A Logic to Reason about Likelihood*

Joseph Y. Halpern
IBM Almaden Research Center, San Jose, CA 95120-6099, U.S.A.

Michael O. Rabin
Department of Mathematics, Hebrew University, Jerusalem, Israel, and Aiken Computation Laboratory, Harvard University, Cambridge, MA 02138, U.S.A.

Recommended by Nils Nilsson and N.S. Sridharan

ABSTRACT

We present a logic LL which uses a modal operator L to help capture the notion of being likely. Despite the fact that likelihood is not assigned quantitative values through probabilities, LL captures many of the properties of likelihood in an intuitively appealing way. We give a possible-worlds style semantics to LL, and, using standard techniques of modal logic, we give a complete axiomatization for LL and show that satisfiability of LL formulas can be decided in exponential time. We discuss how the logic might be used in areas such as medical diagnosis, where decision making in the face of uncertainties is crucial. We conclude by using LL to give a formal proof of correctness of some aspects of a protocol for exchanging secrets.

1. Introduction

Reasoning about likelihood is an important component of decision making in many human endeavours. One way of formalizing and carrying out such reasoning is by means of probability theory. However, there are several problems with this use of probability theory.

One obvious problem is that there are many situations where we want to reason about likelihood when it is not clear that probability theory is even applicable. For example, one might want to assert "It is not likely that a nuclear war will begin tomorrow," but it seems difficult to find an appropriate sample space in which to give this statement probabilistic meaning.

* A preliminary version of this paper appeared in *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, Boston, MA, 1983.

Artificial Intelligence **32** (1987) 379–405
0004-3702/87/$3.50

In other situations, it may be possible in principle to attach probabilities to events, but it may not be reasonable in practice. A typical opinion of researchers studying disease associations and treatment effects is that "... probabilities are virtually useless in medical applications, because the conclusions that one can draw from such probability values almost never justify the expense and inconvenience to the patient necessary to obtain them" [13].

Finally, even in situations where a probabilistic analysis could be carried out, decision makers seem to be quite uncomfortable with this approach. Indeed, there is convincing evidence to suggest that people are very poor at probabilistic reasoning [22]. Moreover, although people seem quite prepared to say that one outcome is more likely than another, they are unwilling to give precise numerical probabilities to outcomes. As Szolovits and Pauker point out [21]:

> They [doctors] are often extremely reluctant to engage in any numerical computation involving the likelihood of a diagnosis or the prognosis for a treatment. Even when official blessing is bestowed upon Bayesian techniques, we have seen both experienced and novice physicians acknowledge and then ignore them.

(Similar remarks apply to other schemes, such as *belief functions* [20] or *fuzzy set theory* [23] which attempt to attach numbers to the likelihood of various outcomes, although we will not consider them here.)

Expert systems designers have also expressed discomfort with the use of certainty factors and numerical probabilities in expert systems. Interestingly, it has been observed that "the performance of most systems remains constant under all sorts of small (<30%) perturbations in the precise values used" [1]. This suggests that a qualitative, nonnumerical notion of likelihood may be useful.

In this paper we introduce a modal logic which we call LL, the logic of likelihood, where we use a modal operator L to help us capture the notion of being likely. Thus there will be formulas in the language of the form Lp, which roughly translates to "p is reasonably likely to be a consistent hypothesis." We should stress here that the degree of confidence indicated by the symbol L depends on the user. In addition, it is quite possible that Lp and $L\neg p$ hold simultaneously.

Although likelihood is not assigned quantitative values through probabilities, we can capture many properties of likelihood in an intuitively appealing way. For example, if p_2 is reasonably likely given p_1, and p_3 is reasonably likely given p_2, then we can deduce that p_3 is somewhat likely given p_1. The longer a chain of reasoning, the less confidence we tend to have in the conclusion. In LL, we can show that from the statements $p_1 \Rightarrow Lp_2$ and $p_2 \Rightarrow Lp_3$, we can infer $p_1 \Rightarrow LLp_3$, which we abbreviate as $p_1 \Rightarrow L^2p_3$. As we shall see, "powers" of L denote a dilution of likelihood.

In order to increase the expressive power of our language, we have added

two additional modal operators, L^* and G. L^*p is used to denote the limit of the chain Lp, L^2p, $L^3p, \ldots$, while G is used to denote necessity. Gp intuitively says that p is necessarily the case.

It is hoped that LL will be useful in actual decision making, for example in management or in medical diagnosis. Our semantics for LL is intended to serve beyond providing a setting for a completeness proof in the usual style. The user is expected to operate with this semantic model in developing his reasoning to uncover the "real" state of affairs in the situation he is studying.

The basic semantic notion of a *state* s represents a complete and consistent set of "working hypotheses" concerning the situation under consideration, adopted by the decision maker at some point during his deliberations. A state s may have several successor states $s', s'', \ldots$, of which some are *likely* successors and others are *conceivable* successors. The logic LL is *not* temporal and a successor s' of s does not represent a likely or conceivable future development from the present state s. Rather, s' is one of the complete sets of hypotheses that the decision maker may move to, perhaps as a result of getting more information about the true state of affairs, if he adopts s as his current view of the state of the world. We remark that in this way we can view the models of LL as allowing belief revision. Thus we can capture some of the reasoning that goes on in nonmonotonic logics (cf. [11, 12]) in a perfectly monotonic framework.

Of course, our work is far from being the first attempt to capture reasoning about likelihood and qualitative reasoning about probability in a modal setting. An early paper by Rescher (cf. [17, Ch. IV]) gives axioms for a logic with a modal operator which is essentially like our L, but does not provide a semantics. Moore and Hutchins [14] consider a possible-worlds model for reasoning about certainty levels (which essentially correspond to our "power of L") for medical decision making and suggest the use of a modal logic for medical reasoning, but do not provide such a logic. More recently, Segerberg [19] and Gärdenfors [4] have described modal logics which allow formulas of the form $p \gtrsim q$, which can be interpreted as "p is more likely than q." However, there is no direct way of saying "p is likely" in their logics. The semantics required to capture such a binary comparison operator is, perhaps not surprisingly, much more complicated than that of LL, as is the axiomatization; we suspect the decision procedure will be more complex as well. On the other hand, the logics of Segerberg and Gärdenfors, by their design, do capture probability theory more directly than LL. We do not view this necessarily as a shortcoming of LL. In fact, one of the main theses of the present work is that probability theory is not the only way of reasoning about likelihood!

An important area in which the need for reasoning about likelihood arises and where, in general, there is no sound probabilistic basis for doing so, is in analyzing the credibility actions and threats. A head of state declares: "If you

cross this line we shall go to war." Or a mother leaves a note to her (somewhat undependable) child saying "Turn off the oven at noon." How do we express statements about the possible occurrence of these actions? Being able to do so is obviously important when arguing about the behavior of states, organizations, or individuals. The logic LL does not tell us how to assign likelihood to such actions, but does give us a formalism for expressing our beliefs and experiences on the subject, and a semantics for interpreting the formal statements that we make.

Nevertheless, a reasonable question to ask is to what extent LL can capture probability theory. One way of interpreting "reasonably likely" is as meaning "with probability greater than α" (for some user-defined α). It turns out, however, if we simply translate "P holds with probability greater than α" by LP, we quickly run into inconsistencies. We make some remarks on a solution to this problem in Sections 3 and 4. This issue is studied in more detail in a companion paper [5], where it is shown that there is a way of translating probability statements into LL in such a way that inferences made in LL are *sound* with respect to this interpretation of likelihood.

The rest of this paper is organized as follows. In the next section, we give the syntax and semantics of LL. In Section 3, we apply standard techniques of modal logic to show that validity of LL formulas is decidable in exponential time and give a complete axiomatization for LL. In Section 4, we discuss how to translate English sentences about likelihood into LL, taking medical decision making as our example.

This work was motivated by the need to find a logic in which to prove that certain security protocols and cryptographic protocols are correct. In Section 5, we sketch a protocol (due to the second author) for exchanging secrets, and use LL to give a formal proof of correctness of some aspects of the protocol. The analysis involves a careful formalization of the credibility of certain actions and threats. We exclude in Section 6 with a discussion of the relationship of LL to nonmonotonic logic and some directions for further research.

2. Syntax and Semantics

2.1. Syntax

Starting with a set $\Phi_0 = \{P, Q, R, \ldots\}$ of *primitive propositions*, we build more complicated LL formulas using the propositional connectives $\neg$ and $\wedge$ and the modal operators G, L, and L^*. Thus, if p and q are formulas, then so are $\neg p$, $(p \wedge q)$, Gp ("necessarily p"), Lp, and L^*p. We omit parentheses if they are clear from context. We also use the abbreviations $p \vee q$ for $\neg(\neg p \wedge \neg q)$, $p \Rightarrow q$ for $\neg p \vee q$, $p \equiv q$ for $(p \Rightarrow q) \wedge (q \Rightarrow p)$, Fp ("possibly p") for $\neg G \neg p$, and $L^i p$ for $L \cdots Lp$ (i Ls). (Our G and F correspond to the $\Box$ and $\Diamond$ often used in the temporal logic literature [10].) We will occasionally restrict attention to LL^-, the sublanguage of LL which consists of formulas with no mention of L^*.

The *size* of a formula p, written $|p|$, is its length as a string over the alphabet $\Phi_0 \cup \{\neg, \wedge, G, L, L^*,), (\}$.

