PART FOUR

Reasoning Under Uncertainty

10

Uncertainty and Evidential Support

As we began developing the first few rules for MYCIN, it became clear that the rules we were obtaining from our collaborating experts differed from DENDRAL's situation-action rules in an important way—the inferences described were often uncertain. Cohen and Axline used words such as "suggests" or "lends credence to" in describing the effect of a set of observations on the corresponding conclusion. It seemed clear that we needed to handle probabilistic statements in our rules and to develop a mechanism for gathering evidence for and against a hypothesis when two or more relevant rules were successfully executed.

It is interesting to speculate on why this problem did not arise in the DENDRAL domain. In retrospect, we suspect it is related to the inherent complexity of biological as opposed to artificial systems. In the case of DENDRAL we viewed our task as hypothesis generation guided by rule-based constraints. The rules were uniformly categorical (nonprobabilistic) and were nested in such a way as to assure that contradictory evidence was never an issue.[1] In MYCIN, however, an overall strategy for nesting categorical rules never emerged; the problem was simply too ill-structured. It was possible to tease out individual inference rules from the experts working with us, but the program was expected to select relevant rules during a consultation and to accumulate probabilistic evidence regarding the competing hypotheses.

In response to these observations we changed the evolving system in two ways. First, we modified the rule structure to permit a conclusion to be drawn with varying degrees of certainty or belief. Our initial intent was to represent uncertainty with probabilistic weights on a 0-to-1 scale. Second, we modified the data structures for storing information. Rather than simply recording attribute-object-value triples, we added a fourth element to represent the extent to which a specific value was believed to be true. This meant that the attribute of an object could be associated with multiple competing values, each associated with its own certainty weight.

[1]In the model of mass spectrometry used by DENDRAL, the statistical nature of events is largely ignored in favor of binary decisions about occurrence or nonoccurrence of events.

It was logical to turn to probability theory in our initial efforts to define the meaning of these certainty values. Bayes' Rule (or Bayes' Theorem) the traditional evidence-combining technique used in most medical diagnosis programs, provided a model for how the weights could be manipulated if they were interpreted as probabilities. For reasons that are discussed in detail in the next chapter, we were gradually led to consider other interpretations of the numerical weights and to reject a purely probabilistic interpretation of their meaning.

Shortliffe was encouraged by Buchanan, as well as by Professors Patrick Suppes and Byron Brown, who were on his thesis committee, to attempt to formalize the numerical weights rather than to define and combine them in a purely *ad hoc* fashion. There ensued many months of reading the literature of statistics and the philosophy of science, focusing on the theory of confirmation and attempting to understand the psychological issues underlying the assignment of certainty weights. Chapter 11, originally published in 1975, summarizes the formal model that ultimately emerged from these studies. The concept of certainty factors (CF's) was implemented and tested in MYCIN and became a central element of other EMYCIN systems that have been developed in the ensuing years.

Another source of uncertainty in a knowledge base is the imprecision in language. Even though the vocabulary of medicine is technical, it is not without ambiguity. For example, one question asks whether the dosage of a drug given previously was "adequate." Rules use the answers given in response to such questions with the assumption that the user and the expert who wrote the rules agree on the meanings of such terms. What do we do to help satisfy this assumption? Rule writers are encouraged to anticipate the ambiguities when formulating their questions. They write the English forms of the TRANS and PROMPT values. Also, they can supply further clarification in the REPROMPT value, which is printed when the user types a question mark. MYCIN (and EMYCIN) provides facilities for experts to clarify their use of terms, but cannot guarantee the elimination of ambiguity.[2]

10.1 Analyses of the CF Model

Although the motives behind the CF model were largely pragmatic and we justified the underlying assumptions by emphasizing the system's excellent performance (see, for example, Chapter 31), several theoretical ob-

[2]Fuzzy logic (Zadeh, 1978) quantifies the degree to which imprecise concepts are satisfied, thus adding another level of detail to the reasoning. For our purposes, it is sufficient to ask the user whether a concept, such as "adequateness," is satisfied—where an appropriate response may be "Yes (0.7)." In fuzzy logic, a possibility distribution for the user's understanding of the concept "adequate" would be matched against a corresponding distribution for the rule writer's understanding. We believe this is an unnecessary layer of detail for the precision we want to achieve (or feel is justified by the precision of the information).

jections to the model were subsequently raised. Professor Suppes had been particularly influential in urging Shortliffe to relate CF's to the rules of conventional probability theory,[3] and the resulting definitions of MB's and MD's did help us develop an intuitive sense of what our certainty measures might mean. However, the probabilistic definitions also permitted formal analyses of the underlying assumptions in the combining functions and of limitations in the applicability of the definitions themselves.

For example, as we note in Chapter 11, the source of confusion between CF(h,e) and $P(h|e)$ becomes clear when one sees that, for small values of the prior probabilities $P(h)$, CF(h,e) $\approx P(h|e)$. Our effort to ignore prior probabilities was largely defended by observing that, in the absence of all information, priors for a large number of competing hypotheses are uniformly small. For parameters such as organism identity, which is the major diagnostic decision that MYCIN must address, the assumption of small priors is reasonable. The same model is used, however, to deal with *all* uncertain parameters in the system, including yes-no parameters for which the prior probability of one of the values is necessarily greater than or equal to 0.5.

The significance of the 0.2 threshold used by many of MYCIN's predicates (see Chapter 5) was also a source of puzzlement to many observers of the CF model. This discontinuity in the evaluation function is not an intrinsic part of the CF theory (and is ignored in Chapter 11) but was added as a heuristic for pruning the reasoning network.[4] If *any* small positive CF were accepted in evaluating the premise of a rule, without a threshold, two undesirable results would occur:

1. Very weak evidence favoring a condition early in the rule premise would be "accepted" and would lead to consideration of subsequent conditions, possibly with resulting backward-chained reasoning. It is wasteful to pursue these conditions, possibly with generation of additional questions to the user, if the evidence favoring the rule's premise cannot exceed 0.2 (recall that $AND uses min in calculating the TALLY—see Chapters 5 and 11 for further details).
2. Even if low-yield backward chaining did not occur, the rule would still have limited impact on the value of the current subgoal since the TALLY for the rule premise would be less than 0.2.

[3]Suppes pressed us early on to state whether we were trying to model how expert physicians *do* think or how they *ought to* think. We argued that we were doing neither. Although we were of course influenced by information regarding the relevant cognitive processes of experts [see, for example, the recent books by Elstein et al. (1978) and Kahneman et al. (1982)], our goals were oriented much more toward the development of a high-performance computer program. Thus we sought to show that the CF model allowed MYCIN to reach good decisions comparable to those of experts and intelligible both to experts and to the intended user community of practicing physicians.

[4]Duda et al. (1976) have examined this discontinuity and the relationship of CF's to their Bayesian updating model used in the PROSPECTOR system.

Thus the 0.2 threshold was added for pragmatic reasons and should not be viewed as central to the CF model itself. In later years questions arose as to whether the value of the threshold should be controlled dynamically by the individual rules or by meta-rules (rather than being permanently bound to 0.2), but this feature was never implemented.

