In Armin Laux and Heinrich Wansing (Eds.), *Knowledge and Belief in Philosophy and Artificial Intelligence* (pp. 199–229). Berlin: Akademie-Verlag, in press.

# 9 Logic Is Not Enough: Why Reasoning About Another Person's Beliefs Is Reasoning Under Uncertainty

Anthony Jameson[1]

ABSTRACT
A system that reasons about the beliefs of a person must in general be able to ascribe a good deal of general background knowledge to that person, often in the absence of reliable evidence that the person possesses that knowledge. In some cases it is possible to make the necessary inferences within a logical framework, often as default inferences, using as premises information about the person's membership in some group or information about the system's own knowledge. But these approaches do not in general adequately deal with the uncertainty that pervades the ascription of general background knowledge. An explicitly probabilistic framework, *intuitive psychometrics*, was developed to provide a more adequate normative and descriptive account of this ascription process. Implemented using Bayesian networks, intuitive psychometrics can be used in conjunction with a modal-logic-based system for epistemic reasoning, so that the power of both the logical and the probabilistic approaches can be exploited.

## 9.1 The Problem of Uncertainty About Background Knowledge

As other papers in this volume show, a great deal of creative and subtle thought has been devoted to the development of logical formalisms for reasoning about the beliefs of agents. As is the case with all logical formalisms, their inference methods can only produce interesting specific

[1]Department of Computer Science, University of Saarbrücken, PO Box 151150, D-66041 Saarbrücken, Federal Republic of Germany, e-mail: `jameson@cs.uni-sb.de`. This research is being supported by the German Science Foundation (DFG) in its Special Collaborative Research Program on Artificial Intelligence and Knowledge-Based Systems (SFB 314), project N1, PRACMA. Thanks are due to the members of the TICS project and the ESPRIT project MEDLAR II for making MOTEL available for use within the system PRACMA, and in particular to Ullrich Hustadt for comments on the present manuscript. Stephan Baldes implemented the part of PRACMA which makes use of MOTEL's knowledge representation and inference facilities. An anonymous reviewer made valuable suggestions concerning the presentation of the present chapter.

conclusions if particular types of information are available. With epistemic logics, there is a particular type of information that is often systematically unavailable or uncertain.

### 9.1.1 A Simple Example

This problem can be illustrated with a simple example. Suppose that a person or system $\mathcal{S}$ (for "self" or "system") is trying to sell a used car $\mathcal{C}$ to a person $\mathcal{O}$ ("other").[2] $\mathcal{S}$ might have the following meta-level beliefs about $\mathcal{O}$:[3]

$$\Box_{(\text{believe},\mathcal{S})}\Box_{(\text{believe},\mathcal{O})}\mathcal{C} \in \text{BMW};$$
$$\Box_{(\text{believe},\mathcal{S})}\Box_{(\text{believe},\mathcal{O})}\text{German\_car} \sqsubseteq \text{well-engineered\_car}.$$

The first belief could result from $\mathcal{S}$'s having stated the fact in question and the second from an explicit statement by $\mathcal{O}$. A natural inference for $\mathcal{S}$ to make would be:

$$\Box_{(\text{believe},\mathcal{S})}\Box_{(\text{believe},\mathcal{O})}\mathcal{C} \in \text{well-engineered\_car}.$$

But an attempt to infer this automatically will fail unless some proposition like the following is contained in $\mathcal{S}$'s knowledge base:

$$\Box_{(\text{believe},\mathcal{S})}\Box_{(\text{believe},\mathcal{O})}\text{BMW} \sqsubseteq \text{German\_car}.$$

We will refer to knowledge such as "BMWs are German cars" as *general background knowledge*. These are factual beliefs that many people acquire in the course of everyday experience and that they can therefore use to make inferences on the basis of specific facts. The problem is that such beliefs cannot in general be assumed to be *universally* known. Therefore, the question of whether an $\mathcal{S}$ can ascribe them often introduces uncertainty into an attempt to make inferences about another person's beliefs. This problem will be referred to as the *problem of uncertainty*

---

[2]For clarity in exposition, we arbitrarily use feminine pronouns to refer to $\mathcal{S}$ and masculine pronouns for $\mathcal{O}$.

[3]The notation used in the examples throughout this paper is that of a subset of the language $\mathcal{FALCM}$, a description (or terminological) logic with epistemic operators (see [Hustadt, 1995] in this volume). Regarding the syntax and semantics of this subset of $\mathcal{FALCM}$, see the Appendix to this chapter.

*about background knowledge.*[4]

### 9.1.2 A More Complex Example

Suppose that in a dialog about another car $\mathcal{C}$, $\mathcal{S}$ at some point wonders whether $\mathcal{O}$ knows that $\mathcal{C}$ has high gas consumption—a fact about which $\mathcal{S}$ has so far said nothing specific. Suppose further that $\mathcal{S}$ *has* stated facts that might lead $\mathcal{O}$ to *infer* high gas consumption, namely that $\mathcal{C}$ is large and that it has high horsepower. Here again, general background knowledge plays a role: the knowledge that each of these two properties of a car implies high gas consumption. Since $\mathcal{S}$ cannot be sure whether $\mathcal{O}$ possesses part or all of this knowledge, $\mathcal{S}$ cannot be sure whether $\mathcal{O}$ has inferred high gas consumption. Similarly, if $\mathcal{O}$ spontaneously expresses the knowledge that $\mathcal{C}$ has high gas consumption, $\mathcal{S}$ cannot be certain how $\mathcal{O}$ arrived at this conclusion.

### 9.1.3 Overview of Solutions

It is not immediately obvious that ascribing background knowledge requires numerical techniques for reasoning under uncertainty. Several approaches to the problem have been used which can be formalized within a nonquantitative logical framework:

1. Assume that particular knowledge is possessed by *every person* within a particular group.

   For example, $\mathcal{S}$ might believe that all adult German males know that BMWs are German cars.

2. Assume that (some particular part of) the knowledge that $\mathcal{S}$ herself possesses is shared by $\mathcal{O}$.

   If it is a familiar fact to $\mathcal{S}$ that BMWs are German, she may expect other people to know the fact as well.

[4] For concreteness, the examples discussed in this chapter concern mainly true propositions which are either known or not known to the persons in question—hence the frequent use in the text of the term *knowledge* instead of *belief*. The basic techniques and arguments presented here are also applicable, with some extensions, to incorrect factual beliefs, probabilistic factual beliefs, and nonfactual beliefs such as value judgments (see, e.g., [Jameson *et al.*, 1994]).

3. Ascribe particular factual beliefs to $\mathcal{O}$ without being certain that $\mathcal{O}$ possesses them, but view these ascriptions as *default* inferences which can be retracted given contrary evidence.

   This approach is often used in conjunction with the first two, as a nonquantitative way of taking into account the fact that the first two types of inference are not entirely reliable.

### 9.1.4 Preview of the Rest of the Chapter

Section 9.2 will examine each of these three approaches in turn. It will be argued that each is based on a kernel of truth but that this truth cannot in general be adequately captured within a logical framework. Section 9.3 will briefly introduce the theory of *intuitive psychometrics*, which captures the insights of these approaches within a probabilistic framework. Finally, Section 9.4 will illustrate how the probabilistic approach can be used in combination with a modal-logic-based system.

## 9.2 Solutions Within a Logical Framework

### 9.2.1 Ascription of Background Knowledge on the Basis of Group Membership

In some contexts it may be reasonable to ascribe to $\mathcal{O}$ all or some of the general background knowledge that is relevant, on the grounds that every person within some category to which $\mathcal{O}$ belongs must possess this knowledge. For example, in a game-playing context, the only relevant background knowledge might concern the most basic rules of the game, which can be assumed to be known to any player of the game. Similarly, given that $\mathcal{O}$ is a native speaker of a given language, there is a large set of conceptual relationships that he can safely be assumed to know. When the domain in question involves everyday objects and actions, the relevant background knowledge may consist solely of the sort of common-sense knowledge that people need simply to get by in the physical world (i.e., the kind of knowledge that is being formalized in the CYC project—see, e.g., [Guha and Lenat, 1994]).

Even in these relatively straightforward situations, there is a problem of determining the boundary between the knowledge that can safely be ascribed and the knowledge which is possessed by only some of the group members. For example, which rules of a game are the "basic" ones? What conceptual relationships do all native speakers know? And where does everyday common-sense knowledge end and more specialized knowledge begin? These questions can be avoided only if all of the background knowledge relevant to $\mathcal{S}$'s meta-level reasoning clearly falls within the category that can be assumed to be generally shared.

Other cases where this type of inference may be valid are those where the particular beliefs to be ascribed are associated with some more specifically delineated group, for example:

- $\mathcal{O}$ is a Quaker $\rightarrow$ $\mathcal{O}$ believes that war should be avoided at all costs.
- $\mathcal{O}$ is a logician $\rightarrow$ $\mathcal{O}$ knows the most basic results of logic.

In such cases, where a particular group of persons is involved, one often speaks of a *stereotype* with which particular beliefs are associated (cf. [Ballim and Wilks, 1991; Hustadt, 1995; Kobsa and Pohl, 1995], among many others). The above-mentioned problem of determining which beliefs can safely be ascribed to all group members is just as acute when a specific group like these is involved. As is well known, there are Quakers who do not hold the above-mentioned belief. And while there may well be some "basic results of logic" that all logicians know, it would not be easy to list a large number of results that definitely fall into this category.

