ARTIFICIAL INTELLIGENCE DIALECTS OF THE BAYESIAN BELIEF REVISION LANGUAGE

by

Shimon Schocken
Information Systems Department
Leonard N. Stern School of Business
New York University
90 Trinity Place
New York, New York 10006

and

Paul R. Kleindorfer
Decision Sciences Department
The Wharton School
University of Pennsylvania

Revised September 1989

Please direct all inquiries and comments to Shimon Schocken, 624 Tisch Hall, 40 West 4th Street, New York, NY 10003, or by electronic mail, bitnet, to sschocken@nybvx1.

Artificial Intelligence Dialects of the Bayesian Belief Revision Language

SHIMON SCHOCKEN AND PAUL R. KLEINDORFER

Abstract —Rule-based expert systems must deal with uncertain data, subjective expert opinions, and inaccurate decision rules. Computer scientists and psychologists have proposed and implemented a number of belief languages designed to cope with these uncertainties. These languages are widely used in applied systems, and their normative validity is clearly an important question, both on practical as well on theoretical grounds. Several well-known belief languages are reviewed, and both previous work and new insights into their Bayesian interpretations are presented. In particular, the authors focus on three alternative belief-update models: the certainty factors calculus, Dempster–Shafer simple support functions, and the descriptive contrast/inertia model. Important "dialects" of these languages are shown to be isomorphic to each other and to a special case of Bayesian inference. Parts of this analysis were carried out by other authors; these results were extended and consolidated using an analytic technique designed to study the kinship of belief languages in general.

I. BELIEF LANGUAGES IN ARTIFICIAL INTELLIGENCE

DURING the past decade, computer-based expert systems have emerged to become the most applied facet of artificial intelligence (AI). To date, expert systems have proven to be particularly effective in diagnostic tasks, e.g., antimicrobial selection and interpretation of geological data (Duda and Shortliffe [15]). Such "domains of expertise" are typically characterized by uncertain field data, subjective expert opinions, and inexact decision rules. The challenge of dealing with these uncertainties has stirred a lively debate among developers of expert systems on the one hand, and faithful followers of the Bayesian religion on the other. First, the drive to experiment with real expert systems has led AI researchers to implement *ad hoc* uncertainty mechanisms that are rather limited on normative grounds. This, in turn, has drawn criticism from Bayesian writers who, nonetheless, were forced to admit that the classical methods they preached did not always scale up to realistic applications. As the reference section of this paper indicates, the result was an inspiring exchange of ideas that is currently going strong in many academic circles in computer science, statistics, and psychology.

Just like any other formal model, an expert system imposes a rigid structure on the problem it attempts to support. In particular, rule-based systems make the assumption that expertise can be captured through a modular set of inference rules. These rules are supposed to represent "objective" knowledge as well as subjective expert opinions. To illustrate, consider the following familiar problem.

A tenured professor (hereafter referred to as a "recruiter") attempts to guess the academic potential of a candidate to a junior faculty position. The information available to the recruiter is the typical mix of resume, papers, and recommendation letters, along with his own past recruiting experience. The review process is complicated by the fact that many young Ph.D.'s are competing for the same slot; therefore, the goal of the recruiter is to rank-order the candidates in terms of their prospective academic potential. The overall criterion for academic success is taken to be the perceived likelihood of the proposition, "the candidate will be offered a tenured position in our department within the next decade."

A recruiter is said to be an "expert" if his predictions consistently exhibit a great deal of external validity. To make an extreme case, let us assume that our recruiter is a perfect predictor: all the candidates that he has recommended in the past were subsequently offered tenure, and all the people that he has rejected were refused tenure in similar departments. Furthermore, our expert is willing to describe his proven recruiting rationale in terms of a set of inference rules. These rules represent, in his mind, the perceived academic significance of various credentials. Given this and other relevant information, is there a plausible model that can credibly synthesize this rule-base into a belief about the academic potential of a particular candidate? Can this model be further implemented in a computer-based system designed to carry out routine screening of candidates? These questions must be addressed by developers and users of rule-based expert systems.

For the sake of simplicity, we assume that the credentials of prospective candidates can be enumerated through a finite set of dichotomous propositions $\{e_1, \cdots, e_n\}$. For example, e_1 might say that the candidate has an undergraduate degree in mathematics, and e_2 that he has extensive consulting experience. In a rule-based system, these clues are related to various hypotheses through a set of

rules, elicited from a recruiting expert. For example, our recruiter might suggest that, more often than not, a math degree e improves tenure prospects h. This inexact implication might be represented as

if e then h with degree of belief $\mathrm{bel}(h, \{e\})$.

The meaning of the quantitative degree of belief $\mathrm{bel}(h, \{e\})$ in this logical context is an open question, tracing back to Carnap's work on inductive logic [7]. It is tempting to give these degrees of belief a probabilistic interpretation; nonetheless, scores of writers have demonstrated that logic and probability do not mix very well. This difficulty, combined with the pressure to develop rule-based systems that can handle uncertainty, has led to a proliferation of incompatible quasi-probabilistic belief languages. In this paper we review and analyze these languages from a Bayesian perspective. Our general approach is as follows: given a particular belief language, we wish to address three questions, related to the syntax, calculus, and semantics, respectively, of the underlying language.

1) What is the Bayesian interpretation of the number $\mathrm{bel}(h, \{e\})$?
2) Given a set of degrees of belief $\mathrm{bel}(h, \{e_1\}), \cdots,$ $\mathrm{bel}(h, \{e_n\})$ and a belief combination rule $C(\mathrm{bel}(h, \{e_1\}), \cdots, \mathrm{bel}(h, \{e_n\}))$, what is the Bayesian interpretation of this rule?
3) Do the answers to 1) and 2) place any constrains on the scope of inference problems that rule-based systems can credibly solve?

It goes without saying that proper knowledge engineering requires that degrees of belief be elicited, computed, and interpreted in a credible manner. Beside the normative significance of this objective, degrees of belief should be handled cautiously since the plausibility of the final set of hypotheses generated by an expert system is typically a function of the algebraic ranking of their posterior degrees of belief. For example, the medical diagnosis program internist (Pople [43]) uses the disease with the highest ranking degree of belief as an anchor, around which a fuzzy subset of plausible hypotheses is defined. Thus, both the external as well as the internal validity of an expert system are directly related to the validity of its underlying belief language.

Nonetheless, it seems by now that no single belief language exists that will be effective, efficient, and plausible for all possible applications and, at the same time, satisfy all knowledge engineers, experts, users, and researchers. This realization has led to a proliferation of new belief languages in the last decade. At the foundation of any of these languages lies either a descriptive or a normative argument about human judgement under uncertainty. A *descriptive* language attempts to capture the way experts actually reason; hence its measure of performance is its capability of simulating human judgment. In contrast, a *normative* model is based on the premise that human

reasoning under uncertainty is often suboptimal; rather than attempting to replicate it blindly, the normative approach is more concerned with the proximity of the system's judgment to such rigorous standards as logic or probability theory. It is well-known by now that normative models are not necessarily consistent with descriptive models, as has been demonstrated by numerous studies in cognitive psychology and in decision sciences. For a good discussion of this dichotomy, the reader is referred to Baron [4].

The classical method for representing and updating degrees of belief is the Bayesian language, which is consistent with the axioms of subjective probability. Researchers who attempted to implement this method in expert systems, however, were quick to discover three major limitations: a) standard manipulations of discrete joint distribution functions are computationally complex (Pearl [40]); b) the classical Bayesian language does not lend itself easily to representing ambiguity and fuzzy expert opinions (Shafer [50])[1]; and c) human reasoning under uncertainty is systematically inconsistent with Bayesian inference (Tversky and Kahneman [57]).

Efforts to curb the inherent limitations of a complete Bayesian design have taken several directions. Early expert systems such as MYCIN (Shortliffe [53]) and Prospector (Duda et al. [14]) employed *ad hoc* belief languages based on *certainty factors* and subjective *likelihood ratios*, respectively. The resulting belief-update mechanisms were only partially consistent with the axioms of subjective probability. These "first generation" systems gave way to a renewed interest in probability theory, led by Howard's [28] work on *influence diagrams* and Pearl's [40] *belief networks*. Pearl's methods of propagating probabilities through a network of propositions are consistent with standard probability theory. In the case of singly connected networks, the run-time of Pearl's algorithm is polinomial with the size of the network. The problem of probabilistic belief-update in a general network has been shown to be *NP*-hard (Copper [11]).

Perhaps the most important development in the quest to "extend" the Bayesian language is Shafer's work on *belief functions* and the corresponding Dempster rule for combining them (Shafer [50]). The theory of belief functions has a rigorous mathematical foundation based on a relaxation of the additivity axiom of probability theory. The resulting Dempster–Shafer language provides explicit tools for dealing with ambiguity and "uncommitted belief." There has recently been a surge of interest in the Dempster–Shafer language within the AI community, and a considerable number of expert systems and expert system shells already employ this technique (e.g., Baldwin and Monk [3]).

Finally, there have been several attempts in decision sciences and cognitive psychology to specify descriptive or

---

[1] There have been attempts to deal with second-order probability within the Bayesian language, e.g., Pearl [41], Baron [5], and Kyburg [31].

"behavioral" belief languages. For example, Einhorn and Hogarth [19] proposed a pragmatic anchoring and adjustment belief-update model called the *contrast/inertia* language. Such experimental works are typically based on trying to fit a mathematical model that best explains the behavior of subjects in a controlled experiment. To the extent that this empirical effort can be extended to model the reasoning of successful experts rather than naive subjects, this approach clearly merits attention from practitioners of expert systems.

In the 1970's, the development of new belief languages was carried out. primarily by people outside the mainstream of the statistics community. This line of research went by and large unnoticed until the 1980's. Since then, a number of faithful Bayesians have begun to defend their turf by arguing that every new belief language can be matched with a plausible probabilistic interpretation. As Kyburg [31] puts it, "it is appropriate to examine the formal relations between various Bayesian and non-Bayesian approaches... in order to explore the question of whether the new techniques are really more powerful than the old, and the question of whether, if they are, this increment of power is bought at too high a price." The present work is motivated by this line of thought. We will review previous work and present new results on the Bayesian interpretations of important versions of the certainty factors, contrast/inertia, and Dempster–Shafer languages.

The paper draws on previous work by Heckerman [26], Hajek [25], and Grosof [23]. Most of what we currently know about the normative validity of the certainty factors language is due to Heckerman's interpretation of certainty factors as transformed likelihood ratios. Hajek has also investigated the normative validity of alternative belief revision calculi, but his point of departure was not probabilistic at all. He proposed an axiomatic calculus consisting of four plausible combination functions and went on to analyze the algebraic structure of the set of degrees of belief induced by these functions. This analysis led to his conclusion that the certainty factors language and the *ad hoc* Bayesian language used in Prospector induce isomorphic sets of degrees of belief. This isomorphism, however, was based on nonprobabilistic algebraic mappings. In this paper we take a different approach, as follows.

