Chapter 3

Incompleteness and Uncertainty

What we know is not much. What we do not know is immense.[1]

Marquis Simon de Laplace

The goal of this chapter is twofold: (i) to present the concept of incompleteness and (ii) to demonstrate how incompleteness is a source of uncertainty.

3.1 Observing a water treatment unit

The uncertainty of a phenomenon has two causes: (i) inaccuracies in the model and (ii) ignored variables. In this section, we demonstrate this fact by a second experiment: the water treatment unit. This experiment consists of two stages. In the first stage, we describe the complete model of the water treatment unit, giving all the variables and functions involved in the model. In the

[1] "Ce que nous connaissons est peu de chose, ce que nous ignorons est immense." Reported as his nearly last words by Joseph Fourier in his "Historical praise for M. le Marquis de Laplace" in front of the French Royal Academy of Science in 1829 [Fourier, 1829]. The complete citation is the following: "Les personnes qui ont assisté à ses derniers instants lui rappelaient les titres de sa gloire, et ses plus éclatantes découvertes. Il répondit: "Ce que nous connaissons est peu de chose, ce que nous ignorons est immense." C'est du moins, autant qu'on l'a pu saisir, le sens de ses dernières paroles à peine articulées. Au reste, nous l'avons entendu souvent exprimer cette pensée, et presque dans les mêmes termes. Il s'éteignit sans douleur."

second stage, we pretend that some of the variables and functions of the model are not available. In other words, we generate a synthetic incompleteness of our model. The goal is to show the consequences of this incompleteness and to present a first step toward Bayesian modeling.

3.1.1 The elementary water treatment unit

We now describe the complete model of the water treatment unit. Figure 3.1 is a schematic representation.

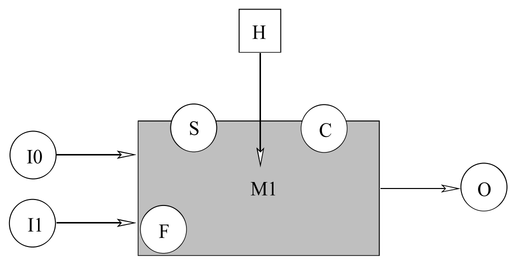

FIGURE 3.1: The treatment unit receives two water streams of quality I_0 and I_1 and generates an output stream of quality O. The resulting quality depends on I_0, I_1, two unknown variables H and F, and a control variable C. An operator regulates C, while the value of F is estimated by a sensor variable S.

The unit takes two water streams as inputs with respective water qualities I_0 and I_1. Two different streams are used because partly purified water is recycled to dilute the more polluted stream, to facilitate its decontamination.

The unit produces an output stream of quality O.

The internal functioning state of the water treatment unit is described by the variable F. This variable F quantifies the efficiency of the unit but is not directly measurable. For instance, as the sandboxes become more loaded with contaminants the purification becomes less and less efficient and the value of F becomes lower and lower.

A sensor S helps to estimate the efficiency F of the unit.

A controller C is used to regulate and optimize O, the quality of the water in the output stream.

Finally, some external factor H may disturb the operation of the unit. For instance, this external factor could be the temperature or humidity of the air.

For didactic purposes, we consider that these seven variables may each take 11 different integer values ranging from 0 to 10. The value 0 is the worst value for I_0, I_1, F, and O, and 10 is the best.

When all variables have their nominal values, the ideal quality Q of the output stream is given by the equation:

$$Q = Int\left(\frac{I_0 + I_1 + F}{3}\right) \tag{3.1}$$

Where $Int\,(x)$ is the integer part of x.

The value of Q never exceeds the value O^*, reached when the unit is in perfect condition, with:

$$O^* = Int\left(\frac{I_0 + I_1 + 10}{3}\right) \tag{3.2}$$

The external factor H may reduce the ideal quality Q and the control C may try to compensate for this disturbance or the bad condition of the treatment unit because of F. Consequently, the output quality O is obtained according to the following equations:

$$\alpha = Int\left(\frac{I_0 + I_1 + F + C - H}{3}\right) \tag{3.3}$$

$$O = \begin{cases} \alpha & \text{if } (0 \leq \alpha \leq O^*) \\ (2O^* - \alpha) & \text{if } (\alpha \geq O^*) \\ 0 & \text{Otherwise} \end{cases} \tag{3.4}$$

We consider the example of a unit directly connected to the sewer: $[I_0 = 2]$, $[I_1 = 8]$.

