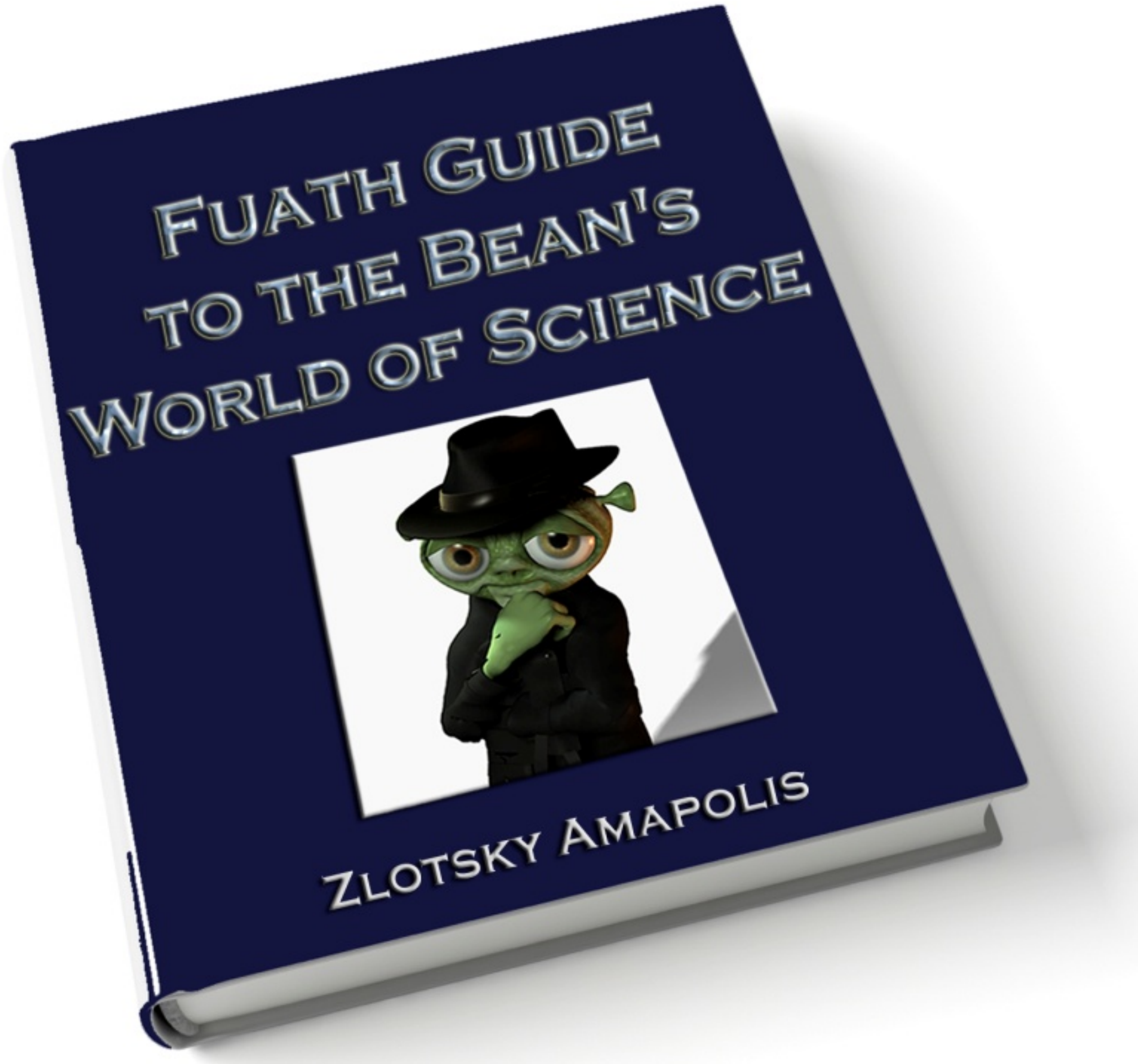


FUATH GUIDE
TO THE BEAN'S
WORLD OF SCIENCE
ZLOTSKY AMAPOLIS

Dear Reader:

You are about to read something that is literally "out of this world." It is called the "Fuath Guide to the Bean's World of Science". You may not know it (yet) but Fuaths are alien creatures that live on a planet called Thoth. As in any advanced society, Fauths discovered the scientific method.

