-

APPROXIMATE REASONING WITH CREDIBLE SUBSETS

Eliezer L. Lozinskii
Institute of Computer Science
The Hebrew University
Jerusalem 91904, Israel
e-mail: lozinski@cs.huji.ac.il

Short title: Reasoning with credible subsets

Abstract. The complexity of logical inference in knowledge systems grows with the system size determined, in particular, by the number of its clauses. This growth is so rapid, that reasoning in sizeable systems becomes intractable. However, given a system S and a query Q, should we be able to detect a subset of S informative enough to provide a correct answer to Q, and small enough to fit into the range of tractable computation, the answer to Q could be produced efficiently. This paper presents a way of implementing this idea.

1. Introduction.

Consider a Knowledge System S describing a real world W. A typical query addressing the system is whether certain statement about the state of the world is true. Assume that the knowledge stored in S is expressed in a First Order language, and a query is presented by a First Order formula F.

If F (or $\neg F$) is a logical consequence of S, then the answer to the query is that F is true (false, respectively) in W. If neither F nor $\neg F$ follows from S, then no answer can be as certain as in case of logical consequence. However, the system may suggest that F is more likely to be true than false in W. In this case, if the likelihood of the truth (or falsehood) of F is sufficiently high, we adopt a *belief* that F is true (false, respectively). This way of reasoning is prone to errors, as with any likelihood weaker than a full certainty, our beliefs may contradict the reality. Such uncertain beliefs represent *nonmonotonic reasoning* in the sense that new additions to the knowledge of S may change our previous beliefs. So, in many cases S provides only *approximate* answers. This phenomenon reflects the fact that our knowledge about the world is generally incomplete, such that S presents actually an approximation to the reality, and the goal of extending our knowledge is to make this approximation as close as possible.

Another problem of reasoning in knowledge systems is a high complexity of this process. Existing approaches to semantics of knowledge systems (McCarthy 1980, Reiter 1980, Genesereth and Nilsson 1987) require checking satisfiability of S, and computing its models. At the present state of the art of the complexity of computations, the algorithms implementing these processes require in general a time that is at least exponential in the size of S (Cook 1971, Valiant 1979). Assuming for the purpose of this discussion that S is presented in a CNF form, its size is determined by the number and size of its clauses, and the size of the set of its ground atomic formulae.

Though automated reasoning is formidably complex in general, its specific instances often turn out to be quite tractable. While checking satisfiability is NP-complete, it has been shown by an extensive research (Goldberg 1979, 1982, Purdom and Brown 1985, Gallo and Urbani 1989, Iwama 1989, Lozinskii 1993) that for many practical cases, algorithms based on the famous Davis-Putnam procedure (Davis and Putnam 1960) have a polynomial average time complexity. While determining the number of models of a logic system is #P-complete (Valiant 1979), Luby and Velickovic (1991) have presented an algorithm for approximating the number of models of a DNF propositional formula F

with m clauses on n variables, such that running time of the algorithm is estimated by a polynomial in m and n multiplied by $n^{O(\frac{1}{\delta}(\log^2 m + \log^2 \frac{1}{\varepsilon})(\log\log m + \log\frac{1}{\varepsilon} + \log\frac{1}{\delta}))}$, where ε, δ are the approximation errors. Linial and Nisan (1990) have presented an algorithm for computing the size of a union of a family of sets, that can be used for approximating the number of models. The algorithm runs in time $O(m^k)$, where k grows with the precision. It has been shown in (Lozinskii 1992) that under reasonable assumptions the average running time of computing the exact number of models of a propositional formula is $O(m^c n)$ where $c = O(\log m)$. For all that, the size of a system in which reasoning can be performed efficiently (presently, up to 100 variables and 100 clauses, in propositional terms) remains rather limited, well under the size of many practical knowledge collections.

Thus, given a system S and a query F, the amount of knowledge needed to be processed for answering F determines the very existence of a possibility to produce such an answer. So, an important question is whether the entire content of S or its sizable part is needed for answering a query. Should this be the general case, we would face a paradoxical situation, in which the larger is the system, the fewer queries can be answered, or the more we know the less we are able to decide. Although this proposition may remind an aphorism of ancient philosophers, human experience is different, namely, that a wider knowledge is beneficial as there usually exists a relatively small amount of knowledge *most relevant* to a given query that provides a reasonable answer.

So, given S and F, should we be able to detect a subset of S informative enough to provide a correct answer to F, and small enough to fit into the range of tractable computation, the answer to F could be produced efficiently. In the sequel we describe a way of implementing this idea.

2. Reasoning by evidence.

2.1. Evidence *versus* probability.

Given a system S representing an incomplete knowledge about a world W, and a query F, such that neither F nor $\neg F$ follows from S, the best answer to the query would be the probability that F is true in W.

A widely adopted approach to probabilistic reasoning is based on the *Bayes' Theorem* that allows computing unknown probability of some event given corresponding probabilities of other ones (Lee et al. 1987, Pearl 1988). Let $p(A)$, $p(A|B)$ denote, respectively, the prior probability that a statement (formula) A is true, and the posterior probability that A is true given that B is true. If $p(A) \neq 0$ and $B_1, B_2, \ldots, B_n$ are mutually exclusive and jointly exhaustive statements such that $\sum_{i=1}^{n} p(B_i) = 1$, then

$$p(B_i|A) = \frac{p(B_i) \cdot p(A|B_i)}{p(A)} = \frac{p(B_i) \cdot p(A|B_i)}{\sum_{i=1}^{n} p(B_i) \cdot p(A|B_i)} .$$

To apply the Bayesian approach to computing of $p(B_i|A)$ one needs to know $p(B_i)$ and $p(A|B_i)$ for all $i = 1, 2, \ldots, n$. This requirement of a full specification of probability distribution is often difficult to meet in real systems.

Several methods have been developed to cope with a lack of statistical information. A celebrated one is the *Principle of Maximum Entropy* (Jaynes 1957, 1979, Tribus 1961). The notion of entropy, widely used in physics, was considered by Hartley (1928) and Wiener (1948) in connection with information transmission, and introduced into Information Theory by Claude Shannon (Shannon 1948). Let P be a probability distribution assigning probabilities $p_1, \ldots, p_n$ to n mutually exclusive and jointly exhaustive events (or sentences) of a set R. Then the *entropy* $H(P)$ of P is defined by

$$H(P) = -\sum_{i=1}^{n} p_i \log p_i \ , \tag{2.1}$$

where $p_i \log p_i$ is assigned zero if $p_i = 0$.

$H(P)$ is a measure of uncertainty associated with R given the distribution P. In particular, $H(P)$ reaches its maximum value of $\log n$ if all the events of R are equiprobable,

$$p_1 = \cdots = p_n = 1/n \ . \tag{2.2}$$

This distribution is, indeed, most uncertain, since it does not allow to expect one event more than any other, or to prefer the truth of some statement of R to that of another one.

Now suppose that some knowledge K is available about real relationships existing among the elements of R. This knowledge can be regarded as a set of constraints on R. As the available knowledge about the real world is usually incomplete, there exists a set $\{ P_K \}$ of different distributions satisfying the constraints. Then an important question is, which one of the various P_K is most adequate to the reality? This *preferred distribution*, P_K^*, must reflect the fact that there is no other knowledge of R beyond that presented by K. So, besides satisfying the constraints, P_K^* must maximize the uncertainty associated with $R \cup K$. And this is just the decision suggested by the Principle of Maximum Entropy: Given R and K, prefer a probability distribution P_K^* satisfying K such that for all P_K, $H(P_K^*) \geq H(P_K)$. This principle, that can be traced some 100 years back to the works of Boltzmann, Maxwell, Gibbs, and Planck, and has been introduced in its modern form by Jaynes (1957) and Tribus (1961), proved its efficiency in various fields of Physics, Statistics, Information Theory, Pattern Recognition, Signal Processing, etc. (Aczel and Forte 1986, Goldszmidt 1990, Jaynes 1957, 1979, 1985, Justice 1986, Levine and Tribus 1979, Pearl 1989, Skyrms 1985, Smith and Erickson 1983, Smith and Grandy 1985, Tribus 1961, Williams 1980).