Another important limitation of MYCIN's control scheme was noted in the mid-1970s but was never changed (although it would have been easy to do so). The problem results from the requirement that the premise of a rule be a conjunction of conditionals with disjunctions handled by multiple rules. As described in Chapter 5, A V B V C → D was handled by defining three rules: A → D, B → D, and C → D. If all rules permitted conclusions with certainty, the three rules would indeed be equivalent to a single disjunctive rule with certain inference (CF = 1). However, with CF's less than unity, all three rules might succeed for a given case, and then each rule would contribute incremental evidence in favor of D. This evidence would be accumulated using the CF combining function, that is, $CF_{COMBINE}$, and might be very different from the CF that the expert would have given if asked to assign a weight to the single disjunctive rule. This problem could have been handled by changing the rule monitor to allow disjunctions in a rule premise, but the change was never implemented because a clear need never arose.

The rule interpreter does not allow rules to be written whose primary connective is disjunction ($OR). We have encouraged splitting primary disjunctions into separate rules for this reason. Thus

[1] ($OR ABC) → D

would be written as three separate rules:

[2] A → D
[3] B → D
[4] C → D

Conceptually this is simple and straightforward. In some cases, however, the disjuncts are better understood as a set, and [1] would be a clearer expression than [2], [3], and [4]. In these cases, Carli Scott has pointed out that [1] can be rewritten as a primary conjunction wih only one clause:

[5] ($AND ($OR A B C)) → D

This uncovers a limitation on the CF model, however. While [5] should give the same results as [2], [3], and [4] together, the resulting CF's on conclusion D will differ. The reason is that in [5] the CF on the rule will be multiplied by the MAX of the CF's of the disjunction A, B, or C, while in [2], [3], and [4] the cumulative CF associated with D will be the result of combining three products according to the combining function.[5]

[5]It is possible to force them to give the same result by adjusting the CF's either on [5] or on [2], [3] and [4]. We would not expect a rule writer to do this, however, nor would we think the difference would matter much in practice.

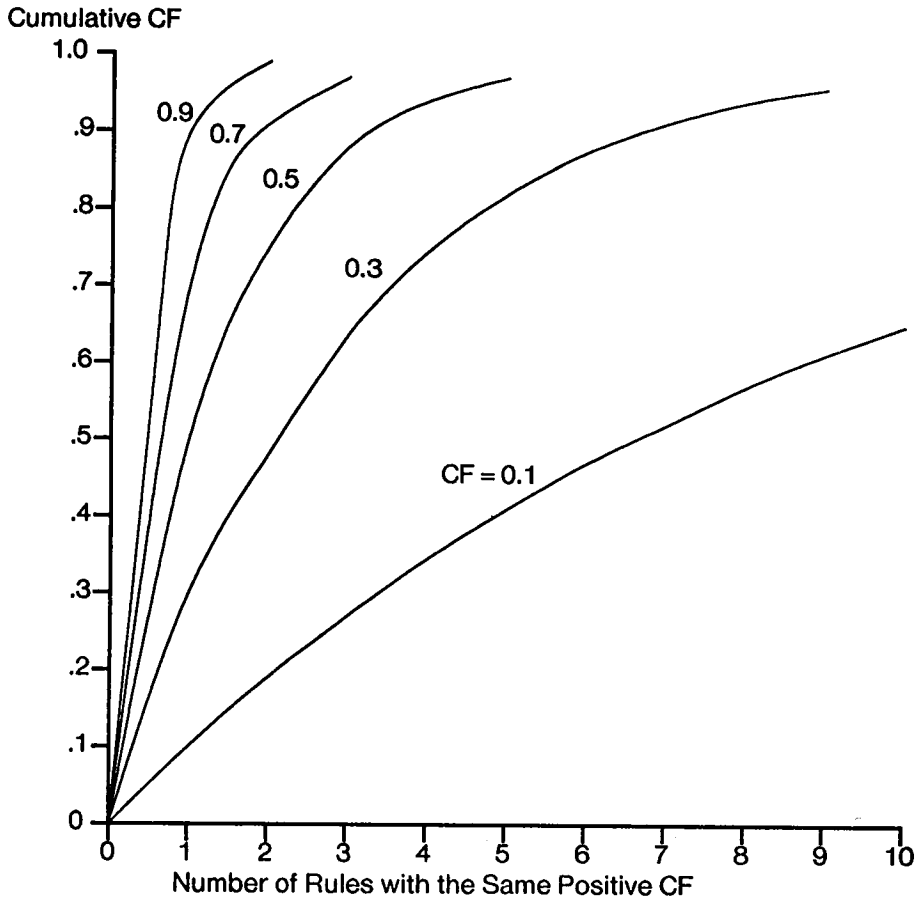


FIGURE 10-1 Family of curves showing how rapidly MYCIN's CF combining function converges for rules with the same CF.

Another limitation for some problems is the rapidity with which CF's converge on the asymptote 1. This is easily seen by plotting the family of curves relating the number of rules with a given CF, all providing evidence for a hypothesis, to the resulting CF associated with the hypothesis.[6] The result of plotting these curves (Figure 10-1) is that $CF_{COMBINE}$ is seen to converge rapidly on 1 no matter how small the CF's of the individual rules are. For some problem areas, therefore, the combining function needs to be revised. For example, damping factors of various sorts could be devised

[6]This was first pointed out to us by Mitch Model, who was investigating the use of the CF model in the context of the HASP/SIAP program (Nii et al., 1982).

(but were not) that would remedy this problem in ways that are meaningful for various domains. In MYCIN's domain of infectious diseases, however, this potential problem never became serious. In PROSPECTOR this problem does not arise because there is no finite upper limit to the likelihood ratios used.

As we were continuing to learn about the CF model and its implications, other investigators, faced with similar problems in building medical consultation systems, were analyzing the general issues of inexact inference (Szolovits and Pauker, 1978) and were in some cases examining shortcomings and strengths of CF's. Later, Schefe analyzed CF's and fuzzy set theory (Schefe, 1980). Dr. Barclay Adams, a member of the research staff at the Laboratory of Computer Science, Massachusetts General Hospital, responded to our description of the MYCIN model with a formal analysis of its assumptions and limitations (Adams, 1976), included in this book as Chapter 12. The observations there nicely specify the assumptions that are necessary if the CF's in MYCIN's rules are interpreted in accordance with the probabilistic definitions from Chapter 11. Adams correctly notes that there may be domains where the limitations of the CF model, despite their minimal impact on MYCIN's performance, would seriously constrain the model's applicability and success. For example, if MYCIN had required a single best diagnosis, rather than a clustering of leading hypotheses, there would be reason to doubt the model's ability to select the best hypothesis on the basis of a maximal CF.

