Likelihood, probability, and knowledge

JOSEPH Y. HALPERN
IBM Research Division, Almaden Research Center, San Jose, CA 95120, U.S.A.

AND

DAVID A. MCALLISTER
MIT Artificial Intelligence Laboratory, Cambridge, MA 02139, U.S.A.

Received August 13, 1986
Revision accepted September 25, 1989

The modal logic LIL was introduced by Halpern and Rabin as a means of doing qualitative reasoning about likelihood. Here the relationship between LIL and probability theory is examined. It is shown that there is a way of translating probability assertions into LIL in a sound manner, so that LIL in essence can capture the probabilistic interpretation of likelihood. However, the translation is subtle; several more obviously attempts are shown to lead to inconsistencies. We also extend LIL by adding modal operators for knowledge. This allows us to reason about the interaction between knowledge and likelihood. The propositional version of the resulting logic LILK is shown to have a complete axiomatization and to be decidable in exponential time, probably the best possible.

Key words: qualitative reasoning about likelihood, relating probability and likelihood, combining knowledge and likelihood, modal logic.

La logique modale LIL a été proposée par Halpern et Rabin comme moyen de procéder à un raisonnement qualitatif à propos de la vraisemblance. Dans cet article, la relation entre la logique modale LIL et la théorie des probabilités est examinée. Les auteurs démontrent qu’il existe une façon de traduire des assertions probabilistes en logique modale LIL de façon à ce que cette dernière puisse s’interpréter probabiliste de la vraisemblance. Cependant, cette traduction est subtile; plusieurs tentatives plus évidentes ont entraîné des inconsistencies. Des opérateurs modaux ont été ajoutés à la logique modale LIL afin de permettre un raisonnement sur l’interaction de la connaissance et de la vraisemblance. On a constaté que la version propositionnelle de la logique résultant possédait une axiomatisation complète et avait un facteur décompté en temps exponentiel.

Mots clés: raisonnement qualitatif à propos de la vraisemblance, lien probabilité et vraisemblance, combinaison connaissance et vraisemblance, logique modale.

[Traduit par la revue]


1. Introduction

Reasoning in the presence of incomplete knowledge plays an important role in many AI expert systems. One way of representing partially constrained situations is with sentences of first-order logic (cf. McCarthy and Hayes 1969; Lipski 1977; Reiter 1984). A set A of first-order sentences specifies a set of possible worlds (first-order models). Intuitively, this is the set of all possible worlds that satisfy the constraints in the first-order statements.

While such assertions can deal with partial knowledge, they cannot adequately represent knowledge about relative likelihood. This would suggest that we should attach probabilities to sentences and/or possible worlds. Indeed, that approach has been taken by many authors, going back to Carnap (1950). More recently, this approach has been renewed study (Bacchus 1988; Fagin et al. 1988; Halpern 1989; Moss 1986).

However, the use of probability is not always appropriate. Philosophers have spent years debating the situation (see Nutter (1987) and Shafer (1976) for some interesting discussion of this issue, and Chomsky (1985) for a spirited defense of probability). The epistemological problems with the use of probability in AI were first noted by McCarthy and Hayes (1969), who made the following comment:

"We agree that the formalism will eventually have to allow statements about the probabilities of events, but attaching probabilities to all statements has the following objections:
1. it is not clear how to attach probabilities to statements containing quantifiers in such a way that corresponds to the amount of conviction that people have.
2. the information necessary to assign numerical probabilities is not ordinarily available. Therefore, a formalism that required numerical probabilities would be epistemologically inadequate.

There have been a number of proposals for numerical representations of likelihood where a numerical estimate, or certainty factor, is assigned to each bit of information and to each conclusion drawn from that information (see Davis et al. (1977)). Shafier (1976), and Zadeh (1978) for example), but none of these proposals have been able to adequately satisfy the objections raised by McCarthy and Hayes. It is never quite clear where the numerical estimates are coming from, nor do these proposals seem to capture how people approach such reasoning. While people seem quite prepared to give qualitative estimates of likelihood, they are often notoriously unwilling to give precise numerical estimates to outcomes (cf. scoIovis and Pauker 1978).

Halpern and Rabin (1987) introduce a logic LIL that is designed to allow qualitative reasoning about likelihood without the requirement of assigning precise numerical probabilities to outcomes. Indeed, numerical estimates and probability do not enter anywhere in the syntax or semantics of LIL.

This is an expanded version of a paper with the same title that appears in the Proceedings of the National Conference on Artificial Intelligence, Austin, TX, 1984.
Despite the fact that no use is made of numbers, LL is able to capture many properties of likelihood in an intuitively appealing way by using a modal operator $L$ to capture the notion of likelihood. For example, consider the following chain of reasoning: if $P_1$ holds, then it is reasonably likely that $P_2$ holds, and if $P_2$ holds, it is reasonably likely that $P_1$ holds. Hence, if $P_1$ holds, it is somewhat likely that $P_2$ holds. (Clearly, the longer the chain, the less confidence we have in the likelihood of the conclusion.) In LL, this essentially becomes: \[ (P_1 \equiv L P_2) \] conclude $P_1 \equiv L P_2$.\footnote{The choice of the modal operator $L$ is perhaps somewhat unfortunate, since it has been used in other papers (e.g., McDermott 1982) to denote necessity, and still others (e.g., Levesque 1984) to denote implicit belief. Nevertheless, we stick with $L$ for consistency with Halpern and Rabin (1987) and because it suggests likelihood.} Note that the powers of $L$ denote dilution of likelihood.

While IL is meant to capture a nonprobabilistic approach to reasoning about likelihood, given the prevalent usage of probability theory, it is important to understand relationship between LL and probability theory. This is especially so, since one reasonable way of understanding likelihood is via probability theory. To quote Halpern and Rabin (1987), "We can think of likely (the modal operator $L$) as meaning "with probability greater than $\alpha"$ (for some user-defined $\alpha$).\footnote{Halpern and Rabin (1987), besides likely successors there were also conceivable successors. For ease of exposition, we have omitted "conceivable" relation here, thus identifying the operator $L$ of Halpern and Rabin (1987) with the $F$ operator, which is the dual of $G$. We leave it to the reader to check that all our results also hold if we reiterate the conceivable relation.} The exact relationship between LL and probability theory is not studied in Halpern and Rabin (1987). However, a close examination shows that it is not completely straightforward.