This book is their manual for the scientific method. What is remarkable is that it shows the scientific method to be truly universal. (It has been translated and verified by the few Fuathologists that exist – us.) On Earth, we would call this a "textbook" most suitable for a class on research methods.

We began to be aware of Fauths when our research group started collaborating some years ago. We were interested in how best to teach research methods to Human students. Then our phones rang. It was the Institute for Educational Sciences within the United States Department of Education which had decided to give us the funds to do just that – teach research methods (R305B070349). Then we started getting other phone calls that were simply weird. One in particular changed our lives. It was from someone deep inside the U.S. Government that told us that they already had a book on research methods, but the author was an alien being. Would we be interested?

What we subsequently uncovered was beyond belief. Through a series of Freedom of Information Act (FOIA) requests, we discovered that the United States along with some other governments (e.g., Israel), had (previously disguised) alien agents held in captivity. These alien "agents" were arrested for implanting intentionally bad science reports in our popular press in order to confuse readers about science. (Fun fact: this was the origin of "fake news.") The most accepted reason why they were doing this was that they wanted the People of Earth to be ignorant of the scientific method so that they would hold an advantage in space.

This book is a manual about how the scientific method works, and was used by their agents on Earth, disguised as "Beans", to write and publish poor research here on Earth. Their motto is "in order to deceive, we must first learn the truth."

As of this writing, the war against the Fauths is still hidden to the public, but with mixed results. On the one hand, the United States had officially recognized the United States Space Force in December of 2019 based on the so-called "Fuath situation." But on the other hand, because of social media (an alien invention), poor research and other fake news abounds on the Internet.

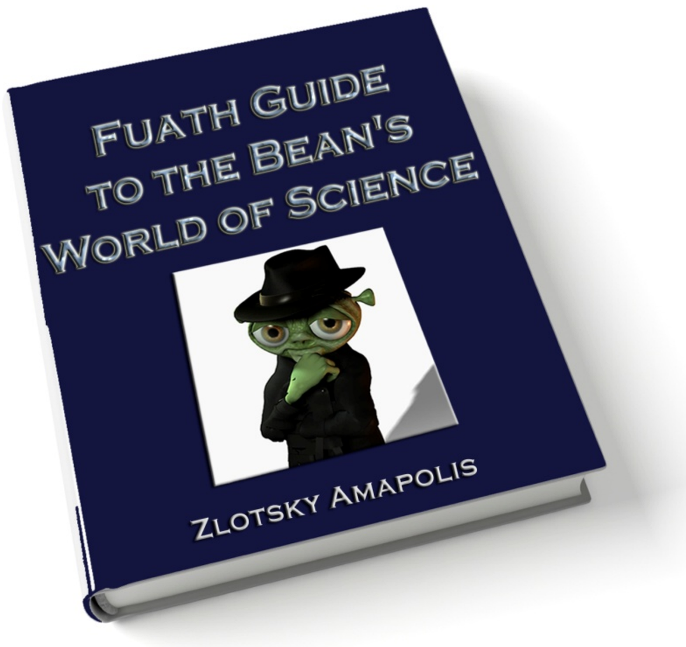
This book can be useful for learning research methods. Except for some alien words and examples, the ideas are sound and applicable for learning the scientific method here at home.

Notes: These materials are provided free of charge. Please use them with your students and colleagues and forward them to anyone who might find them useful. Always use the citation showing our names as authors and agree to never sell or make money in any other way with these materials. We hope you will find them useful and that your students will enjoy reading and learning.

Note: Operation Aries was a result of a grant (R305B070349) awarded to Northern Illinois University from the Institute of Education Sciences, U.S.

Keith Millis, Chief Fuath Scientist
Diane Halpern, Chief Fuath Translator of Zlotsky's "Guide"
Arthur Graesser, Chief Fuath Scientist
Patricia Wallace, Fuath Linguist
Heather Butler, Fuath Linguist

Fuath Guide to the Bean's World of Science



Acknowledgements

I have so many wonderful Fuath to thank for this marvelous book. First, there are all the Fuath agents who helped me live undercover like a Human Bean on their miserable planet Earth. Several times, I was almost detected as a Fuath, but the other agents distracted the Beans every time they became suspicious about me.