Even before the Adams paper appeared in print, many of the same limitations were being noted within the MYCIN project. For example, in January of 1976 Shortliffe prepared an extensive internal memo that made several of the same observations cited by Adams.[7] He was aided in these analyses by Dana Ludwig, a medical student who studied the CF model in detail as a summer research project. The Shortliffe memo outlined five alternate CF models and argued for careful consideration of one that would require the use of *a priori* probabilities of hypotheses in addition to the conventional CF's on rules. The proposed model was never implemented, however, partly due to time constraints but largely because MYCIN's decision-making performance was proving to be excellent despite the theoretical limitations of CF's. Some of us felt that a one-number calculus was preferable in this domain to a more theoretically sound calculus that requires experts to supply estimates of two or more quantities per rule. It is interesting to note, however, that the proposals developed bore several similarities to the subjective Bayesian model developed at about the same time for SRI's PROSPECTOR system (Duda et al., 1976). The CF model has been used successfully in several EMYCIN systems (see Part Five) and in the IRIS system (Trigoboff, 1978) developed at Rutgers University for diagnosing glaucomas.

[7]This is the file CF.MEMO referred to by Clancey in the exchange of electronic messages at the end of this chapter.

There is an additional element of uncertainty in rules that is also bound up in the CF's. Besides capturing some measure of increased probability associated with the conclusion after the premises are known and some measure of the utility associated with the conclusion, the CF also includes some measure of how "flaky" the rule is. That is, a CF of 0.2 can indicate that the probability increases by 20% (rather precisely) or that the rule writer felt there was a positive association between premises and conclusion but was only 20% certain of it. Some rule writers would be able to quantify their degree of doubt about the CF's (e.g., "I am about 90% certain that this strength of association is 0.5"), but there is no provision in our CF model for doing so. In most cases where increased precision is possible, rule writers would have prior and posterior probabilities and would not need a one-number calculus.

Despite the shortcomings of the CF model, it must be recognized that the issues we were addressing reflected a somewhat groping effort to cope with the limitations of probability theory. It has therefore been with considerable interest that we have discovered in recent years the work of Dempster and Shafer. Shafer's book, *The Mathematical Theory of Evidence,* appeared in 1976 and proposed solutions to many of the same problems being considered in the MYCIN work. Several aspects of the CF model appear as special cases of their theory. Interestingly, Bayesian statistics is another special case. Our recent attempt to understand the Dempster-Shafer model and its relevance to MYCIN is described in Chapter 13. This work, the most recent in the book, was largely done by Jean Gordon, a mathematician who recently joined our group when she came to Stanford as a medical student. Because of new insights regarding the topics underlying CF's and the relationships to probabilistic reasoning, we have chosen to include that analysis in this volume even though we have not implemented the ideas in the program.

10.2 Evolution of the CF Model

Although the model described in Chapter 11 has persisted to the present for the MYCIN program, and for other EMYCIN systems (see Part Five), a few revisions and additional observations have been made in the intervening years. The only major change has been a redefinition of the combining function by Bill van Melle. This was undertaken for two reasons:

1. the potential for a single piece of negative evidence to overwhelm several pieces of positive evidence (or vice versa); and
2. the computational expense of storing both MB's and MD's (rather than cumulative CF's) in order to maintain commutativity.

The second of these points is discussed briefly in Chapter 11, but the first may require clarification. Consider, for example, eight or nine rules all supporting a single hypothesis with CF's in the range 0.4 to 0.8. Then the asymptotic behavior of the cumulative MB would result in a value of about 0.999. Suppose now that a single disconfirming rule were to succeed with CF = 0.8. Then the net support for the hypothesis would be

$$\mathrm{CF} = \mathrm{MB} - \mathrm{MD} = 0.999 - 0.8 = 0.199$$

This behavior was counterintuitive and occasionally led MYCIN to reach incorrect inferences, especially in situations where the final CF after tracing became less than 0.2. This would drop the final belief below the established threshold. Hence a single piece of negative evidence could overwhelm and negate the combined evidence of *any number* of supporting rules.

As a result, we changed both the definition of a CF and the corresponding combining function to soften the effect:

$$\mathrm{CF} = \frac{\mathrm{MB} - \mathrm{MD}}{1 - \min(\mathrm{MB},\mathrm{MD})}$$

$$\mathrm{CF}_{\mathrm{COMBINE}}(X,Y) = \begin{cases} X + Y(1 - X) & X, Y \text{ both} > 0 \\ \dfrac{X + Y}{1 - \min(|X|, |Y|)} & \text{one of } X, Y < 0 \\ -\mathrm{CF}_{\mathrm{COMBINE}}(-X,-Y) & X, Y \text{ both} < 0 \end{cases}$$

Note that the definition of CF is unchanged for any single piece of evidence (where either MD or MB is zero by definition) and that the combining function is unchanged when both CF's are the same sign. It is only when combining two CF's of opposite sign that any change occurs. The reader will note, for example, that

$$\mathrm{CF}_{\mathrm{COMBINE}}(0.999, -0.80) = 0.199/0.2 = 0.99$$

whereas

$$\mathrm{CF}_{\mathrm{COMBINE}}(0.55, -0.5) = 0.05/0.5 = 0.1$$

In addition, the change in $\mathrm{CF}_{\mathrm{COMBINE}}$ preserved commutativity without the need to partition evidence into positive and negative weights for later combination. Thus, rather than storing both MB and MD for each hypothesis, MYCIN simply stores the current cumulative CF value and combines it with new evidence as it becomes available. Beginning in approximately 1977 these changes were incorporated into all EMYCIN systems.

10.3 Assessing the CF Model

Even before the change in the combining function was effected, we had observed generally excellent decision-making performance by the program and therefore questioned just how sensitive MYCIN's decisions were to the CF's on rules or to the model for evidence accumulation. Bill Clancey (then a student on the project) undertook an analysis of the CF's and the sensitivity of MYCIN's behavior to those values. The following discussion is based in large part on his analysis and the resulting data.

The CF's in rules reflect two kinds of knowledge. In some cases, such as a rule that correlates the cause of meningitis with the age of the patient, the CF's are statistical and are derived from published studies on the incidence of disease. However, most CF's represent a mixture of probabilistic and cost/benefit reasoning. One criticism of MYCIN's rules has been that utility considerations (in the decision analytic sense) are never made explicit but are "buried" in a rule's CF. For example, the rule that suggests treating for *Pseudomonas* in a burned patient is leaving out several other organisms that can also cause infection in that situation. However, *Pseudomonas* is a particularly aggressive organism that often causes fatal infections and yet is resistant to most common antibiotics. Thus its "weight" is enhanced by rules to ensure that it is adequately considered when reaching therapy decisions.[8] Szolovits and Pauker (1978) have also provided an excellent discussion of the issues complicating the combination of decision analytic concepts and categorical reasoning in medical problems.

Figure 10-2 is a bar graph showing how frequently various CF values occur in MYCIN's rules. All but about 60 of the 500 rules in the most recent version of the system have CF's.[9] The cross-hatched portion of each bar shows the frequency of CF's in the 1975 version of MYCIN, when there were only 200 rules dealing with bacteremia. The open portion of each bar refers to the CF's of incremental rules since that time, most of which deal with meningitis. The overall pattern is about the same, although the more recent system has proportionally more small positive CF's. This makes sense because the newer rules often deal with softer data (clinical evidence) in contrast to the rules for bacteremia, which generally interpret

[8]Self-referencing rules, described in Chapter 5, were often used to deal with such utility considerations. As mentioned in Chapter 3, they allowed dangerous organisms, initially suggested with only minimal certainty, to be reconsidered and further confirmed by special evidence. For example: *if* you are already considering *Pseudomonas and* the patient has ecthyma gangrenosum skin lesions, *then* there is even greater importance to the conclusion that the pathogen is *Pseudomonas*.

