation of a black raven would logically "confirm" the hypothesis that "all ravens are black" (where "confirm" means "lends credence to"), we also want the fact that an organism is gram-positive to "confirm" the hypothesis that it is a *Streptococcus*, even though the conclusion is based on world knowledge and not on logical analysis.

Carnap (1950) makes a useful distinction among three forms of confirmation, which we should consider when trying to characterize the needs of our decision model. He calls these classificatory, comparative, and quantitative uses of the concept of confirmation. These are easily understood by example:

a. classificatory: "the evidence e confirms the hypothesis h"
b. comparative: "e_1 confirms h more strongly than e_2 confirms h" or "e confirms h_1 more strongly than e confirms h_2"
c. quantitative: "e confirms h with strength x"

In MYCIN's task domain, we need to use a semiquantitative approach in order to reach a comparative goal. Thus, although our individual decision criteria might be quantitative (e.g., "gram-positive suggests *Streptococcus* with strength 0.1"), the effort is merely aimed at singling out two or three identities of organisms that are approximately equally likely and that are "comparatively" much more likely than any others. There is no need to quote a number that reflects the consulting expert's degree of certainty regarding his or her decisions.

When quantitative uses of confirmation are discussed, the degree of confirmation of hypothesis h on the basis of evidence e is written as $C[h,e]$. This form roughly parallels the familiar P-function notation for conditional probability, $P(h|e)$. Carnap has addressed the question of whether it is reasonable to quantify degree of confirmation (Carnap, 1950). He notes that, although the concept is familiar to us all, we attempt to use it for comparisons of relative likelihood rather than in a strict numerical sense. In his classic work on the subject, however, he suggested that we all know how to use confirmation as a quantitative concept in contexts such as "predictions of results of games of chance [where] we can determine which numerical value [others] implicitly attribute to probability$_1$, even if they do not state it explicitly, by observing their reactions to betting proposals." The reason for our reliance on the opinions of experts is reflected in his observation that individuals with experience are inclined to offer theoretical arguments to defend their viewpoint regarding a hypothesis; "this shows that they regard probability$_1$ as an objective concept." However, he

was willing to admit the subjective nature of such concepts some years later when, in discussing the nature of inductive reasoning, he wrote (Carnap, 1962, p. 317):

> I would think that inductive reasoning should lead, not to acceptance or rejection [of a proposition], but to the assignment of a number to the proposition, viz., its value (credibility value) . . . . This rational subjective probability . . . is sufficient for determining first the rational subjective value of any act, and then a rational decision.

As mentioned above, quantifying confirmation and then manipulating the numbers as though they were probabilities quickly leads to apparent inconsistencies or paradoxes. Carl Hempel presented an early analysis of confirmation (Hempel, 1965), pointing out as we have that $C[h,e]$ is a very different concept from $P(h|e)$. His famous Paradox of the Ravens was presented early in his discussion of the logic of confirmation. Let h_1 be the statement that "all ravens are black" and h_2 the statement that "all nonblack things are nonravens." Clearly h_1 is logically equivalent to h_2. If one were to draw an analogy with conditional probability, it might at first seem valid, therefore, to assert that $C[h_1,e] = C[h_2,e]$ for all e. However, it appears counterintuitive to state that the observation of a green vase supports h_1, even though the observation does seem to support h_2. $C[h,e]$ is therefore different from $P(h|e)$ for it seems somehow wrong that an observation of a vase could logically support an assertion about ravens.

Another characteristic of a quantitative approach to confirmation that distinguishes the concept from probability was well-recognized by Carnap (1950) and discussed by Barker (1957) and Harré (1970). They note that it is counterintuitive to suggest that the confirmation of the negation of a hypothesis is equal to one minus the confirmation of the hypothesis, i.e., $C[h,e]$ is not $1 - C[\neg h,e]$. The streptococcal decision rule asserted that a gram-positive coccus growing in chains is a *Streptococcus* with a measure of support specified as 7 out of 10. This translates to $C[h,e] = 0.7$ where h is "the organism is a *Streptococcus*" and e is the information that "the organism is a gram-positive coccus growing in chains." As discussed above, an expert does not necessarily believe that $C[\neg h,e] = 0.3$. The evidence is said to be *supportive* of the contention that the organism is a *Streptococcus* and can therefore hardly also support the contention that the organism is *not* a *Streptococcus*.

Since we believe that $C[h,e]$ does not equal $1 - C[\neg h,e]$, we recognize that disconfirmation is somehow separate from confirmation and must be dealt with differently. As Harré (1970) puts it, "we need an independently introduced D-function, for disconfirmation, because, as we have already noticed, to confirm something to ever so slight a degree is not to disconfirm it at all, since the favorable evidence for some hypothesis gives no support whatever to the contrary supposition in many cases." Our decision model

must therefore reflect this distinction between confirmation and disconfirmation (i.e., confirmatory and disconfirmatory evidence).

The logic of confirmation has several other curious properties that have puzzled philosophers of science (Salmon, 1973). Salmon's earlier analysis on the confirmation of scientific hypotheses (Salmon, 1966) led to the conclusion that the structure of such procedures is best expressed by Bayes' Theorem and a frequency interpretation of probability. Such an assertion is appealing because, as Salmon expresses the point, "it is through this interpretation, I believe, that we can keep our natural sciences empirical and objective." However, our model is not offered as a solution to the theoretical issues with which Salmon is centrally concerned. We have had to abandon Bayes' Theorem and the *P*-function simply because there are large areas of expert knowledge and intuition that, although amenable in theory to the frequency analysis of statistical probability, defy rigorous analysis because of insufficient data and, in a practical sense, because experts resist expressing their reasoning processes in coherent probabilistic terms.

11.3.3 Other Approaches

There are additional approaches to this problem area that bear mentioning, even though they are peripheral to confirmation and probability as we have described them. One is the *theory of fuzzy sets* first proposed by Zadeh (1965) and further developed by Goguen (1968). The theory attempts to analyze and explain an ancient paradox paraphrased by Goguen as follows:

> If you add one stone to a small heap, it remains small. A heap containing one stone is small. Therefore (by induction) every heap is small.