Indeed, as shown below, if we simply translate "$P$ holds with probability greater than $\alpha$" by $LP$, we quickly run into inconsistencies. Nevertheless, we confirm the sentiment in the quote above by showing that there is a way of translating numerical probability statements into LL in such a way that inferences made in LL are sound with respect to this interpretation of likelihood. Roughly speaking, this means that if we have a set of probability assertions about a certain domain, translate them (using the suggested translation) into LL, and then reason in LL, any conclusions we draw will be true when interpreted as probability assertions about the domain. However, our translation is somewhat subtle, as is the proof of its soundness; several more obvious attempts fail. These subtleties also shed some light on nonmonotonic reasoning.

In many situations, it does not suffice to reason about likelihood alone. We also have reason to reason about the subtask interplay between knowledge, belief, and likelihood. Work in the modal logic of knowledge and belief goes back to Hintikka (1962); more recent work can be found, for example, in Moore (1985), Fagin et al. (1984), and Halpern and Moses (1985) (see Halpern (1986) for an overview). It is clearly important to be able to reason simultaneously about knowledge and likelihood; there are many cases in which knowledge is heuristic or probabilistic. For example, suppose I know that Mary is a woman, but I have never met her and therefore do not know how tall she is. Under such circumstances, I consider it unlikely that she is over six feet tall. Recently, a logic for reasoning simultaneously about knowledge and probability has been proposed (Fagin and Halpern 1983). Here we provide a logic for reasoning simultaneously about knowledge and likelihood, by enriching LL with modal operators for knowledge to get the modal logic IL. LLK is shown to have a complete axiomatization, which is essentially obtained by combining the complete axiomatization of LL with that of the modal logic of knowledge. In addition, we show that there is a procedure for deciding validity of LLK formulas which runs in deterministic exponential time, the same as that for LL. This is probably the best possible.

The rest of the paper is organized as follows. In the next section, we review the syntax and semantics of LL. In Sect. 3, we discuss the translation of English sentences into LL and show that there is a translation which is sound with respect to the probabilistic interpretation of $L$. In Sect. 4, we add knowledge to the system to get the logic LLK, and state some technical results on decision procedures and axiomatizations for LLK. We conclude in Sect. 5 with some further directions for research.

Syntax and semantics

We briefly review the syntax and semantics of LL here, referring the reader to Halpern and Rabin (1987) for more details.

LL is a logic which extends standard propositional logic by the addition of two modal operators, $L$ and $G$. Roughly speaking, a formula of the form $Lp$ should be viewed as saying "$p$ is likely," while $Gp$ should be viewed as saying "necessarily $p$." The syntax of LL is quite straightforward. Starting with a set of variables $\Phi = \{p, q, r, \ldots\}$, or primitive propositional formulas, we build more complicated LL formulas using the propositional connectives $\neg$ and $\land$ and the modal operators $G$ and $L$. Thus, if $p$ and $q$ are formulas, then so are $\neg p$, $p \land q$, $Gp$ (necessarily $p$), and $Lp$. We omit parentheses if they are clear from context. We also use the abbreviations $p \lor q$ for $\neg \neg p \land q$, $p \land q$ for $\neg p \lor q$, $Gp$ for $(\neg p \land q) \land Gq$, and $Lp$ for $(Lp \lor q) \land Lq$. Thus, a typical LL formula is $LQLq$, which can be read "is it somewhat likely that G is necessarily the case." Note that the syntax allows arbitrary nesting and alternations of $L$s and $Gs$.

We give semantics to LL formulas by means of Kripke structures. An LL model is a triple $M = (S, \Sigma, \lambda)\footnote{In Halpern and Rabin (1987), besides likely successors there were also conceivable successors. For ease of exposition, we have omitted "conceivable" relation here, thus identifying the operator $L$ of Halpern and Rabin (1987) with the $F$ operator, which is the dual of $G$. We leave it to the reader to check that all our results also hold if we reiterate the conceivable relation.} \in S \subseteq \Sigma$, where $S$ is a set of states, $\Lambda$ is a reflexive binary relation on $S \cup \{\Lambda\}$ for all $s \in S$, we have $(s, e) \in \Lambda$, and $\Sigma = \Phi \times \Sigma \times \{\text{true}, \text{false}\}$. Thus $\lambda$ tells us for each primitive proposition $P \in \Phi$ and each state $s \in S$ whether $P$ is true in $s$.

Intuitively, a state is a complete and consistent set of "working hypotheses" concerning the situation under consideration, which we take to be "true for now." The likely successors of a state $s$ (i.e., those states $t$ such that $(t, s) \in \Lambda$) are those states that describe a set of hypothesis that is reasonably likely, given our current hypotheses.\footnote{The choice of the modal operator $L$ is perhaps somewhat unfortunate, since it has been used in this context (M., McDermott 1982) to denote necessity, and still others (e.g., Levesque 1984) to denote implicit belief. Nevertheless, we stick with $L$ for consistency with Halpern and Rabin (1987) and because it suggests likelihood.}

We can think of $(S, \Sigma, \lambda)$ as a graph with vertices $S$ and edges $\Lambda$. If $(s, t) \in \Lambda$ then we say that $t$ is an $\lambda$-successor of $s$. We will say $t$ is reachable ($\lambda$-steps) from $s$ if, for
some finite sequence \( s_0, \ldots, s_k \), we have \( q_0 = s_0, s_k = t \), and
\( (s_i, s_{i+1}) \in \Sigma \) for all \( i < k \).

We define \( M, s = p \), read \( p \) is satisfied in state \( s \) of model \( M \), by induction on the structure of \( p \):
- \( M, s = p \) for \( P \in \mathcal{E} \) if \( \tau(p, s) = \text{true} \),
- \( M, s = \neg p \) if \( M, s \not= p \),
- \( M, s = p \land q \) if \( M, s = p \) and \( M, s = q \),
- \( M, s = Gp \) if \( M, t = p \) for all \( t \) reachable from \( s \),
- \( M, s = LP \) if \( M, t = p \) for some \( t \) with \( (s, t) \in \Sigma \).

Definitions

A formula \( p \) is valid (resp. satisfiable) iff for all (resp. some) \( M = (S, \Sigma, \tau) \) and all (resp. some) \( s \) we have \( M, s = p \). It is easy to check that \( p \) is valid iff \( \neg p \) is not satisfiable. If \( \Sigma \) is a set of L L formulas, we write \( M, s \models \Sigma \) iff \( M, s = p \) for every formula \( p \in \Sigma \).

Note while that we require \( \Sigma \) reflexive guarantees us that \( p \land Lp \) is valid. This says that if \( p \) is true, then it is likely to be true. On the other hand, if \( \Sigma \) were transitive, then it is easy to see that \( Lp \land \neg Lp \) would be valid for all \( t > 0 \). This would not be desirable, since we would like to capture the diminution of likelihood using powers of \( L \). If \( \Sigma \) were symmetric, then \( Lp \land \neg Lp \) would be valid; again, this does not seem to be an appropriate property for likelihood.