[9]The rules without CF's do not associate evidence with hypotheses but make numerical computations or save a text string to be printed later. Note also that some rules, particularly tabular rules, make many conclusions and thus account for the fact that there are more CF's than rules.

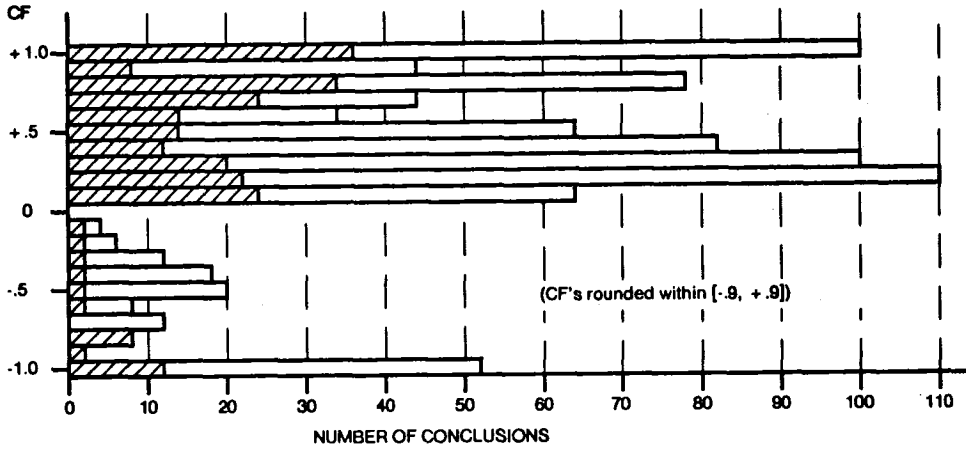


FIGURE 10-2 Frequency of CF's in MYCIN's rules. Cross-hatched bars indicate frequencies for the 1975 version of MYCIN. Open bars show frequencies since then.

more concrete laboratory results. The bimodal distribution with peaks at 0.8 and 0.2 (ignoring for a moment those rules, often definitional, that reach conclusions with certainty) suggests that experts tend to focus on strong associations (+0.8, a number that might seem less binding than 0.9) and many weak associations (+0.2, the minimum CF that will allow the inferred parameter to exceed the threshold for partial belief). In contrast there are relatively few rules with negative CF's. We suspect this reflects the natural tendency to state evidence in a positive way.

Analysis of MYCIN's reasoning networks suggests that the program should not be very sensitive to changes in rule CF's. This conclusion is based on two observations about how CF's are actually used in the program. First, inference chains are short, and premises often pass a TALLY of 1.0 to the conclusion (see Chapter 5), so the effect of multiplying CF's from one step in the chain to the next is minimal. Second, conclusions are frequently made by only a single rule, thereby avoiding the use of $CF_{COMBINE}$ for all but a few key parameters. Observe that the first effect deals with combination of CF's from goal to goal (by passing a value from a rule premise to the conclusion) and the second deals with combination of evidence for a single goal.

Intrigued by observations such as those outlined above, Clancey enlisted the assistance of Greg Cooper, and in 1979 they undertook an experiment to determine quantitatively how sensitive MYCIN is to changes in rule CF's. The ten cases used in the formal evaluation of the meningitis rule set (see Chapter 31) were used for this study. The cases were run in batch mode using systematic variations of the CF's in MYCIN's rules. For

Number of intervals	Number of cases (out of 10): Same organisms and therapy	Different organisms	Different organisms and therapy
10	9	1	0
5	7	3	0
4	8	2	1
3	5	5	1
2	1	9	3

FIGURE 10-3 Results of CF sensitivity experiment.

each run, rules were modified by mapping the existing rule CF's onto a new, coarser scale. The original CF scale has 1000 intervals from 0 to 1000.[10] Trials were run using ten, five, four, three, and two intervals. Thus, when there are five intervals, all rule CF's are mapped onto 0, 200, 400, 600, 800, and 1000, rounding as necessary. When there are two intervals, only the numbers 0, 500, and 1000 are used.

CF's were combined using the usual combining function (the revised version that was in use by 1979). Thus intermediate conclusions mapped onto arbitrary numbers from 0 to 1000. Clustering the final organism list was done in the normal way (cutting off at the largest gap). Finally, negative CF's were treated analogously, for example, mapping onto 0, −333, −666, and −1000 when there were three intervals.

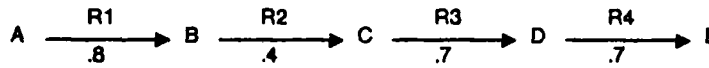
In examining results, we are interested primarily in three possible outcomes: (1) no change to the item list (and hence no change in therapy); (2) different organisms, but the same therapy; and (3) new therapy (and therefore different organisms). Figure 10-3 summarizes the data from the ten cases run with five different CF scales.

Degradation of performance was only pronounced when the number of intervals was changed to three (all rule CF's mapped onto 0, 333, 666, and 1000). But even here five of the ten cases had the same organism list and therapy. It wasn't until CF's were changed to 0, 500, and 1000 that a dramatic change occurred; and even with nine new organism lists, we find that seven of the ten cases had the same therapy. The fact that the organism list did not change radically indicates that MYCIN's rule set is not "fine-tuned" and does not need to be. The rules use CF's that can be modified by ±0.2, showing that there are few deliberate (or necessary) interactions in the choice of CF's. The observed stability of therapy despite changing organism lists probably results because a single drug will cover for many organisms, a property of the domain.

[10]CF's are handled internally on a 0 to 1000 scale to avoid floating-point arithmetic, which is more expensive in Interlisp than is integer arithmetic.

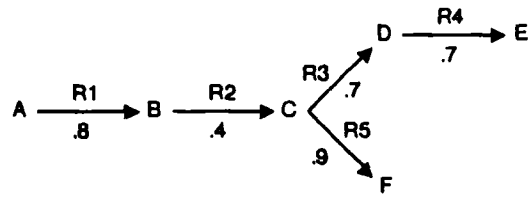
10.4 Additional Uses of Certainty Factors

By the early 1980s, when much of our research was focusing on issues other than EMYCIN systems, we still often found CF's to be useful computational devices. One such example was the work of Jerry Wallis, described in detail in Chapter 20. His research modeled causal chains with rules and used CF's to represent the uncertainty in the causal links. Because his system reasoned both from effects to causes and from causes to effects, techniques were needed to prevent fruitless searching of an entire connected subgraph of the network. To provide a method for search termination, the concept of a subthreshold path was defined, i.e., a path of reasoning whose product of CF's can be shown to be below the threshold used to reject a hypothesis as unknown. For example, if there is a linear reasoning path of four rules (R1, R2, R3, and R4) where A can be asked of the user and E is the goal that initiated a line of backward-chained reasoning:



then if B were known with certainty, E would be known only with a CF of (0.4)(0.7)(0.7) = 0.19. This is less than the conventional cutoff of 0.2 used in EMYCIN systems, so the line of reasoning from B to E would be considered a subthreshold path. There is no need to invoke rule R1 and ask question A in an effort to conclude B because the result cannot affect the final value for the variable E. If the product of CF's is tabulated during the backward-chaining process, the accumulated value provides a method for limiting the search space that needs to be investigated.