Special thanks to my good colleague Orama who quickly helped me make up a story about discovering penicillin when a group of Beans got suspicious because I was inhaling some lovely bacteria. Thanks also go to Zxxyaala for traveling cross-county at a moment's notice to pick me up after I crashed my aircraft. I was on a secret mission taking aerial scans of a place called New Mexico when my flux crystals ran out and my craft crashed into the desert near Roswell.

And again, I almost got caught when I showed my outrage at the movie ET-- so silly and so sad. I mean really, what self-respecting "alien" needs a machine to communicate with his family. All "aliens," including Fuaths, know that Beans are the ONLY creatures that are not evolved enough to have developed the power of telepathy. The Bean depiction of an alien species as helpless so infuriated me that I started to change color. Thankfully, another fine colleague intervened and told the Beans that I was overcome with emotion (so true--but not in the way they thought). I could go on, but you get the idea. Lots of fine friendly Fuaths helped me to pass as a Bean so that I could learn their science and pass that knowledge on to you.

Learn well the science ways of the Beans.

About the Author

Zlotsky Amapolis is a Supreme Commander of Fuath Espionage. He studied at several of our most prestigious universities including Fuath State University and University of the Thoth Hinterlands.

This knowledge is critical for our Mission.

Zlotsky is the proud parent of Nomma and Chess, both of whom are following in their parent's zlot-steps and studying Fuath Espionage. Zlotsky lived undercover as a Human Bean on Planet Earth so that he could learn their best science. His sacrifices are sincerely appreciated by all the Fuaths on the Planet Thoth. He received a Medal of Espionage in recognition of his contributions to the future of Thoth.

Book Reviews--

The Fuath Guide to the Bean's World of Science is a must read for all Thoths who want our beautiful planet to survive and thrive.
----- Professor and Head, Fuathburg University

Zlotsky has done us all a great service in putting together everything he learned about the Bean World of Science. We must use his knowledge to further our missions.
--------- Commander and President, Fuath Federation.

Chapter 1: The Experimental Method: Theories, Hypotheses, and Variables

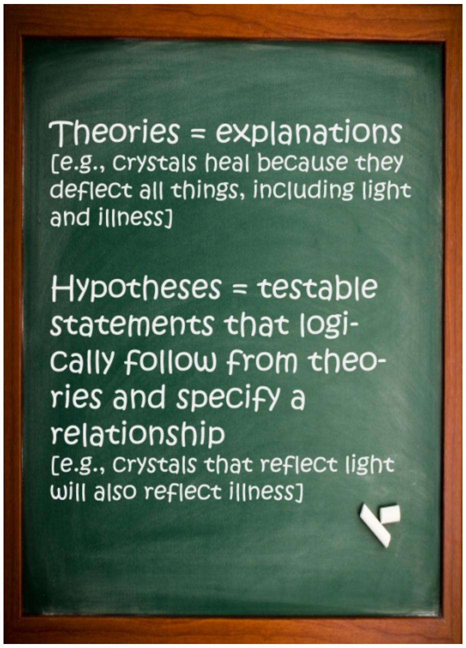
We all want to understand things in our universe—like why did that cute Qkbbi completely ignore me after I said hello, or how can we cure purple chickel-pox, or why do some of those smelly Human Beans prefer paper to plastic bags? One important method that can be used to understand these things is called the **experimental method**. It is a complicated method, so we will teach you about it in different stages as you progress through this training course.

The starting point for understanding something is an explanation, which is called a **theory** in science. If I wanted to know, for example, why those Beans get red, runny noses in the winter (they call it getting a "cold"—but they are not cold, they actually get hotter—oh those silly Beans), I might have an explanation like this: Maybe tiny germy things get inside their huge hairy noses and when these germy things sweat the Beans' noses get all watery. Germy-things love to live in noses where they multiply and make Human Beans feel really crummy, which is what they call "having a cold." This is an example of a theory. A theory is one possible answer to "why" questions. Keep in mind, though, that a theory is more than just an explanation for why something has occurred. **A theory is an organized set of principles that are a plausible explanation for something; a theory allows for the creation of new predictions.**

Having a theory is a starting point in the scientific method. Theories are important to science because they lead to **hypotheses**, which are statements about relationships that can be tested with experiments.

Here is an example:

Suppose the Human Beans believe that crystals heal their many illnesses by reflecting the illness-vapors away from the person wearing the crystal. If this was their theory, then a hypothesis would be that there is a relationship between reflectance and illness. If the crystal reflects light, then it will also reflect illness. This hypothesis could actually be tested with an experiment.



