are also truth-functional; for example, the degree of belief in $A \lor B$ is a function of the belief in $A$ and the belief in $B$.

The bad news for rule-based systems is that the properties of locality, detachment, and truth-functionality are simply not appropriate for uncertain reasoning. Let us look at truth-functionality first. Let $H_1$ be the event that a fair coin flip comes up heads, let $T_1$ be the event that the coin comes up tails on that same flip, and let $H_2$ be the event that the coin comes up heads on a second flip. Clearly, all three events have the same probability, 0.5, and so a truth-functional system must assign the same belief to the disjunction of any two of them. But we can see that the probability of the disjunction depends on the events themselves and not just on their probabilities:

$$
P(A) = 0.5 \quad P(B) = 0.5 \quad P(A \lor B) = 0.75
$$

It gets worse when we chain evidence together. Truth-functional systems have rules of the form $A \rightarrow B$ that allow us to compute the belief in $B$ as a function of the belief in the rule and the belief in $A$. Both forward- and backward-chaining systems can be devised. The belief in the rule is assumed to be constant and is usually specified by the knowledge engineer—for example, as $A \rightarrow B$.

Consider the wet-grass situation from Figure 14.12(a) (page 529). If we wanted to be able to do both causal and diagnostic reasoning we would need the two rules

$$\text{Rain} \leftarrow \text{WetGrass} \quad \text{and} \quad \text{WetGrass} \leftarrow \text{Rain}.$$  

These two rules form a feedback loop: evidence for Rain increases the belief in WetGrass, which in turn increases the belief in Rain. Clearly, uncertain reasoning systems need to keep track of the paths along which evidence is propagated.

Intercausal reasoning (or explaining away) is also tricky. Consider what happens when we have the two rules

$$\text{Sprinkler} \leftarrow \text{WetGrass} \quad \text{and} \quad \text{WetGrass} \leftarrow \text{Rain}.$$  

Suppose we see that the sprinkler is on. Chaining forward through our rules, this increases the belief that the grass will be wet, which in turn increases the belief that it is raining. But this is ridiculous: the fact that the sprinkler is on explains away the wet grass and should reduce the belief in rain. A truth-functional system acts as if it also believes $\text{Sprinkler} \rightarrow \text{Rain}$.

Given these difficulties, how can truth-functional systems be made useful in practice? The answer lies in restricting the task and in carefully engineering the rule base so that undesirable interactions do not occur. The most famous example of a truth-functional system for uncertain reasoning is the certainty factors model, which was developed for the MYCIN medical diagnosis program and was widely used in expert systems of the late 1970s and 1980s. Almost all uses of certainty factors involved rule sets that were either purely diagnostic (as in MYCIN) or purely causal. Furthermore, evidence was entered only at the "roots" of the rule set, and most rule sets were singly connected. Beckerman (1986) has shown that,
under these circumstances, a minor variation on certainty-factor inference was exactly equiv-
alent to Bayesian inference on polytrees. In other circumstances, certainty factors could yield
disastrously incorrect degrees of belief through overcounting of evidence. As rule sets be-
came larger, undesirable interactions between rules became more common, and practitioners
found that the certainty factors of many other rules had to be "tweaked" when new rules were
added. For these reasons, Bayesian networks have largely supplanted rule-based methods for
uncertain reasoning.

14.7.2 Representing ignorance: Dempster—Shafer theory

The Dempster—Shafer theory is designed to deal with the distinction between uncertainty
and ignorance. Rather than computing the probability of a proposition, it computes the
probability that the evidence supports the proposition. This measure of belief is called a
belief function, written Bel(X).

We return to coin flipping for an example of belief functions. Suppose you pick a
coin from a magician’s pocket. Given that the coin might or might not be fair, what belief
should you ascribe to the event that it comes up heads? Dempster—Shafer theory says that
because you have no evidence either way, you have to say that the belief Bel(Heads) = 0
and also that Bel(¬Heads) = 0. This makes Dempster—Shafer reasoning systems skeptical
in a way that has some intuitive appeal. Now suppose you have an expert at your disposal
who testifies with 90% certainty that the coin is fair (i.e., he is 90% sure that P(Heads) = 0.5).
Then Dempster—Shafer theory gives Bel(Heads) = 0.9 x 0.5 = 0.45 and likewise
Bel(¬Heads) = 0.45. There is still a 10 percentage point "gap" that is not accounted for by
the evidence.

The mathematical underpinnings of Dempster—Shafer theory have a similar flavor to
those of probability theory; the main difference is that, instead of assigning probabilities
to possible worlds, the theory assigns masses to sets of possible world, that is, to events.
The masses still must add to 1 over all possible events. Bel(A) is defined to be the sum of
masses for all events that are subsets of (i.e., that entail) A, including A itself. With this
definition, Bel(A) and Bel(¬A) sum to at most 1, and the gap—the interval between Bel(A)
and 1 Bel(¬A)—is often interpreted as bounding the probability of A.

As with default reasoning, there is a problem in connecting beliefs to actions. Whenever
there is a gap in the beliefs, then a decision problem can be defined such that a Dempster—
Shafer system is unable to make a decision. In fact, the notion of utility in the Dempster—
Shafer model is not yet well understood because the meanings of masses and beliefs them-
Celves have yet to be understood. Pearl (1988) has argued that Bel(A) should be interpreted
not as a degree of belief in A but as the probability assigned to all the possible worlds (now
interpreted as logical theories) in which A is provable. While there are cases in which this
quantity might be of interest, it is not the same as the probability that A is true.

A Bayesian analysis of the coin-flipping example would suggest that no new formalism
is necessary to handle such cases. The model would have two variables: the Bias of the coin
(a number between 0 and 1, where 0 is a coin that always shows tails and 1 a coin that always
shows heads) and the outcome of the next Flip. The prior probability distribution for Bias