We remark that in Halpern and Reinh (1967) a complete axiomatization is provided for \( L \), which completely characterizes the properties of the \( \sigma \) operator. (The axioms are described in Sec. 4, where we give a complete axiomatization for \( L \).

3. The probabilistic interpretation of likelihood

\( Lp \) is supposed to represent the notion that \( p \) is reasonably likely. Certainly one way of interpreting this statement is \( "p \) holds with probability greater than or equal to \( \sigma \)." How- ever, as already noted in Halpern and Reinh (1967), there are problems with this interpretation of \( Lp \). Suppose we take \( \sigma = \frac{1}{2} \), and consider a situation where we toss a fair coin twice. If \( \mathcal{P} \), we represent the "coin will land heads both times," then we clearly have \( Lp \in \mathcal{Q} \), as well as \( L\neg p \in \mathcal{Q} \). But, for any L L model, \( Lp \in \mathcal{Q} \) is true iff \( Lp \in \mathcal{Q} \) is true, giving us a contradiction.

We solve this problem by changing the way we translate statements of the form \( p \) is reasonably likely into L L. Note that if a state satisfies the formula \( p \) then \( M_s = p \).

this does not imply that \( p \) is necessarily true at \( s \), but simply that \( p \) is one of the hypotheses that we are taking to be true at this stage. We must use \( Gp \) to capture the fact that \( p \) is necessarily true at \( s \), since \( p \), \( \tau(p, s) = \text{true} \), \( p \) is true for all \( t \) reachable from \( s \), and thus in no state reachable from \( s \) it is the case that \( \neg p \) is true. The English statement "The coin is likely to land heads twice in a row" should be interpreted as \( "It is likely to be (necessarily) the case that the coin lands heads twice in a row" \) and thus should be translated as \( LGp \) rather than \( LPp \). Similarly, "the coin is likely to land tails twice in a row" is \( LG\neg p \), while "it is likely that the coin lands either heads or tails" is \( LGp \land LG\neg p \). With these translations, we do not run into the problem described above, for \( LGp \) being translated to \( LQp \). These observations suggest that the only L L formulas that describe real-world situations are (Boolean combinations of) formulas of the form \( LQ \), where \( C \) is a Boolean combination of primitive propositions. We will return to this point later.

Having successfully dealt with that problem, we next turn our attention to translating statements of conditional predictability: if \( \mathcal{P} \), then \( \mathcal{Q} \) is reasonably likely. Certainly it is \( Gp \), but the obvious translation, \( Gp \land LQp \), runs into trouble.

Consider a doctor making a medical diagnosis. His view of the world can be described by primitive propositions, which stand for diseases, symptoms, and test results. The relationship between these formulas can be represented by a joint probability distribution, or as a Venn diagram where the area of each region indicates its probability, and the basic regions correspond to the primitive propositions.

For example, the Venn diagram shows in Fig. 1 might represent parts of the doctor's view, where \( P_1 \) and \( P_2 \) represent diseases and \( P_3 \) and \( P_4 \) represent symptoms. The diagram shows (among other things) that (i) disease \( P_1 \) is reasonably likely given symptom \( P_3 \); (ii) \( P_3 \) is always a symptom of \( P_1 \); (iii) if a patient has \( P_1 \), then it is not reasonably likely that he also has \( P_4 \); (iv) and \( P_4 \) and \( P_2 \) never occur simultaneously.

The second statement is clearly \( GP_3 \land P_1 \), from which we can deduce \( GP_3 \land GP_1 \). Now suppose that we represented the first and third statements, as suggested above, by \( GP_1 \land GP_3 \) and \( GP_1 \land GP_2 \), respectively. Then simply using propositional reasoning, we could deduce that \( GP_1 \land GP_3 \land GP_2 \), surely a contradiction.

The problem is that when we make such English statements as \( "P_1 \) is reasonably likely given \( P_3 \); or \( "the conditional probability of \( P_3 \) given \( P_1 \) is greater than one half," we are implicitly saying "\( P_1 \) and all else being equal" or "\( P_1 \) and no other information." Indeed, it is not quite clear precisely what this statement means (cf. Halpern and Moses 1984). However, we can say "in the absence of any information about the formulas \( P_1, \ldots, P_4 \), which would cause us to conclude otherwise," and this suffices for our applications. In our present example, \( P_1 \) is reasonably likely given \( P_1 \), as long as we are not given \( P_3 \) or \( P_2 \) or both. So, a better translation of \( P_1 \) is reasonably likely given \( P_1 \) is

\[ \neg GP_1 \land GP_1 \land GP_3 \land GP_4 \land GP_1 \]

Similarly, if a patient has \( P_1 \), then it is unlikely that he has \( P_3 \) can be expressed by
In general, we must put all the necessary caveats into the precondition to avoid contradictions.

This translation seems to avoid the problem mentioned above, but how can we be sure that there are no further problems lurking in the bushes? We now show that, in a precise sense, there are not.

Fix a finite set of primitive propositions \( \Phi = \{ \Phi_1, \ldots, \Phi_n \} \). An atom of \( \Phi \) is any conjunction \( Q_1 \wedge \ldots \wedge Q_n \), where each \( Q_i \) is either \( P_i \) or \( \neg P_i \). Note that there are \( 2^n \) such atoms. Let \( \text{LET}(\Phi) \) be the set of atoms of \( \Phi \), and let \( \text{LET}(\Phi) = \{ \Phi_i, \neg \Phi_i \mid \Phi_i \in \text{LET}(\Phi) \} \). Thus \( \text{LET}(\Phi) \) consists of all the primitive propositions and their negations. Let \( \text{ Conj}(\Phi) \) be the set of all possible conjunctions of literals in \( \text{LET}(\Phi) \). We identify the empty conjunction with the form \( \bot \); thus, \( \bot \) is a member of \( \text{ Conj}(\Phi) \). Finally, let \( \text{PROP}(\Phi) \) be the set of all propositional formulas that can be formed using the propositions of \( \Phi \); if \( C, C' \) is an element of \( \text{PROP}(\Phi) \), we write \( C \leq C' \) if \( C' \) is a propositional tautology.

We say a function \( Pr : \text{LET}(\Phi) \to P(\{0, 1\}) \) is a probability assignment on \( \Phi \) if \( \forall \Phi \in \text{LET}(\Phi) \), \( Pr(\Phi) = 1 \), intuitively, \( Pr \) assigns a probability to all the atoms in \( \text{LET}(\Phi) \) in such a way that the total probability is 1. We can extend \( Pr \) to \( \text{PROP}(\Phi) \) by defining \( Pr(C) = \sum_{A \subseteq \Phi} \prod_{\Phi_i \in A} \text{Pr}(\Phi_i) \). Since \( A \subseteq \Phi \) for every atom \( A \), this means that \( Pr(\bot) = 1 \), as expected.