The paper commences with a review of rule-based inference, the context in which belief languages are used in artificial intelligence. A Bayesian language is then presented, and its underlying (extraprobabilistic) rationality is demonstrated. This is done to motivate our choice of the Bayesian language, as opposed to another *desiredata* (e.g., Hajek's), as a standard against which other belief languages will be compared. We then present an analytic methodology, which, lacking a better name, is termed *dialectal analysis*. This set of tools, designed to investigate the kinship to two or more belief language, is then used to describe the implicit relationships among the certainty factors, contrast/inertia, and Dempster–Shafer languages. Each of these languages is discussed in a separate section

which has three parts. First, a brief overview of the language is given. Second, a Bayesian interpretation is presented and justified. Third, the mathematical implications of the interpretation are discussed. A discussion section integrates the preceding results and suggests future research directions.

II. RULE-BASED INFERENCE UNDER UNCERTAINTY

The two major building blocks of an expert system are the knowledge base and the inference engine. The *knowledge-base* is conceptually a directed graph consisting of propositional nodes and inferential arcs. The boundary of the graph consists of a set of competing hypotheses (e.g., diseases in a medical diagnosis application) and a set of observable clues (e.g., diagnostic symptoms). Inner nodes represent subhypotheses (e.g., clinical syndromes). The directed arc connecting nodes e and h represents a direct inferential relationship between e and h, and the arc's label is the strength of this relationship, or the *degree of belief*, bel (h, e).[2]

The *inference* engine is a search algorithm that prunes this evidential network and applies modus ponens repetitively. One difference between this and standard theorem proving stems from the uncertainty associated with rules: as the inference engine prunes rules that ultimately imply a hypothesis, a *belief calculus* is applied to update the posterior belief in this hypothesis. This noncategorical reasoning process terminates when the belief in one or more hypotheses exceeds a certain predefined cutoff value. At least in theory, this value should be based on the cost of gathering additional evidence and on the consequences of committing type I and II errors.

The preceding paragraph emphasizes the central role that belief calculi play in noncategorical rule-based inference. According to Shafer and Tversky [51], the building blocks of a belief language are syntax, calculus, and semantics. In the context of this paper, we define *syntax* to be the set of all degrees of belief that are relevant to a particular hypothesis h. Typically, a set of atomic degrees of belief bel$(h, \{e_1\}), \cdots,$ bel$(h, \{e_n\})$ is elicited directly from a human expert, while compound degrees of belief are computed *ad hoc* through a set of operators collectively known as a belief calculus.

A completely specified rule-based belief calculus must consist of five combination functions: parellel combination (combining the degrees of belief rendered by two or more independent rules), sequential combination (combining the rule's degeree of belief with the uncertainty associated with the rule's antecedent), and three logical combination functions (for negations, dijunctions and conjunctions of uncertain pieces of evidence). In this paper we focus on parallel combination only.

With regard to *semantics*, we take the position that the semantics of a particular belief language is given in terms

---

[2] From now on, the notation bel(h, e) is shorthand for bel$(h, \{e\})$.

of a mapping from the syntax and calculus dimensions of the language onto the theory of subjective probability. To the extent that the latter theory is taken to be a norm for rationality, this mapping provides normative validity to the syntax and calculus dimensions of the belief language in question.

It is reasonable to assume that any rule-based belief calculus with be founded on some parameterized variant of the following model:

$$\text{bel}(h, \{e_1, \cdots, e_n\})$$

$$= C(\text{bel}(h), \text{bel}(h, e_1), \cdots, \text{bel}(h, e_n)). \quad (1)$$

The single-place degree of belief $\text{bel}(h)$, which is a shorthand of $\text{bel}(h, \varnothing)$, represents prior belief, i.e., unconditional belief in a hypothesis that is not supported by any evidence. For example, $\text{bel}(h)$ might be the "objective" belief in the tenure prospects (in one particular department) of a random draw from the typical annual crop of Ph.D's. The degree of belief $\text{bel}(h, e_i)$ measures the degree of support that the clue e_i renders to the hypothesis h. Depending on the underlying belief language, this evidential relationship might take two forms. Under a *diagnostic* mode of inference, $\text{bel}(h, e)$ parameterizes the rule $\langle$IF e THEN $h\rangle$ and represents the expert's belief in the likelihood of h in light of the supporting evidence e. Under a *causal* interpretation, $\text{bel}(h, e)$ parameterizes the rule $\langle$IF h THEN $e\rangle$, representing the degree of belief in the effect e occurring given its retrospective cause h. The direction of inference in rules varies across belief languages, and we therefore leave it unspecified at the abstract level of (1). Finally, given a set of evidence $E = \{e_1, \cdots, e_n\}$, the posterior belief in h in light of E is computed by the belief synthesis operator C.

Depending on the properties of C, it is sometimes possible to rewrite bel recursively, making it an anchoring-and-adjustment belief-update model:

$$\text{bel}(h, \varnothing) = \text{bel}(h)$$

$$\text{bel}(h, \{e_1, \cdots, e_n\})$$

$$= C'(\text{bel}(h, \{e_1, \cdots, e_{n-1}\}), \text{bel}(h, e_n)). \quad (2)$$

Focusing on some intuitive properties of belief-update, we might require that bel be commutative, i.e.,

$$\text{bel}(h, \{e_1, e_2\}) = \text{bel}(h, \{e_2, e_1\}),$$

and associative, i.e.,

$$\text{bel}(h, \{e_1, \{e_2, e_3\}\}) = \text{bel}(h, \{\{e_1, e_2\}, e_3\}).$$

These properties are quite plausible on rational grounds; we wouldn't like a physician to change his diagnosis simply because the order by which information is presented to him is altered. In general, one is free to construct an

axiomatic desiderata regarding the intuitive properties of bel, using it further to evaluate formal belief languages. Such an approach was undertaken by Cox [12], Savage [47], Popper [44], and, most recently, Horvitz et al, [27]. For example, Cox enumerated seven intuitive properties of belief-update and proceeded to prove that the resulting bel function is a probability. By augmenting Cox's framework with three additional intuitive properties, Horvitz et al. have shown that any belief-update measure that satisfies the extended framework must be equal to some monotonic transformation of a likelihood-ratio.

Note that the belief-update model (1) is neither "good" in any philosophical sense, nor does it reflect any plausible desiderata. At the same time, the general form of (1) is largely dictated by the rule-based architecture, which assumes that a) wholistic expert knowledge can be decomposed into a finite series of discrete observations (rules), and b) that subsets of this knowledge-base can be synthesized or "rolled back" into posterior beliefs. Suppose that we accept a) and b) as reasonable restrictions on the subset on inference problems to be studied. Can we define this subset more precisely? Is it true that different belief languages are capable of modeling different subsets of inference problems? These questions will be addressed in what follows.

The term *inference problem* introduced in the last paragraph refers to an ordered set of propositions, say $\langle h, e_1, \cdots, e_n \rangle$, in which h and $\{e_1, \cdots, e_n\}$ are interpreted as a hypothesis and relevant body of evidence, respectively. Given this terminology, the "solution" of an inference problem refers to the correct Bayesian computation of the posterior belief in h in light of $\{e_1, \cdots, e_n\}$. From a probabilistic standpoint, $\{h, e_1, \cdots, e_n\}$ is viewed as a space of dichotomous random variables, characterized by a joint distribution function $P: W \rightarrow [0,1]$ where W is the set of 2^{n+1} boolean permutations defined over the space. Under this Bayesian interpretation, the ultimate goal of the inference process is to compute the posterior probability $P(h|e_1, \cdots, e_n)$. It is easy to see, however, that any brute-force attempt to compute this conditional probability from the joint distribution function $P(h, e_1, \cdots, e_n)$ is bound to be exponential in n.

The preceding paragraph assumes that P is known. In reality, this is clearly not the case. For example, the inference problem $\langle h, e_1, \cdots, e_n \rangle$ might represent the inferential relationships that exist between, say, a hidden oil deposit h and its geological manifestations $\{e_1, \cdots, e_n\}$. Clearly, complete knowledge of the joint distribution function $P(\cdot)$ will rarely be credibly available and $P(\cdot)$ will have to be generated piecemeal using partial data and subjective expert judgment. At the same time, the elusive P might serve as a mechanism for defining classes of inference problems which vary in terms of their computational complexity. This is motivated by the notion that the cognitive complexity of an inference problem has something to do with the mathematical modularity of its underlying P. With that in mind, the following P characterization of inference problems is of special interest, as will be seen

shortly:

Ratio-Form Conditionally Independent Problems: An inference problem $\langle h, e_1, \cdots, e_n \rangle$ is said to be ratio-form conditionally independent (Grosof [23]) if

$$\frac{P(e_1, \cdots, e_n | h)}{P(e_1, \cdots, e_n | \bar{h})} = \frac{P(e_1 | h)}{P(e_1 | \bar{h})} \cdot , \cdots, \cdot \frac{P(e_n | h)}{P - (e_n | \bar{h})} . \quad (3)$$

Many writers have haphazardly read (3) to imply that $e_1, \cdots, e_n$ are conditionally independent given both h and $\bar{h}$.[3] Clearly, (3) reflects a weaker notion of modularity which is necessary but not sufficient for the latter assertion.

We now turn to describe a well-known version of the Bayesian language that is limited to ratio-form conditionally independent problems. This language is used as a standard against which other belief languages will be compared. This strategy (as opposed to formulating a desiredata) is based on the premise that the Bayesian belief-update model has extraprobabilistic tenants of rationality that cannot be debated by a reasonable person. We hope that the following section will help to convince skeptical readers that this is indeed true.

Before delving into this discussion, we take the liberty of confining the analysis to inference problems $\langle h, \{e_1, e_2\} \rangle$ consisting of one hypothesis and two pieces of evidence. All the findings reported here can be easily extended to any finite number of pieces of evidence, as long as they are all directly connected to the single hypothesis in question. The inclusion of intermediate hypotheses presents computational problems which normally require heuristics and/or very restrictive assumptions on the underlying joint distribution function. Bayesian inference in complex networks is a challenging and active research area and the interested reader is referred to Cheeseman [9], Lemmer [33], Cooper [10], Pearl [40], and Shenoy and Shafer [52].