The term *fuzzy set* refers to the analogy with set theory whereby, for example, the set of tall people contains all 7-foot individuals but may or may not contain a man who is 5 feet 10 inches tall. The "tallness" of a man in that height range is subject to interpretation; i.e., the edge of the set is fuzzy. Thus, membership in a set is not binary-valued (true or false) but is expressed along a continuum from 0 to 1, where 0 means "not in the set," 1 means "in the set," and 0.5 means "equally likely to be in or out of the set." These numbers hint of statistical probability in much the same way that degrees of confirmation do. However, like confirmation, the theory of fuzzy sets leads to results that defy numerical manipulation in accordance with the axioms of the *P*-function. Although an analogy between our diagnostic problem and fuzzy set theory can be made, the statement of diagnostic decision criteria in terms of set membership does not appear to be a natural concept for the experts who must formulate our rules. Fur-

thermore, the quantification of Zadeh's "linguistic variables" and the mechanisms for combining them are as yet poorly defined. Fuzzy sets have therefore been mentioned here primarily as an example of another semi-statistical field in which classic probability theory fails.

There is also a large body of literature discussing the *theory of choice*, an approach to decision making that has been reviewed by Luce and Suppes (1965). The theory deals with the way in which personal preferences and the possible outcomes of an action are considered by an individual who must select among several alternatives. Tversky describes an approach based on "elimination by aspects" (Tversky, 1972), a method by which alternatives are ruled out on the basis of either their undesirable characteristics (aspects) or the desirable characteristics they lack. The theory thus combines preference (utility) with a probabilistic approach. Shackle suggests a similar approach (Shackle, 1952; 1955), but utilizes different terminology and focuses on the field of economics. He describes "expectation" as the act of "creating imaginary situations, of associating them with named future dates, and of assigning to each of the hypotheses thus formed a place on a scale measuring the degree of belief that a specified course of action on our own part will make this hypothesis come true" (Shackle, 1952). Selections among alternatives are made not only on the basis of likely outcomes but also on the basis of uncertainty regarding expected outcomes (hence his term the "logic of surprise").

Note that the theory of choice differs significantly from confirmation theory in that the former considers selection among mutually exclusive actions on the basis of their potential (future) outcomes and personal preferences regarding those outcomes, whereas confirmation considers selection among mutually exclusive hypotheses on the basis of evidence observed and interpreted in the present. Confirmation does not involve personal utilities, although, as we have noted, interpretation of evidence may differ widely on the basis of personal experience and knowledge. Thus we would argue that the theory of choice might be appropriately applied to the selection of therapy once a diagnosis is known, a problem area in which personal preferences regarding possible outcomes clearly play an important role, but that the formation of the diagnosis itself more closely parallels the kind of decision task that engendered the theory of confirmation.

We return, then, to confirmation theory as the most useful way to think about the medical decision-making problem that we have described. Swinburne suggests several criteria for choosing among the various confirmation theories that have been proposed (Swinburne, 1970), but his reasons are based more on theoretical considerations than on the pragmatics of our real-world application. We will therefore propose a technique that, although it draws closely on the theory of confirmation described above, is based on desiderata derived intuitively from the problem at hand and not from a formal list of acceptability criteria.

11.4 The Proposed Model of Evidential Strength

This section introduces our quantification scheme for modeling inexact medical reasoning. It begins by defining the notation that we use and describing the terminology. A formal definition of the quantification function is then presented. The remainder of the section discusses the characteristics of the defined functions.

Although the proposed model has several similarities to a confirmation function such as those mentioned above, we shall introduce new terms for the measurement of evidential strength. This convention will allow us to clarify from the outset that we seek only to devise a system that captures enough of the flavor of confirmation theory that it can be used for accomplishing our computer-based task. We have chosen *belief* and *disbelief* as our units of measurement, but these terms should not be confused with their formalisms from epistemology. The need for two measures was introduced above in our discussion of a disconfirmation measure as an adjunct to a measure for degree of confirmation. The notation will be as follows:

- MB[h,e] = x means "the measure of increased belief in the hypothesis h, based on the evidence e, is x"
- MD[h,e] = y means "the measure of increased disbelief in the hypothesis h, based on the evidence e, is y"

The evidence e need not be an observed event, but may be a hypothesis (itself subject to confirmation). Thus one may write MB[h_1,h_2] to indicate the measure of increased belief in the hypothesis h_1 given that the hypothesis h_2 is true. Similarly MD[h_1,h_2] is the measure of increased disbelief in hypothesis h_1 if hypothesis h_2 is true.

To illustrate in the context of the sample rule from MYCIN, consider e = "the organism is a gram-positive coccus growing in chains" and h = "the organism is a *Streptococcus*." Then MB[h,e] = 0.7 according to the sample rule given us by the expert. The relationship of the number 0.7 to probability will be explained as we proceed. For now, let us simply state that the number 0.7 reflects the extent to which the expert's belief that h is true is increased by the knowledge that e is true. On the other hand, MD[h,e] = 0 for this example; i.e., the expert has no reason to increase his or her disbelief in h on the basis of e.

In accordance with subjective probability theory, it may be argued that the expert's personal probability $P(h)$ reflects his or her belief in h at any given time. Thus $1 - P(h)$ can be viewed as an estimate of the expert's disbelief regarding the truth of h. If $P(h|e)$ is greater than $P(h)$, the observation of e increases the expert's belief in h while decreasing his or her

disbelief regarding the truth of h. In fact, the proportionate decrease in disbelief is given by the following ratio:

$$\frac{P(h|e) - P(h)}{1 - P(h)}$$

This ratio is called the measure of increased belief in h resulting from the observation of e, i.e., MB[h,e].

Suppose, on the other hand, that $P(h|e)$ were less than $P(h)$. Then the observation of e would decrease the expert's belief in h while increasing his or her disbelief regarding the truth of h. The proportionate decrease in belief in this case is given by the following ratio:

$$\frac{P(h) - P(h|e)}{P(h)}$$

We call this ratio the measure of increased disbelief in h resulting from the observation of e, i.e., MD[h,e].