If \( Pr(C) > 0 \), we define the conditional probability of \( C \) given \( D \), written \( Pr(C \mid D) \), as \( Pr(C \mid D) = \frac{Pr(C \wedge D)}{Pr(D)} \). Note that \( Pr(\bot \mid D) = Pr(D) \).

Define the propositional probability space \( W = \{ Pr : \text{PROP}(\Phi) \to P(\{0, 1\}) \} \) to be a pair \( (\Phi, Pr) \), where \( Pr \) is a probability assignment on \( \Phi \). We now consider a restricted class of probability statements about \( W \).

Fix \( w \in \text{PROP}(\Phi) \). A probability assertion about \( W \) is a formula of the form \( Pr(C) = \alpha \) or \( Pr(C) < \alpha \), where \( \alpha \geq 0 \), \( C \in \text{PROP}(\Phi) \), and \( Pr(\Phi) > 0 \). (Closure under negation is built into these formulas since, for example, \( \neg Pr(C) = \alpha \) if \( Pr(C) < \alpha \). The language is powerful enough to express assertions such as \( Pr(C) \geq \alpha \), where \( \alpha \) is a constant.

It may seem at first that taking \( D \in \text{PROP}(\Phi) \) in a statement such as \( Pr(C) \geq \alpha \) is a rather powerful restriction, but this is not so. If we wish to talk about the conditional probability of \( C \) with respect to an arbitrary formula \( D \), we can simply extend \( Pr \) by adding one more primitive proposition, say \( P_D \), extend \( Pr \) so that \( Pr(P_D \equiv D) = 1 \) (this can be done easily), and write \( Pr(C \mid P_D) \) instead of \( Pr(C \mid D) \). In fact, this translation shows that we could have restricted to conditional probability statements of the form \( Pr(C \mid D) \), where \( D \) is a primitive proposition. There are two reasons not to do so. The first is pragmatic: we would like to keep \( \Phi \), the set of primitive propositions, small since, as we shall see, this keeps the set of possible "caveats" small. Second, we think of \( D \) as representing the fact that the agent has learned or observations that the agent has made so far; in practice, this can often be represented as a conjunction of literals.

Corresponding to these probability assertions about \( W \), we consider the logic \( L \) of formulas over \( \Phi \). These are formed by taking formulas of the form \( L(C \wedge \neg D) \) and \( \neg L(G) \), \( i \geq 0 \), where \( C \in \text{PROP}(\Phi) \), and closing off under conjunction and disjunction. By the observations above, these are, in some sense, exactly those \( L \) formulas that describe a "real world" situation involving the primitive propositions of \( \Phi \).

We want to translate probability assertions about \( W \) into standard \( L \) formulas over \( \Phi \). As discussed above, a conditional probability assertion of the form \( Pr(C \mid D) \geq \alpha \) will be translated into a form of the sort \( \neg G_0 \wedge \ldots \wedge \neg G_i \wedge GD \leq L(G) \), where \( Q_1, \ldots, Q_n \) are the "necessary caveats." We now make the notion of a "necessary caveat" precise. Given \( C \in \text{PROP}(\Phi), D \in \text{CON}(\Phi) \), and \( Q \in \text{LET}(\Phi) \), we say \( Q \) has negative (resp. positive) impact on \( C \) given \( D \) with respect to \( (w.r.) Pr \) if \( Pr(D \wedge Q) > \alpha \) and \( Pr(C \wedge D \wedge Q) < Pr(C \wedge D) \) (resp. \( Pr(C \wedge D \wedge Q) > Pr(C \wedge D) \)).

Thus \( Q \) has negative (resp. positive) impact on \( C \) given \( D \) w.r.t. \( Pr \) if discovering \( Q \) lowers (resp. increases) the probability of \( C \) given \( D \). We say \( Q \) has potential negative (resp. positive) impact on \( C \) given \( D \) w.r.t. \( Pr \) if for some \( D' \leq D \) with \( Pr(D') > 0 \), \( Q \) has negative (resp. positive) impact on \( C \) given \( D' \) w.r.t. \( Pr \). (In the sequel, we omit the phrase "w.r.t. \( Pr \)" if \( Pr \) is clear from context.)

Intuitively, if \( D' \leq D \), then \( D' \) represents more information than \( D \). Thus, if \( Q \) does not have potential negative impact on \( C \) given \( D \), then once we know \( D \), no matter what extra information we get, finding out \( Q \) will not lower the probability that \( C \) is true. Similar remarks hold for potential positive impact. We define

\[
\begin{align*}
\text{PNI}(C, D) &= \{ Q \in \text{LET}(\Phi) \mid \text{Q has potential negative impact on \( C \) given \( D \)} \}
\end{align*}
\]

Now the idea of potential positive and negative impact, we give a translation \( \gamma : \text{PROP}(\Phi) \to \text{PROP}(\Phi) \) of probability assertions about \( W \) to standard formulas over \( \Phi \). We define

\[
\begin{align*}
\text{Pr}(C) &\geq \alpha \\
\text{Pr}(C)^{> \alpha} &= \left( \text{let } \text{PROP}(\Phi) \rightarrow \text{PROP}(\Phi); \neg G_0 \wedge \ldots \wedge \neg G_i \wedge GD \leq L(G) \right)
\end{align*}
\]

Again we note that the term \( \text{PROP}(\Phi) \rightarrow \text{PROP}(\Phi); \neg G_0 \wedge \ldots \wedge \neg G_i \wedge GD \leq L(G) \) is intended to capture the idea of "putting in all the necessary caveats in order to avoid contradictions." With these definitions in hand, we can now state the theorem which asserts that there is a translation from probability assertions about \( W \) into \( L \) which is sound.

\[\text{Theorem I.} \quad \text{Let} \quad D' \quad \text{be} \quad \text{a} \quad \text{finite} \quad \text{set} \quad \text{of} \quad \text{probabilities} \quad \text{assertions} \quad \text{true} \quad \text{in} \quad W \quad \text{and} \quad \text{D} \quad \text{the} \quad \text{conjunction} \quad \text{of} \quad \text{all} \quad \text{standard} \quad \text{formulas} \quad \text{resulting} \quad \text{from} \quad \text{translating} \quad \text{the} \quad \text{formulas} \quad \text{in} \quad D \quad \text{into} \quad L \quad \text{via} \quad \gamma \quad \text{p} \quad \gamma \quad \text{is} \quad \text{true} \quad \text{in} \quad W \quad \text{and} \quad \text{the} \quad \text{translation} \quad \text{of} \quad \text{the} \quad \text{formulas} \quad \text{in} \quad D \quad \text{into} \quad L \quad \text{via} \quad \gamma \quad \text{is} \quad \text{true} \quad \text{in} \quad W \]