Here is another hypothesis: taking quizzes while learning helps students learn better. All hypotheses can be broken down into two parts, called **independent** and **dependent variables**. An **independent variable** is the potential cause of the **dependent variable** which is the outcome. In experiments, the independent variable is used to group people (or plants or whatever the experimenter is studying) so that we can compare different groups on some outcome measure. In this example, taking quizzes is the independent variable. To test our hypothesis that taking quizzes while learning helps students learn better, an experimenter might assign some students to take quizzes and some students not to take quizzes. The independent variable is the one that the experimenter manipulates. In this example, the experimenter creates two different groups and assigns students to one of them. With this manipulation, we could compare the two groups (those who take quizzes and those who don't) to see whether they learned differently.

The second part of the hypothesis is the dependent variable. The **dependent variable** is the outcome you are interested in. In this example, the dependent variable is the measure of learning. Like independent variables, dependent variables can be defined many ways. For our hypothesis, let's define learning as scores on a math exam. Technically, "learning math" would be the dependent variable and the "scores on the math exam" would be the dependent measure. Scores on some test related to what was studied seems like a reasonable dependent measure for this study. What if the experimenter decided to test her hypothesis by measuring how much each student weighs? Is weight a reasonable dependent measure of learning? Weight is a ridiculous dependent measure for this hypothesis because it is not a measure of learning. We would say that weight is an invalid (not valid) measure of learning. **A valid measure actually measures what it is supposed to measure.**

This variable business is difficult when you first learn about it, but gets easier over time. Let's do a few more examples. Here's an example from a Bean's newspaper.

The advertisement for the *Smoke-Out* ™ Subliminal Audio Tapes makes quite a promise. How would we know if these tapes really worked? You know I have been meaning to lose a few pounds, and there is a subliminal tape for that too. It would be amazing if playing a message that was so soft that no one could hear it actually made people stop smoking. How would we know if it works?

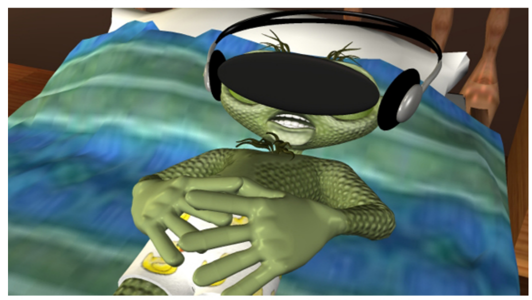
First, let's think back to what was already learned. Do we have a *theory*? Is there an explanation of how a subliminal tape might work? The message recorded on the tape – that you can't consciously hear – gets into your brain, and it causes you to stop smoking. I guess it gets in through your ears, even though you can't hear it. The idea that your brain makes you do something is also odd because we don't exist separate from our brain, but let's not deal with that problem. Let's assume that the people who are selling these tapes have some theory or explanation as to how these tapes help people stop smoking.

Do we have a testable *hypothesis*? Sure we do. If the theory is plausible, then listening to the tape should be associated with stopping smoking. But, we can't just give some people who smoke these tapes and see if they stop smoking because we won't know how many people who want to stop smoking do so without the tape. To determine whether the tapes are effective in getting listeners who want to stop smoking actually do stop smoking, we would need to find out if more of these people stop smoking after they listened to the subliminal tape than another group of people who also wanted to stop smoking but did not listen to the subliminal messages.

Can you think of an *independent variable* for this study?
Try it. There is more than one possible correct answer. When you are finished thinking about an independent variable for a study of the effectiveness of subliminal messages on audio tapes in getting people to stop smoking, flip to the next page.

One obvious way to study if *listening to subliminal messages recorded on tapes actually get listeners to stop smoking* is to create two groups/ One group receives the subliminal messages and the other does not. We then see if there is a difference between the two groups in regards to quiting smoking For the group that does not receive the subliminal messages, we can't just give them nothing to do at all because that would not be comparable to people who do listen to the tapes. Therefore, we could create "dummy tapes" that look like the subliminal message tapes in every way, except that they don't contain a subliminal message. Some people who want to stop smoking would be given the dummy tape and others would be given the tape with the subliminal message. Everyone in both groups would get the same instructions—play the tapes every night as you fall asleep. This is one way to design an experiment to test the theory that subliminal messages can cause Human Beans to stop smoking.

