1.4 Tests and their design

We have discussed how hypotheses can lead to predictions, and comparing predictions to observations is ultimately how we test hypotheses. If the battery is dead, we predict that the radio will not work, and if the battery is not dead the radio will work. Those predictions provided a simple test of our hypothesis. Most tests of scientific hypotheses are more complex. Hypotheses make many predictions, but which predictions will be most helpful for a test? What other things could produce the observations that we predict for our hypothesis, and how can we discriminate between causes related to our hypothesis and other kinds of causes? What do we measure and how do we measure it in order to know whether or not our predictions have been met? Even if we can measure something, how do we know if it has changed in the way we predicted or not? The answers to these questions are all part of designing a good test of a hypothesis.

We typically call these well-designed tests **experiments**, but note that there are different kinds of experiments. The word "experiment" often conjures up images of antiseptic laboratories, glassware with graduated measurements, and people in white coats. Certainly, many experiments are performed in labs, because labs provide environments amenable to good experiments. But experiments can—and sometimes must—occur outside of the lab.

Good experiments have a number of qualities. First and foremost, the results of the experiment will be different depending on whether one hypothesis or another is true. In other words, in a good experiment, the null hypothesis will predict different results than what is predicted by the alternate hypothesis. If the hypotheses make the same predictions, then the test will not allow us to determine whether one hypothesis fits the data better than the other. We can then ask whether the actual results match the predictions of one hypothesis or the other, or perhaps even of neither.

Good tests also provide a way to minimize the effects of factors other than the one about which we are making predictions. This is the function of a **controlled experiment**, where we control all of the relevant variables in an experiment, whether they pertain to the cause we are testing or to other factors. Controlled experiments also typically include **control treatments**. Control treatments differ from **experimental treatments** in that they have not been manipulated in terms of the variable that is being tested. For instance, in our car example, we could control for other causes of a radio not working. We could start the engine and turn on the radio of another car, which would allow us to know what we should hear on the radio if it is working. We could replace the car's battery with a new one (or one known to be functioning) and see if the radio is working.

As another example, let's say you wanted to test whether a particular drug has a particular effect on mice. You could administer the drug to mice in your experimental treatments but not to those in your control treatments, and all mice in all treatments would otherwise be as similar as possible. This way, if the mice in the experimental treatments exhibit effects that differ from those in the control treatments, we can have the greatest confidence that those differences are due to the effect of the drug.

In essence, the control treatments give us a point of comparison for our experimental treatments. In the plant experiment, our experimental treatments were the plants we placed in the dark. We predict that these plants would not grow as well as our control plants, which were the ones in the sunlight. Without the control plants, we would have no way of evaluating whether what happened to the plants in the dark was really affected by the lack of sunlight.

Activity

Using the three hypotheses and experiments in the previous activity, do the following:

Identify the control treatment(s).

Use the results given in the tables below to determine whether the data were consistent with the predictions of the alternate or null hypotheses.

1. *Acetylcholine experiment*

Type of treatment	*Percentage contracting*
Acetylcholine added	*100*
Solvent only	*0*
No acetylcholine or solvent	*0*

2. *Water movement in plants*

Type of treatment	*Movement of water (as percentage of plant height)*
All stomata covered	*0*
Half of stomata covered	*50*
No stomata covered	*100*

3. *Amphibian metamorphosis*

Type of treatment	*Percentage metamorphosing*
Thyroxin added	*85*
No thyroxin added	*2*

There are additional qualities of a good experiment. Good experiments include more than one subject, to ensure that the results are not biased by the peculiarities of a single individual; the number of subjects is often what we refer to as the **sample size**. Most experiments strive to have a large enough sample size to make comparisons meaningful and to enhance the utility of any statistics that are applied to the data.

Good experiments also account for other **confounding variables** that could be affecting the results, particularly those pertaining to **environmental variation** and **individual variation**. Environmental variation includes a number of conditions pertaining to the