We prove the theorem by constructing, for every propositional probability space \( W \), an \( L \) model \( M = (\Phi, \Gamma, \gamma) \) which we call the canonical model corresponding to \( W \). The set of states consists of countably many copies of each \( C \in \text{PROP}(\Phi) \) with \( Pr(C) > 0 \). Successive copies are connected by \( \gamma \), as well as state you are likely to move to as your knowledge increases. More formally,
\[ S = \{ (C_i) \mid c_i \in C, \; \exists \in \text{PROP}(\Phi), \; Pr(C) > 0 \} \]

We define \( r \) as follows. For each \( \exists \in \text{PROP}(\Phi) \) such that \( Pr(C) > 0 \), choose some atom \( A \in \text{AT}(\Phi) \) such that \( A \neq C \) and \( Pr(A) > 0 \) (such an atom must exist, since \( Pr(C) = \text{SAGG}_C, A \) \( \exists \)), Call this atom \( A \). Then we define \( r(A, C) = \text{true} \) if \( F \) is one of the conjuncts in \( \text{AT}(\Phi) \). Note that this definition guarantees that \( M_{\Phi}, C \subset C \). In fact, for any propositional formula \( C \), we have \( M_{\Phi}, C \subset C \) if \( \text{AT}(\Phi) \subset C \).

The result now follows from these lemmas. We just state the lemmas here, leaving their proof to the appendix.

Lemma 1

Suppose \( C, C' \in \text{PROP}(\Phi) \) and \( Pr(C) > 0 \). Then

(a) \( \exists k \geq 1, \) then \( C \) is reachable from \( C \) in \( k \) steps if \( C' \subset C \) and \( Pr(C') > 0 \).

(b) \( M_{\Phi}, C \subset C' \) if \( Pr(C') > 0 \) and \( \text{AT}(\Phi) \subset C \).

Lemma 2

If \( q \) is a probability assertion true about \( W \), then \( M_{\Phi}, D_q = \emptyset \) for all \( D \in \text{CON}(\Phi) \).

Lemma 3

If \( q \) is of the form \( \text{Pr}(C(D)) \geq a \) or \( \text{Pr}(C(D)) < a \) and \( M_{\Phi}, D_q = \emptyset \), then \( q \) is true in \( W \).

Proof of Theorem 1

Suppose \( S \) is a finite set of probability assertions true in \( W \), \( M_{\Phi} \) is the canonical model for \( W \) constructed above, and \( q \) is a probability assertion about \( W \) such that \( \exists q' \) is valid. Assume that \( q \) is of the form \( \text{Pr}(C(D)) > a \). (The case that \( q \) is of the form \( \text{Pr}(C(D)) < a \) is similar.) By Lemma 2, since each formula \( q' \in S \) is true in \( W \), we have \( M_{\Phi}, D_q = \emptyset \). Since \( \exists \) is valid, we have \( M_{\Phi}, D_q = q \), now by Lemma 3, it follows that \( q \) is true in \( W \).

Roughly speaking, Theorem 1 says that if an agent reasoning about a situation has some probabilistic information about how likely certain events are and translates this information into LL using the translation described above, then every conclusion can be traced in LL that can be given a probabilistic interpretation will be true under the underlying situation. Thus LL can be viewed as a soundness result, in that it indicates that doing probabilistic reasoning in LL will not lead to any contradictory results.

The following example should give the reader a feel for how the translation might work.

Example

Suppose that we are told that a randomly selected Stanford student is likely (with probability at least 0.9) to be both intelligent and athletic. This statement intuits that a randomly selected student is likely to be both intelligent and athletic. More specifically, if \( S \) stands for the proposition "\( S \) is a Stanford student," \( I \) stands for the proposition "\( I \) is intelligent," and \( A \) stands for the proposition "\( A \) is athletic." We are told that \( Pr(I \wedge A) \geq 0.9 \). We leave it to the reader to check that the laws of probability allow us to conclude \( Pr(I \wedge A) \geq 0.9 \).

We now show that this conclusion can be derived in LL.

We must first translate the assertion \( Pr(I \wedge A) \geq 0.9 \). To translate this into LL we need to consider a particular probability space. Assume that \( \Phi \) includes the primitive propositions \( I \) and \( A \), and we fix some proposition \( \Phi \) of probability space \( W \). (\( \Phi \)) Assume that \( Pr(I \wedge A) > 0 \), otherwise the assertions we are interested in are not legitimate probability assertions about \( W \). We want to show that \( \Omega(1) \) is valid.

To compute these translations, we first need to compute \( \Phi = \text{Pr}(I \wedge A) \geq 0.9 \). And we have \( Pr(I \wedge A) \geq 0.9 \).

Now, to see that \( \Phi \) is valid, we must show that \( \Phi \) is true in \( W \). That is, we must show that \( Pr(I \wedge A) \geq 0.9 \) is valid. And we have \( Pr(I \wedge A) \geq 0.9 \).

Remains

Note that the translation given in Theorem 1 depends on \( W \), the underlying propositional probability space. Thus, if we are told that a randomly selected student is likely to be both intelligent and athletic, we can translate this information into LL using the translation described above, and then every conclusion can be traced in LL that can be given a probabilistic interpretation will be true under the underlying situation. Thus LL can be viewed as a soundness result, in that it indicates that doing probabilistic reasoning in LL will not lead to any contradictory results.

The following example should give the reader a feel for how the translation might work.

Example

Suppose that we are told that a randomly selected Stanford student is likely (with probability at least 0.9) to be both intelligent and athletic. This statement intuits that a randomly selected student is likely to be both intelligent and athletic. More specifically, if \( S \) stands for the proposition "\( S \) is a Stanford student," \( I \) stands for the proposition "\( I \) is intelligent," and \( A \) stands for the proposition "\( A \) is athletic." We are told that \( Pr(I \wedge A) \geq 0.9 \). We leave it to the reader to check that the laws of probability allow us to conclude \( Pr(I \wedge A) \geq 0.9 \).

We now show that this conclusion can be derived in LL.

We must first translate the assertion \( Pr(I \wedge A) \geq 0.9 \). To translate this into LL we need to consider a particular probability space. Assume that \( \Phi \) includes the primitive propositions \( I \) and \( A \), and we fix some proposition \( \Phi \) of probability space \( W \). (\( \Phi \)) Assume that \( Pr(I \wedge A) > 0 \), otherwise the assertions we are interested in are not legitimate probability assertions about \( W \). We want to show that \( \Omega(1) \) is valid.