The problem can be further illustrated with the help of a quotation from an influential paper by Clark and Marshall [1981], in which they emphasize *community membership* as a clue to the background knowledge that other persons possess: "In the broad community of educated Americans, ... people assume that everyone knows such *generic* things as these: Cars drive on the right; senators have terms of six years and representatives terms of two years; and steak costs more than hamburger" (p. 35). Empirical studies of the distribution of general knowledge reveal that there is surprisingly little general background knowledge of this sort that is really known by "everyone", even within a delimited community such as that of "educated Americans". For example, Nelson and Narens [1980] compiled difficulty norms for a pool of 300 general information items, using as subjects students at two U.S. colleges. Only 7 of these 300 items were known to more than 90% of the students (e.g., 91% could name the capital of France). About 65% could name the "man who proposed the theory of relativity", and 26% knew that the "navigation instrument used at sea to plot position by the stars" is a sextant. The specific percentages of course depend strongly on the sample of persons and items investigated; but a general conclusion is that the knowledge items of this sort that can be safely ascribed to (almost) everybody in a given group make up a small proportion of the general knowledge that any group member possesses—too small a proportion to be of much use in making nontrivial inferences about the beliefs of other persons.

### 9.2.2 Ascription of Background Knowledge on the Basis of One's Own Knowledge

It is a common observation that people tend to use their own beliefs as a starting point when ascribing beliefs to others. Empirical research has documented this phenomenon abundantly (see, e.g., [Ross *et al.*, 1977; Nickerson *et al.*, 1987; Jameson *et al.*, 1993]), and normative analyses have shown that there is often considerable justification for this tendency (see, e.g., [Hoch, 1987; Dawes, 1989] and Section 9.3.2 below).

In a nonprobabilistic form, this basic idea can be used by a system to ascribe general background knowledge (see, e.g., [Ballim and Wilks, 1991, chap. 3]). In our example domain of used cars, $\mathcal{S}$ would assume (most plausibly by default, as discussed in 9.2.3) that $\mathcal{O}$ has the same general knowledge about cars that $\mathcal{S}$ has. This approach suffers from the same fundamental problem as the approach that uses group membership as evidence, namely the only partly predictable distribution of general knowledge among people. In particular, given a fact that is known to only some of the members of any given group, $\mathcal{S}$ can only ascribe the fact to any given $\mathcal{O}$ with high confidence if $\mathcal{S}$ has some evidence about *this particular* $\mathcal{O}$ which indicates that $\mathcal{O}$ knows this fact; yet $\mathcal{S}$'s own possession of the knowledge is not evidence that specifically concerns $\mathcal{O}$.

The relevant empirical research shows that people interpret their own possession of a fact as one piece of evidence suggesting that another person might possess it as well, a piece of evidence which is combined with evidence of other types. For example, in the study by Nickerson *et al.* [1987], students estimated the percentage of other students who knew particular facts from the pool of items developed by Nelson and Narens [1980] (cf. 9.2.1). The estimates, which covered the whole range from 0%

to 100%, were in general higher in cases where the predictor ($\mathcal{S}$) herself knew the fact in question—even when the overall actual difficulty of the items was controlled for. But this relationship between $\mathcal{S}$'s own knowledge and $\mathcal{S}$'s predictions for others was loose: An $\mathcal{S}$ often gave a low estimate even for a fact that she herself knew with 100% confidence. This pattern is normatively appropriate, given that $\mathcal{S}$ can have reasons to believe that a fact that she herself knows is difficult for people in general to acquire.

Even to the extent to which $\mathcal{S}$'s own knowledge constitutes useful evidence about $\mathcal{O}$'s knowledge, this is only the case if $\mathcal{S}$ has the knowledge typical of a human being who more or less resembles $\mathcal{O}$. In systems whose knowledge is entered or acquired in a way that has little to do with the way people acquire their knowledge, the applicability of this approach is especially limited.

### 9.2.3 Ascription of Background Knowledge by Default

Granted that the two approaches discussed in the previous sections cannot in general give rise to knowledge ascriptions which are certain, a way of applying the approaches without resorting to numerical handling of uncertainty is to treat the knowledge ascriptions as default inferences (see, e.g., [Ballim and Wilks, 1991; Kobsa and Pohl, 1995]).

There are several different senses in which the idea of a *default inference* has been used. Examining each in turn, we can see that most have some degree of applicability to the ascription of background knowledge but that none is sufficiently applicable to serve as a general way of dealing with the problem of uncertainty about background knowledge.

**Extremely high probability.** Sometimes the concept of a default is used when a proposition has a probability so high that it is not worth the cognitive effort to consider the possibility that it might not be true (e.g., the proposition that no bomb will fall onto your roof during the next 5 minutes). In the knowledge ascription context, such high probabilities can be found for the ascription of generally shared linguistic or common-sense world knowledge; but as we have seen, often much of the relevant background knowledge can be ascribed only with a probability much lower than 100%.

**Normality.** Sometimes the concept of a default has nothing inherently to do with probability but rather concerns the extent to which a state of affairs is normal, as opposed to abnormal. For example, it is normal for an elevator door to close when the *Close* button is pressed; if this does not happen, there must be some malfunction. Similarly, one might say that it is normal for a human elevator operator to possess knowledge of how to get the elevator door to close; otherwise he could not perform his function. This example illustrates that this use of the concept of a default has some applicability to knowledge ascription, but it also shows that the applicability is limited: If an $\mathcal{S}$ tries to predict whether an elevator operator $\mathcal{O}$ could perform some more esoteric operation with the elevator, or how he would react in an unusual situation, the concept of "normally possessed knowledge" will not by itself be adequate.

**Autoepistemic reasoning.** A further justification for assuming a fact if you don't know that it is untrue may be the assumption that if it *were* untrue you would be aware of its untruth (see, e.g., [Moore, 1985]). For example, you may assume that the guest you are talking to is *not* a Nobel Prize winner, because if he were, this interesting fact would have come to your attention by now. This reasoning schema can also be applied to the ascription of knowledge: You may assume that your guest does not believe that you are a Nobel Prize winner, because if he did, this belief would surely have manifested itself somehow in his behavior. But note that the application of this schema itself requires the ascription of background knowledge and/or other beliefs—for example, the ascription of beliefs about what behavior is appropriate in an interaction with a Nobel Prize winner.

**Content-specific communication conventions.** The idea of a default can be applied to specific conventions governing human dialog. For example, when a traveler at a train station asks for a ticket, he means a second-class ticket by default (i.e., if he doesn't specify otherwise). It is not easy to think of corresponding content-specific conventions that could apply to the ascription of background knowledge.

**Conversational implicature.** In a roughly similar way as content-specific communication conventions, general dialog conventions can sometimes warrant inferences about the speaker's beliefs on the basis of the fact that the speaker has not stated contrary beliefs. For example, when asked if he supports a given military intervention, $\mathcal{O}$ may reply "I'm a Quaker". Even given the existence of nonpacifist Quakers, the listener is justified in ascribing to $\mathcal{O}$ the typical Quaker attitude toward war, because a cooperative speaker would not have produced this answer if he didn't have the typical attitude. This type of belief ascription occurs frequently and merits careful analysis; but the schema is applicable only under certain conditions and cannot serve as a general way of conceptualizing the uncertainty inherent in the ascription of background knowledge.

## 9.3 Basic Concepts of Intuitive Psychometrics

The previous section has shown that the ascription of background knowledge requires the integration of various types of evidence and even then often yields conclusions that are by no means certain. A conceptualization which takes these points into account is *intuitive psychometrics* (see, e.g. [Jameson, 1990; Jameson, 1992]).

As the name suggests, this theory is based on an analogy between everyday knowledge ascription by humans and formal techniques of psychological testing. In particular, a basic assumption of the branch of psychological testing called *Item Response Theory* (see, e.g., [Hambleton and Swaminathan, 1985]) is that for a given knowledge *item* a probabilistic relationship exists between:

- the *difficulty* of the item,
- the *knowledgeability* of a person, and
- the likelihood that the person will give a correct *response* to that item.

To employ these concepts to model everyday knowledge ascription, it is useful to implement them within the framework of a *Bayesian network* (see, e.g., [Pearl, 1988], whose notation is used in this chapter, or [Neapolitan, 1990]). Figure 9.1 gives a simple example of how this can be done. The network contains three *nodes*, each of which represents the belief of $\mathcal{S}$ about a variable that is relevant to her prediction of whether $\mathcal{O}$ possesses the knowledge that BMWs are German cars (cf. the first example in Section 9.1). The two arrows indicate that the upper two nodes are the *parents* of the third node—i.e., that the variables in the upper two nodes can be viewed as causes of the variable in the third node.