2.2. Semantics

We give semantics to LL formulas by means of Kripke structures. An *LL model* is a quadruple $M = (S, \mathscr{L}, \mathscr{C}, \pi)$, where S is a set of *states*, $\mathscr{L}$ and $\mathscr{C}$ are binary relations on S with $\mathscr{L}$ reflexive (i.e., for all $s \in S$, we have $(s, s) \in \mathscr{L}$) and $\pi: \Phi_0 \to 2^S$. Intuitively, π associates with each primitive proposition the set of states of which it is true.

The *size* of a model $M = (S, \mathscr{L}, \mathscr{C}, \pi)$ is $|S|$. We can think of $(S, \mathscr{L}, \mathscr{C})$ as a graph with vertices S and edges $\mathscr{L} \cup \mathscr{C}$. If $(s, t) \in \mathscr{L}$ (respectively $\mathscr{C}$), then we say that t is an *$\mathscr{L}$-successor* (respectively *$\mathscr{C}$-successor*) of s. Informally, a state s consists of a set of hypotheses that we take to be "true for now." An *$\mathscr{L}$-successor* of s describes a set of hypotheses that is reasonably likely given our current hypotheses, while a *$\mathscr{C}$-successor* describes a set of hypotheses which is conceivable, but not necessarily reasonably likely. (If I buy a ticket to the Irish sweepstakes, it is conceivable, although not reasonably likely, that I will win.)[1] We will say t is *reachable* (resp. *$\mathscr{L}$-reachable*) from s if, for some finite sequence $s_0, \ldots, s_k$, we have $s_0 = s$, $s_k = t$, and $(s_i, s_{i+1}) \in \mathscr{L} \cup \mathscr{C}$ (resp. $(s_i, s_{i+1}) \in \mathscr{L}$) for $i < k$.

We extend π to a mapping $\pi: \{\text{LL formulas}\} \to 2^S$ as follows:

$$\pi(\neg p) = S - \pi(p),$$

$$\pi(p \wedge q) = \pi(p) \cap \pi(q),$$

$$\pi(Gp) = \{s \mid \text{for all } t \text{ reachable from } s, t \in \pi(p)\},$$

$$\pi(Lp) = \{s \mid \text{for some } t \text{ with } (s, t) \in \mathscr{L}, t \in \pi(p)\},$$

$$\pi(L^*p) = \{s \mid \text{for some } t \text{ which is } \mathscr{L}\text{-reachable from } s, t \in \pi(p)\}.$$

As usual we write $M, s \models p$ instead of $s \in \pi(p)$.

Note that $M, s \models Fp$ iff $M, t \models p$ for some state t reachable from s, while $M, s \models L^*p$ iff $M, t \models p$ for some state t $\mathscr{L}$-reachable from s. Thus, in strictly decreasing order of strength we have

$$Gp, \quad \neg L^*\neg p, \quad \ldots, \quad \neg L^N\neg p, \quad \ldots, \quad \neg L\neg p, \quad p,$$
$$Lp, \quad \ldots, \quad L^Np, \quad \ldots, \quad L^*p, \quad Fp.$$

[1] For technical reasons, we have decided to take the $\mathscr{C}$ relation to be "conceivable but not necessarily reasonably likely," rather than just "conceivable." Thus we have not postulated any semantic relationship between $\mathscr{C}$ and $\mathscr{L}$. The real conceivability relation can thus be viewed as the reflexive transition closure of $\mathscr{L} \cup \mathscr{C}$; this view of the conceivability relation is enforced by the semantics of the modal operator G.

We remark that we could have omitted the $\mathscr{C}$ relation from our semantics and essentially taken L^*p to be equivalent to Fp (this in fact is done in [5]). The resulting development would not have differed greatly from that we give here. In fact, none of our examples make use of the L^* operator. We include the L^* operator and a conceivable successor here because we feel that it might be helpful in modelling the way people actually seem to discuss likelihood.

Definition 2.1. A formula p is *satisfiable* iff for some model $M = (S, \mathscr{L}, \mathscr{C}, \pi)$ and some $s \in S$, we have $M,s \models p$; p is *valid* iff for all models $M = (S, \mathscr{L}, \mathscr{C}, \pi)$ and all $s \in S$, we have $M,s \models p$. It is easy to check that p is valid iff $\neg p$ is not satisfiable.

3. Finite Models, Decision Procedures, and Axiomatization

LL can be viewed essentially as a propositional dynamic logic (PDL) (cf. [3]) with two primitive programs L and C. The formula Lp in LL corresponds to $\langle L \rangle p$, L^*p corresponds to $\langle L^* \rangle p$, and Gp corresponds to $[(L \cup C)^*]p$. The only thing that prevents us from obtaining a decision procedure for LL by immediate translation into PDL in this way is our requirement that L be reflexive. Nevertheless, it is straightforward to adapt the standard techniques used to obtain the finite model property, a complete axiomatization, and an exponential time decision procedure for PDL and other modal logics (cf. [2, 3, 8, 15]) in order to show that the same results hold for LL. In particular we have the following theorems:

Theorem 3.1. *An LL formula p is satisfiable iff it is satisfiable in a model of size $\leq 2^{|p|}$.*

Theorem 3.2. *For some $c > 0$, there is a procedure for deciding if a formula p is satisfiable (respectively valid) which runs in deterministic time $O(2^{c|p|})$.*

Theorem 3.3. *The problem of deciding satisfiability (validity) of LL formulas is complete for deterministic exponential time.*

Theorem 3.4. *The following axiom system is sound and complete for LL.*
Axiom schemes:

All (substitution instances of) tautologies of propositional logic. (AX1)

$Gp \Rightarrow p$. (AX2)

$Gp \Rightarrow GGp$. (AX3)

$$Gp \Rightarrow \neg L \neg p . \tag{AX4}$$

$$p \Rightarrow Lp . \tag{AX5}$$

$$L(p \vee q) \equiv (Lp \vee Lq) . \tag{AX6}$$

$$G(p \Rightarrow q) \Rightarrow (Gp \Rightarrow Gq) . \tag{AX7}$$

$$G(p \Rightarrow q) \Rightarrow (Lp \Rightarrow Lq) . \tag{AX8}$$

$$G(p \Rightarrow q) \Rightarrow (L^*p \Rightarrow L^*q) . \tag{AX9}$$

$$L^*p \equiv (p \vee LL^*p) . \tag{AX10}$$

Rules of inference:

$$\frac{p}{Gp} \quad (\textit{generalization}) . \tag{R1}$$

$$\frac{p,\ p \Rightarrow q}{q} \quad (\textit{modulus ponens}) . \tag{R2}$$

$$\frac{\neg p \Rightarrow \neg Lp}{\neg p \Rightarrow L^*p} . \tag{R3}$$

Not surprisingly, Theorems 3.1, 3.2, and 3.3 hold for LL^- as well. And if we omit (AX9), (AX10), and (R3) we obtain a complete axiomatization for LL^-. The proofs of Theorems 3.1–3.4 for LL are somewhat technical, using well-known ideas from corresponding proofs for dynamic logic and temporal logic. We refer the reader to [2, 3, 8, 15] for the details. The proofs for LL^- are somewhat easier, and have much the same flavor. To give the reader an idea of how these proofs proceed, we prove the analogues of Theorems 3.1, 3.2, and 3.4 for LL^- below.

However, before we do this, perhaps a few words of discussion regarding the axioms are in order. We do not view the axiom system as giving a technique for proving valid formulas. Optimal proof procedures using tableau methods can be obtained from the proof of Theorem 3.2 (cf. [2]). Rather, a sound and complete axiomatization gives us a complete characterization of the properties of a system. Once we have such a characterization we can check whether our proposed semantics really captures properties of likelihood in a reasonable way.

It is easy to check that all the axioms and inference rules are sound, and for the most part, they are in accord with our intuitions about likelihood. (AX5), for example, says that if p is true, then it is likely to be true, while (AX7) says if it is necessarily the case that p implies q, then if p is likely, then so is q. The one conspicuous exception to the intuitive plausibility of the axioms is (AX6). Indeed, if we think of L as meaning "with probability $\frac{1}{2}$," then (AX6) does not

hold. To see this, consider a situation where we toss a fair coin twice, and let P represent "the coin lands heads both times," while Q represents "the coin lands tails both times." It is easy to see that the event $(P \vee Q)$ holds with probability $\geq \frac{1}{2}$, although neither P or Q individually does. Thus, if we interpret L to mean "with probability $\geq \frac{1}{2}$," we have $L(P \vee Q)$, but not $LP \vee LQ$.

Clearly this example suggests that it is inappropriate to think of L as meaning "with probability $\geq \frac{1}{2}$." As we show in the next section, Lp is best thought of as saying "p is reasonably likely to be a consistent hypothesis," which is a much weaker statement than "p holds with probability $\geq \frac{1}{2}$." One way to capture the stronger statement is to use LGp instead of Lp. We motivate this translation in the next section; it is also discussed in much greater detail in [5]. Note that $LG(P \vee Q)$ is not equivalent to $LGP \vee LGQ$, so that with this translation the problem mentioned above does not arise.

We defer further discussion of this issue to the next section, and conclude this section with a sketch of the proofs of Theorems 3.1, 3.2, and 3.4 in the case of LL^-.