Group 1 - Listens to the Subliminal Tape.



Group 2 - "Listens" to the Blank Tape.

With this design, the independent variable is whether the tapes have the message or not [the tapes *vary* in whether they have a message or are blank--nothing is recorded on them]. This variable is manipulated by the experimenter who decides to have two kinds of tapes and assigns different people, all of whom want to stop smoking, to one of these two groups. Of course, no one knows if they are in the group with the subliminal message recorded on their tape or in the blank tape group because people are influenced by their expectations, a topic that we will learn more about later in this book.

Now that we have the independent variable set-up, we need to think about the variable associated with the second part of the hypothesis.

What does the experimenter really want to know about? The main question for this experiment concerns whether people stop smoking or continue smoking. This is another variable. In this case, it is the **dependent variable**, which involves what the participant does – or outcome behavior – that gets measured. If the theory is plausible, then whether someone stops smoking or not **depends** on whether they are in the group that gets a tape with a subliminal message or the group that gets a blank tape, which is the independent variable.

The dependent variable depends on the independent variable when the theory is likely to be true. So, in this example the dependent variable is smoking behavior. It is measured a certain number of times after the subliminal tapes are received by the smokers, by checking to see whether they are still smoking or not. Is this a valid dependent variable? Sure—measuring whether someone is still smoking or stopped smoking is a good measure of whether or not the tapes worked.

Dependent variables depend (what else would you expect them to do) on the independent variable. Let's try another example.

Have you ever noticed that the ones they call women smear a waxy colored stick over the darkish area around their mouths—the part they call lips? Don't you think it is strange that some women engage in this ritual behavior before they go off to work or school, but the men don't? Some Beans think the waxy colored stick smeared on women's lips will make them more popular. Ok, you are thinking, Ugh! You could never think of a Smelly One as being popular, but suppose you were a Human Bean. Do you think you would find that women who smear that stuff on their lips are more popular than those who don't? Can you design an experiment to find out? I know this is a tough question. The independent variable is not so hard. Try to figure it out—we could find female Beans and make some of them smear that reddish-pinkish stick on their lips and find other female Beans and tell them not to smear the colored stick on their lips. So far this is not too hard. Smearing or not smearing colored sticks on their lips would be what kind of variable?

I hope you said the independent variable because it is the variable that might be the cause of being popular—it is the one the experimenter manipulates. But, what is the dependent variable? How would we know if the female Beans that smeared the waxy stick on their lips are more popular than those who didn't? How would we define "popular" in a way that we can measure it consistently? This is a challenge because different people may view what popularity is quite differently. Here is one suggestion for measuring how popular someone is. We could carefully watch each female Bean, both with and without the smeared colored stuff on her lips, and count how often each woman is asked to dance at a dance club. If smearing a waxy stick on their lips makes women more popular (a leap of reasoning to be sure), then on average, the women in the "waxy stick group" would be asked to dance more often than the women in the "no waxy stick" group. Would that work?

This research design would work, although the number of times someone is asked to dance [or asks others to dance] may not be the best dependent measure of popularity. In this example, the number of times each female Bean is asked to dance would vary among the women. It is the dependent variable because our theory suggests that women in the group who smear waxy colored stuff on their lips would be asked to dance more often than the women in the group that does not smear stuff on their lips. Of course, the number of times each woman is asked to dance will vary depending on all sorts of variables, including how well she can dance, so our predictions about the two groups are about group averages. If the theory is plausible, the women in the group who smeared colored stuff on their lips will have more dances, on average, than the women in the group that do not smear waxy stuff on their lips, even though just by chance some of the women in the "no waxy stick" group will probably be asked to dance more often than some of the women in the "waxy stick group" even if the theory is plausible. Is this a valid dependent measure? While there are probably better ways to measure being popular than the number of times someone is asked to dance, you could make a reasonable case for it, so we would say it is valid.

Do you think you understand the concept of dependent variables? Let's think about other ways of measuring popularity. Do you think that using the women's shoe sizes as the dependent variables would be a good way of measuring popularity? Of course not, because there is not any reason to believe that shoe size is directly related to how popular a woman might be. In the language of the scientific method, shoe size is not a **valid** measure of popularity. It does not measure the concept we are interested in—it does not measure popularity, so it is an invalid [not valid] measure of popularity. Let these ideas sink in. Reread the examples on the last few pages if you are not confident in your understanding of dependent variables. Turn the page when you are ready to go on to the next chapter.

