Science in Process

A Journey into the Scientific Method

by Arms and Camp

The scientific method is a way of answering questions about cause and effect. It is a logical way to try to solve problems and similar reasoning is used by each of us in everyday life. Once we understand the reasoning and procedures scientists use, we are better able to decide for ourselves whether the conclusions of the latest scientific study are justified from the data presented. We can ask for further tests if a claim does not appear to be well supported by the evidence, and we can agree or disagree with predictions based on such claims.

When we speak of the scientific method, we usually mean a method of doing science that involves experiments. This experimental method has three main steps (although in practice, scientists work in many different ways). The first, and key, step is to collect **observations,** not only by sight, but perhaps using other senses too (hearing, smell, taste, and touch). Scientists often use instruments to extend human senses or to detect things our senses cannot; examples are microscopes, radar, voltmeters, or Geiger counters. Second, the scientist thinks of several alternative hypotheses (singular: hypothesis), proposed answers to questions about what has been observed. The third step is experimentation, performing tests designed to show that one or more of the hypotheses is more or less likely to be incorrect. As a result of these experiments, the scientist should be able to draw some conclusions about *why* the original observed events occur. Let us see how this works in practice.

Scientists usually start with observations that stimulate questions. Some years ago, one of your authors was part of a group of biologists discussing the clusters of butterflies that seemed to be everywhere that June.

"Today," said one, "I saw about 20 yellow sulfur butterflies by a stream and some black swallowtails on a manure heap. What are they doing?"

"It's called *puddling behavior*," replied another. "You find puddling butterflies in groups in open places such as the edges of drying puddles, or sandbars. I don't think anyone knows what they are doing. Another odd thing is that in many species only the males puddle."

These observations of puddling led us to ask what the butterflies were doing and why. To answer these questions, we had to think of some hypotheses that would account for the observations. That evening, the hypotheses came thick and fast from our armchair scientists.

"An article I read suggested it was a method of population control. Coming together permits the males to count each other. A newcomer can see if there is likely to be enough land for him to set up a territory in the area. Puddling saves them having to fight over territories."

"That sounds wrong to me," replied one of the company. "How can a butterfly figure out the density of males in the area from a group like that? Besides, swallowtails do fight for territories—I've seen them."

"I think it is more likely they're feeding," another contributed. "It was called 'puddling' in the first place because the butterflies often have their proboscises [tongues] out and seem to be sucking something up from the ground."

"I wonder if they are feeding on substances that contain nitrogen. In our lab we've shown that butterfly caterpillars grow faster if you feed them extra nitrogen, and there is lots of nitrogen in a manure pile."

"But not in sand," came the objection. "And if they are after nitrogen, you'd expect females to puddle, not males. The females lay the eggs that hatch into caterpillars, and extra nitrogen in the egg might be very useful, but it's not the females that puddle."

"It sounds to me," chipped in another, "as if they're after salts—perhaps salts containing sodium. All the puddling places contain quite a lot of salts: manure piles have salts from urine, and puddles have salts at the edges, left behind by evaporation of water. Lots of animals that feed on plants are short of sodium because plants contain so little of it. We put out salt blocks for cows and horses and end up attracting deer and rabbits as well. Perhaps male butterflies need more sodium than females do."

We could test these alternative hypotheses only by doing experiments. Some hypotheses are of no use because they cannot be tested. For instance, the hypothesis "puddling butterflies count each other" is probably untestable because it is hard to imagine an experiment that could show us whether or not an animal has counted its neighbors. Even a testable hypothesis usually cannot be tested directly. We must first develop a testable **prediction** from it. From the hypothesis that butterflies sucked up sodium when they puddled, we predicted that if we put out trays containing sodium, butterflies would be attracted to puddle on them. The hypothesis that puddling butterflies suck up nitrogen generated the prediction that butterflies would puddle on trays of amino acids, substances that contain nitrogen. These predictions can be tested and, in this case, both can be tested at the same time, in the same experiment.