3.1. Proof sketch of Theorems 3.1, 3.2, and 3.4 for LL^-

We write $\vdash p$ if the LL^- formula p is provable from (AX1)–(AX8), (R1), and (R2). We say p is *consistent* if it is not the case that $\vdash \neg p$. A finite set Σ of formulas is consistent if the conjunction of the formulas in Σ is consistent.

Clearly, if p_0 is satisfiable, then p_0 is consistent. (This is just a reformulation of the fact that the axioms are *sound*.) We now show that if p_0 is consistent, then p_0 is satisfiable in a model of exponential size which can be constructed in exponential time. This will suffice to prove Theorems 3.1 and 3.2. Standard arguments show that this also suffices to prove Theorem 3.4. For suppose p_0 is valid but not provable. Then by definition, $\neg p_0$ is consistent, and hence satisfiable. But this contradicts the assumption that p_0 is valid.

Lemma 3.5. *Let Σ be a consistent finite set of formulas. Then*:

(a) *if $p \wedge q \in \Sigma$, then $\Sigma \cup \{p, q\}$ is consistent,*

(b) *if $\neg(p \wedge q) \in \Sigma$, then either $\Sigma \cup \{\neg p\}$ or $\Sigma \cup \{\neg q\}$ is consistent,*

(c) *if $\neg\neg p \in \Sigma$, then $\Sigma \cup \{p\}$ is consistent,*

(d) *if $Gp \in \Sigma$, then $\Sigma \cup \{p\}$ is consistent,*

(e) *if $\neg Lq \in \Sigma$, then $\Sigma \cup \{\neg q\}$ is consistent,*

(f) *if $\neg Gq \in \Sigma$, then $\Gamma = \{\neg q\} \cup \{Gp \mid Gp \in \Sigma\}$ is consistent,*

(g) *if $Lr \in \Sigma$, then $\Gamma = \{r\} \cup \{\neg q \mid \neg Lq \in \Sigma\} \cup \{Gp \mid Gp \in \Sigma\}$ is consistent.*

The proofs of parts (a)–(e) are straightforward. For example, to prove (e), let σ be the conjunction of the formulas of Σ, and let σ' be the conjunction of the formulas of $\Sigma \cup \{\neg q\}$. Using (AX5) and propositional reasoning, we can

see that $\vdash \neg Lq \Rightarrow \neg q$, so $\vdash \sigma \equiv \sigma'$. Thus the consistency of Σ implies the consistency of $\Sigma \cup \{\neg q\}$. We sketch a proof of parts (f) and (g) in Appendix A.

From Lemma 3.5, we immediately get:

Lemma 3.6. *If Σ is a consistent finite set, then there exists a consistent finite set of formulas Σ^c such that Σ^c is a minimal set containing Σ satisfying the following properties*:

(a) *if $p \wedge q \in \Sigma^c$, then $p, q \in \Sigma^c$,*
(b) *if $\neg(p \wedge q) \in \Sigma^c$, then $\neg p \in \Sigma^c$ or $\neg q \in \Sigma^c$,*
(c) *if $\neg\neg p \in \Sigma^c$, then $p \in \Sigma^c$,*
(d) *if $Gp \in \Sigma^c$, then $p \in \Sigma^c$,*
(e) *if $\neg Lq \in \Sigma^c$, then $\neg q \in \Sigma^c$.*

We call Σ^c a "completion" of Σ.

Returning now to the proof of the theorem, suppose p_0 is consistent. We will construct a model whose graph looks like a tree such that p_0 is satisfied at the root. We first construct a sequence of trees $T_0, T_1, T_2, \ldots$ such that $T_i \subseteq T_{i+1}$. Each node of T_i is labelled by a pair of consistent finite sets of formulas (Σ, Σ^c), where Σ^c is a completion of Σ as in Lemma 3.6. There will be two types of successor relations in T_i: conceivable successors and likely successors.

T_0 simply consists of one node labelled $(p_0, \{p_0\}^c)$. Suppose we have constructed $T_0, \ldots, T_i$. We construct T_{i+1} by adding conceivable and likely successors to the leaves of T_i as follows. Suppose a leaf s of T_i is labelled by (Σ, Σ^c). For each formula $Lr \in \Sigma^c$ we create a likely successor of s labelled (Γ_r, Γ_r^c), where

$$\Gamma_r = \{r\} \cup \{\neg q \mid \neg Lq \in \Sigma^c\} \cup \{Gp \mid Gp \in \Sigma^c\}$$

(cf. Lemma 3.5(g)) and Γ_r^c is a completion of Γ_r. Similarly, for each formula $\neg Gq \in \Sigma$, we create a conceivable successor labelled (Δ_q, Δ_q^c), where

$$\Delta_q = \{\neg q\} \cup \{Gp \mid Gp \in \Sigma^c\},$$

and Δ_q^c is any completion of Δ_q.

We obtain a model $M = (S, \mathscr{L}, \mathscr{C}, \pi)$ as follows:

$$S = \bigcup_i \{s \mid s \text{ is a node of } T_i\},$$

$$\mathscr{L} = \{(s, s) \mid s \in S\} \cup (\bigcup_i \{(s, t) \mid t \text{ is a likely successor of } s \text{ in } T_i\}),$$

$$\mathscr{C} = \bigcup_i \{(s, t) \mid t \text{ is a conceivable successor of } i \text{ in } T_i\},$$

$$\pi(P) = \{s \in S \mid P \in \Sigma^c, \text{ where } (\Sigma, \Sigma^c) \text{ is the label of } s\}$$

for each primitive proposition P.

It is now a straightforward matter to show (by induction on the structure of p) that $p \in \Sigma^c$ implies $M,s \models p$, and $\neg p \in \Sigma^c$ implies $M,s \models \neg p$, where (Σ, Σ^c) is the label of s. Thus we have constructed a model for p_0.

The model we have just constructed is in general infinite. We get a finite model by first observing that the label on each node of the graph consists essentially of subformulas of p_0. To make this precise, we define the *closure of* p_0, $\mathrm{Cl}(p_0)$, to be the least set H such that:

(a) $p \wedge q \in H \Rightarrow p, q \in H$,

(b) $\neg(p \wedge q) \in H \Rightarrow \neg p, \neg q \in H$,

(c) $\neg\neg p$ or Gp or $Lp \in H \Rightarrow p \in H$,

(d) $\neg Lp$ or $\neg Gp \in H \Rightarrow \neg p \in H$.

An easy proof by induction on $|p_0|$ shows:

Lemma 3.7. $|\mathrm{Cl}(p_0)| \leq |p_0|$.

It is easy to check that if a node is labelled by (Σ, Σ^c) in the construction above, then both Σ and Σ^c are subsets of $\mathrm{Cl}(p_0)$. We can now slightly modify the construction of the model above to get a model of size 2^n, where $n = |p_0|$, by identifying nodes with the same label. More precisely, given the sequence of trees $T_0, T_1, \ldots$, as constructed above, let $M' = (S', \mathscr{L}', \mathscr{C}', \pi')$ be defined as follows:

$$S' = \bigcup_i \{\Sigma^c \mid \text{there is a node labelled } (\Sigma, \Sigma^c) \text{ in } T_i\},$$

$$\mathscr{L}' = \{(\Sigma^c, \Sigma^c) \mid \Sigma^c \in S'\} \cup \Big(\bigcup_i \{(\Sigma^c, \Gamma^c) \mid \text{a node labelled } (\Gamma, \Gamma^c) \text{ is a likely successor of a node labelled } (\Sigma, \Sigma^c) \text{ in } T_i\}\Big),$$

$$\mathscr{C}' = \bigcup_i \{(\Sigma^c, \Gamma^c) \mid \text{a node labelled } (\Gamma, \Gamma^c) \text{ is a conceivable successor of a node labelled } (\Sigma, \Sigma^c) \text{ in } T_i\},$$

$$\pi'(P) = \{\Sigma^c \mid P \in \Sigma^c\} \quad \text{for each primitive proposition } P.$$

Again a straightforward argument by induction on the structure of p shows that

$$p \in \Sigma^c \text{ implies } M', \Sigma^c \models p, \qquad \neg p \in \Sigma^c \text{ implies } M', \Sigma^c \models \neg p.$$

This gives us the desired exponential size model of p_0. To see that we can construct such a model in deterministic exponential time, note that the only source of nondeterminism in the construction above comes from part (b) of Lemmas 3.5 and 3.6. We construct the model deterministically by adding both

of these "or" branches, pruning one later if we discover it leads to an inconsistency (a node labelled by both p and $\neg p$ for some formula p). We refer the reader to the tableau construction described in [2] for further details. □

Definition 3.8. A set of formulas Σ is *finitely satisfiable* if every finite subset of Σ is satisfiable. A logic is *compact* if every finitely satisfiable set of formulas is satisfiable.