III. THE RATIO-FORM BAYESIAN LANGUAGE

Degrees of belief in the ratio-form Bayesian language are expressed in terms of odds and likelihood ratios: the posterior belief in a hypothesis h in light of $E = \{e_1, e_2\}$ is the odds $\text{bel}(h, E) = P(h|E)/P(\bar{h}|E)$, P being a standard subjective probability. The degree of belief in the causal rule $\langle$ if h then $e\rangle$ is represented through the conditional likelihood ratio $P(e|h)/P(e|\bar{h})$. With that in mind, the Bayesian calculus is viewed as a mechanism designed to synthesize a set of causal degrees of belief into a combined posterior belief. The remainder of this section presents the derivation of this calculus. Similar derivations were carried out in numerous papers in decision theory and in AI, e.g., Peirce [42], Edwards *et al.* [17], and Charniak [8].

A. Derivation of the Bayesian Calculus

We begin with Bayes rule, applied to both $P(h|e_1, e_2)$ and $P(\bar{h}|e_1, e_2)$:

$$P(h|e_1, e_2) = \frac{P(e_1, e_2 | h)}{P(e_1, e_2)} \cdot P(h) \quad (4)$$

$$P(\bar{h}|e_1, e_2) = \frac{P(e_1, e_2 | \bar{h})}{P(e_1, e_2)} \cdot P(\bar{h}). \quad (5)$$

Dividing (4) by (5) gives the odds-ratio version of Bayes rule:

$$\frac{P(h|e_1, e_2)}{P(\bar{h}|e_1, e_2)} = \frac{P(e_1, e_2 | h)}{P(e_1, e_2 | \bar{h})} \cdot \frac{P(h)}{P(\bar{h})} . \quad (6)$$

Now, if the problem $\langle h, e_1, e_2 \rangle$ is *ratio-form conditionally independent*, (Grosof, [23]) (6) reduces to the following definition of the Bayesian calculus:

$$\frac{P(h|e_1, e_2)}{P(\bar{h}|e_1, e_2)} = \frac{P(e_1 | h)}{P(e_1 | \bar{h})} \cdot \frac{P(e_2 | h)}{P(e_2 | \bar{h})} \cdot \frac{P(h)}{P(\bar{h})} \quad (C1)$$

or, using a simplified notation,

$$R(h|e_1, e_2) = L(e_1 | h) \cdot L(e_2 | h) \cdot R(h). \quad (7)$$

The elimination of the joint distribution of evidence $P(e_1, e_2)$ in the step from (4) and (5) to (6) has important practical implications. First, the elicitation of any joint distribution function is a painstaking undertaking which should be avoided whenever possible. Second, (6) is completely independent of the degree of uncertainty associated with the evidence $E = \{e_1, e_2\}$: if a clue e is uncertain, i.e., $P(e) < 1$, this uncertainty is as relevant to h as it is to $\bar{h}$, and therefore it can be canceled out. Finally, the elimination of $P(e_1, e_2)$ gives the knowledge engineer a wide choice of elicitation techniques, as any one of the following measures might be used to parameterize the evidential support that e renders to h: the pair of probabilities $\langle P(e|h), P(e|\bar{h}) \rangle$, the likelihood ratio $P(e|h)/P(e|\bar{h})$, or the betting odds $O(h|e)/O(h) = (P(h|e)/P(\bar{h}|e))/(P(h)/P(\bar{h}))$. The latter two expressions are equivalent, and any one of them can be derived from the former. This variety is important, as different applications and different experts might prefer one elicitation method over the other.

We now turn to discuss some nonmathematical properties of rationality that are consistent with the Bayesian calculus (C1).

B. On the Underlying Rationality of the Bayesian Language

The *subjective*[4] school of probability (e.g., Ramsey [46]) is based on the argument that the semantic interpretation of probability can be given in terms of rational human judgment under uncertainty. Proponents of this school of thought argued vigorously and quite convincingly that any

---

[3] $\bar{h}$ hereafter denotes "not h."

[4]Also referred to as *Bayesian* or *personal* school of probability (e.g., Ramsey [46]).

other interpretation of probability is merely a special case of the subjective philosophy (de Finetti [13]). As Savage [47, p. 67] puts it, "the personalistic view incorporates all the universally acceptable criteria for reasonableness in judgment known to me... when any criteria that may have been overlooked are brought forward, they will be welcomed into the personalistic view."

Following Savages's line of thought, we wish to show that, given some plausible interpretations, important "dialects" of new belief languages are indeed special cases of the classical Bayesian language. The latter language is considered a gold standard because, normatively speaking, it is consistent with many intuitive principles of rational judgment. Some of these principles, which are all consistent with (C1), are discussed below. In what follows, a human (or mechanized) agent who behaves in accordance with (C1) is called a *Bayesian judge*. A human agent who entertains an abstract belief-update model which may or may not be related to (C1) is called a *human judge*.

Proper Synthesis of Degrees of Belief: The tendency of human judges to underweight or ignore base-rate information is a well-known manifestation of the *representativeness bias* (Tversky and Kahneman [57]). For example, consider a candidate for a junior faculty position who already has several major publications under his belt. We argue that the fact that the candidate is a prolific writer (e) might cause a human judge to *overestimate* his future tenure prospects (h). This overoptimism occurs either by misinterpreting $P(e|h)/P(e|\bar{h})$ to be $P(h|e)/P(\bar{h}|e)$, or by letting a high and salient diagnostic impact $P(e|h)/P(e|\bar{h})$ to overshadow the "dull" background information that only, say, 20 percent of new recruits are ultimately promoted to tenure.

This bias will not distract a Bayesian judge who adheres to (C1), where the base-rate information $P(h)/P(\bar{h})$ is explicitly represented and carries the same weight as $P(e|h)/P(e|\bar{h})$. Furthermore, (C1) is both commutative and associative, meaning that the evidential impact of clues is independent of order and clustering effects. This is in sharp contrast to human judgment that is prone to such belief synthesis biases as primacy effect (Anderson [2]) recency effect (Lopes [35]), misinterpretation of new evidence (Nisbett and Ross [37]), conservatism (Edwards [16]), and a host of other "averaging" rules that violate (C1) (Slovic and Lichtenstein [54]).

Fusion of Quantitative and Qualitative Evidence: The formation of a posterior belief in an uncertain hypothesis typically requires a joint consideration of factual information as well as subjective opinions. For example, consider the two propositions *prolific researcher* e_1 and *good teacher* (e_2). A Bayesian inference system that evaluates tenure prospects (h) will have to use, among other things, the probabilities $P(e_1|h)$ and $P(e_2|h)$. Where do these numbers come from? The probability that a tenured professor is a prolific researcher can be obtained in a frequentist fashion, counting the number of prolific researchers among a known sample of tenured professors. Teaching ability may be a more elusive property, and the probability

$P(e_2|h)$ can be elicited as a "personal" degree of belief, using a domain expert. According to the subjective school of probability, the scope of rational values that the frequentist $P(e_1|h)$ and the subjective $P(e_2|h)$ may attain is constrained by the same set of axioms. Due to this uniformity, the Bayesian language provides a homogeneous framework in which stochastic and epistemic degrees of belief are combined through the same calculus.

Equal Attention to Positive and Negative Evidence: Human judges who are left to their own devices are known consistently to seek and overweigh supportive evidence at the expense of neglecting or underestimating negative clues (Koriat *et al.* [30]). At the same time, a knowledge engineer who is guided by (C1) is forced to discount supportive belief of the form $P(e|h)$ with its negative complementary belief $P(e|\bar{h})$. For example, the impact of a good record of publications (e) on tenure prospects (h) is automatically discounted by the observation that good research is also produced by nontenured candidates ($P(e|\bar{h}) > 0$). In other words, (C1) forces the Bayesian judge to "dilute" his belief $P(e|h)$ with the complementary belief $P(e|\bar{h})$; said otherwise, if (C1) is transformed into a log-linear scoring formula, both $P(e|h)$ and $P(e|\bar{h})$ will have the same (unit) "coefficient of importance" in the resulting model. Cognitive distortions of the "dilution effect" were discussed by Nisbett *et al.* [38].

The More Information, the Better: A close examination of (C1) reveals that the Bayesian calculus is consistent with the commonly held principle that data (pieces of evidence) should never be thrown away. More precisely, Savage [47, p. 48] has shown that for n sufficiently large, the probability, given that h is true, that the likelihood-ratios product $L(e_1|h)\cdot, \cdots, \cdot L(e_n|h)$ (as in (7)) is greater than any preassigned number is almost one, i.e.,

$$\lim_{n \to \infty} P(L(e_1|h)\cdot, \cdots, \cdot L(e_n|h) > c|h) = 1$$

for $0 < c < \infty$, barring two banal exceptions. Our interpretation of this beautiful theorem is that the Bayesian judge becomes wiser as he goes along in his judicious search for relevant information: as more and more balanced and nonredundant evidence is brought to bear, the posterior belief in any hypothesis approaches certainty (through (C1)), provided that this hypothesis is indeed true. It is trivial to show that this holds for any nonzero prior belief in the hypothesis; in that respect, the theorem also confirms that a rational person is a pragmatic learner. Regardless of how little belief s/he initially holds in an unpopular truth, e.g., that jogging is bad for one's health, s/he is willing to change that opinion freely as new information becomes available.

Explicit Treatment of Conditional Independence: The Bayesian language offers a rich variety of tools for detecting, representing, and dealing with correlated evidence. For example, it is easy to show that any one of the following three assertions implies the other two: 1) $P(e_1|h, e_2) = P(e_1|h)$; 2) $P(e_2|h, e_1) = P(e_2|h)$; 3) $P(e_1, e_2|h) = P(e_1|h) \cdot P(e_2|h)$. The equivalence of 1)

and 2) implies that the assertion "given h, knowing e_1 does not change my belief in e_2" is symmetric with respect to e_1 and e_2 as we would have expected. The equivalence of 1) and 3) implies that either 1) or 2) is consistent with the classical definition of statistical independence. Note, however, that either 1) or 2) is preferred to 3) from a cognitive standpoint: each conveys a better understanding of the notion of independence, and together they offer two cognitively different but mathematically equivalent techniques to elicit the same phenomenon. Hence Bayesian knowledge engineers who seek to detect correlated evidence might use 1) and 2) either separately or in tandem. This provides a richer language for knowledge acquisition and a simple means for cross-verification of human inputs.

In the event that some pieces of evidence are correlated (with respect to either h or to $\bar{h}$), (C1) becomes invalid on normative grounds. Several authors, e.g., Charniak [8] and Pearl [40], proposed heuristic techniques to decompose correlated inference networks into equivalent (but not identical) networks in which some version of (C1) might be applied. As this line of research is quite new, these techniques are somewhat limited. At the same time, statistical dependency is a prevailing feature of nature that cannot be swept under the rug. It is therefore fortunate that the Bayesian language provides a powerful arsenal of tools for expressing this phenomenon and dealing with it explicitly.