To be applied successfully, most of the known approaches to reasoning with uncertain and incomplete information require an extensive statistical knowledge (e.g., prior and posterior probabilities for Bayesian methods, basic probability assignment in Dempster-Shafer theory (Dempster 1967, 1968, Shafer 1976), prior beliefs for computing certainty factors in MYCIN (Heckerman 1986, Shortliffe 1976), partial probability assignments for Nilsson's Probabilistic Logic (Nilsson 1986)). In many practical situations such statistical data are either unavailable or unreliable. These circumstances raise an important question of how reasoning with incomplete information can be performed efficiently in the absence of reliable statistical data.

Although S may not imply the truth or falsehood of F, it provides some *semantic information* about F. For instance, if $S = \{p \vee q\}$, then S suggests that p and q cannot be false together, so, if q is false, then p is certainly true, etc. As a knowledge system S enables reaching certain conclusions regarding the truth of a statement F, we may say that S provides certain *evidence* of F (denoted $E(S, F)$). We would like the value of evidence $E(S, F)$ to present a measure of credibility of a belief in the truth of F.

As pointed out by the anonymous referee, in its spirit $E(S, F)$ is analogous with the conditional probability of F given S, or with the prior probability of F if S is assumed to be true with regard to W (as a knowledge system is usually believed to present a faithful description of the world). However, because of an inevitable incompleteness of the available knowledge about the real world, the value of $E(S, F)$ (as shown in the sequel) usually differs from these probabilities. The latter fact justifies the usage of a different term, such as "evidence", for $E(S, F)$.

To present a measure of credibility of a belief in the truth of F the value of evidence has to possess the following reasonable properties, that are equivalent to the probability axioms:

(1) If F is a logical consequence of S, then the truth of F is most credible, so $E(S, F \mid S \models F) = 1$ (Let the range of evidence be normalized in [0, 1]).

(2) If $S \models \neg F$, then the belief in the truth of F has no credibility. Hence, $E(S, F \mid S \models \neg F) = 0$.

(3) In particular, if ϕ, ψ are a tautology and an unsatisfiable formula, respectively, then for all S, $E(S, \phi) = 1$ and $E(S, \psi) = 0$.

(4) $E(S, \phi \vee \psi) = E(S, \phi) + E(S, \psi) - E(S, \phi \wedge \psi)$.

(5) Properties (3) and (4) imply for all S and ψ, $E(S, \psi) + E(S, \neg\psi) = 1$.

If S represents W faithfully, then every momentary state of W corresponds to a model of S, although, because of an incompleteness of the knowledge presented in S, some models may describe states of W that do not occur in the reality. However, these states appear to be possible from the standpoint of the current knowledge. This observation underlies the widely adopted view of W as a set of states, referred to as *possible worlds*, such that there is a bijection between the set of possible worlds and the set $MOD(S)$ of all models of S (S is supposed to be consistent).

At any moment the world is in one and only one of its possible states. So, $MOD(S)$ provides a set of mutually exclusive and exhaustive particular representations of W. Let $p(\mu)$ denote the probability that W is in a state represented by a model $\mu \in MOD(S)$. Then the probability $p(F)$ that a formula F is true in the current state of W can be expressed as

$$p(F) = \sum_{\mu \in MOD(S)} p(F|\mu) \cdot p(\mu) = \sum_{\mu \in MOD(S) \wedge \mu \models F} p(\mu) , \tag{2.3}$$

since

$$p(F|\mu) = \begin{cases} 1 & if \quad \mu \models F \\ 0 & otherwise. \end{cases}$$

Unfortunately, a full probability distribution over the space of possible states is usually either not available or not satisfactorily reliable, although a partial distribution may be known for some possible worlds represented by a subset of models $M \subset MOD(S)$. If M does not contain all models asserting F , then a question arises of how the probability of the models not contained in M should be estimated.

2.2. Applying the Principle of Maximum Entropy.

Consider a user interested in finding out the actual state of a world W given its description by a knowledge system S . As the system represents every possible world by means of one of its models, a probability distribution over $MOD(S)$ is necessary for reasoning about the actual state of W . Suppose that prior probabilities $\rho(\mu)$ are known for models of a subset M of $MOD(S)$. Given this partial distribution P_M , a rational reasoner should prefer a full distribution P_{MOD} over $MOD(S)$ that (a) is consistent with S , (b) includes P_M , and (c) implies no extra knowledge beyond S and P_M . According to the Principle of Maximum Entropy (Jaynes 1957, Tribus 1961), P_{MOD} must maximize the value of Shannon's entropy $H(P)$ (cf Section 2.1). This implies the following definition of P_{MOD} :

$$p(\mu) = \begin{cases} \rho(\mu) & if \quad \mu \in M \\ \dfrac{1 - \sum_{\mu \in M} \rho(\mu)}{|MOD(S) - M|} & otherwise. \end{cases} \tag{2.4}$$

Now the evidence $E(S, F)$ can be defined as the probability that F is true in the current state of the world given the reasonable, although not necessarily actual, distribution P_{MOD} .

Definition 2.1. *Evidence.* Due to (2.3) and (2.4),

$$E(S, F) = \sum_{\mu \in M \wedge \mu \models F} \rho(\mu) + \sum_{\mu \in (MOD(S) - M) \wedge \mu \models F} p(\mu)$$

$$= \sum_{\mu \in (M \cap MOD(S \cup \{F\}))} \rho(\mu) + (1 - \sum_{\mu \in M} \rho(\mu)) \frac{|MOD(S \cup \{F\}) - M|}{|MOD(S) - M|} , \tag{2.5}$$

where $|MOD(S)|$, $|MOD(S \cup \{F\})|$ denote, respectively, the cardinality of the set $MOD(S)$ of models of S , and that of its subset consisting of all models in which F is true.

If $MOD(S \cup \{F\}) \subseteq M$, then $E(S, F)$ amounts to the probability of F . If no prior probabilities of possible worlds are known, such that $M = \emptyset$, then

$$E(S, F) = \frac{|MOD(S \cup \{F\})|}{|MOD(S)|} , \tag{2.6}$$

that corresponds to the assumption (by the Principle of Maximum Entropy) that all possible worlds are equiprobable. □

It is easy to see that Definition 2.1 is in accord with the properties (1)-(5).

In the sequel we adopt evidence $E(S, F)$ as a measure of credibility of a belief in the truth of F suggested by S . Thus,

(i) if $E(S, F) = 1$ then F is certainly true;

(ii) if $0.5 < E(S, F) < 1$ then F is more likely to be true than false, and the larger is the value of $E(S, F)$ the more credible is the belief in the truth of F ;

(iii) if $E(S, F) = 0.5$ then F is either true or false with the same likelihood, so, both possible beliefs in F have the same, rather low credibility;

(iv) if $0 < E(S, F) < 0.5$ then F is more likely to be false than true, and the smaller is the value of $E(S, F)$ the more credible is the belief in falsehood of F ;

(v) if $E(S, F) = 0$ then F is certainly false.

An implicit assumption underlying Definition 2.1 is that the set of models of S is finite. However, it has been shown in (Lozinskii 1995) how the value of evidence can be calculated in systems with infinite sets of models. Numerous examples presented in (Lozinskii 1994, 1995) illustrate semantic plausibility of reasoning by evidence.