When $[C = 0]$ (no control) and $[H = 0]$ (no disturbance), Figure 3.2 gives the value of the quality O according to F, $(O^* = 6)$.

When the state of operation is not optimal (F different from 10), it is possible to compensate using C. However, if we over-control, then it may happen that the output deteriorates. For instance, if $[I_0 = 2]$, $[I_1 = 8]$, $[F = 8]$, $[H = 0]$, the outputs obtained for the different values of C are shown in Figure 3.3.

The operation of the unit may be degraded by H. For instance, if $[I_0 = 2]$, $[I_1 = 8]$, $[F = 8]$,$[C = 0]$, the output obtained for the different values of H are shown in Figure 3.4.

Finally, the value of the sensor S depends on I_0 and F as follows:

$$S = Int\left(\frac{I_0 + F}{2}\right) \tag{3.5}$$

The outputs of S in the 121 possible situations for I_0 and F are shown in Figure 3.5. Note that, if we know I_0, I_1, F, H, and C, we know with certainty the values of both S and O. At this stage, our water treatment unit is a completely deterministic process. Consequently, a complete model can be constructed. Now consider what happens if we ignore the exact equations that rule the water treatment unit and, of course, the existence of the external factor H. The starting point for constructing our own model is limited to

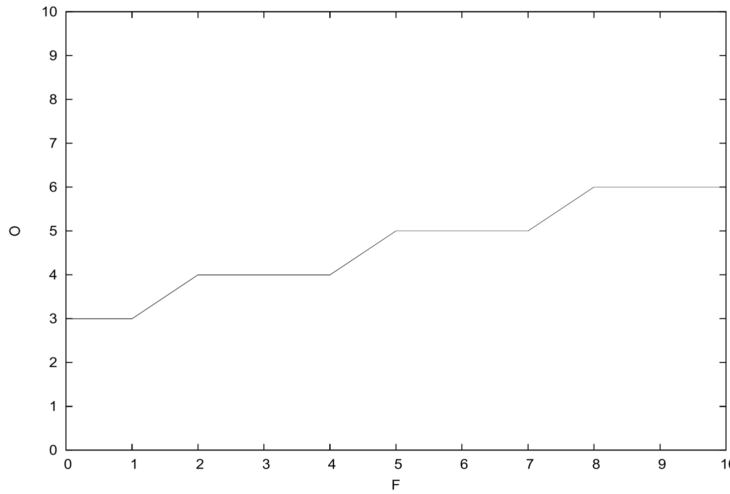

FIGURE 3.2: The output O as a function of the functioning state F with inputs, control, and external factor fixed to: $[I_0 = 2] \wedge [I_1 = 8] \wedge [C = 0] \wedge [H = 0]$.

knowing the existence of the variables I_0, I_1, F, S, C, and O, and that the value of O depends on $I_0 \wedge I_1 \wedge S \wedge C$ and that of S depends on $I_0 \wedge F$.

What do we need to do now? Observe the behavior of the water treatment unit, in particular the quality of the output stream O, for different values of $I_0 \wedge I_1 \wedge S \wedge C$, as well as the sensor value S for different values of I_0 (remember that F cannot be observed). During these observations you will note that there are different situations in which uncertainty appears. The goal of the following section is to discuss this uncertainty.

3.1.2 Experimentation and uncertainty

3.1.2.1 Uncertainty on O because of inaccuracy of the sensor S

A given value of S corresponds to several possible values of I_0 and F. For instance, seven pairs of values of $I_0 \wedge F$ correspond to $[S = 1]$ in Figure 3.5. Worse than this, even knowing I_0 and S, two values of F are possible most of the time (see Figure 3.6). This fact will introduce some "noise" in the prediction of O.

To illustrate this effect let us first experiment with $H = 0$: the operation of the water treatment unit is not disturbed. For $[I_0 = 2]$, $[I_1 = 8]$, $[C = 2]$, we can explore the different possible values of the output O when F varies. However, as F is not directly observable, we can only collect data concerning S and O. These data are presented on Figure 3.7.