Compatibility with Decision Theory: Aside from its compliance with rational judgment, the Bayesian language is syntactically compatible with many models that may be used in the context of expert systems. For example, Raiffa's [45] value-of-information analysis attempts to pursue the "most valuable" clue, i.e., the clue with the (anticipated) maximal diagnostic impact on current belief. In a similar vein, Langlotz *et al.* [32] proposed a decision theoretic extension to rule-based reasoning, in which utility considerations serve to evaluate the merit of potential search paths in a rule-based inference net. There exist numerous other areas in which AI and decision theory might benefit greatly from each other. To reap such benefits, though, both disciplines have to speak in the same language. In this regard it is worth noting that practically all the work in prescriptive decision theory is already cast in terms of the Bayesian language. Hence this language would be a pragmatic starting point for a comparative analysis.

In summary, perhaps the reader is by now convinced that (C1) has more to it than at first appears. From a technical standpoint, Bayes rule is a trivial exercise in set theory, given the axioms of subjective probability. These axioms, however, can be nontrivially derived from rational behavior under uncertainty (Cox [12], de Finetti [13], Savage [47]). Thus the epistemological interpretation of Bayes rule and its implications on judgment, learning, and experience. This dichotomy is clearly a manifestation of the Janus face of probability: "on the one side it is statistical, concerning itself with stochastic laws of chance processes. On the other side it is epistemological, dedicated to assessing degrees of belief in propositions quite devoid of statistical background" (Hacking [24, p. 12]).

We conclude this section with a formal definition of the (ratio-form conditionally independent) Bayesian language, which plays a central role in what follows.

Definition 1: Let L_{Bayes} be the belief language whose syntax consists of likelihood ratios of the form $P(e|h)/P(e|\bar{h})$ and whose calculus is (C1).

Thus we focus on an inference context in which causal rules of the form $h \rightarrow e$ are parameterized by likelihood ratios of the form $P(e|h)/P(e|\bar{h})$. Furthermore, we assume that these degrees of belief are combined through Bayes rule (C1). Since (C1) is restricted to ratio-form conditionally independent inference problems, L_{Bayes} is a special case of a more general Bayesian language. To avoid clutter, though, we will refer below to L_{Bayes} as the Bayesian language. We will comment later on the restrictiveness of this language.

IV. DIALECTAL ANALYSIS

The previous section was meant to serve as a motivation for adopting the Bayesian language L_{Bayes} as a standard against which other belief languages will be judged. This analysis will be carried out within a methodological framework described in this section. The following terminology is used throughout: let $L_i = \langle \text{bel}_i, C_i \rangle$, $i \in \{1, 2\}$, be two belief language. bel_i denotes the syntax (set of degrees of belief) of L_i, and C_i denotes the L_i calculus, i.e., an operator $C_i : \text{bel}_i \rightarrow \text{bel}_i$ defined over and into the L_i degrees of belief.

Definition 2: If the syntax of L_1 is identical to the syntax of L_2 and the calculus of L_1 is a special case of the calculus of L_2, L_1 is said to be a *dialect* of L_2.

Definition 3: Let T be a mapping $T : \text{bel}_1 \rightarrow \text{bel}_2$. If $T(\text{bel}_1) = \text{bel}_2$ implies that $C_2(T(\text{bel}_1)) = C_1(\text{bel}_1)$, then we say that L_2 is *isomorphic* to L_1 under T and denote it $\text{iso}(L_2, L_1)_T$.

Definition 4: In general, we call a mapping $T : \text{bel}_1 \rightarrow \text{bel}_2$ an *interpretation* of L_2 and denote this interpretation $T : L_1 \rightarrow L_2$. If L_1 is the Bayesian language, T is called a Bayesian interpretation of L_2.

Corollary 1: For any three belief languages L_1, L_2, L_3 and interpretations $T1 : L_1 \rightarrow L_2$ and $T2 : L_2 \rightarrow L_3$, the following is true:

1) iso (L_1, L_1) where I is the identity mapping $I(\text{bel}_1) = \text{bel}_1$
2) if iso $(L_1, L_2)_{T1}$ and iso $(L_2, L_3)_{T2}$ then iso$(L_1, L_3)_{T2 \cdot T1}$
3) if T^{-1} exists uniquely then iso $(L_1, L_2)_T$ if and only if iso $(L_2, L_1)_{T^{-1}}$

Definitions 2–4 and Corollary 1 provide a general framework for investigating implicit relationships between two or more belief languages. Given two languages L_1 and L_2, the basic idea is to try to give L_2 a meaning T using the semantics of L_1. If an isomorphism emerges and,

furthermore, the meaning T "makes sense," we argue that L_2 is a special or a "disguised" case of L_1. This exercise is carried out as follows: first, one inspects the degrees of belief of L_2 and tries to define them in terms of the L_1 syntax using some function $T: \text{bel}_1 \rightarrow \text{bel}_2$. In the next stage, one carries out this very interpretation throughout L_2's calculus C_2, replacing every occurrence of $\{\text{bel}_2\}$ in the definition of C_2 with its corresponding degree of belief $\{T(\text{bel}_1)\}$. The objective here is to obtain a mathematical expression which has a calculus-cal meaning within L_2. Ideally, this expression can be shown to be equivalent to the L_2 calculus, in which case L_2 is said to be isomorphic to L_1 under the interpretation T. Multilingual relationships may be established through Corollary 1.

This technique is used extensively in the next three sections to investigate the kinship of the Bayesian, the certainty factors, the contrast/inertia, and the Dempster–Shafer languages.

V. THE CERTAINTY FACTORS LANGUAGE

The certainty factors (CF) language was conceived and implemented as part of the seminal Mycin project, an AI-based medical diagnosis system developed by Shortliffe [53] and his colleagues at Stanford University. Following the subsequent popularity of the general-purpose Emycin (Van Melle [58]) and the M.1 rule-based architectures, the CF language became the *de facto* belief language of most applied expert systems in a survey conducted by Harmon and King [39], seven of the eight expert system "shells" that supported uncertainty management were based on the CF model. Over the past few years, these development tools have generated numerous expert systems, and many of these are thus based on some version of the CF language.

The development of the CF language as an alternative to Bayesian inference was originally justified on normative as well as on practical grounds. From a philosophical standpoint, the certainty factor $CF(h|e)$ is reminiscent of Carnap's [7] confirmation function $C(h, e)$, designed to measure the degree of logical entailment associated with the inexact implication $e \rightarrow h$. In the rule-based architecture, CF's are used to parameterize diagnostic rules of the form $\langle \text{IF } e \text{ THEN } h \rangle$. This type of forward reasoning (i.e., from effects to causes) is widely practiced by knowledge engineers and developers of existing expert systems.

Unlike the Bayesian language, the CF calculus operates on *relative* rather than absolute degrees of belief. It is assumed that, most of the time, there exist several pieces of evidence that increase and decrease one's belief in a hypothesis simultaneously. In the additive CF syntax, $MB(h|e)$ ($MD(h|e)$) represents the increased belief (disbelief) in the hypothesis h rendered by the e, with $0 \leqslant MB, MD \leqslant 1$ and $MB \cdot MD = 0$. If e is irrelevant to h, $MB(h|e) = MD(h|e) = 0$. The extreme case of e being sufficiently convincing to confirm (disconfirm) h in certainty is modeled through $MB = 1$ and $MD = 0$ ($MB = 0$ and $MD = 1$). Normally, a set of atomic MB's and MD's

(jointly referred to as CF's) is elicited directly from a domain expert. The combined, net increased belief in h in light of $E = \{e_1, e_2\}$ is computed through the following calculus:

$$CF(h|e_1, e_2)$$

$$= \begin{cases} MB(h|e_1) + MB(h|e_2) \cdot (1 - MB(h|e_1)), \\ \qquad \text{if both CF's are positive} \qquad \text{(C2a)} \\ MD(h|e_1) + MD(h|e_2) \cdot (1 - MD(h|e_1)), \\ \qquad \text{if both CF's are negative.} \qquad \text{(C2b)} \\ \dfrac{MB(h|e_1) - MD(h|e_2)}{1 - \min\{MB(h|e_1), MD(h|e_2)\}}, \\ \qquad \text{if the CF's have mixed signs} \qquad \text{(C2c)} \end{cases}$$

An inspection of (C2) reveals that the CF calculus is commutative and associative. Hence (C2) might be applied recursively to compute the compound evidential impact of any finite set of clues $\{e_1, \cdots, e_n\}$, using an adjustment and anchoring model similar to (2). Note also that both (C2a) and (C2b) appear to convey an (additive) rational sense: if you open the parentheses of either (C2a) or (C2b), you obtain the sum of $CF(h|e_1)$ and $CF(h|e_2)$ minus some sort of a multiplicative interaction effect:

$$CF(h|e_1, e_2) = CF(h|e_1) + CF(h|e_2)$$
$$- CF(h|e_1) \cdot CF(h|e_2). \quad (8)$$

Now contrast this with the following logarithmic transformation of the Bayesian calculus (7):

$$\log R(h|e_1, e_2) - \log R(h) = \log L(e_1|h) + \log L(e_2|h). \quad (9)$$

We see that the diagnostic CF combination rule (8) is remarkably similar to its Bayesian causal counterpart (9). First, recall that $CF(h|e_1, e_2)$ is defined as the (possibly negative) increased belief in h due to $\{e_1, e_2\}$, which is precisely what we find on the left side of (9). Second, the right sides of (8) and (9) are also similar, excluding the presence of $CF(h|e_1) \cdot CF(h|e_2)$ in the latter. This "interaction factor" can be viewed as and *ad hoc* compensation for double-counting correlated evidence in the CF calculus. Given this rationale, the absence of an interaction effect in L_{Bayes} makes sense, as L_{Bayes} is explicitly confined to ratio-form conditionally independent inference problems which are free of interaction effects. This observation might lead to the deceptive impression that the CF language accounts for interaction effects while L_{Bayes} does not, and consequently, that the CF language is less restricted than the Bayesian language. As will be discussed shortly, this impression is false.

The similarity of the CF and the Bayesian languages is also reflected in their treatment of the *state of insufficient reason* (Savage [47]). In the CF language, a hypothesis h which is backed by no clues is assigned a "prior" certainty factor of zero, i.e., $CF(\bar{h}|\varnothing) = 0$. When a clue e becomes available, the posterior belief in h undergoes a trivial

update, as follows:

$$\text{bel}(h|e) = 0 + \text{CF}(h|e) - 0 \cdot \text{CF}(h|e) = \text{CF}(h|e). \quad (10)$$

In the classical Bayesian language (and following LaPlace), the lack of any prior evidence on h may be modeled by assigning equal odds on h and on $\bar{h}$, yielding $P(h)/P(\bar{h}) = 1.$[5] Hence the Bayesian counterpart of (10) is

$$\text{bel}(h|e) = 1 \cdot \frac{P(e|h)}{P(e|\bar{h})} = \frac{P(e|h)}{P(e|\bar{h})}. \quad (11)$$

Consequently, we see that both languages have an isomorphic interpretation of the state of insufficient reason: the neutral degree of belief in the additive CF language is zero, while its counterpart in the multiplicative odds-ratio Bayesian language is one. The exact nature of this isomorphism will become clearer in the next section.