3. Motivating examples.

Consider a knowledge system S as a set of clauses (with no loss of generality), such that logically S is a conjunction of all its clauses, each clause being a disjunction of literals. Suppose, for simplicity, that a query is concerned with a ground atomic formula A , $F = A$ (Section 6 shows a way of extending this approach to an arbitrary formula F). As we are seeking a subset of S most relevant to A , we have to figure out which clauses of S play a major role in determining our belief in F , that is, make a main contribution to the value of $E(S, F)$. This section is intended to provide just a motivation for the proposed approach, so, in the absence of a specific statistics regarding the possible worlds, we use expression (2.6) for estimating the value of evidence.

To simplify the analysis, let us consider S presented in a propositional form consisting of m clauses over n variables. Let l_i denote the *length* of a clause C_i , that is the number of literals in C_i . Consider different clauses of S such that $A \in C_i$, $\neg A \in C_j$, while C_k contains no occurrence of A , and for each one of these clauses let us compare sets of its models asserting and falsifying A :

$$|MOD(C_i \wedge A)| = 2^{n-1} > |MOD(C_i \wedge \neg A)| = (2^{l_i - 1} - 1)2^{n - l_i} ,$$

$$|MOD(C_j \wedge A)| = (2^{l_j - 1} - 1)2^{n - l_j} < |MOD(C_j \wedge \neg A)| = 2^{n-1} ,$$

$$|MOD(C_k \wedge A)| = (2^{l_k} - 1)2^{n - l_k - 1} = |MOD(C_k \wedge \neg A)| .$$

A clause not involving A has the same number of models that assert and falsify A, while for a clause involving a literal with A, the number of its models asserting the literal is greater than that falsifying it. So, if all these models are equiprobable, then $E(C_i, A) > 0.5$, $E(C_j, A) < 0.5$, $E(C_k, A) = 0.5$. In this case clauses involving A suggest a definite belief in A, while those with no occurrence of A are indifferent w.r.t. A providing no immediate information about A.

Nevertheless, a clause not involving A may influence the evidence $E(S, F)$ if it has a variable in common with some other clause that itself affects the value of $E(S, F)$, as shown by the following example.

Example 3.1. Consider $S = \{C_1, C_2\}$, $C_1 = \{a \vee b \vee c\}$, $C_2 = \{b \vee \neg c\}$.

$$E(S, a) = \frac{3}{5} \ ; \quad E(C_1, a) = \frac{4}{7} \ ; \quad E(C_2, a) = \frac{1}{2} \ .$$

We may say that the contribution to $E(S, a)$ by C_2 over the evidence provided by C_1 is 3/5 - 4/7 = 1/35, while the addition by C_1 to the indecisive evidence provided by C_2 is 3/5 - 1/2 = 1/10. □

However, in the reality the prior probability distribution over the set of models of S may be non-uniform, such that even a clause C_k not involving A may contribute significantly to the value of $E(S, A)$. Suppose that the probability of the models of C_k asserting A is greater than that of the models of C_k falsifying A, such that $E(C_k, A) > 0.5$. But in this case it is likely that the values of $E(C_i, A)$ and $E(C_j, A)$ are also greater than in the case of a uniform distribution, such that the contribution of C_i to $E(S, A)$ is still greater than that of C_k (although the value of $E(C_j, A)$ may become less decisive (closer to 0.5) than under a uniform distribution). Further, if under a non-uniform distribution $E(C_k, A) < 0.5$, then it is likely that the values of $E(C_i, A)$ and $E(C_j, A)$ are less than that of a uniform case, such that again C_k contributes less than C_j. So, it is reasonable to assume that a set of all clauses of S involving A contains clauses making a major contribution to the value of $E(S, A)$. The closer the actual distribution is to a uniform one, the better this assumption conforms to the reality. Although an occurrence of a literal in a clause is a syntactic feature, the abovementioned assumption reflects the semantics of S as it is based on considerations regarding the set of the system models that represent possible states of the real world..

These observations suggest that S can be partitioned into subsets of clauses having different influence on the value of evidence provided by the system.

Definition 3.1. Let $L^{(1)}$ denote a subset of S consisting of all clauses of S containing a literal L or $\neg L$. Call $L^{(1)}$ the *first surrounding* of L. For $i > 1$, let $L^{(i)}$ denote the *i-th surrounding* of L, which is a set of all clauses of S that either belong to $L^{(i-1)}$ or share a common variable with a clause of $L^{(i-1)}$. Denote by $R^{(i)}(S, L)$

the *i-th residue* of S w.r.t. L, such that $R^{(i)}(S, L) = S - L^{(i)}$ □

It appears reasonable to conjecture that the larger is i, the larger is the contribution of $L^{(i)}$ to the evidence $E(S, L)$, and the smaller is the influence of the residue $R^{(i)}(S, L)$ on $E(S, L)$. To check this intuition and motivate<e a further study, let us look at the following example.

Example 3.2. *Chain of clauses.*

Consider a propositional system S, and a literal L such that the n-th surrounding of L in S consists of a chain of clauses displayed in Figure 3.1. Each clause C_i consists of three disjoined segments C_{i1}, C_{i2}, C_{i3} such that for $1 \le i < n$, C_{i3} and $C_{i+1,1}$ are necessarily non-empty and involve the same variables (the latter do not occur in any other clause of S), while C_{i2} is either empty or contains variables shared by no other clause of S. C_{11} contains L. All variables of C_{n3} are shared by the residue $R^{(n)}(S, L)$. Let all clauses of $L^{(n)}$ have the same length such that the length of C_{i1}, C_{i2}, C_{i3} is α, β, α, respectively. □

We wish to compare the evidence of L given by its i-th surrounding, $E(L^{(i)}, L)$, with that provided by the entire system, $E(S, L)$, such that for $1 \le i \le n$, $L^{(i)} = \{C_1, ..., C_i\}$, $S = L^{(i)} \cup R^{(i)}$ (if S and L are obvious, (S, L) may be dropped from the notation $R^{(i)}(S, L)$ of residue).

To do so, let us consider *images* of S such that an i-th image ($1 \le i \le n$), denoted S_i (Figure 3.2), contains the i-th surrounding of L in S, $L^{(i)}$, but the content of the residue $R_i^{(i)}$ of the image is assumed to be unknown, except that $R_i^{(i)}$ shares all variables occurring in C_{i3}, but no other ones occurring in $L^{(i)}$. So, the influence of $R_i^{(i)}$ on the evidence $E(S_i, L)$ is determined by the value of $E(R_i^{(i)}, C_{i3})$. As $R_i^{(i)}$ is undefined, all what is known is that $0 \le E(R_i^{(i)}, C_{i3}) \le 1$. (Occurrence of i in both subscript and superscript of R may be confusing, but please note that a subscript denotes the image of S, while a superscript corresponds to the surrounding of L).

In any image S_i literal L occurs only in C_1, and the only element of C_1 sharing variables with the rest of the image, $R_i^{(1)}$, is C_{13}, so, the influence of the residue $R_i^{(1)}$ on the evidence $E(S_i, L)$ is determined by the value of $E(R_i^{(1)}, C_{13})$, while the latter depends on $E(R_i^{(1)}, C_{21})$ and the relationship between C_{13} and C_{21}. For brevity, denote $G_i = E(R_i^{(1)}, C_{13})$, $H_i = E(R_i^{(1)}, C_{21})$, $M_i = |MOD(R_i^{(1)})|$. Consider any model as a set of literals. Then every model of S_i contains one and only one model of $R_i^{(1)}$.