To compute these translations, we first need to compute \( \Phi = \text{Pr}(I \wedge A) \geq 0.9 \). And we have \( Pr(I \wedge A) \geq 0.9 \).

Now, to see that \( \Phi \) is valid, we must show that \( \Phi \) is true in \( W \). That is, we must show that \( Pr(I \wedge A) \geq 0.9 \) is valid. And we have \( Pr(I \wedge A) \geq 0.9 \).

Remains
set of relevant events, which is something that seems to be frequently done in practice. For example, in medical diagnosis, we could take $G$ as a set of "relevant symptoms, diseases, and possible treatments, where the symptoms are qualitative (his temperature is very high) rather than quantitative this temperature is 104°F.

In any case, if we cannot compute PNI or PPI, we can always "play it safe," by replacing PNIIC, C (resp. PPIIC, C) wherever it occurs in the translation by a superset. It is straightforward to modify the proof of Theorem 1 to show that the resulting translation is still sound. More precisely, given a first set of probability assertions true about $W$, let $\Sigma'$ be the result of replacing each LI in formula $\phi$ by a formula $\phi'$, where instead of PNI or PPI, we use some superset of $\phi$. Then it is still the case that if $\phi' \models q'$, then $\phi$ is true about $W'.

However, since putting extra care could limit the applicability of a rule, this is not always a good strategy to follow.

If, on the other hand, we use a subset of PNI or PPI in the translation, then our reasoning may be unassiduous to explain where the nonmonotonicity comes from in certain natural language situations. People often use a type of informal default reasoning, saying "$\phi'$ is likely given $\phi$" without specifying the situation where the default $Q$ may not obtain. Of course, this means that the conclusion $Q$ may occasionally have to be withdrawn in the light of further evidence.

2. Our translation is somewhat sensitive to the set of allowable probability assertions. We could easily extend the set of probability assertions about $W$ to allow conjunctions and still prove an analogue of Theorem 1. However, if we further extend the set of probability assertions about $W$ so that it is closed under disjunction (so that $Pr(CD) = \alpha \vee Pr(C')D' < \alpha$ would be a typical assertion), and extend the translation $\phi' \models q'$ to deal with disjunctions by taking $\phi \vee \phi'$, then Theorem 1 is no longer valid. For an example, let $\phi = \{P\}$, and take $\Sigma = \{Pr(P) \geq \alpha, Pr(\neg P) \geq \alpha\}$. Let $\phi' \models \{Pr(P \vee \neg P) \geq \alpha\}$. Let $W$ be any probability space such that the assertions in $\Sigma$ are true in $W$. Since $Pr(P \vee \neg P) \geq \alpha$ must both be 0, it is clear that $\phi' \models q'$ in $W$. However, it is easy to see that $\Sigma' = \{\neg (G \vee P) \leq \alpha\} \vee \neg (G \vee \neg P) \leq \alpha\} \vee \neg (G \vee P) \leq \alpha\} \vee \neg (G \vee \neg P) \leq \alpha\}$ from which it follows by propositional reasoning that $\Sigma' \models q' \models q'$ is valid.

3. We have viewed Theorem 1 as a soundness result. It is natural to ask if there is a complementary completeness result. Ideally, we could prove that if $\phi$ is a collection of probability assertions true about $W$ and a probability assertion that follows from $\phi$, then $\Sigma \models q'$ is valid. However, this is too much to ask.

For example, suppose that $\phi = \{P, Q\}$, $\Sigma$ consists of the assertions $Pr(\phitrue) \geq \alpha, Pr(\neg \phitrue) \leq \alpha$, and $Pr(\phitrue) \leq \alpha$, and $Q$ is the assertion $Pr(\neg \phitrue) \geq \alpha$. The first two assertions in $\Sigma$ are true precisely in situations where $\alpha/2 < \alpha$.

4. Theorem 1 shows that care must be taken in order to capture this information in $LL$. In a way that is reasonably consistent with probabilistic intuitions and that, if such care is taken, there is a way of capturing it in $LL$.

2.4. Reasoning about knowledge and likelihood

We can augment $LL$ in a straightforward way in order to accommodate reasoning about knowledge. The syntax of the resulting language, which we call LKK, is the same as that of $LL$ except that we add unary operators $K_1, \ldots, K_n$, one for each of the "agents" or "agents" $u, \ldots, v$, and allow formulas of the form $K_P$ (which is intended to mean "agent $i$ knows $P$". Thus, a typical formula of $LL$ be $K_{KG} \wedge LGP$; agent $i$ knows that $Q$ is actually the case and it is likely that $P$ is the case. Again we allow arbitrary nesting and alternation of $K_i, K_j$, and $\Omega$. It is very difficult for agent $i$ to know that $P$ is likely (i.e., $K_{LG}P$) and for it to be likely that agent $i$ knows $P$ ($K_{LG}P$).

We assume that our knowledge operator satisfies the axioms of the classical modal logic $S5$ (cf. Halpern and Moses 1985). In particular, we know only true things

2Note that in order to ensure soundness, we cannot replace PNI or PPI by a superset in the conclusion $q'$ of the implication. In fact, soundness is preserved only if we use a subset, rather than a superset, of the events specified in the definition of $q'$.

3We say $\models q'$ follows from $\Sigma$, for any interpretation of $a$ with $\alpha < \alpha$, if $\phi$ is true of a propositional probability space $W$ under this interpretation of $a$, then so is $q'$.

4On the other hand, $Pr(P) < \alpha$ does not follow $Pr(Q) > \alpha$ and $Pr(KQ) > \alpha$. It is easy to construct examples to show that this is not a valid probabilistic inference.
we know exactly what we know and do not know. Thus we have

\[ Kp = p \]

\[ Kp = K, Kp \]

\[ \neg Kp = \neg Kp \]

We remark that it is easy to change the semantics for knowledge given below so that none of these axioms can be dropped (and possibly replaced by others). We omit further details here (see Halpern and Moses (1985) for more discussion of this point).

A model for LLK is a tuple \( M = (\mathcal{S}, \mathcal{I}, \mathcal{I}_1, ..., \mathcal{I}_n) \), where, just as before, \( S \) is a set of states, \( \mathcal{I} \) is a reflexive binary relation describing likely successors, and \( \pi \) assigns truth values to the primitive propositions at every state. For each \( \mathcal{I}_i \) is an equivalence relation on \( S \): it is reflexive (\( (s, s) \in \mathcal{I}_i \) for each \( s \in S \)), symmetric (\( (s, s') \in \mathcal{I}_i \) then so is \( (s', s) \)), and transitive (\( (s, s'), (s', s'') \in \mathcal{I}_i \) then so is \( (s, s'') \)). As we shall see, the \( \mathcal{I}_i \) will be used to give semantics to the \( Kp \) formulas; making it an equivalence relation guarantees that it satisfies the three axioms above.