We must design experiments to make their results as clear-cut as possible. For this reason, experiments have to include **control treatments** as well as **experimental treatments**. The two differ only by the factor(s) being investigated. For instance, to test our hypotheses, we had to show that butterflies would puddle on an experimental tray containing amino acids or one containing sodium but would not puddle on control trays that were identical except that they did not contain the amino acids or sodium.

Suppose we put out three trays—one containing sodium, another containing amino acids, and a third containing something butterflies are most unlikely to eat, such as plain sand or sand and water (the control). We would predict that if butterflies are attracted to puddle on sodium, they would come to puddle on the sodium tray but not on the other two. If they are attracted to amino acids, they would puddle only on that tray. If they are attracted to both, they would puddle on both of these but not on the control tray, and if they are attracted neither to amino acids nor to sodium, they would not puddle on the trays at all. Note that there are dozens of other possible reasons for the last result. If no butterflies turned up to puddle on our trays, we would have learned nothing. Butterflies might not puddle on trays because they won't come near trays for some reason, or because they never see the trays, or because they avoid the human watchers nearby, or for any one of a number of other reasons.

So that our experiment would not fail for lack of butterflies, we put our trays on a sandbank by a lake where tiger swallowtail butterflies often puddled in large numbers. We filled the trays with clean sand for the butterflies to stand on, and in each tray we pinned a dead male tiger swallowtail as a decoy, because we thought butterflies might be attracted to puddling places by seeing other butterflies there. We put out ten trays of sand and poured the same volume of solution (substances dissolved in water) into each one. Then we sat nearby, with binoculars, notebooks, and watches, to see what would happen.

Soon dozens of tiger swallowtails were hovering over the trays. Whenever a butterfly landed on a tray, it stuck its proboscis into the sand. At times, as many as 30 butterflies were on a tray together. Most of the butterflies spent a few seconds on every tray, but they puddled (which we defined as staying for more than 15 seconds) on only a few trays: all those containing sodium in any form and those containing amino acids.

We were satisfied that these results were accurate because we had taken another precaution: the people recording the butterflies visits did not know which tray contained which solution (Figure 1-2). Making an experiment "blind" in this way is important. Psychologists have shown that, even in a carefully controlled experiment, experimenters tend to find the results they want to find. This is also why scientists try to form many hypotheses to explain their observations. It's too easy to bend the truth, without even realizing it, to support the only available hypothesis.

Those of us who favored the hypothesis that butterflies puddle in response to sodium were disappointed that they also puddled on amino acids. But prejudice can sometimes be useful, even in science! Not only were we disappointed by the results, we were inclined to think they were wrong. Back we went to our bottle of amino acids. We now made an observation that should have preceded the experiment: the label said, "Prepared in sodium citrate." According to popular myth, scientists are calm and objective, but we were very excited when a technician analyzed our amino acids: they were chock full of sodium! There followed frantic phone calls and special deliveries to obtain amino acids free from sodium. At last came a suspenseful experiment, which showed that butterflies did not puddle on our new, sodium-free amino acids.

We had now conducted a well-controlled scientific experiment. What conclusions could we draw? Had we proved the hypothesis that butterflies puddle so as to obtain sodium? No. We had not even shown that the butterflies actually drank the sodium solution. All we had shown was that male tiger swallowtail butterflies would puddle on sand containing sodium salts but not on sand containing various other solutions. Many more hypotheses and experiments were needed if we were to learn more.