Let us now evaluate the number of models of $R_i^{(1)}$ satisfying C_{13}, and then the number of models of S_i satisfying C_{11}. $G_i M_i$ models of $R_i^{(1)}$ assert C_{13}, and so, satisfy C_1. Hence, if μ' is one of these models, then there are $2^{\alpha + \beta}$ models of S_i containing μ' (since α variables of C_{11} and β variables of C_{12} are shared by no other clause of S_i). Among the latter models there are $(2^\alpha - 1)2^\beta$ ones asserting C_{11}. Further, $(1 - G_i)M_i$ models of $R_i^{(1)}$ falsify C_{13}. Let μ'' be one of these models. Then there are $2^{\alpha + \beta} - 1$ models of S_i containing μ''. Among the latter models there are $(2^\alpha - 1)2^\beta$ ones asserting C_{11}.

-

So,

$$E(S_i, C_{11}) = \frac{(2^\alpha - 1)2^\beta G_i M_i + (2^\alpha - 1)2^\beta (1 - G_i) M_i}{2^{\alpha+\beta} G_i M_i + (2^{\alpha+\beta} - 1)(1 - G_i) M_i} = \frac{(2^\alpha - 1)2^\beta}{2^{\alpha+\beta} - 1 + G_i} . \quad (3.1)$$

As the variables of C_{11} appear in no other element of S_i , the number of models of S_i is a multiple of the number $2^\alpha - 1$ of different assignments to the variables of C_{11} satisfying C_{11} , while $2^{\alpha-1}$ of these assignments assert L . So, in general,

$$E(S_i, L) = \frac{2^{\alpha-1}}{2^\alpha - 1} E(S_i, C_{11}) = \frac{2^{\alpha+\beta-1}}{2^{\alpha+\beta} - 1 + G_i} . \quad (3.2)$$

As G_i depends on H_i , let us evaluate this dependence. C_{13} and C_{21} contain the same variables, but the corresponding literals of these sub-clauses may be either identical or different. If C_{13} is identical to C_{21} , then $G_i = H_i$. If C_{13} differs from C_{21} (i.e., there is at least one literal l such that $l \in C_{13}$, but $\neg l \in C_{21}$), then consider two disjoined subsets of models of $R_i^{(1)}$ such that C_{21} is true in one of them, but false in the other. Observe that in this case of different C_{13} and C_{21} :

(a) C_{13} is true in all models of $R_i^{(1)}$ falsifying C_{21} , that is, in $(1 - H_i)M_i$ models;

(b) Note that $R_i^{(1)}$ contains clauses $C_2, \ldots , C_i$ of the chain, so, no one of α variables of C_{21} is shared by other clauses of $R_i^{(1)}$. Hence, the number $H_i M_i$ of models of $R_i^{(1)}$ satisfying C_{21} is a multiple of the number $2^\alpha - 1$ of different assignments to these variables satisfying C_{21} . Only one of these assignments falsifies C_{13} . Hence, the number of models of $R_i^{(1)}$ satisfying both C_{21} and C_{13} is $H_i M_i (2^\alpha - 2) / (2^\alpha - 1)$.

In summary, if C_{13} differs from C_{21} , then the number of models of $R_i^{(1)}$ satisfying C_{13} is $(1 - H_i)M_i + H_i M_i (2^\alpha - 2) / (2^\alpha - 1) = (1 - H_i / (2^\alpha - 1))M_i$. Hence,

$$G_i = E(R_i^{(1)}, C_{13}) = \begin{cases} H_i \quad \mathit{if} \quad C_{13} = C_{21} \\ 1 - \dfrac{H_i}{2^\alpha - 1} \quad \mathit{otherwise} . \end{cases} \quad (3.3)$$

H_i may vary, so,

$$(G_i)_{max} = max\ [(H_i)_{max} , 1 - \frac{(H_i)_{min}}{2^\alpha - 1}] ; \quad (3.4)$$

$$(G_i)_{min} = min\ [(H_i)_{min} , 1 - \frac{(H_i)_{max}}{2^\alpha - 1}] . \quad (3.5)$$

Consider a *threshold* value H_0 such that

$$H_0 = 1 - \frac{H_0}{2^\alpha - 1} ; \quad H_0 = \frac{2^\alpha - 1}{2^\alpha} . \quad (3.6)$$

Notice that if

$$(H_i)_{min} \geq H_0 \ , \tag{3.7}$$

then

$$(H_i)_{max} \geq (H_i)_{min} \geq H_0 = 1 - \frac{H_0}{2^\alpha - 1} \geq 1 - \frac{(H_i)_{min}}{2^\alpha - 1} \geq 1 - \frac{(H_i)_{max}}{2^\alpha - 1} \ . \tag{3.8}$$

Thus,

$$(G_i)_{max} = (H_i)_{max} \ ; \tag{3.9}$$

$$(G_i)_{min} = 1 - \frac{(H_i)_{max}}{2^\alpha - 1} \ . \tag{3.10}$$

By (3.1),

$$E(S_i, C_{11})_{max} = \frac{(2^\alpha - 1)2^\beta}{2^{\alpha+\beta} - 1 + (G_i)_{min}} = \frac{(2^\alpha - 1)^2 \, 2^\beta}{(2^\alpha - 1)2^{\alpha+\beta} - (H_i)_{max}} \ ; \tag{3.11}$$

$$E(S_i, C_{11})_{min} = \frac{(2^\alpha - 1)2^\beta}{2^{\alpha+\beta} - 1 + (H_i)_{max}} \ . \tag{3.12}$$

Now note that for $1 \leq i < n$, a residue $R_{i+1}^{(1)}$ of an image S_{i+1} becomes an instance of S_i if the following correspondence is maintained: for $1 \leq j \leq i$, C_{j+1} of $R_{i+1}^{(1)}$ corresponds to C_j of S_i (all clauses have the same length), and $R_{i+1}^{(i+1)}$ corresponds to $R_i^{(i)}$ (both are assumed undefined). Due to this correspondence,

$$(H_{i+1})_{max} = E(S_i, C_{11})_{max} \ ; \tag{3.13}$$

$$(H_{i+1})_{min} = E(S_i, C_{11})_{min} \ . \tag{3.14}$$

Lemma 3.1. For all $i \geq 1$,

$$E(S_{i+1}, L)_{max} \leq E(S_i, L)_{max} \ ; \quad E(S_{i+1}, L)_{min} \geq E(S_i, L)_{min} \ .$$

Proof by induction.

Basic step. As $R_1^{(1)}$ is undefined, $0 \leq G_1 \leq 1$, and so, by (3.1),

$$E(S_1, C_{11})_{max} = \frac{(2^\alpha - 1)2^\beta}{2^{\alpha+\beta} - 1} \ ; \tag{3.15}$$

$$E(S_1, C_{11})_{min} = \frac{(2^\alpha - 1)2^\beta}{2^{\alpha+\beta}} = \frac{2^\alpha - 1}{2^\alpha} \ . \tag{3.16}$$

By (3.13) and (3.14),

$$(H_2)_{max} = \frac{(2^\alpha - 1)2^\beta}{2^{\alpha+\beta} - 1} \ ; \tag{3.17}$$

$$(H_2)_{min} = \frac{2^\alpha - 1}{2^\alpha} \ . \tag{3.18}$$

Since the value of $(H_2)_{min}$ satisfies (3.7), we get from (3.11) and (3.12)

$$(H_3)_{max} = E(S_2, C_{11})_{max} = \frac{(2^\alpha - 1)^2\, 2^\beta}{(2^\alpha - 1)2^{\alpha+\beta} - (H_2)_{max}} = \frac{(2^\alpha - 1)(2^{\alpha+\beta} - 1)}{2^{2\alpha+\beta} - 2^\alpha - 1} \ ; \tag{3.19}$$

$$(H_3)_{min} = E(S_2, C_{11})_{min} = \frac{(2^\alpha - 1)2^\beta}{2^{\alpha+\beta} - 1 + (H_2)_{max}} = \frac{(2^\alpha - 1)(2^{\alpha+\beta} - 1)}{2^{2\alpha+\beta} - 2^\alpha - 1 + 2^{-\beta}} \ . \tag{3.20}$$

A comparison between (3.15), (3.16) and (3.19), (3.20) yields

$$E(S_2, C_{11})_{max} \le E(S_1, C_{11})_{max} \ , \quad E(S_2, C_{11})_{min} \ge E(S_1, C_{11})_{min} \ , \tag{3.21}$$

where the equality holds only for $\alpha = 1$, $\beta = 0$.