LL is not compact. To see this, let Σ consist of the formulas L^*p, $\neg p$, $\neg Lp$, $\neg L^2p, \ldots$. Every finite subset of Σ is clearly satisfiable, yet Σ is not. However, the use of L^* in this counterexample is necessary, since we have:

Theorem 3.9. *LL^- is compact.*

Proof. The proof is very similar to the proof of Theorem 3.4 for LL^-. We simply observe that Lemmas 3.5 and 3.6 both hold if we replace "consistent" by "finitely satisfiable" throughout (where now Σ may be an infinite set). If Σ_0 is finitely satisfiable, we can construct a model for it exactly as we constructed a model for the consistent formula p_0 above. We leave details to the reader. □

4. Translating into LL

We now come to the delicate task of using LL to model real-life reasoning about situations involving incomplete and uncertain information. As usual with such modelling tasks, whereas the mathematical structure is uniquely defined, we are sometimes faced with several choices for translating an everyday concept into the mathematical framework. Where appropriate, we shall outline the various possible translations. The actual option to be selected by the user will depend on the application field and on the norms of assurance and safety that he is seeking. What follows is not an application of LL in the sense that we outline a complete system for expressing medical facts within LL and give algorithms for reaching decision concerning such facts, i.e. making a diagnosis. Rather, what we present is an exercise involving medical terminology, to illustrate how a user might translate medical facts and his working hypotheses concerning these facts into LL. The translation brings out the relationship between intuitive concepts such as likelihood, possibly being true, etc., and the corresponding modal operators in LL.

Consider a medical doctor faced with a patient exhibiting a number of symptoms, having a certain medical history, who may have certain diseases. Each basic relevant proposition concerning the situation is expressed by a propositional variable. Thus Y may represent the symptom that the patient's complexion is yellow, DR that he is a heavy drinker, HP that he has hepatitis, D that he will die, TM that he has a tumor, HR that he has heart trouble, etc.

Note that the uncertainty extends to symptoms as well as diseases. Thus the patient may have a high blood count, but it is possible that there is some unreliability in the test. With a new patient, the doctor may suspect a drinking problem, but not be sure about it even after a direct question. Thus the physician may choose to express his current working hypothesis concerning the patient's drinking as $L(\mathrm{DR})$, even before he has strong information supporting this hypothesis.

The logic LL is not temporal. The statement D (patient will die), is not construed as something that the doctor will follow over the coming weeks or months, but rather as a working hypothesis that he may or may not adopt now.

We view a state s as a consistent and complete set of hypotheses, the set of statements each of which is taken to be "true for now." Note that every formula, including complex ones such as $L(\mathrm{D} \vee G(\mathrm{TM}))$, gets a truth value at s. A state s represents the set of hypotheses that the doctor may adopt at a certain point during his diagnosis and decision making process. The assumption that s is complete is an idealization. In practice we imagine that s is a finite set consisting of only the formulas "relevant" to the discussion, and perhaps all their subformulas.

Besides s, there are many other complete consistent sets of hypotheses. Some of these the doctor may consider likely given his current estimate as to the true state of affairs. If s' is considered likely given that s describes the true state of affairs, then this is modelled in the graph structure by having $(s, s') \in \mathscr{L}$. Other states of affairs are *conceivable* (but not necessarily likely). If s'' is considered conceivable given s, then we have $(s, s'') \in \mathscr{C}$. Let s be a state which, among other things, contains the hypotheses Y, HP, and $\neg$TM. The state s may have an $\mathscr{L}$-successor s_1 containing $\neg$Y, $\neg$HP, and HR, and another $\mathscr{L}$-successor s_2 containing Y, DR, and HP. There may also be a $\mathscr{C}$-successor s_3 containing Y, $\neg$DR, $\neg$HP, TM, and D. By our semantic rules, it also follows that s contains $L(\mathrm{Y})$, $L(\neg\mathrm{Y})$, $L(\mathrm{HP})$, $L(\neg\mathrm{HP})$, etc.

As an idealization, we view the semantic model M used in a specific decision making situation as being given in advance. Experts are consulted about the possible successors to any possible state, and the resulting model M is then stored. In practice, the experts will enumerate various rules or extra-logical axioms concerning the situation, these rules embodying their knowledge and past experience. The idealized model M must satisfy these axioms. The decision maker will then use these rules when arguing about specific states that he adopts and the likely and conceivable successors of these states. Thus, for example, medical expertise may be incorporated into the extra-logical axiom $\mathrm{Y} \Rightarrow L(\mathrm{HP})$ (if the patient's complexion is yellow, then his having hepatitis is a likely working hypothesis).

We emphasize that if s contains the formula p (i.e. if $M,s \models p$), this does *not* mean that p is actually true at s, but rather that p is one of the hypotheses that

we are taking to be true at this state. Thus Lp means that p is likely to be a consistent working hypothesis; this falls short of saying that it is likely that p is actually true. However, if in *no* state that is reachable from s is it the case that $\neg p$ is a working hypothesis, then it must be the case that p is actually true at s. Thus, the statement "it is (actually, or necessarily) the case that p" is translated by Gp. Consequently, "it is likely to be the case that p" is translated by LGp. Similarly, "it is not likely to be the case that p" is translated by $\neg LGp$, while "it is likely to be the case that p does not hold" is translated by $LG\neg p$.

Note that the statement "p holds with probability $\geq \frac{1}{2}$" is better captured by "it is likely to be the case that p" rather than "p is likely to be a consistent hypothesis." Indeed, it is shown in [5] that statements of probability theory can be translated to LL in a *sound* manner, using LGp to capture "p holds with probability $\geq \alpha$." That is, if we translate statements of probability theory true of a given probability space appropriately into LL, then any deduction we can draw from the (translated) statements in LL will also be true of that probability space. Thus, in this precise sense, LL can capture notions of probability correctly.

But in some cases Gp might be too strong to capture the true strength of our belief in p. LL gives us a wide range of choice in expressing degree of likelihood. For example, $\neg L^*\neg p$ can be viewed as saying "it is completely unlikely that $\neg p$," since if s contains $\neg L^*\neg p$, then any state $\mathscr{L}$-reachable from s contains p. Similarly, for sufficiently large, user-specified N, $\neg L^N\neg p$ can be viewed as "for all practical purposes p holds." We remark that these translations can also be used to capture a probabilistic notion of likelihood.

Given the wide range of choices in expressing degrees of likelihood, a natural question to ask is at what point should one take action on the basis of likelihood. For example, consider a doctor trying to decide whether or not to operate on a tumor. If the tumor is malignant, then he should certainly operate to prevent the cancer from spreading. On the other hand, the operation is sufficiently dangerous that if the tumor is benign, then he should definitely not operate. He can perform some tests for malignancy, but these are expensive, have undesirable side effects, and are not completely reliable. Let MAL denote that the tumor is malignant, so $\neg$MAL denotes that it is benign, and let OP denote the decision to operate. Clearly MAL $\Rightarrow$ OP, while $\neg$MAL $\Rightarrow \neg$OP. A doctor might certainly feel that if he has reached a state where he has concluded L(MAL) (or perhaps LG(MAL) or $L\neg L^*\neg$(MAL), in the light of our discussion above), then he should operate. But what if he can only conclude L^5(MAL) or L^2(MAL)? This will very much depend on the doctor's philosophy as to what level of risk is acceptable given the potential consequences, and on exactly how he interprets L. It is up to the user of the system to decide which is appropriate here and to incorporate the correspond-

ing extra-logical axioms such as $L(\text{MAL}) \Rightarrow \text{OP}$ or $LG(\text{MAL}) \Rightarrow \text{OP}$, etc., which best express his philosophy concerning treatment. The logic LL provides him with tools for doing this.

5. Using LL to Prove the Correctness of a Protocol

We now show how to apply LL to proving the correctness of some aspects of a protocol to exchange secrets suggested by the second author [16]; other protocols with similar properties have also been developed recently (cf. [9]). It is striking that a relatively short, informal argument of correctness is transformed into a lengthy formal proof. Of course, this phenomenon is not atypical when dealing with formal systems. Actually an important benefit arises from the formalization process in our case. A number of hidden assumptions about credibility of actions and threats must be made explicit in the formal translation for the proof to go through. Thus we gain a better understanding of the situation that we study. Although AI is not in general concerned with analyzing cryptographic protocols, the reader will find that many of the issues that arise here, particularly issues of credibility of actions and threats, also arise in many situations that are of more immediate interest to AI.

The situation is the following. Suppose Alice and Bob have one-bit secrets which they would like to exchange (the multibit case proceeds along similar lines). For definiteness, we assume that the secrets are passwords to certain files. We want the protocol to be self-enforcing; that is, we would like it to work without the necessity of an appeal to a trusted third party to act as an intermediary or a judge to adjudicate disputes. Thus, we assume that the files are booby-trapped in such a way that if someone tries to enter either file with the wrong password, *both* files are destroyed. This assumption prevents either Alice or Bob from just guessing the password. Moreover, when combined with our assumption of *credibility*, which says that it is reasonably likely that Bob (respectively, Alice) will act on the information provided by Alice (respectively Bob) in the course of the protocol, it will prevent cheating. Finally, we assume that Alice can tell if Bob enters her file, and vice versa.

For the reader who feels that our assumptions about mutual destruction (of files) and credibility, and our requirements for the self-enforcing nature of the protocol are far-fetched, we note that similar situations are a serious part of real life. Mutual deterrence between nuclear powers is based on an assurance of mutual destruction if both sides start using their weapons, and on the credibility of the threat that if one side starts direct hostilities, then the other side will be able and willing to use its nuclear weapons. This is, of course, a situation where no trusted third party or sufficiently powerful judge is available. Thus, any arrangement between the powers must be self-enforcing to be valid. Note that in this context, discussion about the participants' possible actions cannot be phrased in probabilistic terms because there are no samples

based on past experience. On the other hand, the methodology and language of LL are suitable for arguing within and about this setup.

The protocol proceeds in three steps which are completely symmetric.