One peculiarity of the scientific method is that a hypothesis can never formally be proved but can only be disproved. A correct hypothesis leads to predictions that are borne out by experiments, but an incorrect hypothesis may also produce correct predictions (that is, the prediction was right, but for the wrong reason). Therefore, if the results of an experiment agree with the prediction, we are still not sure that the hypothesis is correct. For instance, the hypothesis that butterflies puddle to obtain sodium for food is not proven by the experimental finding that butterflies puddle on sodium. They might puddle because wherever there is sodium in nature there is also nitrogen and they really obtain nitrogen from puddling. We have not even disproved the hypothesis that puddling is a means for the butterflies to "count" each other. They might puddle on sodium merely as a convenient rendezvous (although the fact that the butterflies appear to feed when they puddle makes this hypothesis unlikely). The more alternative hypotheses we disprove or cast doubt on, however, the greater the likelihood that the remaining hypothesis is correct.

Scientists also hesitate to accept the results of an experiment until they have tested its repeatability. Repeating an experiment guards against two kinds of errors. The first is **human error** (a polite term for mistakes); we might have inadvertently switched the solutions, written our results in the wrong column of our data notebook, or alarmed the butterflies. (Even in this simple experiment, the possibilities are endless.) Second, any experiment is subject to **sampling error**, error due to using a relatively small number of subjects. Organisms are notoriously variable. Our experiment sampled only a few dozen butterflies on six days. These butterflies might not have been representative of all tiger swallowtails. We could be more confident of our results if we were to repeat the experiment, using more butterflies (that is, a larger sample) and following precisely the same procedure. How many butterflies do we need? The more the better, but we could not possibly test all the butterflies in the world. In practice, we can use statistical tests to tell how "sure" we are of our results with a given sample size.

A hypothesis supported by many different lines of evidence from repeated experiments is generally regarded as a **theory**, and after even further testing it comes to be generally accepted.

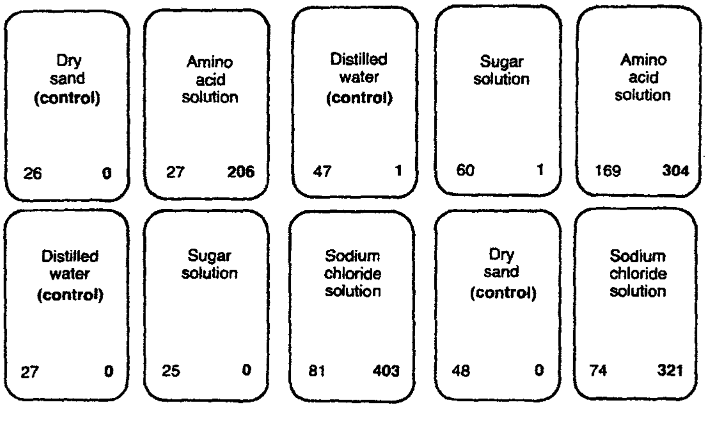

Figure 1-2

Arrangement of trays on one day of the puddling experiment. Each tray contained the same volume of sand. Each of eight trays also contained 1.5 litres of water or solution. Different solutions were placed in different trays on subsequent days. (Sugar was tested because swallowtail butterflies eat sugar-filled nectar from flowers, and therefore we wondered if they might be attracted to puddle on sugar.) The black number on each tray shows the number of "sampling" visits (lasting less than 15 seconds) by butterflies. Colored numbers show the number of butterfly-minutes spent puddling on the tray in visits lasting more than 15 seconds. The numbers make it obvious that butterflies puddled on the trays containing sodium and those containing amino acids but not on any of the other trays.

Science in Process

The Scientific Method

- What practical value can the scientific method play in a person's everyday life?
- What are some ways that scientists collect observations?
- How do control treatments differ from the experimental treatments? Give an example.
- How did the scientists help determine which substance the butterflies were "puddling" for?
- What does it mean for an experiment to be "*blind*"? Why is this important?
- What was wrong with the bottle of amino acids that was used in the first experiment? How was the situation resolved?
- Did the scientists finally prove that the butterflies puddle for sodium?
- Why is *repeatability* important? What does it protect against?
- In science, what does the word *theory* mean? Is this word used differently in everyday conversation?