By (3.2), $E(S_i, L)$ is proportional to $E(S_i, C_{11})$, so, the Lemma holds for $i = 1$.

Inductive step. (a) If for some j , $(H_{j+1})_{min} \ge (H_j)_{min} \ge H_0$, satisfying (3.7), then by (3.11) - (3.14),

$$(H_{j+1})_{max} = \frac{(2^\alpha - 1)^2\, 2^\beta}{(2^\alpha - 1)2^{\alpha+\beta} - (H_j)_{max}} \ ; \tag{3.22}$$

$$(H_{j+1})_{min} = \frac{(2^\alpha - 1)2^\beta}{2^{\alpha+\beta} - 1 + (H_j)_{max}} \ . \tag{3.23}$$

(b) If for this value of j , $(H_{j+1})_{max} \le (H_j)_{max}$, then by (3.22), (3.23),

$$(H_{j+2})_{max} \le (H_{j+1})_{max} \quad \textit{and} \quad (H_{j+2})_{min} \ge (H_{j+1})_{min} \ge H_0 \ . \tag{3.24}$$

Both these conditions, (a) and (b), are satisfied for $j = 2$ in the Basic step, so, inequalities (3.24) hold for all $j \ge 2$. By the same argument as that used in the Basic step, for all $i \ge 1$,

$$E(S_{i+1}, L)_{max} \le E(S_i, L)_{max} \ ; \quad E(S_{i+1}, L)_{min} \ge E(S_i, L)_{min} \ . \quad \square$$

So far we have analyzed images S_i of the sample system S in which the residue $R_i^{(i)}$, although undefined, has been taken into account assuming $E(R_i^{(i)}, C_{i3})_{max} = 1$, $E(R_i^{(i)}, C_{i3})_{min} = 0$. Now we wish to evaluate the evidence $E(L_i^{(i)}, L)$ of L provided solely by its i-th surrounding disregarding the existence of unknown $R_i^{(i)}$. This is equivalent to assuming that $R_i^{(i)}$ is empty. Any interpretation of α variables of C_{i3} is a model of an empty system, but only $2^\alpha - 1$ of them satisfy C_{i3} , so, $E(\varnothing, C_{i3}) = (2^\alpha - 1) / 2^\alpha$. Hence, for $\alpha \ge 1$, and any residue $R_i^{(i)}$,

$$E(R_i^{(i)}, C_{i3})_{min} < E(\varnothing, C_{i3}) < E(R_i^{(i)}, C_{i3})_{max} \ . \tag{3.25}$$

Lemma 3.2. For any image S_i , $E(S_i, L)$ is either monotonously increasing (monotonic) or monotonously decreasing (anti-monotonic) a function of $E(R_i^{(i)}, C_{i3})$.

Proof. By (3.1), $E(S_i, C_{11})$ is anti-monotonic a function of $G_i = E(R_i^{(1)}, C_{13})$, while the latter, by (3.3), is either monotonic or anti-monotonic a function of $H_i = E(R_i^{(1)}, C_{21})$. Note that $R_i^{(1)}$ becomes an instance of S_i if C_2 in $R_i^{(1)}$ corresponds to C_1 in S_i, and $R_i^{(2)}$ in $R_i^{(1)}$ corresponds to $R_i^{(1)}$ in S_i. So, by (3.1), $E(R_i^{(1)}, C_{21})$ is anti-monotonic a function of $E(R_i^{(2)}, C_{23})$. By repeating this recursive argument $i-1$ times, we get that $E(S_i, C_{11})$ is either monotonic or anti-monotonic a function of $E(R_i^{(i)}, C_{i3})$, and so is also $E(S_i, L)$, as the latter is, by (3.2), proportional to $E(S_i, C_{11})$. □

If $L_i^{(i)}$ is considered as a separate system, then an empty system is its *i*-th residue in the same way as $R_i^{(i)}$ is the *i*-th residue of S_i. So, by (3.25), Lemma 3.2 yields

$$E(S_i, L)_{min} < E(L_i^{(i)}, L) < E(S_i, L)_{max} . \tag{3.26}$$

Our goal is to evaluate the *error* ε taking place when the evidence of L provided by the sample system S is approximated by the i-th surrounding of L in S:

$$\varepsilon^{(i)} = E(L^{(i)}, L) - E(S, L) . \tag{3.27}$$

Theorem 3.1. The absolute value of the approximation error $\varepsilon^{(i)}$ is a decreasing function of i.

Proof. Consider an instance of system S described in Example 3.2 (Figure 3.1) containing a chain of n clauses. Given a literal L, we are interested in its i-th ($i \le n$) surrounding $L^{(i)}$ and the corresponding residue $R^{(i)}$, such that $S = L^{(i)} \cup R^{(i)}$. Comparing S to its i-th image $S_i = L_i^{(i)} \cup R_i^{(i)}$, we notice that $L^{(i)} = L_i^{(i)}$, but the residues are different as $R_i^{(i)}$ is undefined, while $R^{(i)}$ is determined by the specific contents of S. So,

$$E(R_i^{(i)}, C_{i3})_{min} \le E(R^{(i)}, C_{i3}) \le E(R_i^{(i)}, C_{i3})_{max} . \tag{3.28}$$

By (3.28) and Lemma 3.2,

$$E(S_i, L)_{min} \le E(S, L) \le E(S_i, L)_{max} . \tag{3.29}$$

As $L^{(i)} = L_i^{(i)}$, we get by (3.26) and (3.29),

$$|\varepsilon^{(i)}| = |E(L^{(i)}, L) - E(S, L)| < E(S_i, L)_{max} - E(S_i, L)_{min} . \tag{3.30}$$

And by Lemma 3.1, the difference $E(S_i, L)_{max} - E(S_i, L)_{min}$ is a decreasing function of i. □

4. Credible subsets.

The analysis presented in Section 3 concerns the sample system S (Example 3.1) containing a chain of n clauses $C_1, \ldots, C_n$ such that for $1 \le i < n$, C_i and C_{i+1} share common variables. It is shown that if an evidence of a literal $L \in C_1$ is derived not from the entire system, but from the i-th surrounding of L, then the error of this approximation decreases with growing i. This suits a common sense intuition that the larger surrounding of L is taken into consideration, the more information relevant to L it contains suggesting a more precise conclusion.

Of course, a real knowledge system has much more complex a structure than the simplistic one of Section 3. However, what they have in common is that in any system, for a literal L and any clause C beyond the i-th surrounding of L, any chain of clauses connecting C to L contains at least i clauses. So, the influence of C on L is supposed to weaken with growing "distance" i. This observation motivates the following approach.

Let KS denote a class of knowledge systems under consideration, and L be a literal of interest. Suppose that given a specific system $T \in KS$, we have decided to base our belief in L on the evidence derived from a proper subset of T, say, the i-th surrounding of L in T.