For a detailed description of the CF language, the reader is referred to Buchanan and Shortliffe [6]. In the present paper we restrict our attention to a dialect of this language, defined as follows.

Definition 5: Let L_{CF} be the language whose syntax consists of pairs $\langle MB, MD \rangle$ and whose calculus consists of (C2a) and (C2b). In what follows, L_{CF} is sometimes referred to as the two-valued CF language.

We see that the major difference between the commonly used CF language and L_{CF} is the exclusion of the combination rule for mixed evidence (C2c) in the latter. Equation (C2c) was added *ad hoc* to the CF language to improve the computational efficiency of Mycin (Van Melle [58]) and is of little interest here. In fact, it is easy to create examples in which (C2c) does not make sense. To illustrate, consider Sue Ellen and Cliff, who were two suspects in shooting J. R. Assume that Sue Ellen and Cliff stand in line to inherit $900 000 and $800 000, respectively, from J. R.'s estate. We represent this information (quite crudely) by assigning MB(Sue Ellen) = 0.8 and MB(Cliff) = 0.9, where $MB(x)$ is the degree of support in the hypothesis "x shot J. R." Now, suppose that we further learn that J. R. had owed Cliff $500 000, and that this loan was to be terminated upon J. R.'s death. This new piece of evidence must decrease our suspicion in Cliff, and we therefore assign MD(Cliff) = 0.5. It turns out, however, that according to (C2c), the pooled evidence suggests that Sue Ellen and Cliff are equal suspects, with CF(Sue-Ellen) = CF(Cliff) = 0.8.

This example is admittedly crude, but it does demonstrate that (C2c) is a problematic evidence-pooling operator. We thus prefer to focus on a "simpler" language L_{CF}, in which the belief in any proposition h in the face of mixed evidence E (consisting of positive and negative pieces of evidence) is represented by a pair of numbers $\text{CF}(h|E) = \langle MB(h|E), MD(h|E) \rangle$ describing a) the in-

creased belief in h due to E, and b) the increased disbelief in h due to E. If E consists of a single fact, at least one of these numbers must be zero. If E is a diverse body of evidence, both numbers may be strictly positive. Defined that way, L_{CF} allows us to simultaneously track the impact of negative and positive evidence on each hypothesis.

A. Bayesian Interpretation to the CF Language

There have been several attempts to justify the CF syntax on Bayesian grounds. The formulae given below were originally proposed by Shortliffe [53] as an *ex-post* Bayesian interpretation of certainty factors. This widely quoted mapping is defined as follows:

$$MB(h|e)$$

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

$$(T1a)$$

$$MD(h|e)$$

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

$$(T1b)$$

Under the (T1) mapping, the CF calculus (C2) is viewed as a belief calculus that operates on normalized differences in probabilities rather than on absolute probabilities. For example, (T1a) represents the increased *belief* in h in light of e as a normalized difference between the posterior $P(h|e)$ and the prior $P(h)$, P taken to be a probability.

Before proceeding with dialectal analysis, we first have to clean up (T1) to make it a well-defined mapping from L_{Bayes} to the two-valued L_{CF}. This mapping ought to map Bayesian likelihood-ratios of the form $P(e|h)/P(e|\bar{h})$ on pairs of numbers $\text{CF}(h|e) = \langle MB(h|e), MD(h|e) \rangle$, one of which must be zero (a single piece of evidence e cannot support and negate h simultaneously). We thus fix (T1) as follows:

$$\text{CF}(h|e)$$

$$= \begin{cases} \left\langle \dfrac{P(h|e) - P(h)}{1 - P(h)}, 0 \right\rangle, & \text{if } P(e|h)/P(e|\bar{h}) > 1 \\ \langle 0, 0 \rangle, & \text{if } P(e|h)/P(e|\bar{h}) = 1 \\ \left\langle 0, \dfrac{P(h|e) - P(h)}{-P(h)} \right\rangle, & \text{if } P(e|h)/P(e|\bar{h}) < 1 \end{cases}$$

$$(T2)$$

Note in passing that (T2) is not unique because it must be parameterized by the Bayesian prior belief $P(h)$. It is

---

[5]This equal spread of belief in the face of insufficient reason was criticized by Shafer [50]. Shafer proposed a subadditive belief function, in which the beliefs in a hypothesis and its negation do not necessarily sum to one. See Section VII.

therefore appropriate to denote the mapping as follows:

$$(T2)_{P(h)}: \frac{P(e|h)}{P(e|\bar{h})} \rightarrow \langle MB(h|e), MD(h|e) \rangle. \quad (12)$$

It is easy to show that (T1) and (T2) are equivalent, excluding the cases of $P(h) = 0$ and $P(h) = 1$ which are undefined in the latter. This can be easily added to (T2), but we leave it out to avoid clutter. Also, for the sake of brevity, when we refer to (T2) from now on, we will only discuss the positive part of the pair $\langle MB, MD \rangle$. Thus (T2) will be technically referred to as a single-valued mapping, remembering that the second value of its image is always zero.

B. Implications

In his probabilistic analysis of certainty factors was carried out by Adams [1], has shown that the CF calculus makes implicit assumptions of conditional independence. Following Adams' work, Heckerman [26] pointed out that the Bayesian transformation of the CF calculus (C2) under the mapping (T1) is sensitive to the order in which evidence is being combined, and Schocken [48] observed that the culprit of this blunder is (C2c). Order effects are clearly an undesirable property of rule-based inference, in which the order that rules "fire" is supposed to have no impact on the posterior belief in the underlying hypothesis. To remedy this and other problems in the CF language, Heckerman proposed a revised interpretation of certainty factors. He then went on to show that the revised interpretation implies that the CF language is equivalent to an odds-ratio version of the Bayesian language.

In this paper we stick to the original interpretation of certainty factors (T1) or, more precisely, (T2), and proceed to show that, under this mapping, L_{CF} is isomorphic to L_{Bayes}. The following lemmas are largely equivalent to Adams' analysis [1], which was rather informal. We feel that recasting these results in the framework of dialectal analysis serves to clarify the implicit relationship between L_{CF} and L_{Bayes}. For the sake of brevity, only the outlines of the proofs are given.

Lemma 1: The Bayesian interpretation (T2) transforms the *MD*-combination rule (C2b) into the following Bayesian belief-update model:

$$P(h|e_1, e_2) = \frac{P(e_1|h)}{P(e_1)} \cdot \frac{P(e_2|h)}{P(e_2)} \cdot P(h). \quad (13)$$

Proof (Adams [1]): Recall that the *MD*-combination rule (C2b) is designed to combine two negative pieces of evidence. Hence the third line of (T2) is used to transform the *MD*'s in (C2b) into their corresponding Bayesian "images." For example, $MD(h|e_1)$ is transformed into $(P(h|e_1) - P(h))/ - P(h)$. The proof proceeds by carrying out this transformation throughout the left side and the right side of (C2b) which, after several algebraic manipulations, becomes Bayes rule (4) for computing the posterior belief in h in light of the ratio-form conditional independent evidence $\langle e_1, e_2 \rangle$.

Lemma 2: The Bayesian interpretation (T2) transforms the *MB*-combination rule (C2a) into the following Bayesian belief-update model:

$$P(\bar{h}|e_1, e_2) = \frac{P(e_1|\bar{h})}{P(e_1)} \cdot \frac{P(e_2|\bar{h})}{P(e_2)} \cdot P(\bar{h}). \quad (14)$$

Proof: It is easy to verify that the Bayesian interpretation (T2) implies that $MB(h|e) = - MD(\bar{h}|e)$. This fact, together with Lemma 1, completes the proof.

Theorem 1: L_{CF} is isomorphic to L_{Bayes} under the Bayesian interpretation (T2).

Proof: From Lemmas 1 and 2, (T2) maps the CF calculus on the joint application of (13) and (14). By dividing the former by the latter, the Bayesian calculus (C1) emerges. Hence L_{CF} is isomorphic to L_{Bayes} under (T2).

VI. THE CONTRAST/INERTIA LANGUAGE

Einhorn and Hogarth [19] proposed a sequential anchoring-and-adjustment model called the contrast/inertia (CI) language. This model is strictly descriptive in the sense that it tries to predict how ordinary people actually behave, rather than prescribe how a rational person *ought* to behave. The model is based on the premise that positive and negative clues are not treated equally by human judges; rather, humans tend to discount the impact of negative evidence and inflate the impact of supportive evidence. These assumptions are consistent with the previous experimental work in cognitive psychology, e.g., Snyder and Swann [55]. In the CI syntax, S_k stands for the posterior belief in a certain hypothesis after k pieces of evidence have been brought to bear. As additional clues become available, this degree of belief is modified through an order-sensitive anchoring and adjustment process.

Suppose that at step $k - 1$, one holds a belief S_{k-1} in a hypothesis h. At step k a new clue relevant to h is brought to bear. A clue which decreases (increases) the belief in h is called negative (positive) evidence and denoted a_k (b_k). Given this nomenclature, the contrast/inertia calculus is defined as

$$S_k = S_{k-1} - w_k \cdot g_1(a_k) \quad (15)$$

$$S_k = S_{k-1} + r_k \cdot g_2(b_k). \quad (16)$$

Equations (15) and (16) are called *discounting* and *accretion* models, respectively. Taken together, (15) and (16) is a sequential anchoring-and-adjustment calculus that can be parameterized to reflect various behavioral assumptions. In particular, the following parameters need to be specified:

w_k adjustment factor for negative evidence at step k,

r_k adjustment factor for positive evidence at step k,

$g_1(a_k)$ subjective evaluation of the kth negative clue,

$g_2(b_k)$ subjective evaluation of the kth positive clue.

Einhorn and Hogarth made the following assumptions: $w_k = \alpha \cdot S_{k-1}$ and $r_k = \beta \cdot (1 - S_{k-1})$. Their rationale is as follows: α and β are constants of proportionality, measuring one's sensitivity to new evidence or, alternatively, one's propensity to change current opinions. This propensity is also influenced dynamically by the current belief S_{k-1} or the *anchor*, as follows: negative evidence is discounted more when the current belief is large, whereas positive evidence is inflated when the current belief is small. Einhorn and Hogarth also specified the functions $g_1(a_k)$ and $g_2(b_k)$, but it turns out that this level of detail is unnecessary for our analysis, so we ignore it for the sake of generality. Thus we will focus on the following belief-update model:

$$S_k = S_{k-1} - \alpha S_{k-1} \cdot g_1(a_k) \tag{C3a}$$

$$S_k = S_{k-1} + \beta \cdot (1 - S_{k-1}) \cdot g_2(b_k). \tag{C3b}$$

Definition 6: Let l_{CI} be the language whose syntax consists of $\{ g_1(a_k), g_2(b_k) \}$ and whose calculus consists of (C3).