To summarize these results in words, we consider the measure of increased belief, MB[h,e], to be the proportionate decrease in disbelief regarding the hypothesis h that results from the observation e. Similarly, the measure of increased disbelief, MD[h,e], is the proportionate decrease in belief regarding the hypothesis h that results from the observation e, where belief is estimated by $P(h)$ at any given time and disbelief is estimated by $1 - P(h)$. These definitions correspond closely to the intuitive concepts of confirmation and disconfirmation that we have discussed above. Note that since one piece of evidence cannot both favor and disfavor a single hypothesis, when MB[h,e] > 0, MD[h,e] $= 0$, and when MD[h,e] > 0, MB[h,e] $= 0$. Furthermore, when $P(h|e) = P(h)$, the evidence is independent of the hypothesis (neither confirms nor disconfirms) and MB[h,e] $=$ MD[h,e] $= 0$.

The above definitions may now be specified formally in terms of conditional and *a priori* probabilities:

$$\text{MB}[h,e] = \begin{cases} 1 & \text{if } P(h) = 1 \\ \dfrac{\max[P(h|e),P(h)] - P(h)}{\max[1,0] - P(h)} & \text{otherwise} \end{cases}$$

$$\text{MD}[h,e] = \begin{cases} 1 & \text{if } P(h) = 0 \\ \dfrac{\min[P(h|e),P(h)] - P(h)}{\min[1,0] - P(h)} & \text{otherwise} \end{cases}$$

Examination of these expressions will reveal that they are identical to the definitions introduced above. The formal definition is introduced, however, to demonstrate the symmetry between the two measures. In addition, we define a third measure, termed a *certainty factor* (CF), that combines the MB and MD in accordance with the following definition:

$$CF[h,e] = MB[h,e] - MD[h,e]$$

The certainty factor is an artifact for combining degrees of belief and disbelief into a single number. Such a number is needed in order to facilitate comparisons of the evidential strength of competing hypotheses. The use of this composite number will be described below in greater detail. The following observations help to clarify the characteristics of the three measures that we have defined (MB, MD, CF):

Characteristics of the Belief Measures

1. Range of degrees:

a. $0 \leq MB[h,e] \leq 1$

b. $0 \leq MD[h,e] \leq 1$

c. $-1 \leq CF[h,e] \leq +1$

2. Evidential strength and mutually exclusive hypotheses:
If h is shown to be certain $[P(h|e) = 1]$:

a. $MB[h,e] = \dfrac{1 - P(h)}{1 - P(h)} = 1$

b. $MD[h,e] = 0$

c. $CF[h,e] = 1$

If the negation of h is shown to be certain $[P(\neg h|e) = 1]$:

a. $MB[h,e] = 0$

b. $MD[h,e] = \dfrac{0 - P(h)}{0 - P(h)} = 1$

c. $CF[h,e] = -1$

Note that this gives $MB[\neg h,e] = 1$ if and only if $MD[h,e] = 1$ in accordance with the definitions of MB and MD above. Furthermore, the number 1 represents absolute belief (or disbelief) for MB (or MD).Thus if $MB[h_1,e] = 1$ and h_1 and h_2 are mutually exclusive, $MD[h_2,e] = 1$.[7]

---

[7]There is a special case of Characteristic 2 that should be mentioned. This is the case of logical truth or falsity where $P(h|e) = 1$ or $P(h|e) = 0$, regardless of e. Popper has also suggested a quantification scheme for confirmation (Popper, 1959) in which he uses $-1 \leq C[h,e] \leq +1$, defining his limits as:

$$-1 = C[\neg h,h] \leq C[h,e] \leq C[h,h] = +1$$

This proposal led one observer (Harré, 1970) to assert that Popper's numbering scheme "obliges one to identify the truth of a self-contradiction with the falsity of a disconfirmed general hypothesis and the truth of a tautology with the confirmation of a confirmed existential hypothesis, both of which are not only question begging but absurd." As we shall demonstrate, we avoid Popper's problem by introducing mechanisms for approaching certainty asymptotically as items of confirmatory evidence are discovered.

3. Lack of evidence:
   a. MB[h,e] = 0 if h is not confirmed by e (i.e., e and h are independent or e disconfirms h)
   b. MD[h,e] = 0 if h is not disconfirmed by e (i.e., e and h are independent or e confirms h)
   c. CF[h,e] = 0 if e neither confirms nor disconfirms h (i.e., e and h are independent)

We are now in a position to examine Paradox 1, the expert's concern that although evidence may support a hypothesis with degree x, it does not support the negation of the hypothesis with degree $1 - x$. In terms of our proposed model, this reduces to the assertion that, when e confirms h:

$$\text{CF}[h,e] + \text{CF}[\neg h,e] \neq 1$$

This intuitive impression is verified by the following analysis for e confirming h:

$$\begin{aligned}
\text{CF}[\neg h,e] &= \text{MB}[\neg h,e] - \text{MD}[\neg h,e] \\
&= 0 - \frac{P(\neg h|e) - P(\neg h)}{-P(\neg h)} \\
&= \frac{[1 - P(h|e)] - [1 - P(h)]}{1 - P(h)} = \frac{P(h) - P(h|e)}{1 - P(h)} \\
\text{CF}[h,e] &= \text{MB}[h,e] - \text{MD}[h,e] \\
&= \frac{P(h|e) - P(h)}{1 - P(h)} - 0
\end{aligned}$$

Thus

$$\begin{aligned}
\text{CF}[h,e] + \text{CF}[\neg h,e] &= \frac{P(h|e) - P(h)}{1 - P(h)} + \frac{P(h) - P(h|e)}{1 - P(h)} \\
&= 0
\end{aligned}$$

Clearly, this result occurs because (for any h and any e) MB[h,e] = MD[$\neg h$,e]. This conclusion is intuitively appealing since it states that evidence that supports a hypothesis disfavors the negation of the hypothesis to an equal extent.