Denote $E = E(T, L)$, $E^{(i)} = E(L^{(i)}, L)$. To evaluate the approximation error $\varepsilon^{(i)} = E^{(i)} - E$, let us consider $\varepsilon^{(i)}$ as a random variable over all systems of KS, with mean value $\mu^{(i)}$ and standard deviation $\sigma^{(i)}$. Generally, the smaller the values of $\varepsilon^{(i)}$, $\mu^{(i)}$, $\sigma^{(i)}$ are, the better is the approximation. Consider an instance of $E = 0.6$ and two approximations: $E_1^{(i)} = 0.4$ and $E_2^{(i)} = 0.99$. The purpose of computing the value of E is producing a credible belief in the truth of L. So, $|\varepsilon_1^{(i)}|$ is almost twice as small as $|\varepsilon_2^{(i)}|$, but on the other hand, $E_2^{(i)}$ suggests the same belief as E (namely, that L is more likely to be true than false), while $E_1^{(i)}$ supports the opposite one. Thus, besides the value of $\varepsilon^{(i)}$, an important question determining the credibility of the approximation is whether both evidences E and $E^{(i)}$ suggest the same belief in the truth of L. Say that an evidence has a *positive*, *zero* or *negative displacement* if it is greater, equal or less than 0.5, respectively. So, E and $E^{(i)}$ suggest the same belief in L iff both of them have the same kind of displacement. A sufficient condition of this event is

$$|\varepsilon^{(i)}| < |E^{(i)} - 0.5| . \tag{4.1}$$

By Chebyshev's inequality, for a random variable X with mean μ and standard deviation σ, and any positive number k, the probability

$$prob[|X - \mu| < k\sigma] \geq 1 - \frac{1}{k^2} . \tag{4.2}$$

Then for the probability $\rho^{(i)}$ that both E and $E^{(i)}$ suggest the same belief in L, we have (by noting that $|\varepsilon^{(i)}| \leq |\varepsilon^{(i)} - \mu^{(i)}| + |\mu^{(i)}|$)

$$\rho^{(i)} = prob[|\varepsilon^{(i)}| < |E^{(i)} - 0.5|] \geq prob[|\varepsilon^{(i)} - \mu^{(i)}| + |\mu^{(i)}| < |E^{(i)} - 0.5|]$$

$$= prob[|\varepsilon^{(i)} - \mu^{(i)}| < |E^{(i)} - 0.5| - |\mu^{(i)}|] \geq 1 - \left(\frac{\sigma^{(i)}}{|E^{(i)} - 0.5| - |\mu^{(i)}|} \right)^2 . \tag{4.3}$$

Definition 4.1. A subset t of a system T is *q-credible* w.r.t. a literal L, and $E(t, L)$ is a *q-credible* evidence if the probability that $E(t, L)$ and $E(T, L)$ suggest the same belief in L is at least q. □

So, given $T \in KS$ and L, the i-th surrounding of L is q-credible if

$$|E^{(i)} - 0.5| \geq \frac{\sigma^{(i)}}{\sqrt{1-q}} + |\mu^{(i)}| \,. \tag{4.4}$$

Let the k-th surrounding be the largest surrounding of L in T. As i grows from 1 to k, $E^{(i)}$ approaches E, and so, the credibility of $L^{(i)}$ tends to 1. Because $L^{(i)} \subset L^{(i+1)}$ for all $i < k$, and since the time of computing $E^{(i)}$ grows very fast (exponentially, in the worst case) with the size of $L^{(i)}$, there is a sharp trade-off between the credibility of evidence and the amount of work for its computing, such that a relatively small decrease of credibility may provide a very large saving of computing time. So, applications that may tolerate certain amount of errors in their decisions, can benefit from reasoning by small surroundings. This benefit is especially significant in cases where a close surrounding of L provides it with a decisive value of evidence satisfying (4.4). As an extreme example, if $E(L^{(1)}, L) = 1$ meaning that L is a logical consequence of $L^{(1)}$, then there is no need to consider any more clauses beyond $L^{(1)}$ (provided that the system is consistent).

5. Experiments.

To gather statistics regarding reasoning by surroundings, we conducted experiments with a class of propositional systems as follows. Each system is a set of m clauses over n variables; each clause is a set of l literals; each of the l variables is chosen randomly out of the n system variables, and appears as a literal with or without negation with the same probability. The system parameters varied in the following ranges: $10 \leq m \leq 100$, $10 \leq n \leq 100$, $3 \leq l \leq 7$. For each tested combination of parameters, a number of system instances between 200 (for large systems) and 3000 (for small ones) were randomly generated. Number of models was computed by an improved version of the algorithm presented in (Lozinskii 1992). Most of the experiments ran on *MOSIX* (Barak 1993), a distributed computer system consisting of 16 Pentium-based processors, designed and implemented by Amnon Barak at the Institute of Computer Science, the Hebrew University of Jerusalem.

By property (5) of evidence (Section 2),

$$E(S, L) = 1 - E(S, \neg L)\,, \quad E(L^{(i)}, L) = 1 - E(L^{(i)}, \neg L)\,. \tag{5.1}$$

In the systems tested in the experiments L and $\neg L$ appeared with the same probability. So, for $0 \leq x \leq 1$,

$$prob[E(S, L) \leq x] = prob[E(S, \neg L) \leq x] = prob[1 - E(S, L) \leq x]$$

$$= prob[E(S, L) \geq 1 - x]\,. \tag{5.2}$$

The same is true of $E(L^{(i)}, L)$. Hence, the mean values of both E and $E^{(i)}$ are 0.5, and so, $\mu^{(i)} = 0$ for all i. (The value of $|\mu^{(i)}|$ calculated from the experimental data was less than 10^{-3}).

Table 5.1 shows an output of a series of tests. The following data are presented:

n, m, l, t are, respectively, the numbers of variables, clauses, clause literals, instances tested;

i is the number of surrounding;

$\mu^{(i)}$ is the mean value of approximation error $\varepsilon^{(i)}$ (expression 3.27) over the t tests;

μ_m is the mean value of the ratio $m^{(i)} / m$, where $m^{(i)}$ is the number of clauses in $L^{(i)}$;

μ_n is the mean value of the ratio $n^{(i)} / n$, where $n^{(i)}$ is the number of variables in $L^{(i)}$;

μ_τ is the mean value of the ratio $\tau^{(i)} / \tau$, where $\tau^{(i)}, \tau$ denote the run time of computing $E^{(i)}$ and E , respectively;

$\sigma^{(i)}, \sigma_\tau$ are standard deviations of $\varepsilon^{(i)}$ and of the ratio $\tau^{(i)} / \tau$, respectively.

A salient feature of all the tests is that computing evidence in the first surrounding is most significantly faster than that in the full system, such that $\tau^{(1)}$ is up to 10^6 times less than τ (see Tables 5.10, 5.11). However, as expected, $\tau^{(i)}$ grows very fast with i becoming comparable to τ already about third surrounding. So, most beneficial would be cases in which already the first surrounding provides a credible approximation.

Tables 5.2, 5.3 show values of $\sigma^{(1)}$ for various system parameters. Typically, $\sigma^{(1)}$ decreases with growing n and l , but increases with m . This behavior of σ , characteristic of all surroundings, agrees with the intuition that growth of n leads to a decrease of the number of variables shared by a pair of clauses, and growth of l is accompanied by an increase of the length β of the middle "separating" section of clauses; both these factors diminish the influence of the system residue on the evidence of L . On the other hand, growth of m increases the number of clauses in the residue, increasing its contribution to the evidence.

Tables 5.4-5.9 show values of a *threshold displacement* (cf. expression (4.4))

$$\Delta^{(1)}(q) = \frac{\sigma^{(1)}}{\sqrt{1-q}} \tag{5.3}$$

such that if $E^{(1)} \geq 0.5 + \Delta^{(1)}(q)$ or $E^{(1)} \leq 0.5 - \Delta^{(1)}(q)$, then the first surrounding suggests the same belief as the entire system with a probability of q at least.

