ities.

(AMPLE 1 ***(Voting)*** A pollster is hired by a presidential candidate to determine his support among the voters of Pennsylvania's two big cities: Philadelphia and Pittsburgh. The pollster designs the following sampling technique: Select one of the cities at random and then poll a voter selected at random from that city. Suppose that in Philadelphia two-fifths of the voters favor the Republican candidate and three-fifths favor the Democratic candidate. Suppose that in Pittsburgh two-thirds of the voters favor the Republican candidate and one-third favor the Democratic candidate.

(a) Draw a tree diagram describing the survey.

(b) Find the probability that the voter polled is from Philadelphia and favors the Republican candidate.

(c) Find the probability that the voter favors the Republican candidate.

(d) Find the probability that the voter is from Philadelphia, given that he favors the Republican candidate.

Solution (a) The survey proceeds in two steps: first, select a city, and second, poll a voter. Figure 3(a) shows the possible outcomes of the first step and the associated probabilities. For each outcome of the first step there are two possibilities for the second step: the person selected could favor the Republican or the Democrat. In Fig. 3(b) we have represented these possibilities by drawing branches emanating from each of the outcomes of the first step. The probabilities on the new branches are actually conditional probabilities. For instance,

$$\tfrac{2}{5} = \Pr(\text{Rep} \mid \text{Phila}),$$

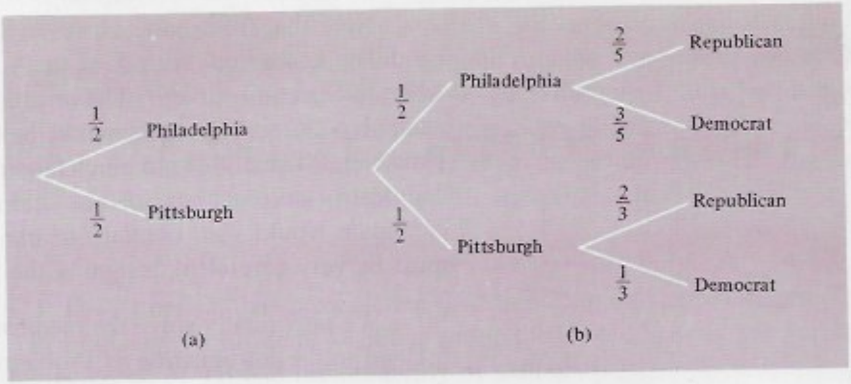

FIGURE 3

the probability that the voter favors the Republican candidate, given that the voter is from Philadelphia.

(b) $\Pr(\text{Phila} \cap \text{Rep}) = \Pr(\text{Phila}) \cdot \Pr(\text{Rep}|\text{Phila}) = \frac{1}{2} \cdot \frac{2}{5} = \frac{1}{5}$.

That is, the probability is $\frac{1}{5}$ that the combined outcome corresponds to the path highlighted in Fig. 4(a). We have written the probability $\frac{1}{5}$ at the end of the path to which it corresponds.

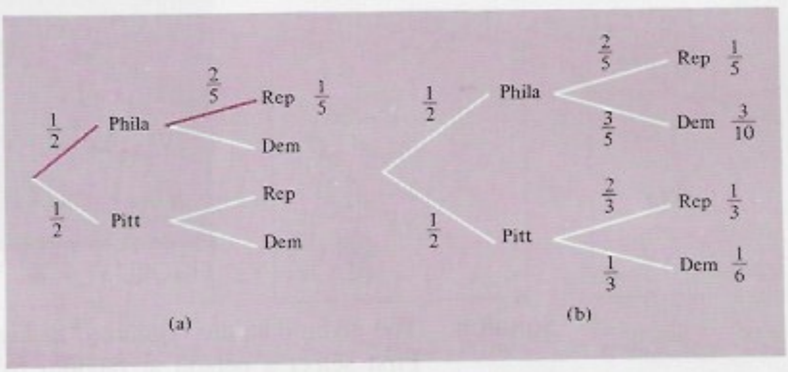

FIGURE 4

(c) In Fig. 4(b) we have computed the probabilities for each path of the tree as in part (b) above. Namely, the probability for a given path is the product of the probabilities for each of its segments. We are asked for Pr(Rep). There are two paths through the tree leading to Republican, namely

Philadelphia–Republican or Pittsburgh–Republican.

The probabilities of these two paths are $\frac{1}{5}$ and $\frac{1}{3}$, respectively. So the probability that the Republican is favored equals $\frac{1}{5} + \frac{1}{3} = \frac{8}{15}$.

(d) Here we are asked for Pr(Phila|Rep). By the definition of conditional probability

$$\Pr(\text{Phila}|\text{Rep}) = \frac{\Pr(\text{Phila} \cap \text{Rep})}{\Pr(\text{Rep})} = \frac{\frac{1}{5}}{\frac{8}{15}} = \frac{3}{8}.$$

Note that from part (c) we might be led to conclude that the Republica candidate is leading, with $\frac{8}{15}$ of the vote. However, we must always be carefu when interpreting surveys. The results depend heavily on the survey design. F example, the survey above drew half of its sample from each of the citie However, Philadelphia is a much larger city and is leaning toward the Democrat candidate—so much so, in fact, that in terms of popular vote the Democrat candidate would win, contrary to our expectations drawn from (c). A pollste must be very careful in designing the procedure for selecting people. ◆

We now finally solve the medical diagnosis problem mentioned in Sectio 6.1 and in the Introduction as Problem 2.