Step 1. (a) Alice sends a random bit R_A to Bob by *oblivious transfer*. (This is a protocol developed by the second author [16] which has the property that with probability $\frac{1}{2}$, Bob knows R_A, but Alice does not know whether or not Bob knows R_A. We could model this by assuming that there is a special channel between Alice and Bob which with probability $\frac{1}{2}$ will transmit R_A correctly and otherwise will transmit garbage. Bob will know whether or not he has gotten R_A, but Alice will not.)

(b) Bob sends a random bit R_B to Alice by oblivious transfer.

Step 2. (a) Alice computes μ_A, where $\mu_A = 1$ if the oblivious transfer succeeded and 0 otherwise. She sends Bob $S_A \oplus \mu_A$, where S_A is her secret and $\oplus$ represents addition mod 2. (This transmission, like all the rest in this short protocol, is sent over a regular transmission line which we assume is error-free.)

(b) Bob computes μ_B and sends Alice $S_B \oplus \mu_B$.

Step 3. (a) Alice sends Bob $S_A \oplus R_A$.

(b) Bob sends Alice $S_B \oplus R_B$.

Assuming there is no cheating, then if Bob received R_A at Step 1, then he will be able to compute S_A at Step 3, and thus enter Alice's file. But when Alice sees Bob entering her file, she will know that $\mu_B = 1$, and so will be able to compute S_B (since she was sent $S_B \oplus \mu_B$ at Step 2) and be able to enter Bob's file. Similar remarks hold with Bob and Alice's role interchanged. Thus, if either Bob receives R_A or Alice receives R_B at Step 1, an event which has probability $\frac{3}{4}$, the secrets will be exchanged.

The main difficulty in proving the correctness of the protocol lies in showing that there is no cheating, i.e. that both Alice and Bob, who have agreed to adopt this protocol, will actually send the prescribed bits. Before we prove that there is no cheating, let us illustrate the need for the credibility assumption by considering simplified situations. From our discussion, it will also be clear why such a proof cannot be conducted within classical logic and why being able to reason about likelihood is required. Assume that Bob's secret password $S_B = 0$. Can we prove that Bob will not tell Alice "my password is 1"? In order to do so, we need to make some assumptions regarding likelihood.

Denote by (B, T, 1) the hypothesis that Bob tells Alice that his password is 1, by (A, E, 1) that Alice enters Bob's file with password 1, and by DS that the files are destroyed. Bob knows that $G(\mathrm{A}, \mathrm{E}, 1) \Rightarrow G(\mathrm{DS})$. We can express the fact that he does not want the files destroyed by $G(\neg \mathrm{DS})$. We now introduce the extra-logical assumption: $G(\mathrm{B}, \mathrm{T}, 1) \Rightarrow L^k G(\mathrm{A}, \mathrm{E}, 1)$ for some k. If we assume that Bob actually tells Alice that his password is 1, then with some diluted likelihood we must assume that Alice actually enters his file with

password 1. This is exactly our assumption of credibility. In this setting, we can prove $G(\neg(B, T, 1))$: Bob will never tell Alice that the password is 1. (Note we have prefixed the primitive propositions by G, since we are concerned here with the situation where they *actually* occur, and not just with the situation where it is a working hypothesis that they occur. The conclusions we could derive by using the alternative translation without the G would be much too weak for our purposes here.)

As we mentioned in the previous section, a more reasonable translation for the assumption that Bob does not want his files destroyed might be $\neg L^* G(DS)$ or even $\neg L^N G(DS)$ for sufficiently large N greater than k—after all, people do not want to die, yet they use car travel, ignoring the conceivable risks involved. In this case, we can show respectively that $\neg L^* G(B, T, 1)$ and $\neg L^{N-k} G(B, T, 1)$. Although there is some degree of latitude here in the translation, the conclusions are quite robust.

By way of contrast, assume that Alice is chained and unable to reach her keyboard. Then an assumption such as $G(B, T, 1) \Rightarrow L^5 G(A, E, 1)$ is not called for. In this situation, it makes sense that we cannot *prove* that Bob will not lie about the password.

The above "proof," which is based on LL, is of course rather simple and not too different from any informal argument that we might give. In fact, why not take the conclusion $G(\neg(B, T, 1))$ as an axiom? The point is that the extra-logical axioms that we adopt are more basic: some facts about the nature of the system and how new facts may be learned, a fact about Bob's aims (at all costs to avoid destruction of the files), and a general statement about the connection between availability of information and action (the credibility assumption). Having adopted these given or "obvious" assumptions, we can proceed to formally derive the particular statement $G(\neg(B, T, 1))$. Note that it is precisely for expressing the credibility assumption that we need the likelihood operator L. It is not reasonable to assume that if Bob tells Alice that $S_B = 1$, then she will necessarily enter his file with the password 1. The logic LL enables us to express the appropriate assumptions concerning Alice's actions. The same remarks apply to the following treatment of the full protocol. As mentioned in the Introduction, the need to express credibility assumptions was one of our initial motivations for creating LL.

We now give an informal proof that there is no cheating, or, more accurately, that with reasonable likelihood there is no cheating, and then formalize it in LL.

There can be no cheating at Step 1 (this is simply a proven property of the oblivious transfer), and it is easy to check that even if there is cheating in Step 2, neither Alice nor Bob has enough information at the end of Step 2 to deduce S_B or S_A, respectively. Suppose Alice cheats at Step 3. Since with probability $\frac{1}{2}$ Bob knows R_A, he will then deduce an incorrect value for S_A. Since we assume that Bob is reasonably likely to act on information provided by Alice (the credibility assumption), he is reasonably likely to enter her file using the wrong

password and thus destroy both files. Therefore Alice does not cheat at Step 3. A similar argument shows that Bob does not cheat at Step 3.

Next we show that it is not to Alice's advantage to cheat at Step 2. For suppose that Alice cheats at Step 2 and deduces S_B at Step 3. She is prevented from using this value unless Bob has already entered her file. Otherwise, Bob will deduce that $\mu_A = 1$ and compute the wrong value of S_A. It is then likely that he will try to enter Alice's file (the credibility assumption again) and thus destroy both files. Therefore Alice does not cheat at Step 2, and by similar arguments, neither does Bob.

Finally, note that if Bob stops the protocol early and does not send Alice $S_B \oplus R_B$, then Alice can still deduce that $\mu_B = 1$ if Bob ever enters her file before she has entered his, and thus she will still be able to deduce S_B.

In order to formalize this reasoning in LL, we will require the following primitive propositions:

(a) four primitive propositions of the form (V, i), where $V \in \{S_A, \mu_A\}$, $i \in \{0, 1\}$ (which intuitively stand for "the value of V is i"),

(b) six primitive propositions of the form (A, S, V, i), where $V \in \{R_A, S_A \oplus \mu_A, S_A \oplus R_A\}$, $i \in \{0, 1\}$ (Alice *sends* $R_A = i$, $S_A \oplus \mu_A = i$, $S_A \oplus R_A = i$),

(c) six primitive propositions of the form (A, D, V, i), where $V \in \{R_B, S_B, \mu_B\}$ (Alice *deduces*—from information provided by Bob—that $R_B = i$, etc.),

(d) two primitive propositions of the form (A, E, i), $i \in \{0, 1\}$ (Alice *enters* Bob's file with password i),

(e) one primitive proposition DS (the files are destroyed).

We get eighteen more primitive propositions by interchanging the roles of A and B.

Besides the logical axioms of LL described in the previous section, we will have other extra-logical axioms which summarize some of the relationships between these primitive propositions. We present the axioms from Alice's point of view. We get another set of axioms by interchanging the roles of A and B.

The first axiom says that S_A and μ_A cannot have two values:

$$\neg(G(V, 0) \wedge G(V, 1)), \quad \text{where } V \text{ is } S_A \text{ or } \mu_A . \tag{1A}$$

The second axiom describes one property of the oblivious transfer: it is likely that Bob will be able to deduce the value of Alice's random bit after the oblivious transfer:

$$G(A, S, R_A, i) \Rightarrow LG(B, D, R_A, i), \tag{2A}$$

All other transmissions proceed over error-free lines, so Bob will deduce something exactly if Alice sends it:

$$(G(\mathrm{A}, \mathrm{S}, S_\mathrm{A} \oplus \mu_\mathrm{A}, i) \equiv G(\mathrm{B}, \mathrm{D}, S_\mathrm{A} \oplus \mu_\mathrm{A}, i))$$
$$\wedge (G(\mathrm{A}, \mathrm{S}, S_\mathrm{A} \oplus R_\mathrm{A}, i) \equiv G(\mathrm{B}, \mathrm{D}, S_\mathrm{A} \oplus R_\mathrm{A}, i)) . \tag{3A}$$

Now we describe how Bob deduces S_A from $S_\mathrm{A} \oplus \mu_\mathrm{A}$ and μ_A (respectively from $S_\mathrm{A} \oplus R_\mathrm{A}$ and R_A), by doing modular arithmetic:

$$[(G(\mathrm{B}, \mathrm{D}, S_\mathrm{A} \oplus R_\mathrm{A}, i) \wedge G(\mathrm{B}, \mathrm{D}, R_\mathrm{A}, j)) \vee$$
$$(G(\mathrm{B}, \mathrm{D}, S_\mathrm{A} \oplus \mu_\mathrm{A}, i) \wedge G(\mathrm{B}, \mathrm{D}, \mu_\mathrm{A}, j))] \Rightarrow$$
$$G(\mathrm{B}, \mathrm{D}, S_\mathrm{A}, i \oplus j) . \tag{4A}$$

If Alice enters Bob's file before he has entered hers, then he will correctly deduce that $\mu_\mathrm{A} = 1$:

$$[\neg G(\mathrm{B}, \mathrm{E}, 0) \wedge \neg G(\mathrm{B}, \mathrm{E}, 1) \wedge (G(\mathrm{A}, \mathrm{E}, 0) \vee G(\mathrm{A}, \mathrm{E}, 1))] \Rightarrow$$
$$G(\mathrm{B}, \mathrm{D}, \mu_\mathrm{A}, 1) \wedge G(\mu_\mathrm{A}, 1) . \tag{5A}$$

We remark that there are many assumptions hidden in this seemingly innocuous axiom. As we shall see, its truth depends on, among other things: (a) that the first step is done by oblivious transfer, (b) that Bob can tell when Alice enters his file, (c) that Alice is rational (so that she would not enter his file without having deduced the value of S_B), and (d) that Alice isn't "blackmailing" Bob (by threatening, for example to enter the file using password 1 within one minute unless Bob tells her in time that the true password is 0). We will look at this axiom more carefully below, and show how it can be reduced to a number of more basic axioms about the system, together with some axioms for reasoning about the other player's beliefs.