The belief about the variable $X$ associated with a node $N_X$ is repre-

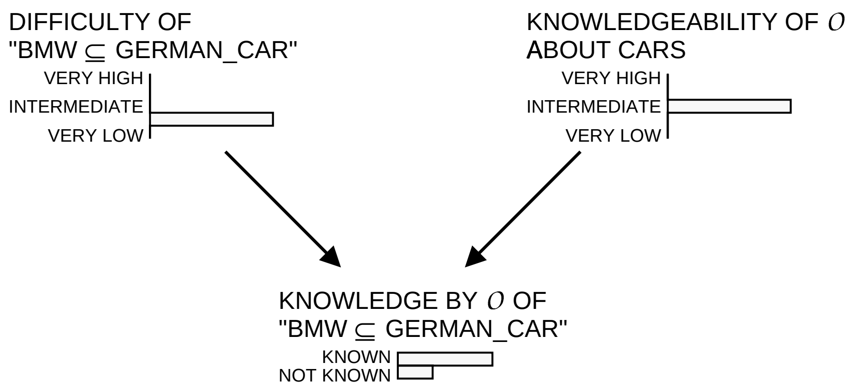


FIGURE 9.1. A simple example of knowledge ascription within intuitive psychometrics.

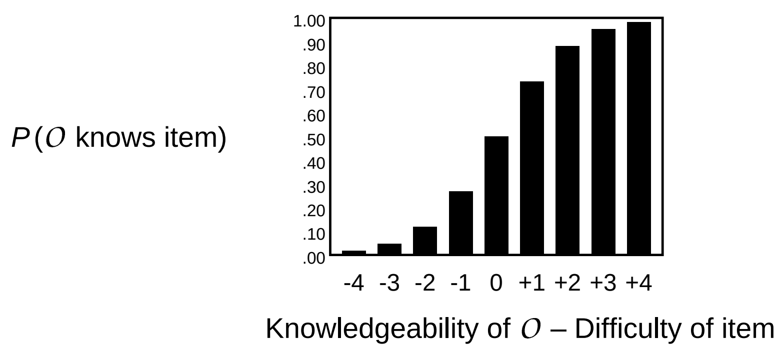


FIGURE 9.2. Graph depicting the probabilistic relation between the parent and child nodes in Figure 9.1. (For the determination of the numbers on the x-axis, the five levels of the variables in the parent nodes are associated with the integers 0, 1, 2, 3, and 4, respectively—cf. Step 5 in Figure 9.3.)

sented by a discrete probability distribution $BEL(x)$.[5] For example, for the node KNOWLEDGE BY $O$ OF "BMW ⊑ GERMAN_CAR", corresponding to the vari-

---

[5]In expressions involving $BEL$ and $P$ (for probability), a capital letter denotes a variable and the corresponding lower-case letter denotes a possible value of that variable. For each possible value $x$ of $X$,

$$BEL(x) \triangleq P(x|\boldsymbol{e}),$$

where the expressions $BEL(x)$ and $P(x|\boldsymbol{e})$ are abbreviations for $BEL(X = x)$ and $P(X = x|\boldsymbol{e})$, respectively, and $\boldsymbol{e}$ represents all of the evidence received so far ([Pearl, 1988, p. 152]).

able $Z$, $\mathcal{S}$ assigns to the value NOT KNOWN a probability of .27 and to the value KNOWN .73; thus $BEL(z)$ corresponds to the vector (.27, .73). The nodes DIFFICULTY OF "BMW $\sqsubseteq$ GERMAN_CAR" and KNOWLEDGEABILITY OF $\mathcal{O}$ ABOUT CARS theoretically should correspond to continuous variables, but for reasons of computational tractability each of these variables is approximated with a discrete variable which has five possible values: VERY LOW, LOW, INTERMEDIATE, HIGH and VERY HIGH. In this simple example, we assume that $\mathcal{S}$ has completely definite beliefs about these two variables: For DIFFICULTY OF "BMW $\sqsubseteq$ GERMAN_CAR" the distribution is (0, 1, 0, 0, 0), and for KNOWLEDGEABILITY OF $\mathcal{O}$ ABOUT CARS it is (0, 0, 1, 0, 0), i.e., $\mathcal{S}$ is convinced that the knowledge item "BMW $\sqsubseteq$ German_car" has the difficulty level LOW and that $\mathcal{O}$ has the knowledgeability level INTERMEDIATE.

In a Bayesian network, a probabilistic relationship between two parent nodes $N_X$ and $N_Y$ and a child node $N_Z$ is represented by a matrix of conditional probabilities $P(z|x, y)$ which contains one probability for each possible combination of values of $Z$, $X$, and $Y$. For these particular three nodes, the matrix is generated using the function depicted in Figure 9.2, which is derived from the Rasch model, or one-parameter logistic model, of Item Response Theory (see, e.g., [Hambleton and Swaminathan, 1985]).[6] Figure 9.3 gives the algorithm for creating a node like KNOWLEDGE BY $\mathcal{O}$ OF "BMW $\sqsubseteq$ GERMAN_CAR" and linking it to its parents.

The belief shown for the child variable $Z$ is derived from those for the parent variables $X$ and $Y$ through the *downward propagation* procedure for singly connected Bayesian networks.[7] In the present example, $\mathcal{S}$ considers it likely that $\mathcal{O}$ knows the fact in question because the difficulty of the fact is lower than $\mathcal{O}$'s level of knowledgeability.

---

[6] If this model were being used to predict empirical data, the appropriate specific form of the function would have to be determined empirically.

[7] See, e.g., [Pearl, 1988, chap. 4]. Where (as here) there is initially no other evidence concerning the child variable $Z$, the resulting belief vector is related to those for the parent nodes as follows:

$$BEL(z) = \sum_{x,y} P(z|x, y) BEL(x) BEL(y).$$

In the present example, where the beliefs about $X$ and $Y$ are both definite, we have:

$$\begin{aligned} BEL(Z = \text{KNOWN}) &= P(Z = \text{KNOWN}|X = \text{LOW}, Y = \text{INTERMEDIATE}); \\ BEL(Z = \text{NOT KNOWN}) &= 1 - BEL(Z = \text{KNOWN}). \end{aligned}$$

### Create Simple Knowledge Node

**Input**

- A sentence of the form $\Box_{(\text{believe},\mathcal{O})}\Phi$, where $\Phi$ is a nonmodal assertional sentence.

**Output**

- A Bayesian network node, linked to its parent nodes, representing $\mathcal{S}$'s belief in $\Box_{(\text{believe},\mathcal{O})}\Phi$.

**Procedure**

1. **If** the required node has already been created, **then** output it and **stop**.
2. **If** there does not yet exist a node $N_X$ representing $\mathcal{S}$'s belief about the variable $X$, the difficulty of $\Phi$, **then**
   - Create $N_X$, where the five possible values for $X$ are VERY LOW, LOW, INTERMEDIATE, HIGH, and VERY HIGH.
   - Initialize $\mathcal{S}$'s belief about $X$ using a default prior distribution.
     ▹ *The default distribution used is the uniform distribution.*
3. **If** there does not yet exist a node $N_Y$ representing $\mathcal{S}$'s belief about the variable $Y$, the knowledgeability of $\mathcal{O}$, **then**
   - Create $N_Y$, where the five possible values for $Y$ are the same as those for $X$.
   - **If** one or more facts are known about $\mathcal{O}$ which have implications concerning $\mathcal{O}$'s knowledgeability, **then** create (or retrieve) the corresponding nodes and link them to $N_Y$ as parent nodes; **otherwise** initialize $\mathcal{S}$'s belief about $Y$ using a default prior distribution.
     ▹ *The linking of $N_Y$ to one or more parent nodes can take various forms, depending on the nature of the facts available.*
4. Create the node $N_Z$ representing $\mathcal{S}$'s belief about the variable $Z$, which has the value KNOWN if $\Box_{(\text{believe},\mathcal{O})}\Phi$ and NOT KNOWN otherwise.
5. Link $N_Z$ to its parents $N_X$ and $N_Y$, defining the matrix of conditional probabilities as follows:

$$\begin{aligned} P(Z = \text{KNOWN}|x,y) &= \frac{e^{f(y)-f(x)}}{1+e^{f(y)-f(x)}}; \\ P(Z = \text{NOT KNOWN}|x,y) &= 1 - P(Z = \text{KNOWN}|x,y), \end{aligned}$$

   where the values of $f(x)$ and $f(y)$ are 0, 1, 2, 3, and 4 for the five possible values of $X$ and $Y$ (cf. Figure 9.2).
6. Output $N_Z$.

FIGURE 9.3. Algorithm for creating a Bayesian network node corresponding to the ascription of a single piece of background knowledge.

### 9.3.1 Probabilistic Inferences on the Basis of Group Membership

The uncertainty associated with a prediction by $\mathcal{S}$ is in general greater than in this first example, because in general $\mathcal{S}$ cannot have a completely definite belief about variables such as DIFFICULTY OF "BMW $\sqsubseteq$ GERMAN_CAR" and KNOWLEDGEABILITY OF $\mathcal{O}$ ABOUT CARS. For example, Figure 9.4 illustrates how the corresponding prediction might be made in the case where $\mathcal{S}$'s only information about $\mathcal{O}$ was that he is a man. Considering only the first histogram shown for each node, we see that $\mathcal{S}$ infers from $\mathcal{O}$'s gender that higher levels of knowledgeability about cars are more likely for $\mathcal{O}$ than lower ones are, although none of the levels is ruled out entirely.[8] Even the uncertain belief of $\mathcal{S}$ about $\mathcal{O}$'s knowledgeability level shown in KNOWLEDGEABILITY OF $\mathcal{O}$ ABOUT CARS allows a prediction about KNOWLEDGE BY $\mathcal{O}$ OF "BMW $\sqsubseteq$ GERMAN_CAR" (cf. Footnote 7). This network illustrates the kernel of truth behind the idea that general background knowledge can be ascribed on the basis of a person's membership in a group. $\mathcal{S}$ is quite confident that $\mathcal{O}$ knows that BMWs are German because $\mathcal{O}$ is a man and because the fact is relatively easy to know. The cases mentioned in Section 9.2.1 in which such ascription can be made with (virtual) certainty are typically those in which membership in a given group (e.g., citizens of Germany) implies at least a moderate level of knowledgeability about some topic (e.g., German politics), and where the difficulty of the item in question with respect to that topic is very low (e.g., "Who is the chancellor of Germany?"). In such cases, the first histogram in the node corresponding to the lowest node in Figure 9.4 shows a probability distribution with an even higher probability for the value KNOWN.