In a branched reasoning tree this becomes slightly more complex. Normally, when a rule is used to conclude a value with a particular CF, that number is stored with the parameter's value in case it is later needed by other rules. In the example above, termination of the search from E back to A (due to the subthreshold condition at B) would have left the value at C "unknown" and might have left a CF of 0 stored at that node. Suppose, though, that another rule, R5, later needed the value of C because of consideration of goal F:



It would be inappropriate to use the unknown value of C stored from the previous inference process, for now it would be appropriate to back-chain further using R1 (the higher CF of 0.9 associated with R5, compared to the composite CF of 0.49 associated with the chaining of rules R3 and R4, keeps the path to A from being subthreshold this time). Thus, if one wants to use previous results only if they are appropriate, it is necessary to store the "vigor" with which a value was investigated along with its CF. Wallis proposed that this be computed by multiplying the CF's from the goal (in this case E) through the value in question. Then, when a node is investigated for a second time via an alternate reasoning chain, this measure of vigor, or *investigation strength,* can be used to determine whether to investigate the node further. If the stored investigation strength is greater than the investigation strength of the new reasoning chain, the old value can be used. Otherwise the backward-chaining process must be repeated over a larger portion of the search space.

Although there is further complexity in these ideas developed by Wallis, the brief discussion here shows some of the ways in which concepts drawn from the CF model have been broadened in other settings. Despite the theoretical limitations discussed above and in the subsequent chapters, these concepts have provided an extremely useful tool for dealing with issues of inexact inference in the expert systems that we have developed.

10.5 An Electronic Exchange Regarding CF's

We close this chapter with a series of informal electronic mail messages that were exchanged by some members of our research group in 1976 (Carli Scott, Bruce Buchanan, Bill Clancey, Victor Yu, and Jan Aikins). Victor was developing the meningitis rule set at the time and was having frequent problems deciding what CF's to assign to individual rules and how to anticipate the ramifications of any decisions made. The messages are included in their entirety. Not only do they provide insight into the way that our ideas about CF's evolved through a collaborative effort over many years, but they are also representative of the kinds of dialogues that occurred frequently among members of the project. Because many of the ideas in this book evolved through such interchanges, we felt it was appropriate to provide one verbatim transcript of a typical discussion. The ideas expressed were fresh at the time and not fully worked out, so the messages (and Clancey's closing memo) should be seen as examples of project style rather than as an exposition of the "last word" on the topics discussed.

Date: 26 Feb 1976
From: Scott
Subject: Summary of discussion
To: MYCIN gang

This is a summary of what I think came out of yesterday's meeting. Please read it and send me comments, objections, etc.

1) Victor [Yu] has assigned certainty factors to his rules based on the relative strengths of the evidence in these rules. While trying to find a numerical scale that would work as he wanted it to with the system's 0.2 cutoff and combining functions, he had to adjust certainty factors of various rules. Now that this scale has been established, however, he assigns certainty factors using this scale, and does NOT adjust certainty factors of rules if he doesn't like the system's performance. Furthermore, he does NO combinatorial analysis before determining what CF to use; he is satisfied that using the scale he has devised, the system's combining function, and the 0.2 cutoff, the program will arrive at the right results for any combination of factors, and if it doesn't, he looks for missing information to add.

2) Assuming that the parameters IDENT and COVERFOR are disambiguated in Victor's set of rules, Ted [Shortliffe] believes the CF's that Victor uses in his rules, and approves of the idea of using a cutoff for COVERFOR since this is what we've been doing with bacteremia (since it is a binary decision, a cutoff makes sense for COVERFOR). Furthermore, this is quite similar to what clinicians do: they accumulate lots of small bits of clinical evidence, then decide if the total is enough to make them cover for a particular organism—independent of what the microbiological evidence suggests.

3) Bruce [Buchanan] and BC [Bill Clancey] still object to Victor's CF's because they seem too precise (since he is working in the 0 to 0.2 range). My claim is that he really isn't making numbers more precise, the difference in CF's from one strength to the next is 0.05 (i.e., the classes of rules he has are assigned CF's 0.05, 0.1, 0.15, 0.2, 0.25, . . .). This is no finer a distinction than we've had in the past—we have rules with CF 0.2, 0.25, 0.3, 0.35. I don't see why the smaller absolute values of the CF's Victor uses makes much difference; the rules have much smaller strengths than any rules we've had before, so they should have smaller CF's.

4) There seems to be concern because Victor believes in his CF's, and relies on them to combine in the right way. In the past, we never dealt with this type of accumulation of small bits of information that would combine to give either enough total info or not (though I believe CF's were designed to handle just such combinations). Since Victor has defined guidelines on deciding how strong the evidence must be in order for a rule to be assigned a certain CF, and since he has tested these guidelines within the framework of MYCIN's combining functions, he believes that it all works as it should. Furthermore, he believes that he can define these "points of reference" so that future medical people can add rules, using the same guideline that Victor has used, and they should fit into the system and work fine with his rules.

5) I am satisfied with what Victor is doing, and would like to try Ted's suggestion of separating COVERFOR from IDENT in Victor's system. I believe the result of this would be that the program would continue to perform very well on meningitis patients, and Ted, Victor, (I believe) Larry [Fagan],

and I would all be happy with the results. I think points (3) and (4) above sum up other people's objections that might remain. If this is so, what are suggestions from people who still aren't happy with the model? Is everyone satisfied with everything now? Are there more objections that I missed? Have I completely misunderstood something? Have I completely misunderstood everything? Please let me know what you think so we can start to work out problems that might remain.

Carli

Date: 27 Feb 1976
From: Buchanan
To: MYCIN gang

Carli,

Thanks for your summary—it appears to be correct in almost every detail. I would like you to try separating COVERFOR and IDENT as soon as possible since that is needed for bacteremia anyway and is a help in clarifying the conceptual basis on which the program makes a recommendation. I also think that everyone will be happy with the results, especially me if it brings the knowledge bases into a common framework.

My concern is I would also like you to begin working on the rerepresentation of the context tree to help us with time relations and the infection-organism link. As Ted described it, you and he have pretty well worked things out. Because it is necessary for the FOREACH[11] mechanism and is desirable for many other reasons, I would like us not to delay it. Do you see problems with this?

As I tried to say yesterday, my reservations with the meningitis system stem from my uneasiness with the CF model, which we all know needs improving (which Pacquerette [a visiting student from France] was starting, but won't finish). I don't want Victor to become dependent on a particular mechanism for combining CF's—because we hope the mechanism will be improved soon. I have no doubt that the rules work well now, and I don't disagree at all with the need for firm reference points for the CF's.