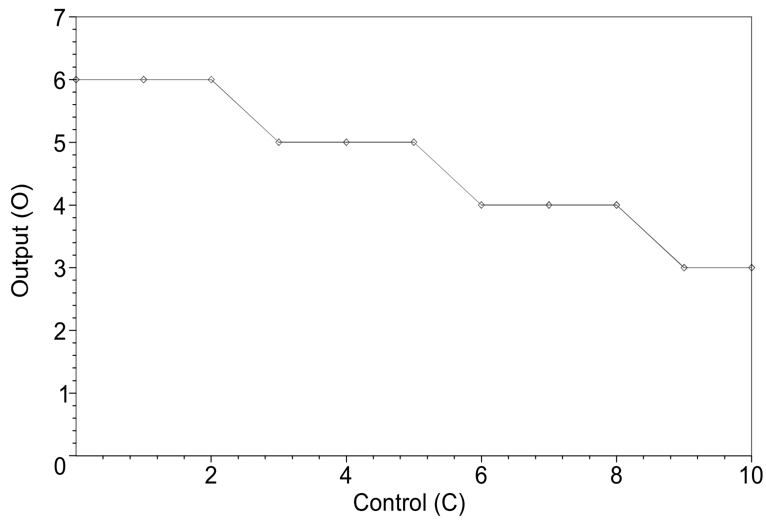

FIGURE 3.3: The output O as a function of control C with inputs, functioning state, and external factor, fixed to: $[I_0 = 2]$, $[I_1 = 8]$, $[F = 8]$, $[H = 0]$.

For some S it is possible to predict exactly the output O:

1. $[S = 1] \Rightarrow [O = 4]$
2. $[S = 3] \Rightarrow [O = 5]$
3. $[S = 4] \Rightarrow [O = 6]$
4. $[S = 6] \Rightarrow [O = 5]$

For some other values of S it is not possible to predict the output O with certainty:

1. If $[S = 2]$, then O may take the value either four or five, with a slightly higher probability for four. Indeed, when $[S = 2]$, then F may be either two or three (see Figure 3.6) and, O will, respectively, be either four or five.

2. If $[S = 5]$, then O may take the value either five or six, with a slightly lower probability for five. When $[S = 5]$, F may be either eight or nine.

3.1.2.2 Uncertainty because of the hidden variable H

Let us now do the same experiment for a disturbed process (value of H drawn at random from the 11 possible values). Of course, we obtain different

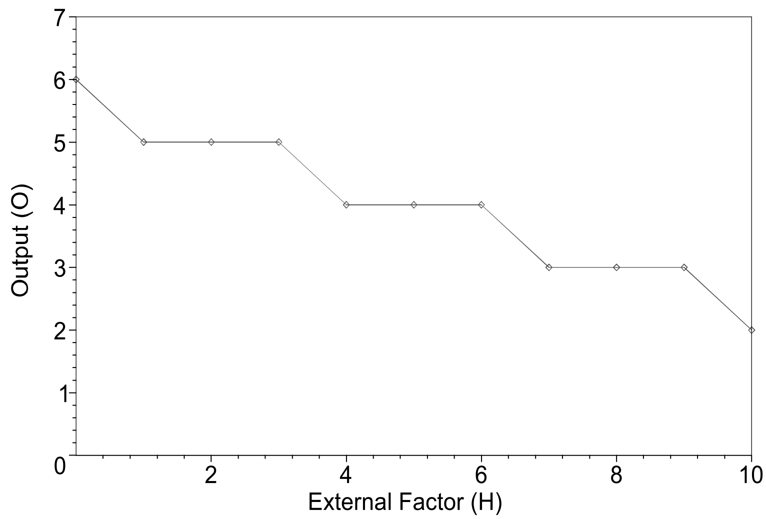

FIGURE 3.4: The output O as a function the external factor H with inputs, functioning state, and control fixed to: $[I_0 = 2]$, $[I_1 = 8]$, $[F = 8]$, $[C = 0]$.

results with more uncertainty due to the effect on the output of the hidden variable H. The obtained data when $[I_0 = 2]$, $[I_1 = 8]$, $[C = 2]$ is presented on Figure 3.8.

In contrast with our previous experiment, this time no value of S is sufficient to infer the value of O exactly.

The dispersion of the observations is the direct translation of the effect of H. Taking into account the effect of hidden variables such as H and even measuring their importance is one of the major challenges that Bayesian Programming must face. This is not an easy task when you are not even aware of the nature and number of these hidden variables!