Suppose now that $\mathcal{O}$ expresses explicitly the knowledge that BMWs are German. The second histogram for KNOWLEDGE BY $\mathcal{O}$ OF "BMW $\sqsubseteq$ GERMAN_CAR" represents $\mathcal{S}$'s resulting definite belief that $\mathcal{O}$ knows this fact. The second histograms shown for the other nodes in the network illustrate another fundamental type of inference within the intuitive psychometrics framework, in addition to prediction, namely the *interpretation* of evidence in $\mathcal{O}$'s behavior. By the general process of *upward propagation* in Bayesian

[8]The matrix of conditional probabilities linking the nodes GENDER OF $\mathcal{O}$ and KNOWLEDGEABILITY OF $\mathcal{O}$ ABOUT CARS represents essentially $\mathcal{S}$'s stereotypes about how knowledgeable men and women are about cars.

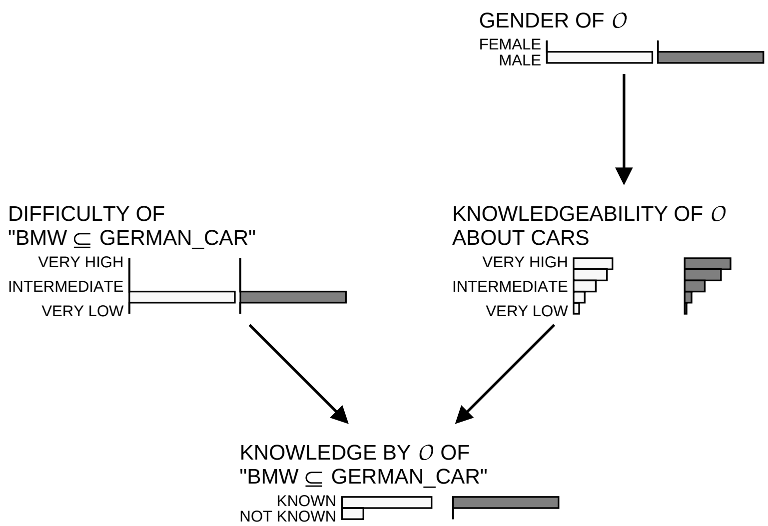


FIGURE 9.4. Example of probabilistic prediction and interpretation of knowledge. (The second histogram for each node represents $\mathcal{S}$'s belief after $\mathcal{S}$ has observed the value of the variable in the lowest node.)

networks, $\mathcal{S}$ revises her belief about KNOWLEDGEABILITY OF $\mathcal{O}$ ABOUT CARS;[9] in this example the revision is minor because $\mathcal{O}$ was expected to know the item in the first place. $\mathcal{S}$'s revised belief about KNOWLEDGEABILITY OF $\mathcal{O}$ ABOUT CARS can now be used for later predictions about $\mathcal{O}$'s knowledge of specific facts. $\mathcal{S}$'s belief about GENDER OF $\mathcal{O}$ could in turn be revised by upward propagation from KNOWLEDGEABILITY OF $\mathcal{O}$ ABOUT CARS, but in this case there is no change, because $\mathcal{S}$'s belief about GENDER OF $\mathcal{O}$ was definite in the first place. For the same reason, there is no change in $\mathcal{S}$'s belief

[9]Upward propagation essentially uses Bayes' Rule to adjust the probability associated with each possible value of a variable in an ancestor node in accordance with the conditional probability of the observed evidence given that value. Although the computations are in general more complex, a simple illustration is offered by the three lowest nodes $N_X$, $N_Y$, and $N_Z$ in Figure 9.4. The updated belief vector $BEL'(y)$ for the parent variable $Y$ in the node KNOWLEDGEABILITY OF $\mathcal{O}$ ABOUT CARS after the observation $Z = \text{KNOWN}$ for the child variable is related to the prior belief vector $BEL(y)$ as follows:

$$BEL'(y) = \frac{BEL(y)P(Z = \text{KNOWN}|X = \text{LOW}, y)}{\sum_y P(Z = \text{KNOWN}|X = \text{LOW}, y)}.$$

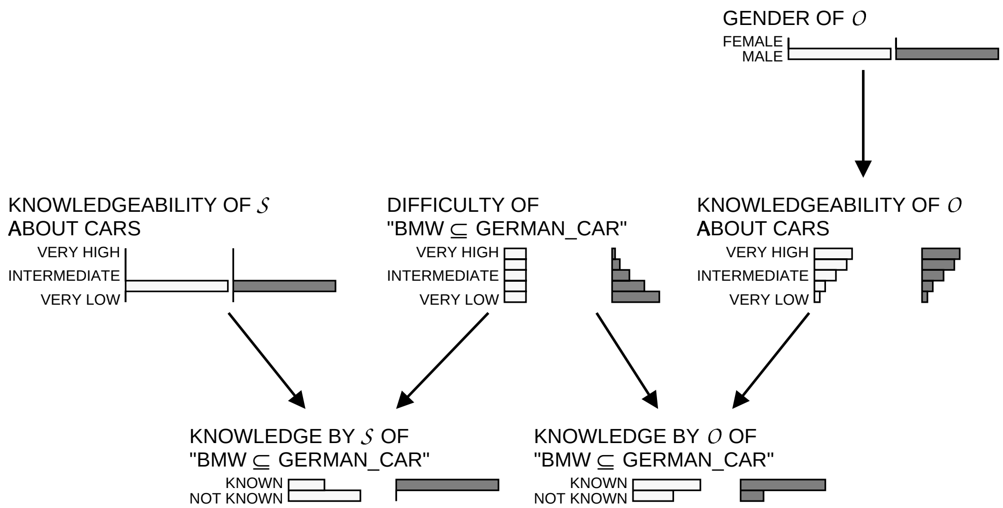


FIGURE 9.5. Example of probabilistic use of one's own knowledge as evidence about the knowledge of another person. (The second histogram for each node represents $\mathcal{S}$'s belief after $\mathcal{S}$ has taken into account her own knowledge of the item in question.)

about DIFFICULTY OF "BMW ⊑ GERMAN_CAR".

### 9.3.2 Probabilistic Use of One's Own Knowledge as Evidence

Figure 9.5 illustrates how the phenomemon of "generalizing from oneself to others" is conceptualized within the intuitive psychometrics framework. It represents basically the line of reasoning "I know that BMWs are German, and I don't know much about cars in general, whereas $\mathcal{O}$, being a man, should be fairly knowledgeable; so $\mathcal{O}$ is likely to know this fact as well".

The first histograms for the four right-most nodes depict a prediction about $\mathcal{O}$'s knowledge like the one in Figure 9.4, except that now $\mathcal{S}$ initially has a completely indefinite belief about DIFFICULTY OF "BMW ⊑ GERMAN_CAR". The three left-most nodes (both histograms) show how $\mathcal{S}$ acquires a more definite belief about DIFFICULTY OF "BMW ⊑ GERMAN_CAR" on the basis of her *own* knowledge of the item (KNOWLEDGE BY $\mathcal{S}$ OF "BMW ⊑ GERMAN_CAR"). This updating involves upward propagation, just as in Figure 9.4. The second histogram for DIFFICULTY OF "BMW ⊑ GERMAN_CAR" shows that $\mathcal{S}$ now considers it most likely that the item is easy, given the fact that she herself knows it. This new belief in turn gives rise to downward propagation, causing a more optimistic belief for KNOWLEDGE BY $\mathcal{O}$ OF "BMW $\sqsubseteq$ GERMAN_CAR" (second histogram).

The account shown in Figure 9.5 is simplified for purposes of exposition: In general, $\mathcal{S}$ has more information about her own knowledge of a particular item than simply whether she herself knows it or not; for example, she can often assess the strength of the *evidence* she has for her knowledge (see, e.g., [Koriat, 1993], [Jameson, 1990, chap. 3]). Such a case can be modeled with a network like the one in Figure 9.5 in which the variable in KNOWLEDGE BY $\mathcal{S}$ OF "BMW $\sqsubseteq$ GERMAN_CAR" is replaced by a more differentiated variable; such a network will often show a more drastic revision of $\mathcal{S}$'s belief about the item's difficulty. But still, cases where $\mathcal{S}$ can ascribe a piece of knowledge to $\mathcal{O}$ with near-certainty solely on the basis of the fact that $\mathcal{S}$ possesses the knowledge herself are extreme cases (e.g., cases where $\mathcal{S}$ considers a given fact completely obvious and believes that $\mathcal{O}$ has a higher knowledgeability level than herself with respect to the relevant topic).