EXAMPLE 2 (***Medical Screening***) Suppose that the reliability of a skin test for active pulmo nary tuberculosis (TB) is specified as follows: Of people with TB, 98% have a positive reaction and 2% have a negative reaction; of people free of TB, 99% have a negative reaction and 1% have a positive reaction. From a large populatio of which 2 per 10,000 persons have TB, a person is selected at random and give a skin test, which turns out to be positive. What is the probability that the perso has active pulmonary tuberculosis?

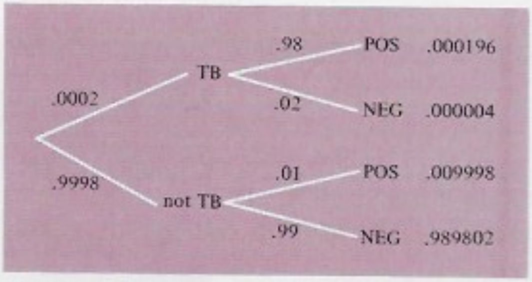

FIGURE 5

Solution The given data are organized in Fig. 5. The procedure called for is as follows First select a person at random from the population. There are two possibl outcomes: The person has TB

$$\Pr(\text{TB}) = \frac{2}{10{,}000} = .0002$$

or the person does not have TB

$$\Pr(\text{not TB}) = .9998.$$

For each of these two possibilities the possible test results and conditional proba bilities are given. Multiplying the probabilities along each of the paths throug the tree gives the probabilities of the different outcomes. The resulting probabili ties are written on the right in Fig. 5. The problem asks for the conditiona probability that a person has TB, given that the test is positive. By definitio

$$\Pr(\text{TB}|\text{POS}) = \frac{\Pr(\text{TB} \cap \text{POS})}{\Pr(\text{POS})} = \frac{.000196}{.000196 + .009998} = \frac{.000196}{.010194} \approx .02.$$

Therefore, the probability is .02 that a person with a positive skin test has TB. In other words, although the skin test is quite reliable, only about 2% of those with a positive test turn out to have active TB. This result must be taken into account when large-scale medical diagnostic tests are planned. Because the group of people without TB is so much larger than the group with TB, the small error in the former group is magnified to the point where it dominates the calculation. ◆

Note: The numerical data presented in Example 2 are only approximate. Variations in air quality for different localities within the United States cause variations in the incidence of TB and the reliability of skin tests.

Tree diagrams come in all shapes and sizes. Three or more branches might emanate from a single point, for example, and some trees may not have the symmetry of those in Examples 1 and 2. Tree diagrams arise whenever an activity can be thought of as a sequence of simpler activities.

EXAMPLE 3 **(*Quality Control*)** A box contains five good light bulbs and two defective ones. Bulbs are selected one at a time (without replacement) until a good bulb is found. Find the probability that the number of bulbs selected is (i) one, (ii) two, (iii) three.

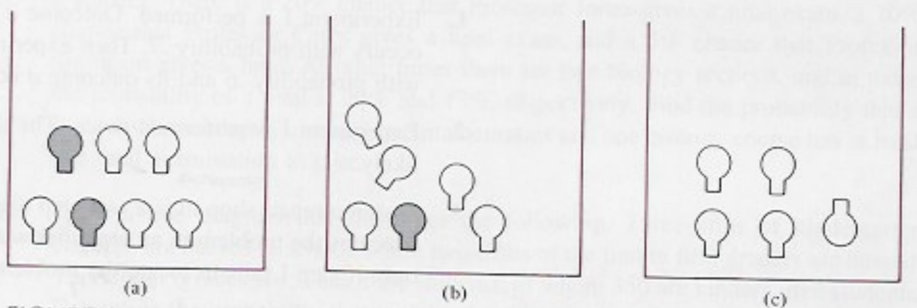

FIGURE 6

Solution The initial situation in the box is shown in Fig. 6(a). A bulb selected at random will be good (G) with probability $\frac{5}{7}$ and defective (D) with probability $\frac{2}{7}$. If a good bulb is selected, the activity stops. Otherwise, the situation is as shown in Fig. 6(b), and a bulb selected at random has probability $\frac{5}{6}$ of being good and probability $\frac{1}{6}$ of being defective. If the second bulb is good, the activity stops. If the second bulb is defective, then the situation is as shown in Fig. 6(c). At this point a bulb has probability 1 of being good.

The tree diagram corresponding to the sequence of activities is given in Fig. 7. Each of the three paths has a different length. The probability associated with each path has been computed by multiplying the probabilities for its branches. The first path corresponds to the situation where only one bulb is selected, the second path corresponds to two bulbs, and the third path to three bulbs. Therefore,

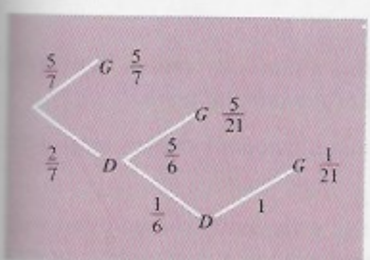

FIGURE 7

$$\text{(i) } \Pr(1) = \tfrac{5}{7}, \quad \text{(ii) } \Pr(2) = \tfrac{5}{21}, \quad \text{(iii) } \Pr(3) = \tfrac{1}{21}.$$ ◆