Chapter 2:
Dependent Variables, Reliability, Accuracy, and Precision,

Dependent variables are the darndest things. (I love that we never cuss on Thoth, unlike those Beans who utter regrettable things when they get upset.) In the last lesson, you learned the difference between independent and dependent variables, and the importance of having dependent measures that are valid.

In this lesson, we will learn about the reliability, accuracy, and precision of dependent variables. Whenever we measure something, we want the results to be reliable. **The reliability of a measure is the *consistency* with which it measures what it's supposed to measure**. If you wanted to know how much you weigh, and you got on a scale that gave you a different number each time, even when you weigh yourself seconds apart and your weight really did not change, the scale would be unreliable. It is like measuring how tall someone is using a rubber ruler that could stretch or shrink, so you'd probably get a different number each time you measured the same person, even when measurements are only a few seconds apart.

Remember that the *dependent variable* is the outcome measurement; therefore, it is the variable that you believe will change as a result of differences in the independent variable. In order to know if the hypothesis is likely to be correct, we need to be concerned with the way we measure dependent variables. Here is an example:

If you believe that men and women are different in how likely they are to exhibit leadership behaviors, you would need to measure leadership behaviors, the dependent variable, in a way that is *valid* and *reliable*. That means that you will need to make sure that you are really measuring leadership behaviors and not something else and that the measurement will give you very similar results every time you use it, assuming that nothing happened to change the way people exhibit leadership behaviors between the first and second [or more] times that you measured these behaviors.

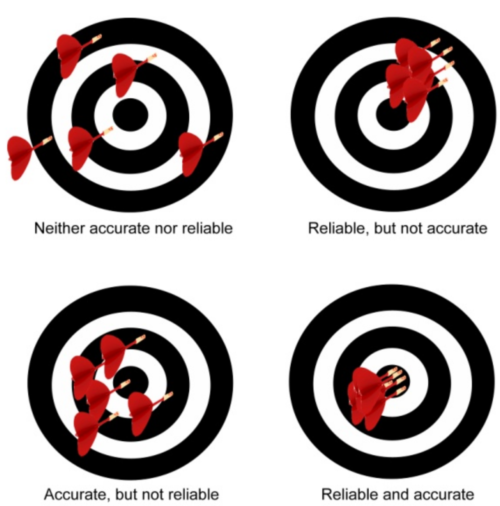
Let's work through this example as a way of clarifying these concepts.

First, what is the independent variable? It is the sex – or if you prefer, gender – of the participants. Of course, the experimenter doesn't actually manipulate which sex each person is—he takes people as they were born. But he creates two comparison groups based on sex, so in that sense he manipulates the independent variable. What do we think depends on the sex of the participants? Asking this type of question is a good way to identify the dependent variable. The answer is that it is "exhibiting leadership behaviors." Now we need an operational definition of "exhibiting leadership behaviors." **An operational definition is a set of procedures or 'operations' that describes how a variable will be measured.** In the last chapter, we operationally defined being popular as the number of times a female Bean was asked to dance. Once we have an operational definition, other people can measure popularity the same way. Suppose for this example, we decide to define "exhibiting leadership behaviors" as how often participants take charge when they are in a group. Will this work as an operational definition? Nope, not yet. It won't because it does not tell us how to recognize "take charge" behaviors.

Suppose we change our definition of the dependent variable to the number of times a participant is elected to be the leader after working in a group for 15 minutes. That measure will work because it will be easy to count the number of times that group members elect someone to be their leader. Is this a valid measure of leadership behaviors? Yes, even though it does not focus on what behaviors potential leaders exhibit, it does reflect a valid measure of leadership behaviors because the group members are deciding if the behaviors shown by the participant are worthy of a leader.