3.2 Lessons, comments, and notes

3.2.1 The effect of incompleteness

We assume that any model of a "real" (i.e., not formal) phenomenon is incomplete. There are always some hidden variables, not taken into account in the model, that influence the phenomenon. Furthermore, this incompleteness is irreducible: for any physical phenomenon, there is no way to build an exact

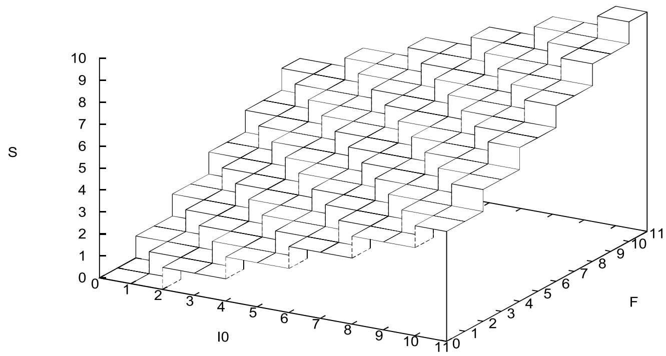

FIGURE 3.5: The sensor S as a function of input I_0 and functioning state F.

model with no hidden variables.[2] The effect of these hidden variables is that the model and the phenomenon never have exactly reproducible behavior. Uncertainty appears as a direct consequence of this incompleteness. Indeed, the model may not completely take into account the data and may not predict exactly the behavior of the phenomenon.[3] For instance, in the above example, the influence of the hidden variable H makes it impossible to predict with certainty the output O given the inputs I_0 and I_1, the reading of the sensor S, and the control C.

3.2.2 The effect of inaccuracy

The above example also demonstrates that there is another source of uncertainty: the *inaccuracy* of the sensors.

By inaccuracy, we mean that a given sensor may read the same value for different underlying situations. Here, the same reading on S may correspond to different values of F.

F is not a hidden variable as it is taken into account by the model. However, F cannot be measured directly and exactly. The values of F can only be inferred indirectly through the sensor S and they cannot be inferred with certainty. It may be seen as a weak version of incompleteness, where a variable

[2]See FAQ/FAM, Section 16.10 "Incompleteness irreducibility" for further discussion of that matter.

[3]See FAQ/FAM, Section 16.12 "Noise or ignorance?" for more information on this subject.

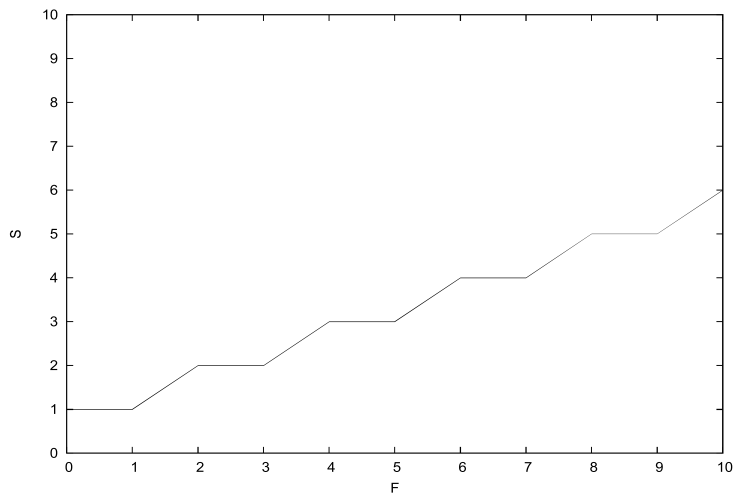

FIGURE 3.6: The sensor reading S as a function of the functioning state F when the input $[I_0 = 2]$.

is not completely hidden but is only partially known and accessible. Even though it is weak, this incompleteness still generates uncertainty.

3.2.3 Not taking into account the effect of ignored variables may lead to wrong decisions

Once the effects of the irreducible incompleteness of models are recognized, a programmer must deal with them either ignoring them and using the incomplete models, or trying to take incompleteness into account using a probabilistic model.

Using a probabilistic model clearly appears to be the better choice as it will always lead to better decisions, based on more information than the nonprobabilistic one.

