It is important to understand that $P(cavity) = 0.2$ is still *valid* after *toothache* is observed; it just isn't especially useful. When making decisions, an agent needs to condition on *all* the evidence it has observed. It is also important to understand the difference between conditioning and logical implication. The assertion that $P(cavity \mid toothache) = 0.6$ does not mean "Whenever *toothache* is true, conclude that *cavity* is true with probability 0.6" rather it means "Whenever *toothache* is true *and we have no further information*, conclude that *cavity* is true with probability 0.6." The extra condition is important; for example, if we had the further information that the dentist found no cavities, we definitely would not want to conclude that *cavity* is true with probability 0.6; instead we need to use $P(cavity \mid toothache \wedge \neg cavity) = 0$.

Mathematically speaking, conditional probabilities are defined in terms of unconditional probabilities as follows: for any propositions a and b, we have

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)}, \tag{12.3}$$

which holds whenever $P(b) > 0$. For example,

$$P(doubles \mid Die_1 = 5) = \frac{P(doubles \wedge Die_1 = 5)}{P(Die_1 = 5)}.$$

The definition makes sense if you remember that observing b rules out all those possible worlds where b is false, leaving a set whose total probability is just $P(b)$. Within that set, the worlds where a is true must satisfy $a \wedge b$ and constitute a fraction $P(a \wedge b)/P(b)$.

The definition of conditional probability, Equation (12.3), can be written in a different form called the **product rule**:

Product rule

$$P(a \wedge b) = P(a \mid b)P(b). \tag{12.4}$$

The product rule is perhaps easier to remember: it comes from the fact that for a and b to be true, we need b to be true, and we also need a to be true given b.

12.2.2 The language of propositions in probability assertions

In this chapter and the next, propositions describing sets of possible worlds are usually written in a notation that combines elements of propositional logic and constraint satisfaction notation. In the terminology of Section 2.4.7, it is a **factored representation**, in which a possible world is represented by a set of variable/value pairs. A more expressive **structured representation** is also possible, as shown in Chapter 15.

Random variable

Variables in probability theory are called **random variables**, and their names begin with an uppercase letter. Thus, in the dice example, *Total* and Die_1 are random variables. Every random variable is a function that maps from the domain of possible worlds Ω to some **range**—the set of possible values it can take on. The range of *Total* for two dice is the set $\{2, \ldots, 12\}$ and the range of Die_1 is $\{1, \ldots, 6\}$. Names for values are always lowercase, so we might write $\sum_x P(X = x)$ to sum over the values of X. A Boolean random variable has the range $\{true, false\}$. For example, the proposition that doubles are rolled can be written as $Doubles = true$. (An alternative range for Boolean variables is the set $\{0, 1\}$, in which case the variable is said to have a **Bernoulli** distribution.) By convention, propositions of the form $A = true$ are abbreviated simply as a, while $A = false$ is abbreviated as $\neg a$. (The uses of *doubles*, *cavity*, and *toothache* in the preceding section are abbreviations of this kind.)

Range

Bernoulli

Ranges can be sets of arbitrary tokens. We might choose the range of *Age* to be the set $\{juvenile, teen, adult\}$ and the range of *Weather* might be $\{sun, rain, cloud, snow\}$. When no