We noted earlier that experts are often willing to state degrees of belief in terms of conditional probabilities but they refuse to follow the assertions to their logical conclusions (e.g., Paradox 1 above). It is perhaps revealing to note, therefore, that when the *a priori* belief in a hypothesis is small (i.e.,

$P(h)$ is close to zero), the CF of a hypothesis confirmed by evidence is approximately equal to its conditional probability on that evidence:

$$\mathrm{CF}[h,e] = \mathrm{MB}[h,e] - \mathrm{MD}[h,e] = \frac{P(h|e) - P(h)}{1 - P(h)} - 0 \approx P(h|e)$$

whereas, as shown above, $\mathrm{CF}[\neg h,e] = -P(h|e)$ in this case. This observation suggests that confirmation, to the extent that it is adequately represented by CF's, is close to conditional probability (in certain cases), although it still defies analysis as a probability measure.

We believe, then, that the proposed model is a plausible representation for the numbers an expert gives when asked to quantify the strength of his or her judgmental rules. The expert gives a positive number ($\mathrm{CF} > 0$) if the hypothesis is confirmed by observed evidence, suggests a negative number ($\mathrm{CF} < 0$) if the evidence lends credence to the negation of the hypothesis, and says there is no evidence at all ($\mathrm{CF} = 0$) if the observation is independent of the hypothesis under consideration. The CF combines knowledge of both $P(h)$ and $P(h|e)$. Since the expert often has trouble stating $P(h)$ and $P(h|e)$ in quantitative terms, there is reason to believe that a CF that weights both the numbers into a single measure is actually a more natural intuitive concept (e.g., "I don't know what the probability is that all ravens are black, but I *do* know that every time you show me an additional black raven my belief is increased by x that all ravens are black.").

If we therefore accept CF's rather than probabilities from experts, it is natural to ask under what conditions the physician's behavior based on CF's is irrational.[8] We know from probability theory, for example, that if there are n mutually exclusive hypotheses h_i, at least one of which must be true, then $\Sigma^n P(h_i|e) = 1$ for all e. In the case of certainty factors, we can also show that there are limits on the sums of CF's of mutually exclusive hypotheses. Judgmental rules acquired from experts must respect these limits or else the rules will reflect irrational quantitative assignments.

Sums of CF's of mutually exclusive hypotheses have two limits—a lower limit for disconfirmed hypotheses and an upper limit for confirmed hypotheses. The lower limit is the obvious value that results because $\mathrm{CF}[h,e] \geq -1$ and because more than one hypothesis may have $\mathrm{CF} = -1$. Note first that a single piece of evidence may absolutely disconfirm several of the competing hypotheses. For example, if there are n colors in the universe and C_i is the ith color, then ARC_i may be used as an informal notation to denote the hypothesis that all ravens have color C_i. If we add the hypothesis ARC_0 that some ravens have different colors from others, we know $\Sigma_0^n P(\mathrm{ARC}_i) = 1$. Consider now the observation e that there is a raven of color C_n. This single observation allows us to conclude that $\mathrm{CF}[\mathrm{ARC}_i,e] = -1$ for $1 \leq i \leq n - 1$. Thus, since these $n - 1$ hypotheses

[8]We assert that behavior is irrational if actions taken or decisions made contradict the result that would be obtained under a probabilistic analysis of the behavior.

are absolutely disconfirmed by the observation e, Σ_1^{n-1} CF[ARC$_i$,e] = $-(n-1)$. This analysis leads to the general statement that, if k mutually exclusive hypotheses h_i are disconfirmed by an observation e:

$$\Sigma_1^k \text{ CF}[h_i,e] \geq -k \text{ [for } h_i \text{ disconfirmed by } e]$$

In the colored raven example, the observation of a raven with color C_n still left two hypotheses in contention, namely ARC_n and ARC_0. What, then, are CF[ARC_n,e], CF[ARC_0,e], and the sum of CF[ARC_n,e] and CF[ARC_0,e]? It can be shown that, if k mutually exclusive hypotheses h_i are confirmed by an observation e, the sum of their CF's does not have an upper limit of k but rather:

$$\Sigma^k \text{ CF}[h_i,e] \leq 1 \text{ [for } h_i \text{ confirmed by } e]$$

In fact, Σ^k CF[h_i,e] is *equal* to 1 if and only if $k = 1$ *and* e implies h_1 with certainty, but the sum can get arbitrarily close to 1 for small k and large n. The analyses that lead to these conclusions are available elsewhere (Shortliffe, 1974).

The last result allows us to analyze critically new decision rules given by experts. Suppose, for example, we are given the following rules: CF[h_1,e] = 0.7 and CF[h_2,e] = 0.4, where h_1 is "the organism is a *Streptococcus*," h_2 is "the organism is a *Staphylococcus*," and e is "the organism is a gram-positive coccus growing in chains." Since h_1 and h_2 are mutually exclusive, the observation that Σ CF[h_i,e] > 1 tells us that the suggested certainty factors are inappropriate. The expert must either adjust the weightings, or we must normalize them so that their sum does not exceed 1. Because behavior based on these rules would be irrational, we must change the rules.

11.5 The Model as an Approximation Technique

Certainty factors provide a useful way to think about confirmation and the quantification of degrees of belief. However, we have not yet described how the CF model can be usefully applied to the medical diagnosis problem. The remainder of this chapter will explain conventions that we have introduced in order to use the certainty factor model. Our starting assumption is that the numbers given us by experts who are asked to quantify their degree of belief in decision criteria are adequate approximations to the numbers that would be calculated in accordance with the definitions of MB and MD if the requisite probabilities were known.

When we discussed Bayes' Theorem earlier, we explained that we would like to devise a method that allows us to approximate the value for $P(d_i|e)$ solely from the $P(d_i|s_k)$, where d_i is the ith possible diagnosis, s_k is the

kth clinical observation, and e is the composite of all the observed s_k. This goal can be rephrased in terms of certainty factors as follows:

> Suppose that MB[d_i,s_k] is known for each s_k, MD[d_i,s_k] is known for each s_k, and e represents the conjunction of all the s_k. Then our goal is to calculate CF[d_i,e] from the MB's and MD's known for the individual s_k's.