For most instances tested in the experiments a threshold displacement of 0.1 (i.e., $E^{(1)} \geq 0.6$ or $E^{(1)} \leq 0.4$) already yields a high credibility of the approximation by the first surrounding. Even a rather strict requirement of $q = 0.99$ is met by most instances except systems consisting of many short clauses ($l = 3$).

Time complexity of the first surrounding approximation is characterized by Tables 5.10, 5.11 displaying the average ratio of $\tau^{(1)} / \tau$. This ratio decreases with growing n and m , but increases with growing l , such that only in small sets of long clauses the ratio exceeds 0.01 ($\tau^{(1)} / \tau = 0.06$ for $n = 20$, $m = 20$, $l = 7$). But, fortunately, these are just the systems with the shortest time of computing evidence, so, $\tau^{(1)}$ is sufficiently short in these cases as well. As τ ranged in the experiments between 12 msec. and 11 hours, $\tau^{(1)}$ did not exceed 40 msec. To compare experiments that ran on different computers, the amount of computing (in the absence of I/O operations) was also measured by the number of memory accesses. Tables 5.12, 5.13 present average values of this measure for sample instances (K, M, B stand for 10^3, 10^6, 10^9, respectively).

6. Conclusion.

The following idea motivated this study. The complexity of logical inference in knowledge systems grows with the system size determined, in particular, by the number of its clauses. This growth is so rapid, that reasoning in sizeable systems becomes intractable. However, if this process could be performed reliably in a relatively small subset of a given system, then the range of efficient reasoning could be extended significantly.

Conclusions suggested by a proper subset of a system may, of course, differ from that derived from the entire system. Nevertheless, it appears possible that if this subset is chosen properly, then even being a small one, it may approximate the right conclusion closely enough.

Given a set S of clauses, and a literal L the truth value of which is to be estimated, we consider a series of growing subsets of S, called *surroundings* of L, such the first surrounding consists of all clauses of S involving L or $\neg L$ (Definition 3.1).

As we wish to produce a credible belief in L based on a proper subset of S rather than on the entire system, we are concerned with a possible error caused by such an approximation. For a sample propositional system with a rather simple structure (Section 3), it is shown that if an instance of the system is approximated by an i-th surrounding of L, then the error in the value of evidence of L is a decreasing function of i.

Given a query Q, if a system S is approximated by its proper subset $\boldsymbol{t}$, then an important question determining the credibility of this approximation is whether both $\boldsymbol{t}$ and S suggest the same belief in Q. If this lucky event has a probability of q at least, we say that $\boldsymbol{t}$ is q-credible. To gather statistics regarding reasoning by surroundings, a series of experiments has been conducted with randomly generated propositional systems (Section 5).

Although only ground atomic queries have been considered in this paper so far, the notion of surrounding can be extended to any query expressed by a First Order formula F by defining the i-th surrounding of F as a union of the i-th surroundings of all literals occurring in F.

This study, even as limited as it is, suggests two conclusions:

(1) Computing evidence in the first surrounding of a given query is significantly faster than in the entire knowledge system;

(2) In many cases the first surrounding provides sufficient bases for a highly credible beliefs.

So, given a query Q, it is worthwhile to begin with a fast computing of an evidence of Q provided by its first surrounding. If the evidence is decisive enough to fit the required credibility, then a proper belief in Q can already be produced. Otherwise, the first surrounding can be extended (gradually) by including clauses of the second surrounding, and the evidence of Q computed again. A relatively short run time of this kind of computation justifies several trials. If the second surrounding does not yield a credible evidence, then an attempt of reasoning in the entire system is justified.

A further research of reasoning with credible subsets should be extended to more classes of knowledge systems described in languages more expressive than the propositional one used in this work.

Acknowledgment.

I am grateful to the anonymous referee for most helpful and, I may say so, benevolent comments. This research has been partly supported by the Israel Science Foundation administrated by the Israel Academy of Sciences and Humanities.

References.

Aczel, J. and Forte, B. (1986) Generalized Entropies and the Maximum Entropy Principle. In *Maximum Entropy and Bayesian Methods in Applied Statistics.* Ed. by J. H. Justice (Cambridge University Press) 95-100.

Barak A., Guday S., and Wheeler R. (1993) *The MOSIX Distributed Operating System, Load Balancing for UNIX.* Lecture Notes in Computer Science, v. 672, Springer-Verlag, ISBN 0-387-56663-5.

Cook, S. (1971) The complexity of theorem-proving procedures. *Proc. 3rd ACM Symp. on Theory of Computing,* 151-158.

Davis, M., and Putnam, H. (1960) A computing procedure for quantification theory. *J. ACM,* v. 7, no. 3, 201-215.

Dempster, A. P. (1967) Upper and lower probabilities induced by a multivariate mapping. *Annals of Mathematical Statistics,* v. 38, 325-339.

Dempster, A. P. (1968) A generalization of Bayesian inference. *Journal of the Royal Statistical Society,* v. 30, 205-247.

Gallo, G., and Urbani, G. (1989) Algorithms for testing the satisfiability of propositional formulae. *J. Logic Programming,* v. 7, 45-61.

Genesereth, M. R., and Nilsson, N. J. (1987) *Logical Foundations of Artificial Intelligence.* Morgan Kaufmann.

Goldberg, A. (1979) Average case complexity of the satisfiability problem. *Proc. 4th Workshop on Automated Deduction,* Austin, TX, 1-6.

Goldberg, A., Purdom, P., and Brown, C. (1982) Average time analyses of simplified Davis-Putnam procedures. *Information Processing Letters,* v. 15, 72-75.

Goldszmidt, M., Morris, P., and Pearl, J. (1990) A maximum entropy approach to non-monotonic reasoning. *Proc. 8th Natl. Conf. on Artificial Intelligence, AAAI'90,* (Boston) 646-652.

Hartley, R. V. (1928) Transmission of information. *Bell Systems Technical Journal,* v.7, 535-563.

Heckerman, D. (1986) Probabilistic interpretation of MYCIN's certainty factors. In *Uncertainty in Artificial Intelligence.* Ed. by L. N. Kanal and J. F. Lemmer (North Holland) 167-196.

Iwama, K. (1989) CNF satisfiability test by counting and polynomial average time. *SIAM. J. Comput.,* v. 18, 385-391.

Jaynes, E. T. (1957) Information Theory and Statistical Mechanics. *Phys. Rev.,* v. 106, 620-630.

Jaynes, E. T. (1979) Where do we stand on maximum entropy? In *The Maximum Entropy Formalism.* Ed. by R. D. Levine and M. Tribus (MIT Press) 15-118.

Jaynes, E. T. (1985) Where do we go from here? In *Maximum-Entropy and Bayesian Methods in Inverse Problems.* Ed. by C. R. Smith and W. T. Grandy (Reidel) 21-58.

Justice, J. H., Ed. (1986) *Maximum Entropy and Bayesian Methods in Applied Statistics* (Cambridge University Press).

Lee, N. S., Grize, Y. L., and Dehnad, K. (1987) Quantitative models for reasoning under uncertainty in Knowledge-Based Expert Systems. *International Journal of Intelligent Systems,* v. 2, 15-38.

Levine, R. D. and M. Tribus, M., Eds. (1979) *The Maximum Entropy Formalism* (MIT Press).

Linial, N., and Nisan, N. (1990) Approximate inclusion-exclusion. *Combinatorica,* v.10, no. 4, 349-365.

Lozinskii, E. (1992) Computing propositional models. *Information Processing Letters,* v. 41, 327-332.