In addition to being able to account plausibly for the general patterns of reasoning discussed above, the theory of intuitive psychometrics has proven useful as a standard of comparison for human judgments, in spite of its basically normative character ([Jameson, 1990, chap. 5]). In particular, some rather counterintuitive patterns of inference which can be observed in human subjects turn out to follow straightforwardly from the basic assumptions of the model. For example, after observing whether a person $\mathcal{O}$ knows a given item, subjects are willing simultaneously to update their beliefs about both $\mathcal{O}$'s knowledgeability and the item's difficulty, if they initially had indefinite beliefs about both variables; this simultaneous updating of two parent variables is in agreement with the Bayesian reasoning embodied in the theory of intuitive psychometrics.

## 9.4 Combination of Epistemic Logic with Intuitive Psychometrics

The previous section showed how the concepts of intuitive psychometrics, implemented using Bayesian networks, can be used to reason about a person's possession of a given piece of general background knowledge. If a system is to handle more complex cases, such as those which occur in the context of natural dialogs, more complex Bayesian networks than those shown in Figures 9.1, 9.4, and 9.5 are required. These networks must be constructed dynamically, because (a) it is impossible for a system designer to foresee all of the specific inferences that $\mathcal{S}$ might conceivably have to make and (b) in any case a large, preconstructed network would require a great deal of irrelevant computation in connection with each necessary inference (cf. [Haddawy, 1994]).

One approach is to use a logic-based system to help construct the relevant Bayesian network and then to allow the probabilistic processing within the Bayesian network to handle the uncertainty involved in the ascription of background knowledge. This is one of the methods for dynamically constructing Bayesian networks that is used in the system PRACMA, which can take the role of the seller ($\mathcal{S}$) in a dialog about a used car.[10] PRACMA makes use of the modal-logic-based knowledge representation system MOTEL ([Hustadt and Nonnengart, 1993]). The subset of MOTEL currently used in PRACMA corresponds to a subset of the language $\mathcal{FALCM}$ (see the Appendix to this chapter and [Hustadt, 1995] in this volume). PRACMA's realization of the hybrid approach will be discussed in this section using the more complex of the two examples introduced in Section 9.1.

### 9.4.1 Representation of the Example Problem

Figure 9.6 shows examples of four types of sentences that might be contained in PRACMA's MOTEL knowledge base.[11] The knowledge shown in the first box is the specific knowledge on the basis of which $\mathcal{O}$ could infer that $\mathcal{C}$ has high gas consumption, if $\mathcal{O}$ had the necessary general background knowledge. $\mathcal{S}$ can confidently ascribe this specific knowledge to $\mathcal{O}$ because $\mathcal{S}$ has stated these facts during the dialog.

One piece of relevant background knowledge is shown in the second

[10] For accounts of the other approaches used in PRACMA, see [Jameson *et al.*, 1995; Jameson *et al.*, 1994; Schäfer, 1994].

[11] Whereas MOTEL is implemented in PROLOG, PRACMA is implemented in COMMON-LISP; it communicates with MOTEL via an interface which performs the necessary syntactic translation.

| Specific knowledge ascribed to $\mathcal{O}$ by $\mathcal{S}$ |
|---|
| $\Box_{(\text{believe},\mathcal{S})}\Box_{(\text{believe},\mathcal{O})}\mathcal{C} \in \text{large\_car}$ |
| $\Box_{(\text{believe},\mathcal{S})}\Box_{(\text{believe},\mathcal{O})}\mathcal{C} \in \text{car\_with\_high\_horsepower}$ |
| $\vdots$ |

| Knowledge assumed by to be shared by everyone |
|---|
| $\Box^{c}_{(\text{believe},\text{all})}\text{car\_with\_powerful\_motor} = \text{powerful\_car}$ |
| $\vdots$ |

| $\mathcal{S}$'s own general knowledge |
|---|
| $\Box_{(\text{believe},\mathcal{S})}\text{old\_car} \sqsubseteq \text{car\_with\_high\_consumption}$ |
| $\Box_{(\text{believe},\mathcal{S})}\text{large\_car} \sqsubseteq \text{car\_with\_high\_consumption}$ |
| $\Box_{(\text{believe},\mathcal{S})}\text{car\_with\_high\_horsepower} \sqsubseteq \text{car\_with\_powerful\_motor}$ |
| $\Box_{(\text{believe},\mathcal{S})}\text{powerful\_car} \sqsubseteq \text{car\_with\_high\_consumption}$ |
| $\vdots$ |

| $\mathcal{S}$'s own specific knowledge |
|---|
| $\Box_{(\text{believe},\mathcal{S})}\mathcal{C} \in \text{car\_with\_high\_consumption}$ |
| $\Box_{(\text{believe},\mathcal{S})}\mathcal{C} \in \text{old\_car}$ |
| $\Box_{(\text{believe},\mathcal{S})}\mathcal{C} \in \text{large\_car}$ |
| $\Box_{(\text{believe},\mathcal{S})}\mathcal{C} \in \text{car\_with\_high\_horsepower}$ |
| $\vdots$ |

FIGURE 9.6. Knowledge in PRACMA's MOTEL knowledge base relevant to the example discussed in the text.

box in Figure 9.6, which contains terminological sentences of the form $\Box^{c}_{(\text{believe},\text{all})}\Phi$. As general empirical knowledge like "Large cars have high gas consumption" cannot be assumed to be shared by all relevant agents, sentences of this form concern only conceptual relationships.

The relevant empirical background knowledge is shown in the third box, which contains terminological sentences characterizing $\mathcal{S}$'s own general knowledge. The knowledge base contains a great deal of general background knowledge of this form which $\mathcal{S}$ possesses but cannot ascribe with confidence to $\mathcal{O}$.

Finally, the fourth box shows the relevant part of the large quantity

Find Relevant General Knowledge

**Input**

- A sentence of the form $\Box_{(\text{believe},\mathcal{O})}\Psi$, where $\Psi$ is a nonmodal assertional sentence.
- A knowledge base of sentences conforming to the syntax specified in the Appendix to this chapter.

**Output**

- A list $(L_1 \ldots L_n)$ of lists of terminological sentences, such that for each list $L_i$,
  - each sentence in $L_i$ is of the form $\Box_{(\text{believe},\mathcal{O})}\Phi$, where $\Phi$ is some nonmodal terminological sentence; and
  - if for each of the sentences in $L_i$ the corresponding sentence $\Box_{(\text{believe},\mathcal{S})}$ $\Box_{(\text{believe},\mathcal{O})}\Phi$ were added to the knowledge base, the sentence $\Box_{(\text{believe},\mathcal{S})}$ $\Box_{(\text{believe},\mathcal{O})}\Psi$ corresponding to the input sentence would be provable.

**Procedure**

1. Use MOTEL's Ask predicate to obtain the list $(\Pi_1 \ldots \Pi_m)$ of possible proofs of $\Box_{(\text{believe},\mathcal{S})}\Psi$, where $\Psi$ is the nonmodal assertional sentence of the input sentence.
   ▷ *$\mathcal{S}$ uses her own knowledge as a source of hypotheses about $\mathcal{O}$'s knowledge.*
2. **For** each $\Pi_i$ in $(\Pi_1 \ldots \Pi_m)$:
   - Given the first assertional sentence $\Box_{(\text{believe},\mathcal{S})}\Xi$ in $\Pi_i$, check whether $\Box_{(\text{believe},\mathcal{S})}\Box_{(\text{believe},\mathcal{O})}\Xi$ is present in the knowledge base; **if** not, terminate the processing of $\Pi_i$.
     ▷ *$\Xi$ is the piece of specific knowledge that $\mathcal{O}$ would need in order to start the chain of inferences represented by $\Pi_i$; thus if $\mathcal{O}$ does not possess it, it is irrelevant whether $\mathcal{O}$ possesses the general knowledge required for the proof.*
   - **For** each terminological sentence in $\Pi_i$ that has the form $\Box_{(\text{believe},\mathcal{S})}\Phi$, add the sentence $\Box_{(\text{believe},\mathcal{O})}\Phi$ to the list of sentences being constructed for $\Pi_i$.
     ▷ *The sentences in this list together represent the general knowledge whose possession would enable $\mathcal{O}$ to infer the conclusion. Sentences of the form $\Box^c_{(\text{believe},\text{all})}\Phi$ are discarded because it is already known that they can be ascribed to $\mathcal{O}$.*
   - Add the resulting list of sentences for $\Pi_i$ to the list of lists of sentences being collected for output.

FIGURE 9.7. Algorithm for determining the background facts on the basis of which a person might make a particular inference.

of specific knowledge that $\mathcal{S}$ has about $\mathcal{C}$, only a small part of which has been acquired by $\mathcal{O}$ during the dialog between $\mathcal{O}$ and $\mathcal{S}$.