As soon as COVERFOR and IDENT are separate, could you try the meningitis patients again, enlisting whatever help you need? Then we'll be able to decide whether that meets all our specs. After that we can be working on the context tree and time problems while Victor continues development on the medical side. I foresee no difficulty in mapping the CF's from existing rules (meningitis as well as bacteremia) into whatever numbers are appropriate for a new CF model when we have one—with firm reference points if at all possible.

Bruce

PS: I think a reference point for defining how strongly suggestive some evidence is for a conclusion is easier when almost all conclusions are about identities of organisms that should be treated for. In bacteremia the rules conclude about so many different things that it is harder—but no less desir-

[11]FOREACH is a quantification primitive in rules.

able—to be precise about what "weakly suggestive" and "strongly suggestive" mean.

Date: 27 Feb 1976
From: Clancey
To: MYCIN gang

Your summarization of the meeting was excellent. Here I will go into more detail about the problem with Victor's choice of certainty factors.

Your claim that Victor's preciseness in selecting CF's is not different from the distinctions made in the past ignores my wariness about the RANGE in which he is being precise. Your examples (0.25, 0.3, 0.35) are greater than 0.2, the range in which I showed that the current system is insensitive to even large variation of the CF's chosen. (That is, a change in the range of ± 0.2 does not affect system performance (rule invocation and success), as long as the numbers are > 0.2.) The area in which Victor is working which is bothersome is <0.2 (your examples: 0.05, 0.1, 0.15, 0.2). What Bruce was saying, I believe, was that accumulation of evidence in this area is going to affect very much the invocation and success of rules. It is in this range that CHANGES to the CF of a rule for purposes of adjusting system performance violate the principle of a rule being a modular, independent chunk of knowledge.

Now, first, Victor tells us that he does not make these adjustments.Rather, he is assigning numbers according to a consistent scale about belief which he has devised in his subdomain. I am very pleased to hear this, and am in full agreement with his claim that such a scale is necessary and should be defined for ALL rules in MYCIN.

What remains disturbing is the certainty factor model itself. Here we have no sure intuition about the performance meaning of 0.05 as opposed to 0.1, yet we are assigning them as if they were significantly different from one another. It is clear to everyone working on the CF model, I believe, that we need a combining function that will make use of these numerical representations of subjective distinctions. For example, I would expect a good model to take as many pieces of 0.1 evidence as Victor deems significant, i.e., makes a condition (parameter value) "true," and bumps the conglomeration above 0.2. The problem here is that I DO NOT expect Victor or anyone to be able to assign facts a weighting that is independent of the entire context. That is, the 0.1 that comes from Rule 371 for CATEGORY FUNGUS may combine (in Victor's mind) with the conclusion in Rule 372 of the same value to give a feeling of the CATEGORY ACTUALLY BEING FUNGUS >0.2, so SAME succeeds. But perhaps the same CF value combination coming from Rule 385 DOES NOT make for belief in the conclusion (NOT >0.2). It seems entirely conceivable in my mind that Victor would find some combination of rule successes to be completely nonsensical. So, he would not know what to make of it at all, and would almost certainly not make the same conclusions as he would if he looked at each set of premise clauses independently.

I am saying here that rules that break observations into many small parts, resulting in CF's <0.2 intended to combine to form an accumulated observation, ignore the total perspective, which says, "Hey, wait a minute, these 6 clauses can't appear together: why was she given corticosteroids if she has

XX? This doesn't mean FUNGUS to me; no, I want to know why that prescription was made." This same criticism does not apply with the same force to many rules with CF >0.2 because they bring together a "more significant set of facts." They do this by capturing (often disjoint) pictures of the world that in themselves MAKE SENSE. I do not at all understand how a rule can be written that can at once stand on its own and yet NOT be significant truth (i.e., believable observation, tangible conclusion). It is my suspicion that Victor has not built a system in which EVIDENCE combines plausibly, but rather a system in which independent rules SUCCEED TOGETHER to make a conclusion that could be expressed as a single rule, and WOULD have to be expressed that way to have a CF > 0.2.

Now, Victor has said that he could have combined these rules to give a body of rules in which these same small observations appear together, thus yielding larger CF's. However, he believes that this would result in far more rules (to allow for the cross product of occurrences), and he would not be sure that he had covered all of the possible cases. Well, certainly, with respect to the latter, we can tell him if the larger set covers all of the various combinations. The question of having far more rules is, I suppose, a valid concern. But at least then we could feel sure that only the PLAUSIBLE observations had been combined.

To summarize, we talk about accumulating "lots of small bits of clinical evidence," but I do not understand how a bit of EVIDENCE could be NOTKNOWN (the definition of CF <=0.2). To me, evidence gathered by a rule should be an all-or-nothing thing—if something more is needed to make the parameter KNOWN [i.e., CF > 0.2], then I expect that there is something to be made explicit in the rule. This is the only way in which I can interpret the notion of a discrete cutoff at 0.2. Above that point I know something; below it I know nothing (NOTKNOWN). The only plausible explanation I have for Victor's small CF's is that they are like tags that record an observation. It would make me much happier to see each of these CF's changed to NOTICEDP, with definite (= 1) CF's. Then these parameters could be combined with evidence garnered from lab rules.

I would be happy to hear other opinions about the 0.2 cutoff and its meaning for rule CF's.

Bill

Date: 28 Feb 1976
From: Aikins
Subject: On Wednesday's meeting
To: MYCIN gang

There are three things that I feel we should consider in our discussions that have not yet been mentioned. The first is a concern about knowledge acquisition. I feel that whatever we decide, the MYCIN acquisition module should be designed so that a recognized medical expert could, without too much difficulty, add a new rule or other piece of knowledge to the MYCIN data base. I wonder if a doctor in Boston would be able to add a meningitis rule to MYCIN without hurting the performance of Victor's system. I got the impression that Victor's system was somewhat fragile in this regard. I doubt that he would want to give up the ability to easily add medical knowl-

edge to MYCIN. I fear that we would be doing just that. (This problem includes the question of maintaining rule modularity.)

My second concern is that even if we can define fairly well what we mean by 0.7, 0.5, anything above 0.2, 0.2, etc., it seems that the next problem will be to define 0.25, 0.225, 0.175, 0.5, etc. We could continue this defining of CF's in smaller and smaller intervals forever. However, I doubt that medical science is exact enough for us to be able to do this.

This brings us to my third concern. In my recent meeting with Dr. Ken Vosti [a professor in Stanford's Division of Infectious Diseases], he stated a problem, already familiar to most of us, that even if we could reach agreement among the infectious disease experts at Stanford as to the "right" CF's to put on our rules, the infectious disease experts on the East Coast and other places would probably not agree with us. Now let's take this one step further. Say we are able to assign fairly straightforward meanings to our CF's. Now we have the problem of a doctor in some other part of the country who doesn't want to use MYCIN because our CF's don't agree with what he would use. In other words, by defining our CF's at all rigorously, we're inviting disagreement. So, concerns two and three are saying that we can never define each number on the 0 to 1.0 scale, and if we could, that might not be such a good idea anyway.