Lozinskii, E. (1993) A simple test improves checking satisfiability. *Journal of Logic Programming,* v. 15, 99-111.

Lozinskii, E. (1994) Information and evidence in logic systems. *Journal of Experimental and Theoretical Artificial Intelligence,* v. 6, 163-193.

Lozinskii, E. (1995) Is there an alternative to parsimonious semantics? *Journal of Experimental and Theoretical Artificial Intelligence,* v. 7, 361-378.

Luby, M., and Velickovic, B. (1991) On deterministic approximation of DNF. *Proceedings of STOC'91,* New Orleans, LA, 430-438.

McCarthy, J. (1980) Circumscription - a form of non-monotonic reasoning. *Artificial Intelligence* **13,** 27-39.

Nilsson, N. J. (1986) Probabilistic logic. *Artificial Intelligence* **28,** 71-87.

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems* (Morgan Kaufmann).

Pearl, J. (1989) Probabilistic semantics for nonmonotonic reasoning: a survey. *Proc. First Intl. Conf. on Principles of Knowledge Representation and Reasoning* (Toronto) 699-710.

Purdom, P., and Brown, C. (1985) The pure literal rule and polynomial average time. *SIAM J. Comput.,* v. 14, no. 4, 943-953.

Reiter, R. (1980) A logic for default reasoning. *Artificial Intelligence* **13,** 81-132.

Shafer, G. (1976) *A Mathematical Theory of Evidence* (Princeton University Press).

Shannon, C. E. (1948) A mathematical theory of communication. *Bell Systems Technical Journal,* v. 27, 379-423, 623-656.

Shortliffe, E. H. (1976) *Computer-Based Medical Consultation: MYCIN* (Elsevier).

Skyrms, B. (1985) Maximum entropy inference as a special case of conditionalisation. *Synthese,* v. 63, 55-74.

Smith, C. R. and G. J. Erickson, Eds. (1983) *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems* (Reidel).

Smith, C. R. and W. T. Grandy, Jr. (1985) *Maximum-Entropy and Bayesian Methods in Inverse Problems* (Reidel).

Tribus, M. (1961) Information Theory as the basis for Thermostatics and Thermodynamics. *Journal of Applied Mechanics,* v. 28, 1-8.

Valiant, L. (1979) The complexity of computing the permanent. *Theoretical Computer Science* **8,** 189-201.

Wiener, N. (1948) *Cybernetics, or Control and Communication in the Animal and the Machine* (Wiley, New York).

Williams, P. M. (1980) Bayesian conditionalisation and the principle of minimum information. *British Journal for the Philosophy of Science,* v. 31, 131-144.

Table 5.1. A summary of a series of tests ($n = 40$, $m = 20$, $l = 3$, $t = 100$).

i	$\mu^{(i)}$	μ_m	μ_n	μ_τ	$\sigma^{(i)}$	σ_τ
1	-.0008	.08	.10	.0001	.0224	.0001
2	-.0004	.31	.30	.0054	.0067	.0134
3	.0000	.65	.55	.1359	.0009	.2403
4	.0000	.86	.69	.4549	.0001	.4193
5	.0000	.93	.78	.5849	.0000	.3324
6	.0000	.95	.86	.6256	.0000	.2940
7	.0000	.97	.92	.7803	.0000	.3094
8	.0000	.98	.95	.8673	.0000	.2574

Table 5.2. Values of $\sigma^{(1)}$, for $l = 5$.

m / n	20	40	60	80	100
20	0.0063	0.0130	0.0206	0.0265	0.0350
30	0.0043	0.0094	0.0150	0.0188	0.0240
40	0.0030	0.0069	0.0089	0.0126	0.0165
100	0.0015	0.0036	0.0052	0.0079	*

* Number of measurements was not enough for a reliable statistics.

Table 5.3. Values of $\sigma^{(1)}$, for $n = 40$.

m / l	20	40	60	80	100
3	0.0461	0.0958	0.1256	0.1620	*
5	0.0030	0.0069	0.0089	0.0126	0.0165
7	0.0004	0.0008	0.0013	0.0020	0.0026

Table 5.4. Threshold displacement $\Delta^{(1)}(q = 0.85)$, for $l = 5$.

n \ m	20	40	60	80	100
20	0.0163	0.0336	0.0532	0.0684	0.0904
30	0.0111	0.0243	0.0387	0.0485	0.0620
40	0.0077	0.0178	0.0230	0.0325	0.0426
100	0.0039	0.0093	0.0134	0.0204	*

Table 5.5. Threshold displacement $\Delta^{(1)}(q = 0.85)$, for $n = 40$.

l \ m	20	40	60	80	100
3	0.1190	0.2474	0.3243	0.4183	*
5	0.0077	0.0178	0.0230	0.0325	0.0426
7	0.0010	0.0021	0.0034	0.0052	0.0067

Table 5.6. Threshold displacement $\Delta^{(1)}(q = 0.95)$, for $l = 5$.

n \ m	20	40	60	80	100
20	0.0282	0.0581	0.0921	0.1185	0.1565
30	0.0192	0.0420	0.0671	0.0841	0.1073
40	0.0134	0.0309	0.0398	0.0563	0.0738
100	0.0067	0.0161	0.0233	0.0353	*

Table 5.7. Threshold displacement $\Delta^{(1)}(q = 0.95)$, for $n = 40$.

l \ m	20	40	60	80	100
3	0.2062	0.4284	0.5000	0.5000	*
5	0.0134	0.0309	0.0398	0.0563	0.0738
7	0.0018	0.0036	0.0058	0.0089	0.0116

Table 5.8. Threshold displacement $\Delta^{(1)}(q = 0.99)$, for $l = 5$.

m / n	20	40	60	80	100
20	0.0630	0.1300	0.2060	0.2650	0.3500
30	0.0430	0.0940	0.1500	0.1880	0.2400
40	0.0300	0.0690	0.0890	0.1260	0.1650
100	0.0150	0.0360	0.0520	0.0790	*

Table 5.9. Threshold displacement $\Delta^{(1)}(q = 0.99)$, for $n = 40$.

m / l	20	40	60	80	100
3	0.4610	0.5000	0.5000	0.5000	*
5	0.0300	0.0690	0.0890	0.1260	0.1650
7	0.0040	0.0080	0.0130	0.0200	0.0260

Table 5.10. Ratio of $\tau^{(1)}/\tau$ in %, for $l = 5$.

m / n	20	40	60	80	100
20	1.0327	0.2432	0.0683	0.0239	0.0104
30	0.2804	0.0208	0.0029	0.0007	0.0001
40	0.1350	0.0041	0.0002	0.0001	0.0000
100	0.0127	0.0001	0.0000	0.0000	*

Table 5.11. Ratio of $\tau^{(1)}/\tau$ in %, for $n = 40$.

m / l	20	40	60	80	100
3	0.0471	0.0013	0.0001	0.0000	*
5	0.1350	0.0041	0.0002	0.0001	0.0000
7	0.4916	0.0768	0.0187	0.0074	0.0031

Table 5.12. Number of memory accesses, for $l = 5$.

m / n	20	40	60	80	100
20	2.66K	0.08M	1.01M	9.18M	62.04M
30	5.41K	0.44M	8.92M	0.11B	1.05B
40	8.64K	1.36M	38.03M	0.61B	7.27B
100	46.19K	23.12M	1.59B	45.24B	*

Table 5.13. Number of memory accesses, for $n = 40$.

m / l	20	40	60	80	100
3	28.52K	8.21M	0.98B	47.34B	*
5	8.64K	1.36M	38.03M	0.61B	7.27B
7	2.67K	0.08M	0.82M	4.75M	23.52M