If Bob deduces that $S_\mathrm{A} = i$, it is reasonably likely he will enter the file with password i (this is the credibility assumption):

$$G(\mathrm{B}, \mathrm{D}, S_\mathrm{A}, i) \Rightarrow LG(\mathrm{B}, \mathrm{E}, i) . \tag{6A}$$

Finally, we need to say that both files get destroyed if the wrong password is used:

$$G(S_\mathrm{A}, i) \wedge G(\mathrm{B}, \mathrm{E}, i \oplus 1) \Rightarrow \mathrm{DS} . \tag{7A}$$

Let (A, C, 3) be an abbreviation for

$$\bigvee_{i=0,1,\ j=0,1} G(\mathrm{A}, \mathrm{S}, R_\mathrm{A}, i) \wedge G(S_\mathrm{A}, j) \wedge G(\mathrm{A}, \mathrm{S}, S_\mathrm{A} \oplus R_\mathrm{A}, i \oplus j \oplus 1)$$

(Alice cheats at Step 3). Then we can show, using axioms (1A)–(7A) and the

logical axioms for LL presented in the previous section, that

$$(\mathrm{A}, \mathrm{C}, 3) \Rightarrow L^2 G(\mathrm{DS}) .$$

If Alice cheats at Step 3, then she must take as an L^2 hypothesis that her files are destroyed. Taking the contrapositive we get:

$$\neg L^2 G(\mathrm{DS}) \Rightarrow \neg(\mathrm{A}, \mathrm{C}, 3) .$$

As long as Alice is not willing to contemplate the risk of having her files destroyed at level L^2, then she will not cheat at Step 3. Of course, by interchanging the roles of A and B to obtain axioms (1B)–(7B), we can prove that this is true for Bob as well. Note that if we weaken the credibility assumption (6A) to $G(\mathrm{B}, \mathrm{D}, S_{\mathrm{A}}, i) \Rightarrow L^k G(\mathrm{B}, \mathrm{E}, i)$, then the L^2 changes to L^{k+1}. The perceived risk involved in cheating depends on the credibility of Bob's threat.

While we cannot prove that Alice does not cheat at Step 2, we can show that cheating is not to her advantage. That is, we can show that if she cheats and she enters Bob's file before he enters hers, then it is likely that the files will be destroyed. Let $(\mathrm{A}, \mathrm{C}, 2)$ be an abbreviation for

$$\bigvee_{i=0,1,\ j=0,1} G(\mu_{\mathrm{A}}, i) \wedge G(S_{\mathrm{A}}, j) \wedge G(\mathrm{A}, \mathrm{S}, S_{\mathrm{A}} \oplus \mu_{\mathrm{A}}, i \oplus j \oplus 1)$$

(Alice cheats at Step 2) and let $(\mathrm{A}, \mathrm{E}, \mathrm{F})$ be an abbreviation for

$$(G(\mathrm{A}, \mathrm{E}, 0) \vee G(\mathrm{A}, \mathrm{E}, 1)) \wedge \neg G(\mathrm{B}, \mathrm{E}, 0) \wedge \neg G(\mathrm{B}, \mathrm{E}, 1)$$

(Alice enters first). Then we can show

$$(\mathrm{A}, \mathrm{C}, 2) \wedge (\mathrm{A}, \mathrm{E}, \mathrm{F}) \Rightarrow LG(\mathrm{DS}) .$$

Similarly we can show that it is not to Bob's advantage to stop the protocol before Step 3. If he enters Alice's files before she enters his, then if he has cheated at Step 2, the same proof as above shows that $LG(\mathrm{DS})$ holds, while if he is honest at Step 2, then it is easy to show that Alice can deduce S_{B}; i.e. we get

$$G(\mu_{\mathrm{B}}, i) \wedge G(S_{\mathrm{B}}, j) \wedge G(\mathrm{B}, \mathrm{S}, S_{\mathrm{B}} \oplus \mu_{\mathrm{B}}, i \oplus j)) \wedge (\mathrm{B}, \mathrm{E}, \mathrm{F}) \Rightarrow$$
$$G(\mathrm{A}, \mathrm{D}, S_{\mathrm{B}}, j)$$

(where $(\mathrm{B}, \mathrm{E}, \mathrm{F})$ says Bob enters first, and is the result of interchanging A and B in $(\mathrm{A}, \mathrm{E}, \mathrm{F})$). Finally, similar reasoning can be used to show that if both

Alice and Bob follow the protocol correctly, then Alice will be able to enter Bob's file iff he can enter hers, and it is likely they will both be able to enter.

We conclude this section by taking a closer look at axiom (5A). In order to do so, it will be helpful to imagine our language as having been augmented by modal operators for knowledge (as is done in [5]). Thus if p is a formula, so are $K_A p$ and $K_B p$ (read "Alice knows p" and "Bob knows p"). The precise details of the semantics of K_A and K_B need not concern us here. The only properties we will need are that if p is a propositional tautology, then both Alice and Bob know it; i.e. if p is a propositional tautology, then both $K_A p$ and $K_B p$ hold. Actually, we will not need this fact for all propositional tautologies, but just the few simple ones we use in our reasoning. We also assume that knowledge is closed under implication, so that $(K_A(p \Rightarrow q) \wedge K_A p) \Rightarrow K_A q$, and similarly for K_B (cf. (AX7)). Again, we will actually need only a few instances of this axiom in our reasoning. From these two assumptions it is straightforward to show that $(K_A p \wedge K_A q) \Rightarrow K_A(p \wedge q)$. (The proof uses the fact that $p \Rightarrow (q \Rightarrow (p \wedge q))$ is a propositional tautology, cf. the proof that $(Gp \wedge Gq) \Rightarrow G(p \wedge q)$ in Appendix A.) Finally, we assume that Alice and Bob know "the rules of the game"; i.e., they both know all logical and extra-logical axioms discussed above and introduced below.

What assumptions do we really need to make for axiom (5A) to hold? Why can Bob deduce that $\mu_A = 1$ if Alice enters Bob's file before he enters hers? Intuitively, Bob's reasoning is the following. He assumes that Alice is rational, so that she would not enter his file without knowing the value of S_B. In particular, she would not just randomly guess a value and enter using that value, since then both files might be destroyed. He also must assume that the only way she can deduce the value of S_B is by having deduced either both $S_B \oplus R_B$ and R_B or both $S_B \oplus \mu_B$ and μ_B. (Note that this is essentially a converse to axiom (3B).) Next he must assume that she cannot deduce the value of μ_B if he does not enter her file. (Note that this would not be true if we did not have an oblivious transfer at Step 1. If Step 1 were done by a regular transmission, then Alice would know that Bob got the value of her random bit, so she would also know that $\mu_B = 1$.) Since Alice cannot deduce the value of μ_B, then the only way she could have deduced the value of S_B is by knowing both $S_B \oplus R_B$ and R_B. Since she knows R_B, we must have $\mu_A = 1$. Of course, this whole chain of reasoning is predicated on the assumption that if Alice enters his file, then Bob knows about it.

We now capture these assumptions axiomatically. First, the assumption that Alice is rational is simply:

$$G(\mathrm{A}, \mathrm{E}, i) \Rightarrow G(\mathrm{A}, \mathrm{D}, S_B, i)\,. \tag{8A}$$

We consider this axiom in greater detail below. Again we will find that it incorporates a number of assumptions about Alice's behavior.