### 9.4.2 Predicting Possible Inferences

One task that $\mathcal{S}$ may wish to perform is to predict whether $\mathcal{O}$ knows that $\mathcal{C}$ has high gas consumption. The first subtask is to determine how

| Proof 1: Because $\mathcal{C}$ has high consumption |
|---|
| $\Box_{(\text{believe},\mathcal{S})}\mathcal{C} \in \text{car\_with\_high\_consumption}$ |

| Proof 2: Because $\mathcal{C}$ is old | |
|---|---|
| | $\Box_{(\text{believe},\mathcal{S})}\mathcal{C} \in \text{old\_car}$ |
| $\cdot$ | $\Box_{(\text{believe},\mathcal{S})}\text{old\_car} \sqsubseteq \text{car\_with\_high\_consumption}$ |
| $\Rightarrow$ | $\Box_{(\text{believe},\mathcal{S})}\mathcal{C} \in \text{car\_with\_high\_consumption}$ |

| Proof 3: Because $\mathcal{C}$ is large | |
|---|---|
| | $\Box_{(\text{believe},\mathcal{S})}\mathcal{C} \in \text{large\_car}$ |
| $\cdot$ | $\Box_{(\text{believe},\mathcal{S})}\text{large\_car} \sqsubseteq \text{car\_with\_high\_consumption}$ |
| $\Rightarrow$ | $\Box_{(\text{believe},\mathcal{S})}\mathcal{C} \in \text{car\_with\_high\_consumption}$ |

| Proof 4: Because $\mathcal{C}$ has high horsepower and thus a powerful motor | |
|---|---|
| | $\Box_{(\text{believe},\mathcal{S})}\mathcal{C} \in \text{car\_with\_high\_horsepower}$ |
| $\cdot$ | $\Box_{(\text{believe},\mathcal{S})}\text{car\_with\_high\_horsepower} \sqsubseteq \text{car\_with\_powerful\_motor}$ |
| $\Rightarrow$ | $\Box_{(\text{believe},\mathcal{S})}\mathcal{C} \in \text{car\_with\_powerful\_motor}$ |
| $\cdot$ | $\Box^{c}_{(\text{believe},\text{all})}\text{car\_with\_powerful\_motor} = \text{powerful\_car}$ |
| $\Rightarrow$ | $\Box_{(\text{believe},\mathcal{S})}\mathcal{C} \in \text{powerful\_car}$ |
| $\cdot$ | $\Box_{(\text{believe},\mathcal{S})}\text{powerful\_car} \sqsubseteq \text{car\_with\_high\_consumption}$ |
| $\Rightarrow$ | $\Box_{(\text{believe},\mathcal{S})}\mathcal{C} \in \text{car\_with\_high\_consumption}$ |

FIGURE 9.8. Proofs yielded by MOTEL, given the knowledge base in Figure 9.6, for the sentence $\Box_{(\text{believe},\mathcal{S})}\mathcal{C} \in \text{car\_with\_high\_consumption}$.

$\mathcal{O}$ could use general background knowledge to infer this fact, given the specific knowledge $\mathcal{O}$ has about $\mathcal{C}$. The algorithm for this subtask is shown in Figure 9.7. Basically, it comprises two steps:

1. $\mathcal{S}$ tries to infer the conclusion in as many ways as possible, using knowledge available to *her*, including generally shared knowledge. This step is motivated by the assumption that the ways in which $\mathcal{S}$ could deduce the conclusion constitute a superset of the ways in which $\mathcal{O}$ could deduce it, as $\mathcal{S}$'s relevant (specific and general) knowledge constitutes a superset of that of $\mathcal{O}$. This step is accomplished using the deduction mechanism of MOTEL, which returns the four proofs shown in Figure 9.8.

| Terminological sentence derived from Proof 3 |
|---|
| $\Box_{(\mathrm{believe},\mathcal{O})}\mathrm{large\_car} \sqsubseteq \mathrm{car\_with\_high\_consumption}$ |

| Terminological sentences derived from Proof 4 |
|---|
| $\Box_{(\mathrm{believe},\mathcal{O})}\mathrm{car\_with\_high\_horsepower} \sqsubseteq \mathrm{car\_with\_powerful\_motor}$ |
| $\Box_{(\mathrm{believe},\mathcal{O})}\mathrm{powerful\_car} \sqsubseteq \mathrm{car\_with\_high\_consumption}$ |

FIGURE 9.9. Lists of sentences yielded by the algorithm Find Relevant General Knowledge given the sentence $\Box_{(\mathrm{believe},\mathcal{O})}\mathcal{C} \in \mathrm{car\_with\_high\_consumption}$ and the knowledge base shown in Figure 9.6.

2. After eliminating the proofs which $\mathcal{O}$ could not have derived for lack of the necessary specific knowledge (here, Proofs 1 and 2), $\mathcal{S}$ reduces each proof to a list of ascriptions of necessary background knowledge (Figure 9.9).

$\mathcal{S}$'s second subtask is to construct a Bayesian network for making inferences about $\mathcal{O}$'s possession of the background knowledge yielded by the first step. The resulting network for the present example is shown in Figure 9.10. The construction begins with the application of the algorithm Create Disjunctive Knowledge Node (Figure 9.11) to the result of the first subtask; this algorithm creates the node labeled in Figure 9.10 as INFERENCE BY $\mathcal{O}$ OF "$\mathcal{C} \in$ CAR_WITH_HIGH_CONSUMPTION", which represents $\mathcal{S}$'s belief that $\mathcal{O}$ could infer $\mathcal{C}$'s high consumption using one of the sets of background facts shown in Figure 9.9. It also in turn invokes the algorithm Create Conjunctive Knowledge Node (Figure 9.12) to create this node's parents, KNOWLEDGE BY $\mathcal{O}$ OF "LARGE_CAR $\sqsubseteq$ CAR_WITH_HIGH_CONSUMPTION" and INFERENCE BY $\mathcal{O}$ OF "CAR_WITH_HIGH_HORSEPOWER $\sqsubseteq$ CAR_WITH_HIGH_CONSUMPTION", each of which corresponds to one of the two sets of sentences shown in Figure 9.9. Create Conjunctive Knowledge Node in turn invokes the previously illustrated algorithm Create Simple Knowledge Node (Figure 9.3), to construct the three nodes with labels of the form KNOWLEDGE BY $\mathcal{O}$ OF $\Phi$. When creating these nodes, Create Simple Knowledge Node finds that the parent nodes representing the difficulty of the facts $\Phi$ already exist as the result of previous inferences (though the other nodes linked to these nodes are not shown in Figure 9.10). The parent node KNOWLEDGEABILITY OF $\mathcal{O}$ ABOUT CARS, on the other hand, is created during the first call to Create Simple Knowledge Node, which initializes it with a uniform distribution.

Given the network constructed in this way, $\mathcal{S}$ can derive a prediction for INFERENCE BY $\mathcal{O}$ OF "$\mathcal{C}$ ∈ CAR_WITH_HIGH_CONSUMPTION" through downward propagation, as illustrated by the first histogram for each node.[12] The prediction assigns a fairly high probability to the value KNOWN for INFERENCE BY $\mathcal{O}$ OF "$\mathcal{C}$ ∈ CAR_WITH_HIGH_CONSUMPTION", which makes sense in that $\mathcal{O}$ might infer the high gas consumption in two different ways.

### 9.4.3 Interpreting the Observed Result of an Inference

A second task that $\mathcal{S}$ can perform using a network constructed in this way is to interpret an utterance by $\mathcal{O}$ expressing a piece of specific knowledge that $\mathcal{O}$ has not been told directly. For example, if $\mathcal{O}$ explicitly expresses the knowledge that $\mathcal{C}$ has high gas consumption, upward propagation is triggered, as shown in the second histogram for each node in Figure 9.10 (cf. the previous examples of upward propagation in Figures 9.4 and 9.5). This upward propagation makes $\mathcal{S}$ almost certain about KNOWLEDGE BY $\mathcal{O}$ OF "LARGE_CAR ⊑ CAR_WITH_HIGH_CONSUMPTION", whereas $\mathcal{S}$'s belief about INFERENCE BY $\mathcal{O}$ OF "CAR_WITH_HIGH_HORSEPOWER ⊑ CAR_WITH_HIGH_CONSUMPTION" changes only minimally. Informally speaking, $\mathcal{S}$ has explained $\mathcal{O}$'s inference mainly by postulating that $\mathcal{O}$ knows that large cars have high consumption; and indeed this explanation is more likely than the alternative explanation, which involves ascribing two pieces of background knowledge, one of which $\mathcal{S}$ judges to be relatively difficult.

Some of the beliefs of $\mathcal{S}$ that are revised through the upward propagation belong to nodes that $\mathcal{S}$ may be able to use later. Most directly, $\mathcal{S}$ revises her beliefs associated with the three nodes with labels of the form KNOWLEDGE BY $\mathcal{O}$ OF $\Phi$. These revised beliefs will be useful if $\mathcal{S}$ later has to make an inference about $\mathcal{O}$'s knowledge in another case where one or more of these background facts is involved. The nodes shown in Figure 9.10 will continue to be used for these new inferences, as the system only constructs a new node if a node representing the variable of interest does not already exist (cf. the first step in the procedure for each of the algorithms in Figures 9.3, 9.11, and 9.12).

[12] The propagation process here is more complex than for the previous examples, because this network is not singly connected: KNOWLEDGEABILITY OF $\mathcal{O}$ ABOUT CARS influences the prediction for INFERENCE BY $\mathcal{O}$ OF "$\mathcal{C}$ ∈ CAR_WITH_HIGH_CONSUMPTION" indirectly along three paths. This complication is handled using the method of *conditionalization* (see [Pearl, 1988, 4.4.2]).