Suppose that $e = s_1$ & s_2 and that e confirms d_i. Then:

$$\mathrm{CF}[d_i,e] = \mathrm{MB}[d_i,e] - 0 = \frac{P(d_i|e) - P(d_i)}{1 - P(d_i)}$$

$$= \frac{P(d_i|s_1 \;\&\; s_2) - P(d_i)}{1 - P(d_i)}$$

There is no exact representation of CF[d_i,s_1 & s_2] purely in terms of CF[d_i,s_1] and CF[d_i,s_2]; the relationship of s_1 to s_2, within d_i and all other diagnoses, needs to be known in order to calculate $P(d_i|s_1 \;\&\; s_2)$. Furthermore, the CF scheme adds one complexity not present with Bayes' Theorem because we are forced to keep MB's and MD's isolated from one another. Suppose s_1 confirms d_i (MB > 0) but s_2 disconfirms d_i (MD > 0). Then consider CF[d_i,s_1 & s_2]. In this case, CF[d_i,s_1 & s_2] must reflect both the disconfirming nature of s_2 and the confirming nature of s_1. Although these measures are reflected in the component CF's (it is intuitive in this case, for example, that CF[d_i,s_2] $\leq$ CF[d_i,s_1 & s_2] $\leq$ CF[d_i,s_1]), we shall demonstrate that it is important to handle component MB's and MD's separately in order to preserve commutativity (see Item 3 of the list of defining criteria below). We have therefore developed an approximation technique for handling the net evidential strength of incrementally acquired observations. The combining convention must satisfy the following criteria (where $e+$ represents all confirming evidence acquired to date, and $e-$ represents all disconfirming evidence acquired to date):

Defining Criteria

1. Limits:

   a. MB[h,$e+$] increases toward 1 as confirming evidence is found, equaling 1 if and only if a piece of evidence logically implies h with certainty

   b. MD[h,$e-$] increases toward 1 as disconfirming evidence is found, equaling 1 if and only if a piece of evidence logically implies $\neg h$ with certainty

   c. CF[h,$e-$] $\leq$ CF[h,$e-$ & $e+$] $\leq$ CF[h,$e+$]

These criteria reflect our desire to have the measure of belief approach certainty asymptotically as partially confirming evidence is acquired, and to have the measure of disbelief approach certainty asymptotically as partially disconfirming evidence is acquired.

2. Absolute confirmation or disconfirmation:

a. If $MB[h,e+] = 1$, then $MD[h,e-] = 0$ regardless of the disconfirming evidence in $e-$; i.e., $CF[h,e+] = 1$

b. If $MD[h,e-] = 1$, then $MB[h,e+] = 0$ regardless of the confirming evidence in $e+$; i.e., $CF[h,e-] = -1$

c. The case where $MB[h,e+] = MD[h,e-] = 1$ is contradictory and hence the CF is undefined

3. Commutativity:

If s_1 & s_2 indicates an ordered observation of evidence, first s_1 and then s_2:

a. $MB[h,s_1 \ \& \ s_2] = MB[h,s_2 \ \& \ s_1]$

b. $MD[h,s_1 \ \& \ s_2] = MD[h,s_2 \ \& \ s_1]$

c. $CF[h,s_1 \ \& \ s_2] = CF[h,s_2 \ \& \ s_1]$

The order in which pieces of evidence are discovered should not affect the level of belief or disbelief in a hypothesis. These criteria assure that the order of discovery will not matter.

4. Missing information:

If $s_?$ denotes a piece of potential evidence, the truth or falsity of which is unknown:

a. $MB[h,s_1 \ \& \ s_?] = MB[h,s_1]$

b. $MD[h,s_1 \ \& \ s_?] = MD[h,s_1]$

c. $CF[h,s_1 \ \& \ s_?] = CF[h,s_1]$

The decision model should function by simply disregarding rules of the form $CF[h,s_2] = x$ if the truth or falsity of s_2 cannot be determined.

A number of observations follow from these criteria. For example, Items 1 and 2 indicate that the MB of a hypothesis never decreases unless its MD goes to 1. Similarly, the MD never decreases unless the MB goes to 1. As evidence is acquired sequentially, both the MB and MD may become nonzero. Thus CF = MB − MD is an important indicator of the *net* belief in a hypothesis in light of current evidence. Furthermore, a certainty factor of zero may indicate either the absence of both confirming and disconfirm-

ing evidence (MB = MD = 0) or the observation of pieces of evidence that are equally confirming and disconfirming (MB = MD, where each is nonzero). Negative CF's indicate that there is more reason to disbelieve the hypothesis than to believe it. Positive CF's indicate that the hypothesis is more strongly confirmed than disconfirmed.

It is important also to note that, if $e = e+ \ \& \ e-$, then CF[h,e] represents the certainty factor for a complex new rule that could be given us by an expert. CF[h,e], however, would be a highly specific rule customized for the few patients satisfying *all* the conditions specified in $e+$ and $e-$. Since the expert gives us only the component rules, we seek to devise a mechanism whereby a calculated cumulative CF[h,e], based on MB[$h,e+$] and MD[$h,e-$], gives a number close to the CF[h,e] that would be calculated if all the necessary conditional probabilities were known.

The first of the following four combining functions satisfies the criteria that we have outlined. The other three functions are necessary conventions for implementation of the model.

Combining Functions

1. Incrementally acquired evidence:

$$\text{MB}[h,s_1 \ \& \ s_2] = \begin{cases} 0 & \text{if MD}[h,s_1 \ \& \ s_2] = 1 \\ \text{MB}[h,s_1] + \text{MB}[h,s_2](1 - \text{MB}[h,s_1]) & \text{otherwise} \end{cases}$$

$$\text{MD}[h,s_1 \ \& \ s_2] = \begin{cases} 0 & \text{if MB}[h,s_1 \ \& \ s_2] = 1 \\ \text{MD}[h,s_1] + \text{MD}[h,s_2](1 - \text{MD}[h,s_1]) & \text{otherwise} \end{cases}$$

2. Conjunctions of hypotheses:

$$\text{MB}[h_1 \ \& \ h_2,e] = \min(\text{MB}[h_1,e], \text{MB}[h_2,e])$$
$$\text{MD}[h_1 \ \& \ h_2,e] = \max(\text{MD}[h_1,e], \text{MD}[h_2,e])$$