We can think of a state and all the states reachable from it via the \( \mathcal{I}_i \) relation as describing a "likelihood distribution,"

Two states are joined via the \( \mathcal{I}_i \) relation iff agent \( i \) views them as possible likelihood distributions (rather than just possible worlds in Hintikka (1962), McCarthy et al. (1979), and Moore (1985) give his/her current knowledge. Thus, once agent \( i \) knows that \( Mary \) is on the women's basketball team, he/she would presumably consider it likely that Mary is over six feet tall, so that in all the likelihood distributions joined by \( \mathcal{I}_i \), where there is no information to the contrary, \( z \in \mathcal{I} \text{Mary is over six feet tall} \) would hold.

As before, we define \( M \rightarrow p \) by induction on the structure of \( p \). The only new clause is

\[ M \rightarrow p \iff M \vdash p \text{ for all } s \text{ such that } (s, s) \in \mathcal{I}_i \]

We now state some technical results for LLK. They essentially parallel the corresponding results for each of LL and SS alone, showing that combining knowledge and knowledge does not lead in any special complications. The proofs use standard techniques of modal logic, so are omitted here. The reader is referred to Emerson and Halpern (1985), Fischer and Ladner (1979), Halpern and Moses (1985), Halpern and Rebin (1987), and Kelemen and Parikh (1984) for further details.

**Definition**

The size of a formula \( \phi \), written \( \text{str} \phi \), is its length as a string over the alphabet \( \{ \land, \lor, \neg, D, L, K_1, ..., K_n, \} \). The size of a model \( M \) is the number of states in \( S \) (and thus could be infinite).

**Theorem 2**

An LLK formula \( p \) is satisfiable iff it is satisfiable in a model of size \( \leq 2^{(\text{str} \phi)} \).

**Theorem 3**

There is a procedure for deciding if a formula \( p \) is satisfiable (respectively, valid) which runs in deterministic time \( O(2^{(\text{str} \phi)}) \) for some \( c > 0 \).

This is the best we can do, as the following shows:

**Theorem 4**

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

**Theorem 5**

The following axiom system is sound and complete for LLK:

**Axiom schemes**

AX1. All substitution instances of tautologies of propositional logic.

AX2. \( Gp \rightarrow p \)

AX3. \( Gp \rightarrow Gp \)

AX4. \( Gp \rightarrow Gp \rightarrow p \)

AX5. \( p \rightarrow Lp \)

AX6. \( L \phi \rightarrow \phi \rightarrow p \)

AX7. \( Gp \rightarrow p \rightarrow (Gp \rightarrow Gp) \)

AX8. \( Gp \rightarrow p \rightarrow (Gp \rightarrow LG) \)

AX9. \( Gp \rightarrow p \rightarrow (Gp \rightarrow Lp) \rightarrow (p \rightarrow Gp) \)

AX10. \( Gp \rightarrow \neg \neg p \)

AX11. \( Kp \rightarrow Kp \)

AX12. \( \neg Kp \rightarrow Kp \rightarrow \neg Kp \)

AX13. \( Kp \rightarrow q \rightarrow (Kp \rightarrow Kq) \)

**Rule of inference**

R1. From \( p \) and \( p \rightarrow q \) infer \( q \) (modus ponens)

R2. From \( q \) infer \( Gp \) (generalization)

R3. From \( p \) infer \( Kp \) (knowledge generalization)

The axiom system given here is simply a combination of the axiom systems for LL (AX1, 9, R1, R2) and SS (AX1, AX10-13, R1, R3). The axioms for \( LL \) presented in Halpern and Rebin (1987) differ slightly due to the presence of the "conceivable" relation in the semantics. Had we reinstated the conceivable relation, we would have to modify AX9 accordingly (cf. Halpern and Rebin 1987).

**5. Conclusions**

We have examined the relationship between the logic LL and probability theory. We have shown that there is a precise sense in which a restricted class of probabilistic assertions about a domain can be captured by LL formulas. However, the translation from probabilistic assertions to LL is subtle; translations more naive than the one we use turn out not to be sound. In particular, in order to correctly deal with statements of conditional probability, we must specifically list all the situations in which the conclusion may or may not hold. The failure to do so in informal human reasoning is frequently the cause of the nonmonotonicity so often observed in such reasoning. (However, we note in passing that a number of the problems raised by McDermott (1982) suggest can be dealt with by nonmonotonic logic can also be dealt with by LL, in a completely monotonic fashion. See Halpern and Rebin (1987) for further discussion on this point). The need to explicitly list all the events can be viewed as a discipline which forces a practitioner to list explicitly all the exceptions to his rules. Of course, this method does not guarantee correctness. If an exception is omitted, the only conclusion made using that rule may be invalid. But, whenever a conclusion is retracted, it should be possible to find the missing exception and correct the rule appropriately.

The fact that we have a translation from probability assertions to LL does not necessarily make LL a good language


Appendix. Proof of lemmas 1, 2, and 3

For easy reference, we repeat the statements of the lemmas here.

**Lemma 1**

Suppose $C, C' \in \text{PROP}(A)$ and $Pr(C) > 0$. Then

(a) If $k \leq 1$, then $C$ is reachable from $C'$ in $k$ steps if and only if $C' \subseteq C$ and $Pr(C') \geq \alpha^{k-1}$.

(b) $M_{\alpha} \subseteq L^{\alpha}$ if $Pr(C') > 0$.

(c) $M_{\alpha} \subseteq L^{\alpha}$ if $Pr(C') \geq \alpha^{0}$.

**Proof**

For part (a) we proceed by induction on $k$. The case $k = 1$ is immediate from the definition of $L$. Suppose $k > 1$. If $C'$ is reachable from $C$ in $k$ steps, then there exists a state $D \subseteq C$ such that $D$ is reachable from $C$ in one step and $C'$ is reachable from $D$ in $k - 1$ steps. By the definition of $L$, $D$ is of the form $C_{1} \cup \ldots \cup C_{n}$, or $E_{0}$, where $E_{0} \subseteq C$ and $Pr(\neg E_{0}) > 0$. If one of the first two possibilities holds, then we get $C' \subseteq C$ and $Pr(C') \geq \alpha^{k-1}$ immediately from the inductive hypothesis. If the third case holds, by the inductive hypothesis we have $C' \subseteq E_{0}$ and $Pr(\neg E_{0}) \geq \alpha^{k-1}$. Since $C' \subseteq E_{0}$ and $E_{0} \subseteq C$, we have $C' \subseteq C$. This also means that for any atom $A$, if $A \in C'$ then $A \in E_{0}$ and $A \in C$. As a consequence, we must have $Pr(C') = Pr(C' \wedge A) = Pr(C' \wedge A) = Pr(E_{0} \wedge A) = Pr(E_{0} \wedge A)$. Easy manipulations show now that $Pr(C') \geq \alpha^{k-1}$. For the converse, if $C' \subseteq C$ and $Pr(C') \geq \alpha^{k-1}$, then $(C_{1} \cup \ldots \cup C_{n}) \subseteq C_{0}$, so it follows immediately from the definition of $L$ that $C_{0}$ is reachable from $C_{1}$ in $k$ steps.