I have no solutions to offer at this time, but I hope everyone will keep these concerns in mind. I feel that CF's are designed to give doctors who read and write the rules a certain "commonsense referent" as to how valid the rule might be. If CF's become more important than that, I fear we will use too much of our medical expertise in deciding on the "right" CF for each rule, time that could be used to add more medical knowledge to the MYCIN data base.

Jan

Date: 29 Feb 1976
From: Yu
Subject: On Wed. meeting and Clancey
To: Clancey, Scott
cc: MYCIN gang

Bill,

1. Why is the system insensitive to CF? Certainly, this is not true for the meningitis rules.

2. Your point about plausible situations is a good one, and deserves further amplification and discussion. The reason I have "separated" the number of premises that in the bacteremia rules would have been combined is that I believe they are independent premises. I don't believe I ever said the reason for separating them is to avoid having too many rules; the reason for separating them is to cover a number of subtle clinical situations that would otherwise not have been considered. More on this later.

3. Finally, I should add that the 0.2 cutoff was selected because it is the one being used for SIGNIFICANCE and I thought it would best mesh with the current system. I must admit that I am surprised at the furor it has

evoked; if you wish to use some other cutoff, that's fine with me—the CF's could be easily adjusted.

4. I didn't understand a few of the points you raised, so I look forward to the next meeting.

Finally, I should say that the system that I have proposed is not meant in any way to replace the current bacteremia rules; it was merely a simple, practical way to handle meningitis. I did not feel the approach used in bacteremia was precise enough to handle meningitis.

Victor

Date: 29 Feb 1976
From: Yu
Subject: On Wed. meeting and Aikins
To: Aikins, Scott
cc: MYCIN gang

Jan,

1. You state that we are giving up the ability to "easily" add rules to MYCIN. Certainly, it is currently "easy" to add new rules to MYCIN; however, it is not so "easy" to rationalize, justify, and analyze these new rules. Furthermore, it becomes "difficult" when the system starts giving incorrect therapy after these new rules have been added.
2. I believe a doctor in Boston would have an "easier" task of adding new meningitis rules, as compared to bacteremia rules. He now has some reference points and definite guidelines on how a rule should be written. Again, the rule is more likely to be compatible with the existing system, since the new rule is written along the same guidelines and same philosophy. This is not the case with the bacteremia rules where it is likely and even probable that any new rule written by a non-MYCIN person could cause the system to malfunction.
3. I have not attempted to specifically define every increment between CF's.
4. I need not remind all of us that we are dealing directly with human lives. If another M.D. on the East Coast disagrees with our CF's and has data (be it strong or weak) as the basis for his disagreement, then we had better know about it. I claim that one of the advantages of specific criteria for CF's is that this "invites disagreement" (or to put it another way—critical analysis of the rules by non-MYCIN experts is possible).
5. What is this mystical "commonsense referent" that you have mentioned? (Likewise, Ted has stated that physicians would PROBABLY agree fairly closely on the CF's currently in MYCIN. If this is true, then my arguments for preciseness are invalid and unnecessary.)
6. Your last point concerning using too much time and effort on the CF question, when we could be adding more medical knowledge—I will merely refer you to Matthew: Chapter 7, verses 24–27.

Cheers,
Victor

Date: 1 Mar 1976
From: Clancey
Subject: More about certainty factors and a reply to your message
To: Yu
cc: MYCIN gang

Thanks for commenting on my remarks on CF's. I am well aware that my observations suffered from vagueness. As you might expect, this was just a first-shot approach to issues that have been bothering me. I am now preparing a paper that discusses rule modularity. I believe that you will find that it clarifies my arguments from last week. Briefly, I see now that the problem is not so much with the CF's you have proposed, but is instead a general issue concerning all rules.

As for the furor, as far as I am concerned, your rules have the precise property I predicted last August would not occur, namely, small CF's. What will come of this discussion, I believe, is primarily a better understanding of rules. More on this later in the week.

I will now briefly reply to your numbered remarks:

1. You will notice that I said the system was insensitive to variations in CF>0.2 in so far as rule success and invocation are concerned. This excludes calculations that use CF's in percentage cutoffs. Do you have other sensitivities in mind?
2. It was Larry who told me that you wanted to form a large rule set from the combinations of these rules. Perhaps this was only the gist of a side argument that centered on allowing for all cases. I look forward to hearing about these "subtle clinical situations that would otherwise not be covered."
3. I have no problem with the 0.2 cutoff, per se.

Bill

---

3 March 1976
From: Clancey
Subject: Modularity of rules
To: Yu
cc: MYCIN gang

I have completed a write-up of my understanding of what we mean by rule independence. I consider this useful as a tutorial to those who perhaps have not fully appreciated the significance of the constraint P(e1 & e2| h) = P(e1|h)*P(e2|h), which is discussed in several of Ted's write-ups on the relation of CF's to probabilities.

For those of you for whom this is old hat by now, I would appreciate it if you would peruse my memo and let me know if I've got it straight.

I've expanded the discussion of plausibility of rule interaction here also. This appears to be an issue worth pursuing.

The memo is CF.MODULAR on my directory. It is about 3 pages long.

Bill Clancey

<CLANCEY>CF.MODULAR.1

I. Introduction

This memo arose from my desire to understand rule CF's of less than the 0.2 threshold. How could such a rule be evidence of something? Does a rule having a CF less than 0.2 pose any problems to the process of combining certainty factors? What does it mean to say that a rule is modular? Must a rule satisfy some property relating to its certainty factor to be considered modular?

After thinking out all of these problems for myself, I re-examined our publications in the light of my new understanding. Alas! The ideas discussed below have long been known and were simply overlooked or undervalued by me. Indeed, I suspect that most of us have to some degree failed to appreciate Ted's thesis, from which I will be quoting below.

II. What Is Modularity?

The following is a restatement of one requirement for rule independence. As Ted discusses in CF.MEMO, it is a necessary assumption for our combining functions to be consistent with probability theory, namely: P(e1 & e2|h) = P(e1|h)*P(e2|h), and the same for ~h (e=premise and h= action of rule).

Let {Ri} be a subset of the UPDATED-BY rules for some parameter P, all of which mention the same value for P in the conclusion, namely VALUEP, though perhaps with different certainty factors. (If P is a yes-no parameter, then this set contains all of the UPDATED-BY rules.) Now let P! be the power set of R, and for every element of P!, let PREMi designate the union of the premises of all rules Rj in the power set element i.

Now for every PREMi that is logically consistent (no subset of premises is unsatisfiable), it must be the case that the CF applied to the new rule PREMi→VALUEP is given by the combining function applied over all rule CF's in the power set element. If so, we can say that these original rules are independent logically and so can contribute evidence incrementally, regardless of the pattern of succession or failure of the set.