3. Disjunctions of hypotheses:

$$\text{MB}[h_1 \text{ or } h_2,e] = \max(\text{MB}[h_1,e], \text{MB}[h_2,e])$$
$$\text{MD}[h_1 \text{ or } h_2,e] = \min(\text{MD}[h_1,e], \text{MD}[h_2,e])$$

4. Strength of evidence:

If the truth or falsity of a piece of evidence s_1 is not known with certainty, but a CF (based on prior evidence e) is known reflecting the degree of belief in s_1, then if MB′[h,s_1] and MD′[h,s_1] are the degrees

of belief and disbelief in h when s_1 is known to be true with certainty (i.e., these are the decision rules acquired from the expert) then the actual degrees of belief and disbelief are given by:

$$\text{MB}[h,s_1] = \text{MB}'[h,s_1] \cdot \max(0, \text{CF}[s_1,e])$$
$$\text{MD}[h,s_1] = \text{MD}'[h,s_1] \cdot \max(0, \text{CF}[s_1,e])$$

This criterion relates to our previous statement that evidence in favor of a hypothesis may itself be a hypothesis subject to confirmation. Suppose, for instance, you are in a darkened room when testing the generalization that all ravens are black. Then the observation of a raven that you think is black, but that may be navy blue or purple, is less strong evidence in favor of the hypothesis that all ravens are black than if the sampled raven were known with certainty to be black. Here the hypothesis being tested is "*all* ravens are black," and the evidence is itself a hypothesis, namely the uncertain observation "*this* raven is black."

Combining Function 1 simply states that, since an MB (or MD) represents a proportionate decrease in disbelief (or belief), the MB (or MD) of a newly acquired piece of evidence should be applied proportionately to the disbelief (or belief) still remaining. Combining Function 2a indicates that the measure of belief in the conjunction of two hypotheses is only as good as the belief in the hypothesis that is believed less strongly, whereas Combining Function 2b indicates that the measure of disbelief in such a conjunction is as strong as the disbelief in the most strongly disconfirmed. Combining Function 3 yields complementary results for disjunctions of hypotheses. The corresponding CF's are merely calculated using the definition CF = MB − MD. Readers are left to satisfy themselves that Combining Function 1 satisfies the defining criteria.[9]

Combining Functions 2 and 3 are needed in the use of Combining Function 4. Consider, for example, a rule such as:

$$\text{CF}'[h,s_1 \;\&\; s_2 \;\&\; (s_3 \text{ or } s_4)] = x$$

Then, by Combining Function 4:

$$\begin{aligned}\text{CF}[h,s_1 \;\&\; s_2 \;\&\; (s_3 \text{ or } s_4)] &= x \cdot \max(0,\text{CF}[s_1 \;\&\; s_2 \;\&\; (s_3 \text{ or } s_4),e]) \\ &= x \cdot \max(0,\text{MB}[s_1 \;\&\; s_2 \;\&\; (s_3 \text{ or } s_4),e] \\ &\qquad - \text{MD}[s_1 \;\&\; s_2 \;\&\; (s_3 \text{ or } s_4),e])\end{aligned}$$

[9]Note that $\text{MB}[h,s_?] = \text{MD}[h,s_?] = 0$ when examining Criterion 4.

Thus we use Combining Functions 2 and 3 to calculate:

$$\begin{aligned} \text{MB}[s_1 \ \& \ s_2 \ \& \ (s_3 \text{ or } s_4),e] &= \min(\text{MB}[s_1,e], \text{MB}[s_2,e], \text{MB}[s_3 \text{ or } s_4,e]) \\ &= \min(\text{MB}[s_1,e], \text{MB}[s_2,e], \\ &\qquad \max(\text{MB}[s_3,e], \text{MB}[s_4,e])) \end{aligned}$$

$\text{MD}[s_1 \ \& \ s_2 \ \& \ (s_3 \text{ or } s_4),e]$ is calculated similarly.

An analysis of Combining Function 1 in light of the probabilistic definitions of MB and MD does not prove to be particularly enlightening. The assumptions implicit in this function include more than an acceptance of the independence of s_1 and s_2. The function was conceived purely on intuitive grounds in that it satisfied the four defining criteria listed. However, some obvious problems are present. For example, the function always causes the MB or MD to increase, regardless of the relationship between new and prior evidence. Yet Salmon has discussed an example from subparticle physics (Salmon, 1973) in which either of two observations taken alone confirms a given hypothesis, but their conjunction disproves the hypothesis absolutely! Our model assumes the absence of such aberrant situations in the field of application for which it is designed. The problem of formulating a more general quantitative system for measuring confirmation is well recognized and referred to by Harré (1970): "The syntax of confirmation has nothing to do with the logic of probability in the numerical sense, and it seems very doubtful if any single, general notion of confirmation can be found which can be used in all or even most scientific contexts." Although we have suggested that perhaps there *is* a numerical relationship between confirmation and probability, we agree that the challenge for a confirmation quantification scheme is to demonstrate its usefulness within a given context, preferably without sacrificing human intuition regarding what the quantitative nature of confirmation should be.

Our challenge with Combining Function 1, then, is to demonstrate that it is a close enough approximation for our purposes. We have attempted to do so in two ways. First, we have implemented the function as part of the MYCIN system (Section 11.6) and have demonstrated that the technique models the conclusions of the expert from whom the rules were acquired. Second, we have written a program that allows us to compare CF's computed both from simulated real data and by using Combining Function 1. Our notation for the following discussion will be as follows:

$\text{CF}^*[h,e]$ = the computed CF using the definition of CF from Section 11.4 (i.e., "perfect knowledge" since $P(h|e)$ and $P(h)$ are known)

$\text{CF}[h,e]$ = the computed CF using Combining Function 1 and the known MB's and MD's for each s_k where e is the composite of the s_k's (i.e., $P(h|e)$ not known, but $P(h|s_k)$ and $P(h)$ known for calculation of $\text{MB}[h,s_k]$ and $\text{MD}[h,s_k]$)