A. Bayesian Interpretation of the CI Language

Before presenting of the Bayesian interpretation of L_{CI}, it is appropriate to step back and reflect on the nature of this analysis. It is important to remember that the whole purpose of the descriptive CI language is to account for systematic violations of the Bayesian belief-update model. It thus seems circular to map L_{Bayes} on L_{CI}. At the same time, this exercise is important because it puts the finger precisely on where the former stops and where the latter takes over. It thus shows explicitly in what ways the CI language extends the Bayesian language.

The Bayesian interpretation of the CI language will be derived indirectly. First, we give a plausible interpretation which grants an isomorphism between the CI and the CF languages. We then use the transitivity part of Corollary 1, combined with the Bayesian interpretation of the CF language, to derive a (compound) Bayesian interpretation for the CI language.

Assume that at stage $k-1$ one holds a prior belief S_{k-2} in a hypothesis h and that a new clue e_{k-1} becomes available, carrying a negative evidential impact a_{k-1}. After the belief S_{k-2} is updated to S_{k-1}, a new negative clue e_k becomes available, carrying a negative impact a_k, and the final belief S_k is established. Now consider the following mapping:

$$MD(h|e_{k-1}) = g_1(a_{k-1})$$

$$MD(h|e_k) = g_1(a_k) \tag{T3a}$$

$$MD(h|e_{k-1}, e_k) = \frac{S_k - S_{k-2}}{-S_{k-2}}.$$

Repeating the previous scenario with two positive pieces of

evidence, consider the following mapping:

$$MB(h|e_{k-1}) = g_2(b_{k-1})$$

$$MB(h|e_k) = g_2(b_k) \tag{T3b}$$

$$MB(h|e_{k-1}, e_k) = \frac{S_k - S_{k-2}}{1 - S_{k-2}}.$$

The justification of the mapping (T3) is as follows: negative and postive impacts in the CF language map precisely on their subjective perceptions in the CI language. Compound CF impacts map on normalized net increases in belief in the CI language. Interestingly, it can be shown that the mapping (T3), along with the alternative, Bayesian interpretation of the CI syntax $S_{k-1} = P(h)$ and $S_k = P(h|e_k)$, jointly imply the Bayesian interpretation of certainty factors (T2) discussed in the previous section. This curious relationship will become clearer in the final discussion section.

We now preceed to show that the mapping (T3): $L_{CI} \rightarrow L_{CF}$ is an isomorphism between the two-valued CF language and L_{CI}.

B. Implications

Lemma 3: The mapping (T3a) transforms the MD-combination rule for negative evidence (C2b) into the CI-discounting rule (C3a) with $\alpha = 1$.

Proof: Carrying out the mapping (T3a) through the MD-combination rule (C2b) gives

$$\frac{S_k - S_{k-2}}{-S_{k-2}} = g_1(a_k) + g_1(a_{k-1}) - g_1(a_k) \cdot g_1(a_{k-1}) \tag{17}$$

which is equivalent (after several algebraic manipulations) to

$$S_k = S_{k-2} \cdot (1 - g_1(a_k) - g_1(a_{k-1}) + g_1(a_k) \cdot g_1(a_{k-1})) \tag{18}$$

which, in turn, is equivalent to

$$S_k = S_{k-2} \cdot (1 - g_1(a_k)) \cdot (1 - g_1(a_{k-1})) \tag{19}$$

and it is easy to show that (19) is the compound impact of a *two-staged* contrast/inertia accretion model (with $\alpha = 1$)

$$S_{k-1} = S_{k-2} - S_{k-2} \cdot g_1(a_{k-1}) \tag{20}$$

$$S_k = S_{k-1} - S_{k-1} \cdot g_1(a_k). \tag{21}$$

Lemma 4: The mapping (T3b) transforms the MB-combination rule (C2a) into the CI accretion rule (C3b) with $\beta = 1$.

Proof: the proof is similar to that of Lemma 3, and we leave it as an exercise for the reader.

Theorem 2: L_{CF} is isomorphic to L_{CI} (with $\alpha = \beta = 1$) under the mapping (T3): $L_{CI} \rightarrow L_{CF}$.

Proof: the theorem is a direct consequence of Lemmas 3 and 4.

Returning to the (T3) mapping, we note that by fixing the prior belief parameter S_{k-2} the mapping can be inverted into what we denote $(T3)_{S_{k-2}}^{-1}: L_{CI} \to L_{CF}$. This gives us the following lemma and theorem which follow from Theorem 2 and Corollary 1

Lemma 5: L_{CF} is isomorphic to L_{CI} under $(T3)_{S_{k-2}}^{-1}: L_{CI} \to L_{CF}$.

Theorem 3: L_{CI} is isomorphic to L_{Bayes} under the compound Bayesian interpretation $((T3)_{S_{k-2}}^{-1} \circ T2): L_{Bayes} \to L_{CI}$.

Note that the requirement that $\alpha = \beta = 1$ under a Bayesian interpretation is not surprising. This is consistent with the standard Bayesian principle that negative and positive evidence should be weighted equally in terms of their impacts on posterior beliefs.

VII. THE DEMPSTER–SHAFER BELIEF LANGUAGE

Given an exhaustive set of mutually exclusive propositions ("frame of discernment") θ, Shafer defines a belief function that assigns degrees of belief to every element in the power set 2^θ, extending the notion of probability which is defined over θ only. Two or more belief functions $bel_1, bel_2: 2^\theta \to [0, 1]$ may be combined through Dempster's rule, yielding a compound belief function $(bel_1 \oplus bel_2)$: $2^\theta \to [0,1]$. Unlike probabilities, belief functions are subadditive. That is, the definition of a belief function is consistent with the property $bel(h) + bel(\bar{h}) \leqslant 1$, and the quantity $1 - bel(h) - bel(\bar{h})$ is defined explicitly as the degree of or uncommitted belief associated with the proposition h. Note that although Shafer describes belief as a single-place function $bel(h)$, he really means this to be a shorthand notation for $bel(h|E)$, E being a relevant body of evidence (which may be empty). We will adopt this abbreviated notation as well.[6]

The Dempster–Shafer (DS) language is well-suited for nondeterministic expert systems for a number of reasons. According to Shafer [49], the theory lends itself nicely to inferential problems in which the prior probabilities of hypotheses are unavailable. Second, Shafer proposed that the *predictive* nature of the degree of belief $bel(h|e)$ is often more intuitively appealing than its *causal* counterpart $P(e|h)$, which is used in most Bayesian inference systems. Third, Dempster's rule is viewed by many practitioners as a belief calculus that operates on belief intervals rather than on point probabilities (Spiegelhalter [56]). Indeed, Shafer has shown that his degree of belief $bel(h)$ corresponds to a belief interval $[bel(h), P^*(h)]$ where $P^*(h) = 1 - bel(\bar{h})$ is the upper probability of h, and, furthermore, if $bel = P^*$ then bel is a probability. Hence the DS language is viewed as a mechanism that enables an expert system to represent, combine, and propagate interval, point, and uncommitted beliefs in hypotheses.

Consider the set of propositions θ. A function $bel: 2^\theta \to [0,1]$ is said to be a Shafer belief function[7] if

1) $bel(\varnothing) = 0$,
2) $bel(\theta) = 1$,
3) for every $m > 0$ and for every collection of subsets $A_1, \cdots, A_m$ of θ, if $A_1 \cap \cdots \cap A_m = \varnothing$, then $bel(A_1 \cup \cdots \cup A_m) \geqslant bel(A_1) + \cdots + bel(A_m)$.

The inference problem $\langle h, e_1, \cdots, e_n \rangle$ discussed in this paper has the following Shaferian interpretation. The hypothesis h corresponds to the dichotomous frame $\theta = \{h, \bar{h}\}$, and $2^\theta = \{\varnothing, \{h\}, \{\bar{h}\}, \{h, \bar{h}\}\}$. The assignment of belief to h is extended to an assignment of degrees of belief to each member of 2^θ. By definition, $bel(\varnothing)$ and $bel(\{h, \bar{h}\})$ are 0 and 1, respectively. Then there exist two numbers $a, b \in [0, 1]$ with $bel(\{h\}) = a, bel(\{\bar{h}\}) = b$, and $a + b \leqslant 1$. The set of clues $\{e_1, \cdots, e_n\}$ corresponds to n independent evidential sources (e.g., expert opinions or pieces of evidence) that support h through the n simple support functions $bel_1, \cdots, bel_n: 2^\theta \to [0, 1]$, defined as follows:

$$bel_i(x) = \begin{cases} 0, & \text{if } x = \varnothing \text{ or } x = \{\bar{h}\} \\ s_i, & \text{if } x = \{h\} \\ 1, & \text{if } x = \{h, \bar{h}\} \end{cases} \quad (22)$$

Simple support functions consist of a special case of belief functions (Shafer [50]). A simple support function bel is focused only on one element in the frame of discernment. That is, bel awards a positive degree of belief to one element in the frame and zero belief to any other element.

belief functions might be combined through Dempster's rule, which is commutative and associative. If there are more then two belief functions, the rule can be applied repetitively. Hence we can resort once again to the simple inference problem $\langle h, e_1, e_2 \rangle$, knowing that the analysis can be extended to any finite $n > 2$. We model this inference problem by two simple support functions defined over $\{h, \bar{h}\}$ and focused on h. Now, in general, Dempster's rule for combining belief functions is quite complex; however, in the case of two (and, in fact, n) simple support functions and a dichotomous frame of discernment, Dempster's rule reduces to, for every $x \in 2^{\{h, \bar{h}\}}$,

$$bel(x) = bel_1(x) + bel_2(x) - bel_1(x) \cdot bel_2(x) \quad (C4)$$

and it is easy to show that the resulting bel is a simple support function defined over $\{h, \bar{h}\}$ and focused on h.

Definition 7: Let L_{DS} be the belief language whose syntax consists of simple support functions and whose calculus consists of Dempster's rule (C4). L_{DS} will be sometimes referred to as the *simple* DS language.

We conclude this section with the definition of an additive *weights of evidence* language discussed by Good [21]

---

[6] The reader should not confuse Shafer's belief function bel with the abstract bel function discussed in Section II.

[7] There are other ways to define belief functions. See Shafer [50].

and referred to by Shafer [50]. The relationship between this language and the present analysis will become clearer shortly.

Definition 8: Let L_W be the belief language whose syntax consists of degrees of belief $bel(h|e_i) = w_i$ with $w_i \in [0, \infty]$ and whose calculus consists of the additive combination rule $W = w_1 + \cdots + w_n$.