For instance, a nonprobabilistic model of our production unit, not taking into account the variable H, would be, for instance[4]:

$$\alpha = Int\left(\frac{I_0 + I_1 + F + C}{3}\right) \tag{3.6}$$

$$O = \begin{cases} \alpha & \text{if } (0 \leq \alpha \leq O^*) \\ (2O^* - \alpha) & \text{if } (\alpha \geq O^*) \\ 0 & \text{Otherwise} \end{cases} \tag{3.7}$$

[4]Note the absence of H.

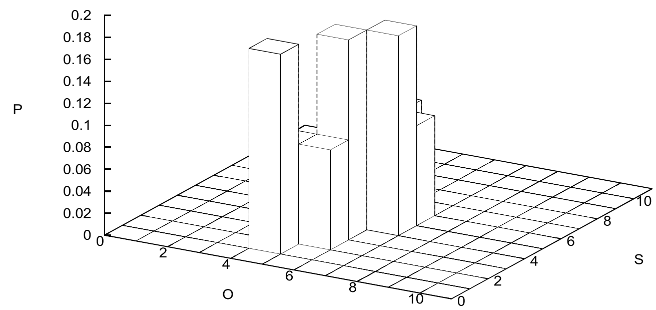

FIGURE 3.7: The histogram of the observed sensor state S and the output O when the inputs, the control, and the external factor are fixed to $[I_0 = 2]$, $[I_1 = 8]$, $[C = 2]$, $[H = 0]$, and the internal function F is generated randomly with a uniform distribution.

$$S = Int\left(\frac{I_0 + F}{2}\right) \tag{3.8}$$

It would lead to false predictions of the output O and, consequently, to wrong control decision on C to optimize this output.

For instance, scanning the 11 different possible values for C when $[I_0 = 2]$, $[I_1 = 8]$, $[F = 8]$ and consequently $[S = 5]$, the above model predicts that indifferently for $[C = 0]$, $[C = 1]$, and $[C = 2]$, O will take its optimal value: six (see Figure 3.3).

The observations depict a somewhat different and more complicated "reality" as shown in Figure 3.9. The choice of C to optimize O is now more complicated but also more informed. The adequate choice of C to produce the optimal output $[O = 6]$ is now, with nearly equivalent probabilities, to select a value of C greater than or equal to two. Indeed, this is a completely different choice from when the "exact" model is used!

3.2.4 From incompleteness to uncertainty

Any program that models a real phenomenon must face a central difficulty: how should it use an incomplete model of the phenomenon to reason, decide, and act efficiently?

The purpose of Bayesian Programming is precisely to tackle this problem

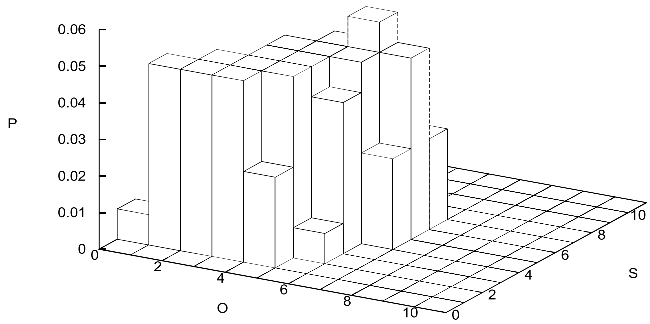

FIGURE 3.8: The histogram of the observed sensor state S and the output O when the inputs and the control are set to $[I_0 = 2]$, $[I_1 = 8]$, $[C = 2]$, and the values of the external factor and the internal functioning $H \wedge F$ are drawn at random.

with a well established formal theory: probability calculus. The sequel of this book will try to explain how to do this.

In the Bayesian Programming approach, the programmer does not propose an exact model but rather expresses a probabilistic canvas in the specification phase. This probabilistic canvas gives some hints about what observations are expected. The specification is not a fixed and rigid model purporting completeness. Rather, it is a framework, with open parameters, waiting to be shaped by the experimental data. Learning is the means of setting these parameters. The resulting probabilistic descriptions come from both: (i) the views of the programmer and (ii) the physical interactions specific of each phenomenon. Even the influence of the hidden variables is taken into account and quantified; the more important their effects, the more noisy the data, and the more uncertain the resulting descriptions.

The theoretical foundations of Bayesian Programming may be summed up by Figure 3.10.

The first step in Figure 3.10 transforms the irreducible incompleteness into uncertainty. Starting from the specification and the experimental data, learning builds probability distributions.