For part (b), first suppose $Pr(C') < 1$ in $W$. Thus $Pr(\neg C') > 0$, so there must be some atom $A$ such that $A \in C'$ and $Pr(A) > 0$. From part (a) it follows that $A_{C}$ is reachable from $C$, from the definition of $*$, it follows that $MA_{C} = A$ and hence that $MA_{C} = \neg C'$. Thus, $MA_{C} \subseteq \neg C'$. For the converse, suppose that $Pr(C') \leq 1$. We want to show that $MA_{C} \subseteq \neg C'$. Suppose not. Then there must be some state $E$ reachable from $C_{0}$ such that $MA_{C} \subseteq \neg C'$. But if $E$ is reachable from $C_{0}$, then $E$ is of the form $C_{0}'$ where $C' \subseteq C$ and $Pr(\neg C') > 0$. Thus, from the definition of $*$, we must have $AT(\neg C') \subseteq \neg C'$. But, again by definition, $AT(\neg C') \subseteq C$, so $AT(\neg C') = C = \neg C'$. Moreover, since $Pr(\neg C') > 0$, we have $Pr(\neg C') > 0$. But this contradicts our assumption that $Pr(C') = 1$. Hence we have $MA_{C} = \neg C'$, as desired.

For part (c), note that if $C_{0} \subseteq L^{\alpha}$, then there exists a state $E$ which is reachable from $C_{0}$ in $k$ steps such that $MA_{C} \subseteq \neg C'$. By part (a), it follows that $Pr(C') \geq \alpha^{k-1}$, and by part (b), we have that $E \subseteq C$ and $Pr(E) \geq \alpha^{k-1}$. Using standard probabilistic reasoning, we get that $Pr(\neg C') \geq \alpha^{k-1}$. But $C' \subseteq E$, so $C' \subseteq C_{0}$ is reachable from $C_{0}$ in $k$ steps. Since $MA_{C} \subseteq \neg C'$ by part (b), it follows that $MA_{C} \subseteq L^{\alpha}$.

**Lemma 2**

Let $q$ be a probability assertion true about $W$, then $MA_{W} = \emptyset$ for all $D \in CON(W)$.

**Proof**

Suppose $q$ is of the form $Pr(C'D) \geq \alpha^{\delta}$ (the proof if $q$ is of the form $Pr(C'D) < \alpha^{\delta}$ is similar and left to the reader.) By definition, we have

$$q' := (Q \cup Q'(\emptyset)) \cap G \vdash D \cap G' \vdash D$$

Our goal is to show that $Pr(C'D) \geq \alpha^{\delta}$ in $W$, for then by part (c) of Lemma 1, we have $MA_{W} = \emptyset$, as desired. To see this, first observe that by part (b) of Lemma 1, we have $Pr(D' \cap G) = 1$. Easy calculations now show that $Pr(D' \cap D) = Pr(D \cap D') = Pr(D' \cap G') = Pr(D')$. Thus it suffices to show that $Pr(D' \cap D) \geq \alpha^{\delta}$. Next observe that none of the conjuncts that make up $D$ can be in $\neg Pr(C'D)$ (for if $Q \in P(W)$, then $D = C_{1} \cup \ldots \cup C_{n}$, or $Q$, where $Q \in P(W)$, $D$). Now, using the definition of $*$, we can show, by induction on $j$,

$$Pr(D' \cap D_{1} \cap \ldots \cap D_{n}) \geq \alpha^{\delta} \Rightarrow Pr(D' \cap D_{1} \cap \ldots \cap D_{n}) \geq \alpha^{\delta} \Rightarrow Pr(D' \cap \alpha^{\delta})$$

Since $Pr(D' \cap D_{1} \cap \ldots \cap D_{n}) \geq \alpha^{\delta} \Rightarrow Pr(D' \cap D_{1} \cap \ldots \cap D_{n}) \geq \alpha^{\delta} \Rightarrow \alpha^{\delta} \Rightarrow Pr(D' \cap D_{1} \cap \ldots \cap D_{n}) \geq \alpha^{\delta}$.

We remark that Lemma 2 does not hold for arbitrary $C \in \text{PROP}(W)$. That is, there exists a probability assertion true $q$ about $W$ and a $C \in \text{PROP}(W)$ such that $MA_{C} = \emptyset$. (We can find a counterexample by taking
\( \Phi = \{ P, Q, R \} \), \( C = (P \land Q) \lor (P \land R) \), and \( q = \Pr(\neg Q | P) \geq \alpha \). We omit details here.

**Lemma 3**

If \( q \) is of the form \( \Pr(C | D) \geq \alpha' \) or \( \Pr(C | D) < \alpha' \) and \( M_{w_2}, D_0 = q' \), then \( q \) is true in \( W \).

**Proof**

Assume that \( q \) is of the form \( \Pr(C | D) \geq \alpha' \) and \( M_{w_2}, D_0 = q' \). (Again, the case where \( q \) is of the form \( \Pr(C | D) < \alpha' \) is similar and left to the reader.) It is easy to check that if \( Q \in \text{PN}(C, D) \) then we must have \( M_{w_2}, D_0 = \neg Q \). (For if \( M_{w_2}, D_0 = Q \), then by part (c) of Lemma 1, we must have \( \Pr(Q | D) = 1 \), from which it follows that if \( D' \leq D \), we also have \( \Pr(QD') = 1 \), and \( \Pr(CD' \land Q) = \Pr(CD') \), so that \( Q \notin \text{PN}(C, D_0) \). Thus we have \( M_{w_2}, D_0 = \langle Q_{\text{every}} \rangle \neg Q \land CD \). Since, by assumption, \( M_{w_2}, D_0 = q' \), it follows that \( M_{w_2}, D_0 = L' \). By part (c) of Lemma 1, we have that \( q = \Pr(C | D) \geq \alpha' \) in \( W \), as desired. \( \blacksquare \)