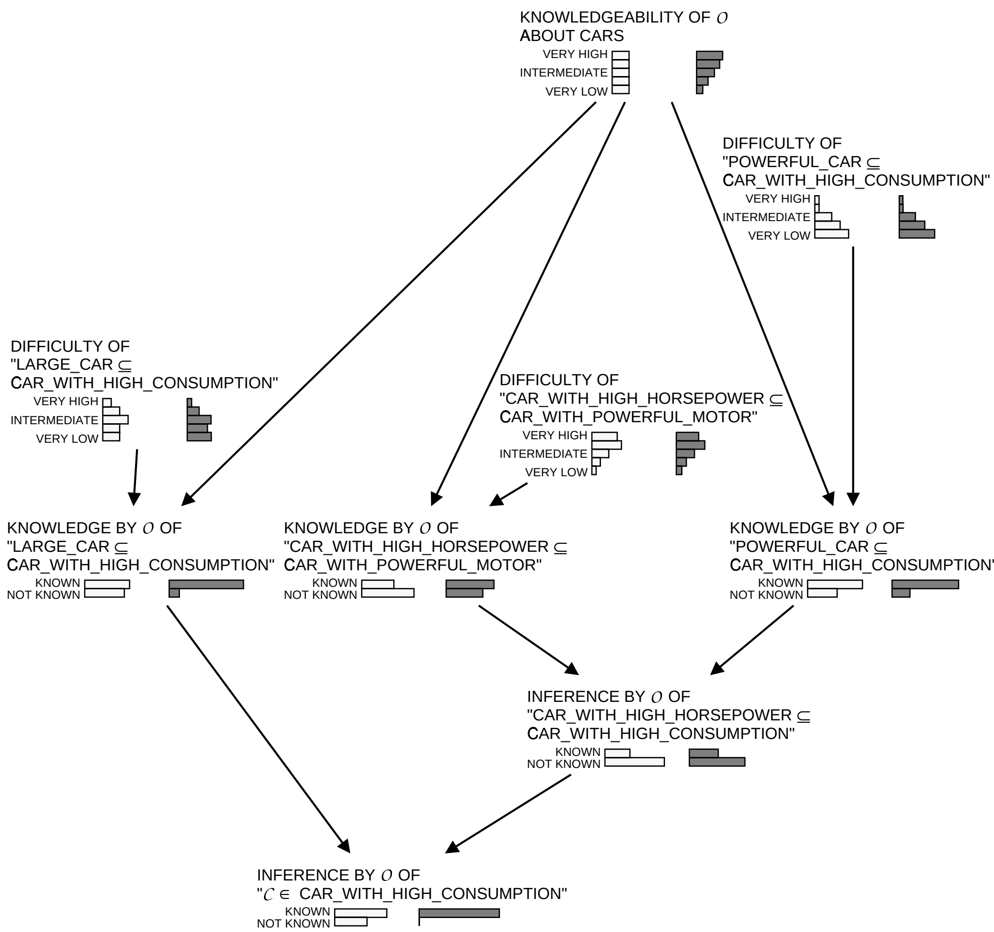


FIGURE 9.10. Bayesian network constructed by the algorithm Create Disjunctive Knowledge Node on the basis of the lists of sentences shown in Figure 9.9. (The second histogram for each node shows $\mathcal{S}$'s belief after $\mathcal{S}$'s observation that $\mathcal{O}$ has inferred that $\mathcal{C}$ has high gas consumption.)

Less directly, $\mathcal{S}$ can revise her belief about KNOWLEDGEABILITY OF $\mathcal{O}$ ABOUT CARS. Even a small revision of this belief can be useful, because this node plays a role in many inferences made by $\mathcal{S}$.

Finally, note that $\mathcal{S}$'s beliefs about the difficulties of the three pieces of general background knowledge have also been revised slightly (this is most visible for DIFFICULTY OF "LARGE_CAR ⊑ CAR_WITH_HIGH_CONSUMPTION").

Create Disjunctive Knowledge Node

**Input**

- A list $(L_1 \ldots L_n)$ of lists of terminological sentences, each list having the form specified for the input to Create Conjunctive Knowledge Node (Figure 9.12).

**Output**

- A Bayesian network node, linked to its parent nodes, representing $\mathcal{S}$'s belief about the disjunction of the conjunctions of the sentences in the lists $(L_1 \ldots L_n)$.
  ▹ *The probability associated with the value* KNOWN *for this node represents the strength of* $\mathcal{S}$*'s belief that* $\mathcal{O}$ *possesses all of the general knowledge ascribed in at least one of these lists of input sentences.*

**Procedure**

1. **If** the required node has already been created, **then** output it and **stop**.
2. Create (or retrieve) the node $N_X$ corresponding to the last list $L_n$, using Create Conjunctive Knowledge Node.
3. **If** $n = 1$ (i.e., there is only one list in the input list of lists), **then** output $N_X$ and **stop**.
4. Recursively create (or retrieve) the node $N_Y$ corresponding to the list $(L_1 \ldots L_{n-1})$.
5. Create the node $N_Z$ representing $\mathcal{S}$'s belief about the variable $Z$, which has the value KNOWN if the disjunction of the conjunctions of the sentences in the lists $(L_1 \ldots L_n)$ is true and NOT KNOWN otherwise.
6. Link $N_Z$ to its parents $N_X$ and $N_Y$, defining the matrix of conditional probabilities as follows:

$$P(Z = \text{NOT KNOWN}|x, y) = \begin{cases} 1, & \text{if } x = y = \text{NOT KNOWN}; \\ 0, & \text{otherwise}; \end{cases}$$

$$P(Z = \text{KNOWN}|x, y) = 1 - P(Z = \text{NOT KNOWN}|x, y).$$

7. Output $N_Z$.

FIGURE 9.11. Algorithm for creating a Bayesian network node for predicting or interpreting an inference made by another person.

$\mathcal{S}$'s new beliefs about these variables will affect her inferences about the beliefs of other persons than $\mathcal{O}$ in cases where these pieces of background knowledge are involved. The extent of these revisions is, incidentally, much smaller than the revision in Figure 9.5 of the belief for DIFFICULTY OF "BMW $\sqsubseteq$ GERMAN_CAR", because in Figure 9.10 a larger number of other variables is involved, about which $\mathcal{S}$ has beliefs which are less definite.

### 9.4.4 Discussion

The approach just described has a certain modularity and simplicity in that the two contrasting types of representation and inference—epistemic

### Create Conjunctive Knowledge Node

**Input**

- A list $(\Psi_1 \ldots \Psi_n)$ of sentences, each of the form $\Box_{(\text{believe},\mathcal{O})}\Phi$, where $\Phi$ is a non-modal terminological sentence.

**Output**

- A Bayesian network node, linked to its parent nodes, representing $\mathcal{S}$'s belief about the conjunction of these sentences.
  ▹ *The probability associated with the value* KNOWN *for this node represents the strength of $\mathcal{S}$'s belief that $\mathcal{O}$ possesses all of the general knowledge represented by these input sentences.*

**Procedure**

1. **If** the required node has already been created, **then** output it and **stop**.
2. Create (or retrieve) the node $N_X$ corresponding to the last sentence $\Psi_n$, using Create Simple Knowledge Node (Figure 9.3).
3. **If** $n = 1$ (i.e., there is only one sentence in the input list), then output $N_X$ and **stop**.
4. Recursively create (or retrieve) the node $N_Y$ corresponding to the list $(\Psi_1 \ldots \Psi_{n-1})$.
5. Create the node $N_Z$ representing $\mathcal{S}$'s belief about the variable $Z$, which has the value KNOWN if the conjunction of the sentences in $(\Psi_1 \ldots \Psi_n)$ is true and NOT KNOWN otherwise.
6. Link $N_Z$ to its parents $N_X$ and $N_Y$, defining the matrix of conditional probabilities as follows:

$$P(Z = \text{KNOWN}|x,y) = \begin{cases} 1, & \text{if } x = y = \text{KNOWN}; \\ 0, & \text{otherwise}; \end{cases}$$

$$P(Z = \text{NOT KNOWN}|x,y) = 1 - P(Z = \text{KNOWN}|x,y).$$

7. Output $N_Z$.

FIGURE 9.12. Algorithm for creating a Bayesian network node corresponding to the hypothesis that a person knows a given set of background facts.

logic and Bayesian networks—are applied in sequence. The approach does not require changes to any aspect of MOTEL, whereas the changes to MOTEL would presumably be drastic if the system were to interleave its probabilistic reasoning with its deductions within the modal logic framework. This modular approach lends itself to applications with other implementations of epistemic logics as well.

The modularity could, however, in some cases lead to counterintuitive and inefficient performance. Suppose, for example, that $\mathcal{S}$ had so much relevant general knowledge that she could deduce a given fact in a great many ways. Then MOTEL would construct a large number of proofs, many of which might make a negligible contribution to the resulting Bayesian network, because the probability of $\mathcal{O}$ possessing the general knowledge involved was very low.

A second limitation is that this approach presupposes that all of $\mathcal{O}$'s relevant knowledge is possessed by $\mathcal{S}$ as well: $\mathcal{S}$ essentially generates hypotheses about inferences $\mathcal{O}$ might make by trying to make relevant inferences herself. But of course $\mathcal{O}$ may use beliefs which are false, or whose truth value is unknown to $\mathcal{S}$, and $\mathcal{S}$ should be able to reason about such cases. A possible extension to cover such cases would require that the algorithm Find Relevant General Knowledge construct its proofs using not $\mathcal{S}$'s own knowledge but rather the beliefs of a hypothetical agent who possesses a larger set of beliefs which $\mathcal{O}$ might conceivably possess.