A. A Bayesian Interpretation to the DS Language

Shafer has argued that 1) belief functions are extensions of Bayesian probabilities, and 2) Bayes' rule is a special case of Dempster's rule. Several Bayesian writers have tried to prove precisely the opposite. Lindley's [34] line of attack is based on the premise that probability can replace belief functions by manipulating the epistemic design of the problem in question. Kyburg [31] has shown that a belief function can be represented through a convex set of classical probability functions. He then went on to show that the probability intervals of the DS language are a special case of the intervals that result from Bayesian updating on uncertain evidence. Baron [5] has shown that the class of simple support functions can be cast in a Bayesian language using the notion of ambiguity (Einhorn and Hogarth [18]) such that "ignorance" on h is represented through a second order probability $Q(P(h))$.

Shafer [50] proposed an intuitively appealing interpretation of simple belief functions based on the proximity between L_{DS} and L_W. More specifically, Shafer has shown that if a) Dempster rule is taken for granted, and b) weights of evidence combine additively, then the following relationship must exist between the belief function bel: $2^\theta \to [0,1]$ and the weight of evidence function $w: \theta \to [0, \infty]$:

$$bel = 1 - e^{c \cdot w}. \tag{23}$$

Since the unit of measurement for the weights of evidence is arbitrary, Shafer proceeded to set $c = -1$ to obtain

$$bel = 1 - e^{-w}. \tag{T4}$$

The Bayesian literature is rife with "weights of evidence" analyses, and Good [20] cites 33 papers on the subject. Turing proposed the following interpretation to the weight of evidence w carried by a clue e to a hypothesis h:

$$w = \log \frac{O(h|e)}{O(h)} = \log \frac{P(h|e)/P(\bar{h}|e)}{P(h)/P(\bar{h})}. \tag{T5}$$

The following section presents a Bayesian interpretation of the *simple* DS language that establishes an isomorphism between Dempster's rule and Bayes' rule. This interpretation is a product of two interpretations, (T4) and (T5). The first half of this analysis is due to Shafer, who proposed the linkage (T4) between L_{DS} and L_W. We go one step further and observe that L_{DS} is isomorphic not only to L_W but also to the Bayesian language L_{Bayes}.

B. Implications

Theorem 4 (Shafer): L_{DS} is isomorphic to L_W under (T4): $L_W \to L_{DS}$.

Proof: Let bel_1 and bel_2 be two simple support functions emanating from two evidential sources e_1 and e_2 and focused on a hypothesis h. Let w_1 and w_2 be the corresponding weights of evidence carried by e_1 and e_2 to h. Carrying out the mapping (T4) through Dempster's rule (C4) gives

$$1 - e^{-w} = 1 - e^{-w_1} + 1 - e^{-w_2} - (1 - e^{-w_1}) \cdot (1 - e^{-w_2}) \tag{24}$$

which after several algebraic manipulations, is equivalent to

$$e^{-w} = e^{-(w_1 + w_2)} \tag{25}$$

which is equivalent to $w = w_1 + w_2$, yielding an isomorphism between L_{DS} and L_W.

Theorem 5: L_W is isomorphic to L_{Bayes} under the Bayesian interpretation (T5): $L_{Bayes} \to L_W$.

Proof: We begin by noting that (T5) is indeed a Bayesian interpretation mapping $P(e|h)/P(e|\bar{h})$ on w by virtue of Bayes rule:

$$w_i = \log \frac{O(h|e)}{O(h)} = \log \frac{P(h|e)/P(\bar{h}|e)}{P(h)/P(\bar{h})} = \log \frac{P(e|h)}{P(e|\bar{h})}. \tag{26}$$

Thus carrying out (T5) through the left and right sides of the additive calculus $W = w_1 + w_2$ gives

$$\log \frac{O(h|e_1, e_2)}{O(h)} = \log \frac{P(e_1|h)}{P(e_1|\bar{h})} + \log \frac{P(e_2|h)}{P(e_2|\bar{h})} \tag{27}$$

which, after an antilogarithmic transformation, gives

$$\frac{O(h|e_1, e_2)}{O(h)} = \frac{P(e_1|h)}{P(e_1|\bar{h})} \cdot \frac{P(e_2|h)}{P(e_2|\bar{h})}. \tag{28}$$

Multiplying both sides by $O(h) = P(h)/P(\bar{h})$ gives the Bayesian calculus:

$$\frac{P(h|e_1, e_2)}{P(\bar{h}|e_1, e_2)} = \frac{P(e_1|h)}{P(e_1|\bar{h})} \cdot \frac{P(e_2|h)}{P(e_2|\bar{h})} \cdot \frac{P(h)}{P(\bar{h})}. \tag{29}$$

Theorem 6: L_{DS} is isomorphic to L_{Bayes} under the compound Bayesian interpretation (T5)∘(T4): $L_{Bayes} \to L_{DS}$.

Proof: The proof follows from Theorems 4, 5, and Corollary 1.

VIII. DISCUSSION

This paper has given an integrated Bayesian analysis of important dialects of the certainty factors, Dempster–Shafer, and the contrast/inertia belief languages. Before proceeding any further, we remind the reader that all the findings reported are restricted to the dialects described

$$L_{CI} \longleftrightarrow L_{CF} \longleftarrow \text{-----} \longrightarrow L_{DS}$$
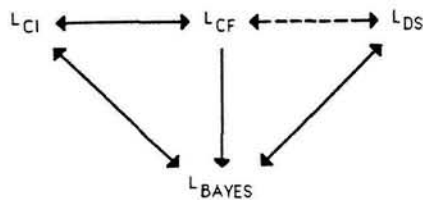
$$L_{BAYES}$$

Fig. 1.

explicitly in Definitions 5, 6, and 7. These dialects are significant in that they form the nucleus of their corresponding languages. At the same time, one cannot extrapolate the Bayesian interpretations presented into a comprehensive statement about the Bayesian rationality of the overall languages. However, these dialects and interpretations provide a good point of departure for a complete analysis of the underlying languages. Having made this disclaimer, we may depict the crux of the paper symbolically (Fig. 1).

The nodes in Fig. 1 represent belief languages, and the arcs connecting nodes L_i and L_j represents the mapping $T_{ij}: L_i \rightarrow L_j$. Each of these arcs corresponds to a mapping discussed in the paper. For example, the node connecting L_{CF} and L_{Bayes} represents the (T2) mapping discussed in Section V-A. The arcs are bidirectional to reflect the observation that each of the mappings discussed in this paper can be made unique and invertible by fixing what is normally termed the prior belief. For example, given the mapping (T2): $L_{Bayes} \rightarrow L_{CF}$, knowledge of the Bayesian prior belief $P(h)$ grants the existence of the inverse mapping $(T2)_{p(h)}^{-1}: L_{CF} \rightarrow L_{Bayes}$. Some authors have argued that the DS and the CF languages do not require the specification of prior beliefs. We see that this assertion does not hold up under a Bayesian interpretation of these languages. In other words, for these languages to be "rational" on Bayesian grounds, one must be able to specify the prior belief in hypotheses.

The dotted arc in Fig. 1 describes one missing piece in the analysis, namely, the direct relationship between the certainty factors and the Dempster–Shafer languages. This linkage was established by Gordon and Shortliffe [22], who have shown that the CF calculus is a special case of Dempster's rule. This mapping between L_{CF} and L_{DS} completes the picture, so to speak, leading to two important observations.

First, we note that by, virtue of Corollary 1 and Fig. 1, there exist (compound) plausible mappings that yield isomorphisms between each pair of the languages L_{CF}, L_{CI}, and L_{DS}, and, furthermore, that all these languages are isomorphic to the Bayesian language L_{Bayes}. This means that at least on normative Bayesian grounds, these languages a) are equivalent and b) achieve no more than the classical belief language L_{Bayes}, which is restricted to ratio-form conditionally independent inference problems. The latter constraint is explicit in L_{Bayes} and in the Dempster–Shafer language but is implicit in the CF and in the contrast/inertia languages. This limitation is quite disconcerting, as real-life inference problems are rarely modular.

Moreover, simple assumptions of conditional independence often make belief-update schemes with multiple hypotheses too restrictive to be useful (Johnson [29]).

Going back to the isomorphism issues, it is intriguing to see that important dialects of different belief languages coming from the quarters of computer science (L_{CF}), psychology (L_{CI}), and statistics (L_{DS}) are all equivalent to a special case of the Bayesian language. This realization might lead to two different reactions. On the one hand, one is tempted to concur with Lindley [34] that "the only satisfactory description of uncertainty is (Bayesian) probability." On the other hand, there is no doubt that each belief language represents a unique cognitive challenge in terms of knowledge elicitation, and, viewed in this way, their tight normative proximity is in fact reassurance. In other words, different but isomorphic belief languages provide knowledge engineers with a rich variety of tools designed to represent the same phenomena.

With that in mind, future research may proceed in three directions. Starting with the normative ground, we reiterate the fact that the dialects L_{CF}, L_{CI} and L_{DS} are all special cases of their respective languages. Having shown that these dialects are remarkably Bayesian, future claims that a belief language can accomplish what L_{Bayes} cannot must be sought within extensions of these dialects. For example, an important normative contribution will be an attempt to describe and deal with correlated evidence outside the Bayesian language.

Efforts to develop and understand belief languages will be incomplete unless their cognitive or descriptive dimension receives serious attention. This experimental line of research can take two forms. Some researchers, e.g., Mitchell [36] and Schocken [48], have compared the validity of the recommendations generated by inference systems that were fed with degrees of belief provided by human experts. By keeping the human subjects constant, and varying the belief language "treatments," one can carry out statistical tests on the external validity of alternative belief languages. A second line of research, e.g. Wise [59] and Yadrick et al. [60], focuses on similar comparisons which were based on simulated, rather than human-supplied, degrees of belief. The former approach offers better proximity to realistic knowledge engineering settings; the latter enables one to generate worst-case data-sets that are difficult to detect with human subjects. Clearly, the two approaches are equally valid and complement each other.

Finally, a precriptive line of research must emerge. This effort will focus on how findings from normative and cognitive studies may be put to use in practice. This may promote a synthetic approach to knowledge engineering, i.e., one that combines attractive descriptive features of one language with the normative rigor of another.

REFERENCES

[1] J. B. Adams, "Probabilistic reasoning and certainty factors," in *Rule-Based Expert Systems*, B. G. Buchanan and E. H. Shortliffe, Ed. Reading, MA: Addison-Wesley, 1984, pp. 263–271.

[2] N. H. Anderson, *Foundations of Information Integration Theory.* New York: Academic, 1981.

[3] J. F. Baldwin and M. R. M. Monk, "SLOP—A system for support logic programming," Univ. of Bristol, I.T.R.C. Res. Rep., 1986.