The next axiom is the converse of (4B) (recall that we get (4B) by

interchanging the role of A and B in (4A)), and says that the only way that Alice can deduce the actual value of S_B is by doing the modular arithmetic described in (4B):

$$\begin{aligned} &G(\mathrm{A}, \mathrm{D}, S_B, i) \Rightarrow \\ &\quad (G(\mathrm{A}, \mathrm{D}, S_B \oplus R_B, i) \wedge G(\mathrm{A}, \mathrm{D}, S_B, R_B, 0)) \vee \\ &\quad (G(\mathrm{A}, \mathrm{D}, S_B \oplus R_B, i \oplus 1) \wedge G(\mathrm{A}, \mathrm{D}, S_B, R_B, 1)) \vee \\ &\quad (G(\mathrm{A}, \mathrm{D}, S_B \oplus \mu_B, i) \wedge G(\mathrm{A}, \mathrm{D}, S_B, \mu_B, 0)) \vee \\ &\quad (G(\mathrm{A}, \mathrm{D}, S_B \oplus \mu_B, i \oplus 1) \wedge G(\mathrm{A}, \mathrm{D}, S_B, \mu_B, 1)) . \end{aligned} \tag{9A}$$

Next we need a partial converse of (5B), which describes under what circumstances Alice can deduce the value of μ_B. In fact, she can never deduce that $\mu_B = 0$ (there is always a possibility that the oblivious transfer worked), and if Bob has not entered her file, then she does not have the information to deduce that $\mu_B = 1$ either:

$$\begin{aligned} &\neg G(\mathrm{A}, \mathrm{D}, \mu_B, 0) \wedge ((\neg G(\mathrm{B}, \mathrm{E}, 0) \wedge \neg G(\mathrm{B}, \mathrm{E}, 1)) \Rightarrow \\ &\quad \neg G(\mathrm{A}, \mathrm{D}, \mu_B, 1)) . \end{aligned} \tag{10A}$$

As we observed above, this axiom makes crucial use of the fact that Step 1 used an oblivious transfer.

Using axioms (8A), (9A), and (10A), we can already show that if Alice enters Bob's file before he enters hers, then Alice must have been able to deduce that $R_B = 0$ or that $R_B = 1$; i.e., the oblivious transfer must have succeeded. That is, we can prove

$$(\mathrm{A}, \mathrm{E}, \mathrm{F}) \Rightarrow (G(\mathrm{A}, \mathrm{D}, R_B, 0) \vee G(\mathrm{A}, \mathrm{D}, R_B, 1)) .$$

But, by definition of μ_A, we must then have $\mu_A = 1$. The next axiom just captures this definition formally:

$$G(\mu_A, 1) \equiv (G(\mathrm{A}, \mathrm{D}, R_B, 0) \vee G(\mathrm{A}, \mathrm{D}, R_B, 1)) . \tag{11A}$$

Provided Bob knows that Alice enters his file before he enters hers and he knows axioms (8A)–(11A), he will know that $\mu_A = 1$. Our assumption that Bob can tell if Alice has entered his file can now be formalized as:

$$G(\mathrm{A}, \mathrm{E}, 0) \vee G(\mathrm{A}, \mathrm{E}, 1) \Rightarrow K_B(G(\mathrm{A}, \mathrm{E}, 0) \vee G(\mathrm{A}, \mathrm{E}, 1)) . \tag{12A}$$

Of course, Bob also knows about his own actions. In particular, he knows whether or not he entered Alice's file:

$$(G(\mathrm{B}, \mathrm{E}, i) \Rightarrow K_{\mathrm{B}} G(\mathrm{B}, \mathrm{E}, i)) \wedge (\neg G(\mathrm{B}, \mathrm{E}, i) \Rightarrow K_{\mathrm{B}} \neg G(\mathrm{B}, \mathrm{E}, i)) . \tag{13A}$$

Using (12A) and (13A), we can easily prove that

$$(\mathrm{A}, \mathrm{E}, \mathrm{F}) \Rightarrow K_{\mathrm{B}}(\mathrm{A}, \mathrm{E}, \mathrm{F}) ;$$

i.e. if Alice enters first then Bob will know about it. From (8A)–(11A), we can prove

$$(\mathrm{A}, \mathrm{E}, \mathrm{F}) \Rightarrow G(\mu_{\mathrm{A}}, 1) .$$

Using the properties of knowledge discussed above, it follows that

$$K_{\mathrm{B}}(\mathrm{A}, \mathrm{E}, \mathrm{F}) \Rightarrow K_{\mathrm{B}} G(\mu_{\mathrm{A}}, 1) .$$

Putting together these observations we get

$$(\mathrm{A}, \mathrm{E}, \mathrm{F}) \Rightarrow (K_{\mathrm{B}} G(\mu_{\mathrm{A}}, 1) \wedge G(\mu_{\mathrm{A}}, 1)) .$$

Now $K_{\mathrm{B}} G(\mu_{\mathrm{A}}, 1)$ says that Bob knows that $\mu_{\mathrm{A}} = 1$. But in this case, $G(\mathrm{B}, \mathrm{D}, \mu_{\mathrm{A}}, 1)$ must hold. Indeed, if we had started with an enriched language, we could have dispensed with a proposition such as $G(\mathrm{B}, \mathrm{D}, \mu_{\mathrm{A}}, 1)$ altogether, identifying it with $K_{\mathrm{B}} G(\mu_{\mathrm{A}} = 1)$. The next axiom makes this identification explicit.

$$K_{\mathrm{B}} G(V, i) \equiv G(\mathrm{B}, \mathrm{D}, V, i), \quad \text{where } V \text{ is } S_{\mathrm{A}} \text{ or } \mu_{\mathrm{A}} . \tag{14A}$$

Using (8A)–(14A) and the properties of knowledge discussed above, we can prove (5A); i.e. we can prove

$$(\mathrm{A}, \mathrm{E}, \mathrm{F}) \Rightarrow (G(\mathrm{B}, \mathrm{D}, \mu_{\mathrm{A}}, 1) \wedge G(\mu_{\mathrm{A}}, 1)) .$$

Note that in doing this deduction, Bob has to reason about Alice's reasoning. Now Alice uses this fact to conclude that it is not worthwhile to cheat at Step 2. Thus, in doing her reasoning, Alice has to reason about Bob reasoning about her! This phenomenon of reasoning about someone else's reasoning is particularly noticeable in negotiations. In doing such reasoning, it is frequently necessary to make assumptions about the rationality of the other party.

Note that if Bob knows that our "axiom of rationality" (8A) is true, then, among other things, he will not succumb to blackmail attempts on the part of Alice. Suppose Alice says to Bob "I will enter your file with password 1 at the end of one minute unless you tell me in time that the password is 0." Bob will

not find Alice's statement credible; rationality precludes her from entering without having deduced the right value. However, if Alice could convince Bob that she was very ill and would die unless she could get information about a cure which was only available in Bob's file, then her blackmail attempt would be quite credible and Bob might very well succumb to it.

Note that what really matters here is not whether the axiom is really true or not, but whether Bob believes it to be true. (This phenomenon holds for many of our axioms that involve credibility.) In a real-world situation, Bob might actually work very hard to convince Alice that he believes it, and Alice might work equally hard to convince Bob that he shouldn't. And indeed, this is precisely that type of maneuvering that one observes in negotiations!

6. Conclusions

We have presented a logic designed to reason qualitatively about likelihood and given examples of how it can be used. As is often the case, the exercise of trying to prove the protocol formally correct using LL was for us a very useful one, helping us to clarify a number of important assumptions that needed to be made, particularly in regard to blackmail threats. This is exactly why we feel it to be so important to develop logics for such purposes, especially when such subtle issues as likelihood, belief, and knowledge are involved.

Of course, we have only scratched the surface here. Work needs to be done to extend LL in order to give it greater scope and applicability. Several extensions suggest themselves. One is to consider the first-order case. Another is to incorporate *cost functions*. There are some outcomes (such as the patient dying in the case of a medical diagnosis) which may not be very likely, but have a high associated cost if they occur. A straightforward way of dealing with cost functions is to simply add primitive propositions, say $C_1, \ldots, C_5$, to represent the range from high cost (C_1) to low cost (C_5), with the relationship $C_1 \Rightarrow C_2 \Rightarrow \cdots \Rightarrow C_5$. The fact that P is an outcome with high associated cost would then be represented by the formula $P \Rightarrow C_1$.

It is also often useful to be able to incorporate knowledge and time into our reasoning. Indeed, we have already seen examples of the usefulness of knowledge in our discussion of the cryptographic protocol in the previous section. Temporal logic—a modal logic for reasoning about time—is already well-known in the literature (cf. [10, 18]), as are modal logics of knowledge and belief (see [6] for an overview). There is no problem augmenting LL with the modal operators discussed in these papers. A logic of likelihood and knowledge is investigated in [5]. It is shown that by combining the separate axiomatizations for knowledge and likelihood we can get a complete axiomatization for the resulting logic, LLK. And, like LL, LLK also has an exponential-time complete decision procedure.

Another avenue worth exploring is the relationship between LL and non-

monotonic logic. Like LL, the nonmonotonic logic of McDermott and Doyle [11, 12] uses modal logic and is designed to make inferences in the presence of uncertainty. It attempts to capture how people leap to conclusions on the basis of, for example, certain heuristic rules of thumb or default rules (the typical example used is "Since Tweety is a bird, and the typical bird flies, then, unless there is information to the contrary, conclude that Tweety flies"). Of course, when extra information is acquired, certain conclusions made using nonmonotonic reasoning may have to be withdrawn. This is exactly why nonmonotonic logic is nonmonotonic: something that can be concluded from a certain set of facts cannot necessarily be concluded from a larger set of facts.

Recall that models for LL also incorporate nonmonotonic reasoning in a very natural way. While p may be a consistent hypothesis at a given state, we might well change our minds and move to another state (possibly as a result of getting further information) in which we take $\neg p$ as a consistent hypothesis. It is thus perhaps not surprising that LL can be used to provide elegant solutions to many of the problems dealt with by nonmonotonic logic. Consider, for example, the "sorites paradox" described in [11]:

> If you remove one grain of sand from a heap of sand, you still have a heap. But if you continue doing this, you will ultimately get to a single grain. Does that mean that a single grain is a heap? If not, is there some number, say 57,895 grains, below which a bunch of grains are not a heap?