This is a requirement for rule modularity. It can also be shown [working from assumption 9 of the memo: P(e1 & e2) = P(e1)*P(e2)] that premises must be independent "for ALL rules dealing with a clinical parameter regardless of the value specified (e.g., all rules that conclude anything about the identity of an organism). This assumption is generally avoided by Bayesians. I have not examined our rules closely with this assumption in mind, but I suspect we may discover several examples of nonindependent PREMISES" (Shortliffe, CF.MEMO). This is a generalization of the above restriction, which I believe is more intuitive.

It is worth reviewing at this time some of the related restrictions on rules and CF's mentioned in Ted's thesis.

A. Given mutually exclusive hypotheses hi for an observation e, the sum of their CF's, CF(hi,e), must not exceed 1. (From CF.MEMO, page 7: "We often find that this rule is broken.")

B. "We must insist that dependent pieces of evidence be grouped into single rather than multiple rules."

C. "The rule acquisition procedure requires a screening process to see if the new rule improperly interacts with other rules in the knowledge base."

Some of the consistency checks Ted discusses are subsumption and rule contradictions.

III. Understanding Modularity

I did not fully appreciate these problems, even after several readings over the past year, until I worked out an example containing nonindependent rules.

Example: Consider the following rules having CF's that I believe to be valid. The rules would be used in a consultation system for deciding whether or not to carry an umbrella.

Rule A: If the weatherman said that there is a 20% chance of rain today, then I expect it to rain today (0.1).

Rule B: If it is summer, then I expect it to rain today (−0.9).

Rule C: If there are many clouds, then I expect it to rain today (0.1).

Now let these rules succeed in various combinations:

Power set element	*Computed CF*	*Preferred CF*	*Evaluation*
A & B & C	−0.71	0.5	wrong
A & B	−0.8	0.21	wrong
A & C	−0.19	?	okay
B & C	−0.8	?	wrong

These rules are not modular—the combined CF does not correspond to what I believe when I form the combination of the premises in my mind. Specifically, I give far more weight to clouds and weatherman's prediction of rain in the summer (when I expect neither) than in the winter (when clouds and 20% chance are common).

Using Webster's definition of belief, "the degree of mental acceptance of an idea or conclusion," I think that it would be fair to say that I DO NOT believe the conclusion of "rain today," given premises A, C, A & C, or B & C. As far as MYCIN's operation is concerned, this corresponds to a CF<0.2. The CF combining function has not worked above because my rules are not independent. (It is also possible for independent rules to combine improperly because the combining function is wrong—more on this later.)

Looking again at the rules I wrote above, I feel that Rule A in particular is a bad rule. It takes a mere fragment of an argument and tries to draw a conclusion. Now admittedly we know something, given that 20% was predicted, but we are being logically naive to think that this fact alone is worth isolating. It depends radically on other information for its usefulness. Moreover, the context in which it is true will radically determine the conclusion we draw from it. We saw above that in summer I am far more inclined to give it weight than in winter. The only thing I AM willing to say given just this clause is that it probably won't be fair (0.3). (Like Wittgenstein, I ask myself, "What do I know now?")

IV. Implications for 0.2 Rules

I see now that the problem I was anticipating in my earlier message will hold if the rules are not modular. My fear was that a rule having a CF<0.2 was more likely to have a premise that was incomplete than was a rule of CF>0.2. I understand now that a 0.2 rule, like any other rule, is acceptable if there is no known argument that involves its premise with that of another

rule, other than one that simply adds the evidence together incrementally according to the combining function. A new argument that is built from the evidence mentioned in the other rules is proof that the individual rules are not modular. (Subsumption is an explicit form of this.) Thus, Victor's claim that he wants to allow for all combinations MUST rest on the inherent independence of his premise sets. Again, no conclusion whatsoever should be drawn from the coincidence of any combination of premise sets, other than that arrived at by the CF combining function. Moreover, every conclusion collected incrementally by the combining function must be one Victor would reach with the same strength, given that union of premise clauses (cf., B & C above). In fact, I am willing to believe now that a rule having a CF<0.2 is perhaps MORE likely to be independent because it wouldn't have been given such a small CF unless the author saw it as minimally useful. That is, it stands on its own as a very weak observation having no other inferential value (I am still wary of calling it "evidence"). If it had a higher CF, it would almost certainly be useful in combination with other observations. Based on Victor's decision to separate meningitis clinical and lab rules, I conclude that doctors do not have the ability to relate the two. Is this correct? I believe that Ted has also questioned Victor's rules in this respect.

V. Plausibility

The problem of plausible combination of rules is difficult to anticipate because it is precisely the unanticipated coincidence of rule success that we are most likely to find objectionable. Suppose that we do find two rules D and E that we can't imagine ever succeeding at the same time, yet there is no logical reason for this not to occur (i.e., the rules are not mutually exclusive; not always easy to determine since all rules that cause these rules to be invoked must be examined). In this case we should try to define a new parameter that explains the connection between these two parameters, which we do not as yet understand. (A method of theory formation: ask yourself "What would I think if these two pieces of evidence were true?" Perhaps the actions are in conflict—why? Perhaps the premises never appear together (usually aren't both true)—why not? Do this for the power set of all evidence under consideration.)

VI. What Does This Say About MYCIN's Rule Set?

(1) They must be disjoint (mutually exclusive) within an UPDATED-BY subset, or (2) the parameters in the premises of rules that succeed together must be logically noninteracting. This means that there must be nothing significant about their coincidence. Their contribution separately must be the same as an inference that considers them together. [In pseudochemical terms, the rule CF is a measure of (logical) force, which binds together the clauses of the premise in a single rule.]

Taking my example, I should rewrite the rules and form a new set including A & B & C and A & B. Rules A and C are incomplete. They say nothing here because they say something when a context is added. Leaving them separate led to a nonsensical result (B & C), which CF theory claims should make sense. This is an example of where plausibility of rule interaction must be made at rule acquisition time. Indeed, I believe now that unless we require our rules to be disjoint within an UPDATED-BY set, it will be very difficult to say whether or not a rule is modular. For too long I have assumed

that because a rule looks like a discrete object it is necessarily modular. I have assumed that it is sufficient to have a CF combining function that models adequately the process of incrementally collecting evidence, forgetting that this evidence MUST be discrete for the function to be valid. Otherwise, a FUNCTION is replacing a logical argument, which a rule unifying the premises would represent.

VII. Making Rules Modular

It remains to detect if MYCIN's rules are modular. We must look for premises that are still "charged" with inference potential, as measured relative to clauses in other rules. Victor has said that his rules are modular (at least the ones having CF<0.2). If so, there is no problem, though we should be wary about the 0.05/0.15 distinctions. (How is it that "evidence" that is too weak to yield an acceptable conclusion nevertheless is definite enough to be put in one of three CF categories: 0.05, 0.10 and 0.15?)

One method for detecting rule modularity is as follows. Given, for example, three rules A, B, and C, where B and C have the same CF (all three mention VALUEP), then if A & B and A & C are determined to have different certainty factors (where & denotes the process of combining the rules into a single rule), then the rules A, B, and C aren't modular.

On the other hand, given two rules A and B known to be modular (our knowledge of the domain cannot yield an argument that combines the premises), then A & B must have a CF given by the combining function (obviously true for disjoint rules). This gives us a way for evaluating a combining function.