Approaches generally analogous to the one proposed here have been used in the realm of student modeling. In that context, the task is typically to explain why the student has answered (or failed to answer) a particular item correctly, or has given an incorrect answer. An explanation may take the form of a set of steps, rules, or facts, analogous to the proofs shown in Figure 9.8. Possible incorrect beliefs that might be possessed by the student are sometimes represented in *bug catalogs*. Recently, student modeling researchers have begun using Bayesian networks to evaluate competing explanations of this sort (see, e.g., [Martin and VanLehn, 1993; Mislevy, 1994]).

The idea of dynamically constructing Bayesian networks on the basis of proofs generated within a logic-based system has recently been explored independently by [Haddawy, 1994], who is interested in the general problem of how to avoid having to code large Bayesian networks directly. But there is a fundamental difference between the two approaches: Note that in the approach presented here, the proofs are derived when $\mathcal{S}$ is simply reasoning on the domain level about the car in question; probabilities are only introduced when reasoning on the meta-level begins, and they are not represented within the logical formalism. By contrast, the approach of [Haddawy, 1994] does not involve reasoning about the beliefs of other persons, and accordingly there is no distinction between domain-level and meta-level reasoning. As the logical and the probabilistic inferences concern the same content, the logical formalism has to express probability information; [Haddawy, 1994] uses probability logic.

In sum, the integration of epistemic logic and probabilistic reasoning presented here represents only one of many ways in which the strengths of these two contrasting types of reasoning can be combined. Because of the recognizable limitations of the present version of this approach, it is perhaps best viewed as an example designed to stimulate further developments. The main lesson of this proposal is that the concepts and techniques of intuitive psychometrics and Bayesian networks can usefully be combined with those of epistemic logic to yield differentiated and realistic solutions to the problem of uncertainty about background knowledge.

## 9.5 References

[Ballim and Wilks, 1991] Afzal Ballim and Yorick Wilks. *Artificial Believers: The Ascription of Belief.* Erlbaum, Hillsdale, NJ, 1991.

[Clark and Marshall, 1981] Herbert H. Clark and Catherine R. Marshall. Definite reference and mutual knowledge. In Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag, editors, *Elements of Discourse Understanding*, pages 10–63. Cambridge University Press, Cambridge, UK, 1981.

[Dawes, 1989] Robyn M. Dawes. Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology*, 25:1–17, 1989.

[Guha and Lenat, 1994] R. V. Guha and Douglas B. Lenat. Enabling agents to work together. *Communications of the ACM*, 37(7):127–142, 1994.

[Haddawy, 1994] Peter Haddawy. Generating Bayesian networks from probability logic knowledge bases. In Ramon Lopez de Mantaras and David Poole, editors, *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, pages 262–269, Seattle, 1994. San Francisco: Morgan Kaufmann.

[Hambleton and Swaminathan, 1985] R. K. Hambleton and H. Swaminathan. *Item Response Theory: Principles and Applications.* Kluwer-Nijhoff, Boston, 1985.

[Hoch, 1987] Stephen J. Hoch. Perceived consensus and predictive accuracy: The pros and cons of projection. *Journal of Personality and Social Psychology*, 53:221–234, 1987.

[Hustadt and Nonnengart, 1993] Ullrich Hustadt and Andreas Nonnengart. Modalities in knowledge representation. In *Proceedings of the Sixth Australian Joint Conference on Artificial Intelligence*, pages 249–254, Sydney, 1993. Singapore: World Scientific Publishing.

[Hustadt, 1995] Ullrich Hustadt. Introducing epistemic operators into a description logic. In Armin Laux and Heinrich Wansing, editors, *Knowledge and Belief in Philosophy and Artificial Intelligence*. Akademie-Verlag, Berlin, 1995.

[Jameson *et al.*, 1993] Anthony Jameson, Thomas O. Nelson, R. Jacob Leonesio, and Louis Narens. The feeling of another person's knowing. *Journal of Memory and Language*, 32:320–335, 1993.

[Jameson *et al.*, 1994] Anthony Jameson, Bernhard Kipper, Alassane Ndiaye, Ralph Schäfer, Joep Simons, Thomas Weis, and Detlev Zimmermann. Cooperating to be noncooperative: The dialog system PRACMA. In Bernhard Nebel and Leonie Dreschler-Fischer, editors, *Proceedings of the Eighteenth Annual German Conference on Artificial Intelligence*, pages 106–117, Saarbrücken, 1994. Berlin: Springer.

[Jameson *et al.*, 1995] Anthony Jameson, Ralph Schäfer, Joep Simons, and Thomas Weis. Adaptive provision of evaluation-oriented information: Tasks and techniques. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, Montreal, 1995. San Mateo, CA: Morgan Kaufmann.

[Jameson, 1990] Anthony Jameson. *Knowing What Others Know: Studies in Intuitive Psychometrics*. PhD thesis, University of Amsterdam, Netherlands, 1990. Available from the author.

[Jameson, 1992] Anthony Jameson. Generalizing the double-stereotype approach: A psychological perspective. In Elisabeth André, Robin Cohen, Winfried Graf, Bob Kass, Cécile Paris, and Wolfgang Wahlster, editors, *Proceedings of the Third International Workshop on User Modeling*, pages 69–83, Dagstuhl, Germany, 1992. Available as Report

DFKI-D-92-17 of the German Research Center for Artificial Intelligence, Saarbrücken.

[Kobsa and Pohl, 1995] Alfred Kobsa and Wolfgang Pohl. The user modeling shell system BGP-MS. *User Modeling and User Adapted Interaction*, 4(2):59–106, 1995.

[Koriat, 1993] Asher Koriat. How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, 100:609–639, 1993.

[Martin and VanLehn, 1993] Joel D. Martin and Kurt VanLehn. OLAE: Progress toward a multi-activity, Bayesian student modeler. In Paul Brna, Stellan Ohlsson, and Helen Pain, editors, *Artificial Intelligence in Education: Proceedings of AI-ED 93*, pages 410–417, Edinburgh, 1993. Charlottesville VA: Association for the Advancement of Computing in Education.

[Mislevy, 1994] Robert J. Mislevy. Evidence and inference in educational assessment. *Psychometrika*, 59:439–483, 1994.

[Moore, 1985] Robert C. Moore. Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25:75–94, 1985.

[Neapolitan, 1990] Richard E. Neapolitan. *Probabilistic Reasoning in Expert Systems: Theory and Algorithms.* Wiley, New York, 1990.

[Nelson and Narens, 1980] Thomas O. Nelson and Louis Narens. Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior*, 19:338–368, 1980.

[Nickerson *et al.*, 1987] Raymond S. Nickerson, Alan Baddeley, and Barbara Freeman. Are people's estimates of what other people know influenced by what they themselves know? *Acta Psychologica*, 64:245–259, 1987.

[Pearl, 1988] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Mateo, CA, 1988. Revised second printing: 1991.

[Ross *et al.*, 1977] Lee Ross, David Greene, and Pamela House. The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13:279–301, 1977.

[Schäfer, 1994] Ralph Schäfer. Multidimensional probabilistic assessment of interest and knowledge in a noncooperative dialog situation. In Christoph G. Thomas, editor, *Proceedings of ABIS-94: GI Workshop on Adaptivity and User Modeling in Interactive Software Systems*, pages 46–62, Sankt Augustin, Germany, 1994.

## 9.6 Appendix: Syntax and Semantics of the Subset of $\mathcal{FALCM}$ Used in Pracma

The syntax and semantics of the complete language $\mathcal{FALCM}$ are specified in [Hustadt, 1995] in this volume. At the time of this writing, PRACMA's integration of epistemic logic with Bayesian networks employs the subset of the knowledge representation system MOTEL ([Hustadt and Nonnengart, 1993]) that corresponds to the following subset of $\mathcal{FALCM}$:

Nonmodal terminological sentences have either the form $D \sqsubseteq E$ or the form $D = E$, where $D$ and $E$ are concept symbols.

Nonmodal assertional sentences have the form $a \in D$, where $a$ is an object symbol and $D$ is a concept symbol.

A modal sentence has one of the following forms:

$$\Box_{(\text{believe},a)}\Phi,$$
$$\Box_{(\text{believe},a)}\Box_{(\text{believe},b)}\Phi,$$
$$\Box^{c}_{(\text{believe},\text{all})}\Phi,$$

where $\Phi$ is a nonmodal (terminological or assertional) sentence and $a$ and $b$ are agent symbols. A modal sentence will be called *terminological* or *assertional*, depending on whether the nonmodal sentence $\Phi$ it contains is terminological or assertional.

MOTEL provides a predicate Ask which, given an assertional sentence $\Phi$, determines whether $\Phi$ is provable given the current knowledge base; if $\Phi$ is provable, Ask returns (if so requested) a list of all possible *proofs* of $\Phi$.

Within the subset of $\mathcal{FALCM}$ used here, a proof has one of the following two forms:

$$\Psi,$$
$$\Pi \cdot \Phi \Rightarrow \Psi,$$

where $\Psi$ is an assertional sentence, $\Phi$ is a terminological sentence, and $\Pi$ is itself a proof (having one of these two forms). A proof of the first form states that $\Psi$ is present in the knowledge base. A proof of the second form states that $\Psi$ is implied by $\Psi' \wedge \Phi$, where $\Psi'$ is the last assertional sentence contained in $\Pi$.