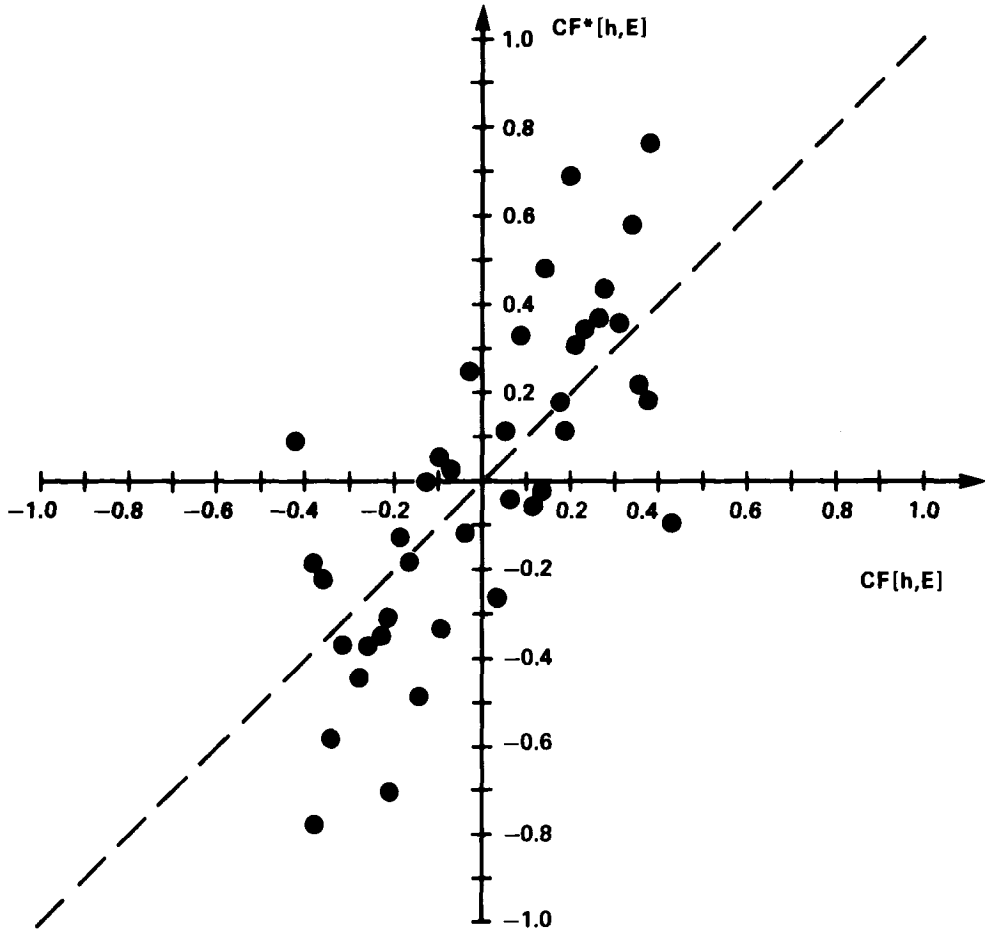

FIGURE 11-1 Chart demonstrating the degree of agreement between CF and CF* for a sample data base. CF is an approximation of CF*. The terms are defined in the text.

The program was run on sample data simulating several hundred patients. The question to be asked was whether CF[*h*,*e*] is a good approximation to CF*[*h*,*e*]. Figure 11-1 is a graph summarizing our results. For the vast majority of cases, the approximation does not produce a CF[*h*,*e*] radically different from the true CF*[*h*,*e*]. In general, the discrepancy is greatest when Combining Function 1 has been applied several times (i.e., several pieces of evidence have been combined). The most aberrant points, however, are those that represent cases in which pieces of evidence were strongly interrelated for the hypothesis under consideration (termed *con-*

ditional nonindependence). This result is expected because it reflects precisely the issue that makes it difficult to use Bayes' Theorem for our purposes.

Thus we should emphasize that we have not avoided many of the problems inherent with the use of Bayes' Theorem in its exact form. We have introduced a new quantification scheme, which, although it makes many assumptions similar to those made by subjective Bayesian analysis, permits us to use criteria as rules and to manipulate them to the advantages described earlier. In particular, the quantification scheme allows us to consider confirmation separately from probability and thus to overcome some of the inherent problems that accompany an attempt to put judgmental knowledge into a probabilistic format. Just as Bayesians who use their theory wisely must insist that events be chosen so that they are independent (unless the requisite conditional probabilities are known), we must insist that dependent pieces of evidence be grouped into single rather than multiple rules. As Edwards (1972) has pointed out, a similar strategy must be used by Bayesians who are unable to acquire all the necessary data:

> An approximation technique is the one now most commonly used. It is simply to combine conditionally non-independent symptoms into one grand symptom, and obtain [quantitative] estimates for that larger more complex symptom.

The system therefore becomes unworkable for applications in which large numbers of observations must be grouped in the premise of a single rule in order to ensure independence of the decision criteria. In addition, we must recognize logical subsumption when examining or acquiring rules and thus avoid counting evidence more than once. For example, if s_1 implies s_2, then CF[h,s_1 & s_2] = CF[h,s_1] regardless of the value of CF[h,s_2]. Function 1 does not "know" this. Rules must therefore be acquired and utilized with care. The justification for our approach therefore rests not with a claim of improving on Bayes' Theorem but rather with the development of a mechanism whereby judgmental knowledge can be efficiently represented and utilized for the modeling of medical decision making, especially in contexts where (a) statistical data are lacking, (b) inverse probabilities are not known, and (c) conditional independence can be assumed in most cases.

11.6 MYCIN's Use of the Model

Formal quantification of the probabilities associated with medical decision making can become so frustrating that some investigators have looked for ways to dispense with probabilistic information altogether (Ledley, 1973). Diagnosis is not a deterministic process, however, and we believe that it

should be possible to develop a quantification technique that approximates probability and Bayesian analysis and that is appropriate for use in those cases where formal analysis is difficult to achieve. The certainty factor model that we have introduced is such a scheme. The MYCIN program uses certainty factors to accumulate evidence and to decide on likely identities for organisms causing disease in patients with bacterial infections. A therapeutic regimen is then determined—one that is appropriate to cover for the organisms requiring therapy.