As McDermott points out, if we try to capture the property that removing one grain from a heap leaves a heap in first-order logic via

$$\mathrm{HEAP}(n+1) \Rightarrow \mathrm{HEAP}(n)\,,$$

we run into a problem when we add the observations that

$$\neg\mathrm{HEAP}(1)$$
$$\mathrm{HEAP}(1000000000)\,.$$

On the other hand, if we replace the first implication by

$$\mathrm{HEAP}(n+1) \Rightarrow \neg L^N \neg \mathrm{HEAP}(n) \quad \text{for some } N \text{ sufficiently large}\,,$$

the problem disappears. A similar solution can be given for the "lottery paradox" of [11].

Given its simple and intuitively appealing syntax and semantics, LL could be quite useful in practice to reason about situations where decisions must be

made under uncertainty, in the absence of quantitative data. But before LL can be used with complete confidence, it will be necessary to understand better the impact of using different translations to capture notions of likelihood, and the relationship between LL and other methods of capturing likelihood.

Ultimately the success or failure of a system of reasoning depends on how well it captures what people intend to say, and how easy it is to use. While we feel that LL scores well on both counts, empirical research will be necessary to bolster that feeling.

Appendix A. Proof Sketch of Lemma 3.5(f) and 3.5(g)

To prove part (f) we first show:

(i) $\vdash (Gp \wedge Gq) \equiv G(p \wedge q)$,

(ii) $\vdash (Gp \wedge \neg Gq) \Rightarrow \neg G \neg (Gp \wedge \neg q)$.

To prove (i), we first note that by propositional reasoning

$$\vdash p \Rightarrow (q \Rightarrow (p \wedge q)). \tag{1}$$

Then using (R1) we get

$$\vdash G(p \Rightarrow (q \Rightarrow (p \wedge q))). \tag{2}$$

By repeated uses of (AX7) and (R2), we get

$$\vdash Gp \Rightarrow (Gq \Rightarrow (G(p \wedge q))). \tag{3}$$

Now $\vdash (Gp \wedge Gq) \Rightarrow G(p \wedge q)$ follows by propositional reasoning. The proof that $\vdash G(p \wedge q) \Rightarrow Gp \wedge Gq$ follows by similar arguments using (AX7), (R1), (R2), and propositional reasoning, and is left to the reader.

To prove (ii), we proceed as follows. Using (AX3) and propositional reasoning, we get that

$$\vdash Gp \wedge \neg Gq \wedge G(\neg(Gp \wedge \neg q)) \Rightarrow GGp \wedge \neg Gq \wedge G(\neg(Gp \wedge \neg q)). \tag{4}$$

Now using (i), we see that

$$\vdash GGp \wedge \neg Gq \wedge G(\neg(Gp \wedge \neg q)) \Rightarrow G(Gp \wedge \neg(Gp \wedge \neg q)) \wedge \neg Gq. \tag{5}$$

By propositional reasoning we can show

$$\vdash Gp \wedge \neg(Gp \wedge \neg q) \Rightarrow q\,, \tag{6}$$

so using (R1), (R2), (AX7), (5), and (6) we get

$$\vdash G(Gp \wedge \neg(Gp \wedge q)) \Rightarrow Gq\,. \tag{7}$$

From (4)–(7) we see that

$$\vdash Gp \wedge \neg Gq \wedge G(\neg(Gp \wedge \neg q)) \Rightarrow Gq \wedge \neg Gq\,. \tag{8}$$

Of course, by propositional reasoning, $Gq \wedge \neg Gq$ is inconsistent, so we get, as desired

$$\vdash Gp \wedge \neg Gq \Rightarrow \neg G\neg(Gp \wedge \neg q)\,. \tag{9}$$

Now suppose that $Gp_1, \ldots, Gp_k$ are all the formulas of the form Gp in Σ. By using (i), (ii), (AX7), and propositional reasoning, we can easily show

$$\vdash Gp_1 \wedge \cdots \wedge Gp_k \wedge \neg Gq \Rightarrow \neg G\neg(Gp_1 \wedge \cdots \wedge Gp_k \wedge \neg q)\,. \tag{10}$$

Now suppose Σ is inconsistent, but $\{Gp_1, \ldots, Gp_k, \neg q\}$ is not. By definition, we then have

$$\vdash \neg(Gp_1 \wedge \cdots \wedge Gp_k \wedge \neg q)\,. \tag{11}$$

By (R1), it follows that

$$\vdash G(\neg(Gp_1 \wedge \cdots \wedge Gp_k \wedge \neg q))\,. \tag{12}$$

Finally, from (10) and (12) and propositional reasoning, we get that

$$\vdash \neg(Gp_1 \wedge \cdots \wedge Gp_k \wedge \neg Gq)\,. \tag{13}$$

Thus $\Delta = \{Gp_1, \ldots, Gp_k, \neg Gq\}$ is inconsistent. Since Δ is a subset of Σ, we have shown that Σ must also be inconsistent, and this contradicts our initial assumption.

The proof of part (g) is very similar to that of (f). The crucial observation we need here is that by using (AX4), (AX6), and (AX8) we can show

$$\vdash Gp_1 \wedge \cdots \wedge Gp_k \wedge \neg Lq_1 \wedge \cdots \wedge \neg Lq_m \wedge Lr \Rightarrow$$
$$L(Gp_1 \wedge \cdots \wedge Gp_k \wedge \neg q_1 \wedge \cdots \wedge \neg q_m \wedge r)\,. \tag{14}$$

We leave details to the reader. □

ACKNOWLEDGMENT

We would like to acknowledge Ed Wimmers for pointing out a missing axiom in the axiom system. The first author would like to thank David McAllester and Ron Fagin for many stimulating conversations on the subject of likelihood and probability, and Moshe Vardi and Hector Levesque for pointing out some useful references.

REFERENCES

1. Doyle, J., Methodological simplicity in expert system construction, Tech. Rept. CMU-CS-83-114, Carnegie-Mellon University, Pittsburgh, PA, 1983.
2. Emerson, E.A. and Halpern, J.Y., Decision procedures and expressiveness in the temporal logic of branching time, *J. Comput. Syst. Sci.* **30** (1) (1985) 1–24.
3. Fischer, M.J. and Ladner, R.E., Propositional dynamic logic of regular programs, *J. Comput. Syst. Sci.* **18** (2) (1979) 194–211.
4. Gärdenfors, P., Qualitative probability as an intensional logic, *J. Philos. Logic* **4** (1975) 171–185.
5. Halpern, J.Y. and McAllester, D.A., Likelihood, probability, and knowledge, in: *Proceedings AAAI-84*, Austin, TX (1984) 137–141.
6. Halpern, J.Y. and Moses, Y.O., A guide to the modal logics of knowledge and belief, in: *Proceedings IJCAI-85*, Los Angeles, CA (1985) 480–490.
7. Hughes, G.E. and Cresswell, M.J., *An Introduction to Modal Logic* (Methuen, London, 1968).
8. Kozen, D. and Parikh, R., An elementary proof of the completeness of PDL, *Theor. Comput. Sci.* **14** (1) (1981) 113–118.
9. Luby, M., Micali, S. and Rackoff, C., How to simultaneously exchange a secret bit by flipping a symmetrically-biased coin, in: *Proceedings of 24th Annual Symposium on Foundations of Computer Science* (1983) 11–22.
10. Manna, Z. and Pnueli, A., The modal logic of programs, in: *Proceedings 6th International Conference on Automata, Languages and Programming* (1979) 385–410.
11. McDermott, D.V., Nonmonotonic logic, II: Nonmonotonic modal theories, *J. ACM* **29** (1) (1982) 33–57.
12. McDermott, D.V. and Doyle, J., Nonmonotonic logic I, *Artificial Intelligence* **13** (1980) 41–72.
13. Moore, G.W., Private communication, School of Medicine, The Johns Hopkins University, 1984.
14. Moore, G.W. and Hutchins, G.M., A Hintikka possible worlds model for certainty levels in medical decision making, *Synthese* **48** (1981) 87–119.
15. Pratt, V.R., Models of program logics, in: *Proceedings 20th Annual Symposium on Foundations of Computer Science* (1979) 115–122.
16. Rabin, M.O., How to exchange secrets by oblivious transfer, unpublished manuscript, 1981.
17. Rescher, N., *Topics in Philosophical Logic* (Reidel, Dordrecht, 1968).
18. Rescher, N. and Urquhart, A., *Temporal Logic* (Springer, Berlin, 1971).
19. Segerberg, K., Qualitative probability in a modal setting, in: E. Fenstad (Ed.), *Proceedings 2nd Scandinavian Logic Symposium* (North-Holland, Amsterdam, 1971).
20. Shafer, G., *A Mathematical Theory of Evidence* (Princeton University Press, Princeton, NJ, 1976).
21. Szolovits, P. and Pauker, S.G., Categorical and probabilistic reasoning in medical diagnosis, *Artificial Intelligence* **11** (1978) 115–144.
22. Tversky, A. and Kahneman, D., Judgement under uncertainty: Heuristics and biases, *Science* **185** (1974) 1124–1131.
23. Zadeh, L.A., Fuzzy sets, *Inf. Control* **8** (1965) 338–353.

Received July 1986; revised version received November 1986