The maximum entropy principle is the theoretical foundation of this first step. Given some specifications and some data, the probability distribution that maximizes the entropy is the distribution that best represents the com-

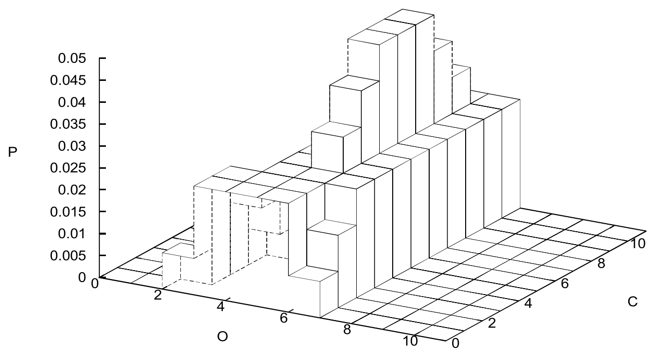

FIGURE 3.9: The histogram of the observed output O and the control C when the inputs are set to $[I_0 = 2]$, $[I_1 = 8]$, and the internal functioning F is set to $[F = 8]$ with H drawn at random.

bined specification and data. Entropy gives a precise, mathematical, and quantifiable meaning to the quality of a distribution.[5]

Two extreme examples may help to understand what occurs:

1. Suppose that we are studying a formal phenomenon. There may not be any hidden variables. A complete model may be proposed. The phenomenon and the model could be identical. For instance, this would be the case if we take the equations of Section 3.1.1 as the model of the phenomenon described in that same section. If we select this model as the specification, any data set will lead to a description made of Diracs. There is no uncertainty; any question may be answered either by true or false. Logic appears as a special case of the Bayesian approach in that particular context (see Cox [1979]).

2. At the opposite extreme, suppose that the specification consists of very poor hypotheses about the modeled phenomenon, for instance, by ignoring H and also the inputs I_0 and I_1 in a model of the above process. Learning will only lead to flat distributions, containing no information. No relevant decisions can be made, only completely random ones.

Specifications allow us to build general models where inaccuracy and hidden

[5]See FAQ/FAM, Section 16.11 "Maximum entropy principle justifications" for justifications for the use of the maximum entropy principle.

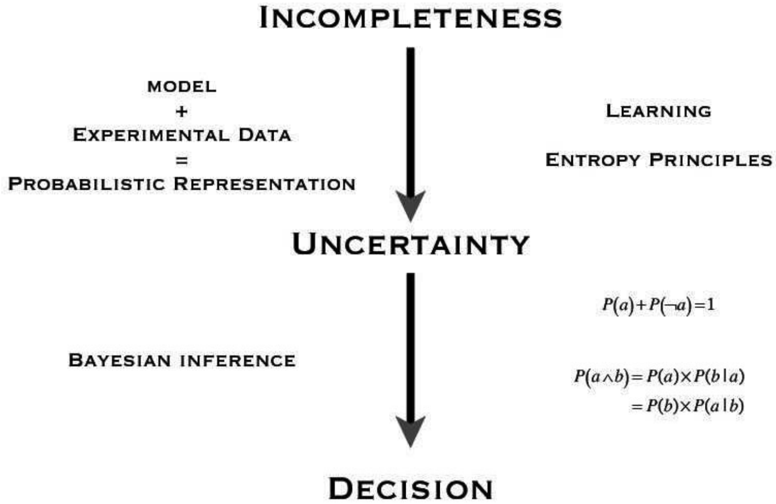

FIGURE 3.10: Theoretical foundation: from Incompleteness to Uncertainty and from Uncertainty to Decision.

variables may be explicitly represented. These models may lead to good prediction and decision. The formalism also allows us to take into account missing variables. In real life, such models are in general poorly informative and may not be useful in practical applications. They give no certitudes, although they provide a means of taking the best possible decision according to the available information. This is the case here when the only hidden variable is H.

The second step in Figure 3.10 consists of reasoning with the probability distributions obtained by the first step. To do so, we only require the two basic rules of Bayesian inference presented in Chapter 2. These two rules are to Bayesian inference what the resolution principle is to logical reasoning (see Robinson [1965], Robinson [1979], Robinson and Silbert [1982a], and Robinson and Silbert [1982b]). These inferences may be as complex and subtle as those usually achieved with logical inference tools, as will be demonstrated in the different examples presented in the sequel of this book.