MYCIN remembers the alternate hypotheses that are confirmed or disconfirmed by the rules for inferring an organism's identity. With each hypothesis is stored its MB and MD, both of which are initially zero. When a rule for inferring identity is found to be true for the patient under consideration, the action portion of the rule allows either the MB or the MD of the relevant hypothesis to be updated using Combining Function 1. When all applicable rules have been executed, the final CF may be calculated, for each hypothesis, using the definition CF = MB − MD. These alternate hypotheses may then be compared on the basis of their cumulative certainty factors. Hypotheses that are most highly confirmed thus become the basis of the program's therapeutic recommendation.

Suppose, for example, that the hypothesis h_1 that the organism is a *Streptococcus* has been confirmed by a single rule with a CF = 0.3. Then, if e represents all evidence to date, MB[h_1,e] = 0.3 and MD[h_1,e] = 0. If a new rule is now encountered that has CF = 0.2 in support of h_1, and if e is updated to include the evidence in the premise of the rule, we now have MB[h_1,e] = 0.44 and MD[h_1,e] = 0. Suppose a final rule is encountered for which CF = −0.1. Then if e is once again updated to include all current evidence, we use Function 1 to obtain MB[h_1,e] = 0.44 and MD[h_1,e] = 0.1. If no further system knowledge allows conclusions to be made regarding the possibility that the organism is a *Streptococcus*, we calculate a final result, CF[h_1,e] = 0.44 − 0.1 = 0.34. This number becomes the basis for comparison between h_1 and all the other possible hypotheses regarding the identity of the organism.

It should be emphasized that this same mechanism is used for evaluating *all* knowledge about the patient, not just the identity of pathogens. When a user answers a system-generated question, the associated certainty factor is assumed to be +1 unless he or she explicitly modifies the response with a CF (multiplied by ten) enclosed in parentheses. Thus, for example, the following interaction might occur (MYCIN's question is in lower-case letters):

14) Did the organism grow in clumps, chains, or pairs?
** CHAINS (6) PAIRS (3) CLUMPS (−8)

This capability allows the system automatically to incorporate the user's uncertainties into its decision processes. A rule that referenced the growth conformation of the organism would in this case find:

$$\begin{array}{llll} \text{MB[chains,}e\text{]} & = 0.6 & \text{MD[chains,}e\text{]} & = 0 \\ \text{MB[pairs,}e\text{]} & = 0.3 & \text{MD[pairs,}e\text{]} & = 0 \\ \text{MB[clumps,}e\text{]} & = 0 & \text{MD[clumps,}e\text{]} & = 0.8 \end{array}$$

Consider, then, the sample rule:

$$\text{CF}[h_1, s_1 \ \& \ s_2 \ \& \ s_3] = 0.7$$

where h_1 is the hypothesis that the organism is a *Streptococcus*, s_1 is the observation that the organism is gram-positive, s_2 that it is a coccus, and s_3 that it grows in chains. Suppose gram stain and morphology were known to the user with certainty, so that MYCIN has recorded:

$$\text{CF}[s_1, e] = 1 \qquad \text{CF}[s_2, e] = 1$$

In the case above, however, MYCIN would find that

$$\text{CF[chains,}e\text{]} = \text{CF}[s_3, e] = 0.6 - 0 = 0.6$$

Thus it is no longer appropriate to use the rule in question with its full confirmatory strength of 0.7. That CF was assigned by the expert on the assumption that all three conditions in the premise would be true with certainty. The modified CF is calculated using Combining Function 4:

$$\begin{aligned} \text{CF}[h_1, s_1 \ \& \ s_2 \ \& \ s_3] &= \text{MB}[h_1, s_1 \ \& \ s_2 \ \& \ s_3] - \text{MD}[h_1, s_1 \ \& \ s_2 \ \& \ s_3] \\ &= 0.7 \cdot \max(0, \text{CF}[s_1 \ \& \ s_2 \ \& \ s_3, e]) - 0 \end{aligned}$$

Calculating $\text{CF}[s_1 \ \& \ s_2 \ \& \ s_3, e]$ using Combining Function 2 gives:

$$\begin{aligned} \text{CF}[h_1, s_1 \ \& \ s_2 \ \& \ s_3] &= (0.7)\,(0.6) - 0 \\ &= 0.42 - 0 \end{aligned}$$

i.e., $\text{MB}[h_1, s_1 \ \& \ s_2 \ \& \ s_3] = 0.42$

and $\text{MD}[h_1, s_1 \ \& \ s_2 \ \& \ s_3] = 0$

Thus the strength of the rule is reduced to reflect the uncertainty regarding s_3. Combining Function 1 is now used to combine 0.42 (i.e., $\text{MB}[h_1, s_1 \ \& \ s_2 \ \& \ s_3]$) with the previous MB for the hypothesis that the organism is a *Streptococcus*.

We have shown that the numbers thus calculated are approximations at best. Hence it is not justifiable simply to accept as correct the hypothesis with the highest CF after all relevant rules have been tried. Therapy is therefore chosen to cover for all identities of organisms that account for a sufficiently high proportion of the possible hypotheses on the basis of their

CF's. This is accomplished by ordering them from highest to lowest and selecting all those on the list until the sum of their CF's exceeds z (where z is equal to 0.9 times the sum of the CF's for *all* confirmed hypotheses). This *ad hoc* technique therefore uses a semiquantitative approach in order to attain a comparative goal.

Finally, it should be noted that our definition of CF's allows us to validate those of our rules for which frequency data become available. This would become increasingly important if the program becomes a working tool in the clinical setting where it can actually be used to gather the statistical data needed for its own validation. Otherwise, validation necessarily involves the comments of recognized infectious disease experts who are asked to evaluate the program's decisions and advice. Evaluations of MYCIN have shown that the program can give advice similar to that suggested by infectious disease experts (see Part Ten). Studies such as these have allowed us to gain confidence that the certainty factor approach is robust enough for use in a decision-making domain such as antimicrobial selection.