[4] J. Baron, *Rationality and Intelligence.* Cambridge Univ. Press, 1985.

[5] ——, "Second-order probabilities and belief functions," *Theory Decision*, vol. 22, 1987.

[6] B. G. Buchanan and E. H. Shortliffe, "Uncertainty and evidential support," in *Rule-Based Expert Systems*, B. G. Buchanan and E. H. Shortliffe, Eds. Reading, MA: Addison-Wesley, 1984.

[7] R. Carnap, *Logical Foundations of Probability.* Chicago, IL: Univ. of Chicago Press, 1954.

[8] E. Charniak, "The Bayesian basis of common sense medical diagnosis," in *Proc. Nat. Conf. Artificial Intelligence*, 1983, pp. 70–73.

[9] P. Cheeseman, "A method of computing generalized bayesian probability values for expert systems," in *Proc. 8th Int. Joint Conf. Artificial Intelligence*, 1983.

[10] G. F. Cooper, "NESTOR: A computer-based medical diagnosis aid that integrates casual and probabilistic knowledge," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1984.

[11] ——, "Probabilistic inference using belief networks is NP-hard," Medical Computer Science Group, Knowledge Systems Lab. Stanford Univ., Stanford, CA, Tech. Rep. KSL-87-27, 1987.

[12] R. Cox, "Probability, frequency and reasonable expectation," *Amer. J. Phys.*, vol. 14, no. 1, pp. 1–13, 1946.

[13] B. de Finetti, *Probability, Induction and Statistics: The Art of Guessing.* New York: Wiley, 1972.

[14] R. O. Duda, P. E. Hart, and N. J. Nilsson, "Development of a computer-based consultant for mineral exploration," Stanford Res. Inst., Tech. Rep., International Projects 5821 and 6415, 1977.

[15] R. O. Duda and E. H. Shortliffe, "Expert systems research," *Science*, vol. 220, 261–268, 1981.

[16] W. Edwards, "Conservatism in human information processing," in *Formal Representation of Human Judgement*, B. Kleinmuntz, Ed. New York: Wiley, 1968.

[17] W. Edwards, H. Lindman, and L. J. Savage, "Bayesian statistical inference for psychological research," *Psychol. Rev.*, vol. 70, pp. 193–242, 1963.

[18] H. J. Einhorn and R. M. Hogarth, "Ambiguity and uncertainty in probabilistic inference," *Psychol. Rev.*, vol. 92, pp. 433–461, 1985.

[19] ——, "A contrast/surprise model for updating beliefs," Univ. Chicago, Chicago, IL, Tech. Rep., 1987.

[20] I. J. Good, *Good Thinking.* University of Minnesota Press, 1983.

[21] ——, *Probability and the Weighing of Evidence.* New York: Hafner, 1950.

[22] J. Gordon and E. H. Shortliffe, "The Dempster Shafer theory of evidence," in *Rule-Based Expert Systems*, B. G. Buchanan and E. H. Shortliffe, Ed. Reading, MA: Addison-Wesley, 1984, pp. 272–294.

[23] B. N. Grosof, "Evidential information as transformed probability," in *Uncertainty in Artificial Intelligence*, J. F. Lemmer and L. Kanal, Eds. Amsterdam, The Netherlands: North Holland, 1986, pp. 272–294.

[24] I. Hacking, *The Emergence of Probability.* Cambridge Univ. Press, 1975.

[25] P. Hajek, "Combining functions for certainty degrees in consulting systems," *Int. J. Man–Machine Studies*, vol. 22, pp. 59–76, 1985.

[26] D. E. Heckerman, "Probabilistic interpretation for Mycin's certainty factors," in *Uncertainty in Artificial Intelligence*, J. F. Lemmer and L. Kanal, Eds. Amsterdam, The Netherlands: North Holland, 1986.

[27] E. J. Horvitz, D. E. Heckerman, and C. P. Langlotz, "A framework for comparing alternative formalisms for plausible reasoning," in *Proc. AAAI Conf.*, Philadelphia, PA, 1986, pp. 210–214.

[28] R. A. Howard and J. E. Matheson, "Influence diagrams," in *Reading on the Principles and Applications of Decision Analysis*, R. A. Howard and J. E. Matheson, Eds. Strategic Decisions Group, Menlo Park CA, 1981, pp. 721–762.

[29] R. Johnson, "Independence and bayesian updating methods," in *Uncertainty in Artificial Intelligence*, L. Kanal and J. Lemmer, Eds. New York: North Holland, 1986, pp. 197–201.

[30] A. Koriat, S. Lichtenstein, and B. Fischhoff, "Reasons for confidence," *J. Experimental Psychol. Human Learning and Memory*, vol. 6, pp. 107–118, 1980.

[31] H. E. Kyburg, "Bayesian and non-bayesian evidential updating," *Artificial Intelligence*, vol. 31, pp. 271–293, 1987.

[32] C. Langlotz, E. Shortliffe, and L. Fagan, "Using decision theory to justify heuristics," in *Proc. AAAI Conf.*, Philadelphia, PA, 1986, pp. 215–219.

[33] J. Lemmer, "Generalized bayesian updating of incompletely specified distributions," in *Large Scale Systems 5*, New York: Elsevier, 1983.

[34] D. V. Lindley, "The probability approach to the treatment of uncertainty in artificial intelligence and expert systems," *Statist. Sci.*, vol. 2, pp. 17–24, 1987.

[35] L. L. Lopes, "Averaging rules and adjustment processes: The role of averaging in inference," Dept. of Psychol. Univ. of Wisconsin, Tech. Rep., 1982.

[36] D. H. Mitchell, "The shape experiment," Northwestern Univ., Tech. Rep., 1986.

[37] R. Nisbett and L. Ross, *Human Inference: Strategies and Shortcomings of Social Judgment.* Englewood Cliffs, NJ: Prentice-Hall, 1980.

[38] R. E. Nisbett, H. Zukier, and R. E. Lemley, "The dilution effect: nondiagnostic information weakens the implications of diagnostic information," *Cognitive Psychol.*, vol. 13, pp. 248–277, 1981.

[39] P. Harmon and D. King, *Expert Systems.* New York: Wiley, 1985.

[40] J. Pearl, "Fusion, propagation and structuring in belief networks," *Artificial Intell.*, Sept., 1986.

[41] ——, *Probabilistic Reasoning in Intelligent Systems.* Morgan-Kauffman, 1988.

[42] C. S. Peirce, "The probability of induction," *World Math.*, vol. 2, pp. 1341–1354, 1956.

[43] H. E. Pople, Jr., "The formation of composite hypotheses in diagnostic problem solving: an exercise in synthetic reasoning," in *Proc. 5th Int. Joint Conf. Artificial Intelligence*, Cambridge, MA, 1987, pp. 1030–1037.

[44] K. R. Popper, "Corroboration, the weight of evidence," in *The Logic of Scientific Discovery*, New York: Scientific Editions, 1959, pp. 387–419.

[45] H. Raiffa, *Decision Analysis.* Reading, MA: Addison-Wesley, 1968.

[46] F. P. Ramsey, "Truth and probability," in *The Foundations of Mathematics and Other Logical Essays*, R. B. Braithwaite, Ed. New York: Scientific Editions, 1931.

[47] L. J. Savage, *The Foundations of Statistics.* New York: Wiley, 1954.

[48] S. Schocken, *Quasi-Probabilistic Inference in Expert Systems: A Critical Review* (working title). Washington, D.C.: ICIT Press, in press.

[49] G. Shafer, "Jeffrey's rule of conditioning," *Phil. Sci.*, pp. 337–362, Sept. 1981.

[50] ——, *A Mathematical Theory of Evidence.* Princeton, NJ: Princeton Univ. Press, 1976.

[51] G. Shafer and A. Tversky, "Languages and designs for probability judgment," *Cognitive Sci.*, vol. 9, pp. 309–339, 1985.

[52] P. Shenoy and G. Shafer, "Propagating belief functions with local computations," *IEEE Expert*, vol. 1, pp. 43–52, 1986.

[53] E. H. Shortliffe, *Computer-Based Medical Consultation: MYCIN.* New York: American Elsevier, 1976.

[54] P. Slovic and S. Lichtenstein, "Comparison of bayesian and regression approaches to the study of information processing in judgment," *Organizational Behavior and Human Performance*, vol. 6, pp. 649–744, 1971.

[55] M. Snyder and W. B. Swann, "Hypothesis testing processes in social interaction," *J. Personality and Social Psychol.*, vol. 36, pp. 1202–1212, 1978.

[56] D. J. Spiegelhalter, "Rejoiner to probabilistic expert systems in medicine: practical issues in handling uncertainty," *Statist. Sci.*, vol. 2, pp. 43–44, 1987.

[57] A. Tversky and D. Khaneman, "Judgment under uncertainty: Heuristics and biases," *Science*, vol. 185, pp. 1124–1131, 1974.

[58] W. van Melle, E. H. Shortliffe, and B. G. Buchanan, "Emycin: A knowledge-engineer's tool for constructing rule-based expert systems," in *Rule-Based Expert Systems*, E. H. Shortliffe and B. G. Buchanan, Eds. Reading, MA: Addison-Wesley, 1984, pp. 302–313.

[59] B. P. Wise, "Experimentally comparing uncertain inference systems in probability," in *Uncertainty in Artificial Intelligence 2*, J. F. Lemmer and L. Kanal, Eds. Amsterdam, The Netherlands: North Holland, 1988, pp. 89–102.

[60] R. M. Yadrick, B. M. Perrin, D. S. Vaughan, P. D. Holden, and K. G. Kempf, "Evaluation of uncertain inference models I prospector, in *Uncertainty in Artificial Intelligence 2*, J. F. Lemmer and L. Kanal, Eds. Amsterdam, The Netherlands: North Holland, 1988, pp. 77–88.

Shimon Schocken received the Ph.D. degree in decision sciences from the Wharton School, University of Pennsylvania, Philadelphia, in 1987.

He is an Assistant Professor of Information Systems at the Leonard N. Stern School of Business of New York University, New York. His research interests focus on the nexus of decision theory and artificial intelligence.

Dr. Schocken received the Best Doctoral Dissertation Award in the information systems field from the International Center of Information Technology in 1987.

Paul R. Kleindorfer received the Ph.D. degree from Carnegie–Mellon University, Pittsburgh, PA, in 1970.

He was on the faculty of the Massachusetts Institute of Technology, the International Institute of Applied Systems Analysis, Vienna, Austria, and the University of Pennsylvania. He is currently Vice Dean and Professor of Decision Sciences and Economics at the Wharton School of the University of Pennsylvania, Philadelphia. He has published more than 50 papers and six books in the areas of decision theory, operations research, and public policy. He is Associate Editor of *Management Science* and the *Naval Research Logistics Quarterly*.