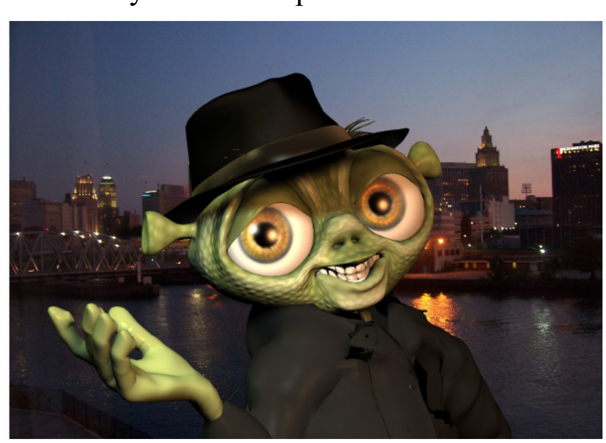
Is the dependent variable reliable? It is difficult to know how often the same behaviors by the same participant would lead to that person being elected the leader. We could test its reliability by coaching different people how to act like a leader and then see if the participant is elected to be the group leader with many different groups. This is what we call "an empirical question" because we can answer it by collecting data that provide an answer. What about using the dependent variable, or outcome measure, “taking charge”? Is this likely to be a reliable dependent variable? It probably would not be reliable because different people would view different outcome behaviors as “taking charge,” and what looks like “taking charge” to one person may look aggressive or rude to someone else.

Why is it important for dependent measures to be reliable? **When a dependent measure is not reliable, it is difficult to know if the results of the experiment are due to the independent variable.** So, if we measured leadership with an operational definition of “taking charge” that cannot be reliably measured across participants, the results of our experiment on gender and leadership will be heavily influenced by measurement error rather than the conditions set up in the independent variable. So, the flip side of this duva (what the Beans call a coin) is that **reliable dependent measures help experimenters make the claim that their results are due to what is being manipulated or contrasted with the independent variables.**



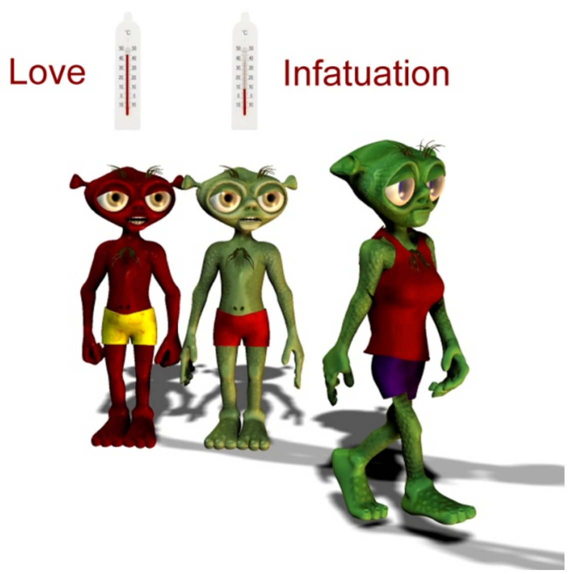
In addition to being reliable, dependent measures have to be **accurate** and **precise.** When we measure something, we systematically assign a number to it for the purposes of quantification. Someone who is heavier than you is assigned a higher number of pounds – or kilograms or stones – in weight than you are. If not, the concept of weight would be meaningless. **An *accurate* measure is one that is done carefully so that it is as close as possible to a “true” value. A measure that is *precise* uses units that are small enough to measure what you are studying.** Sometimes the word *sensitive* is used instead of precise—they mean the same thing.

It is easy to mix-up these concepts because reliability, accuracy, and precision are related. Here is an example that separates them. Suppose you are testing a hypothesis that something, some independent variable, affects heart rate where heart rate is the dependent variable. If you are using a heart rate monitor that is very unreliable, being careful in how you place the monitors that record heart rate and in how you read the printouts will not make



it reliable. You will get inconsistent readings. The reverse is not true, however. You need to be accurate to get reliable results. If you were sloppy in placing the monitors or in reading and recording the data, you will not have reliable measurement. If the monitor records large changes in heart rate, but the effect is small, then the measurement is imprecise, and this will be true regardless of whether they are reliable or accurate.

Again, let's turn to another example to make these concepts easier to understand.



A researcher proposed the hypothesis that being in love makes those Beans ill. Suppose we operationally define the dependent variable "being ill" as an increased body temperature. To test this hypothesis, we sort participants into two groups of people—in love and not in love, the independent variable – and measure their body temperature. As the researcher, you decide to use a thermometer that measures body temperature by placing it under your arm [in your armpit] for 3 minutes. You would want to take these measurements of body temperature carefully, leaving the thermometer under each person's arm for a full 3 minutes, being careful that they keep their arm down the entire time.

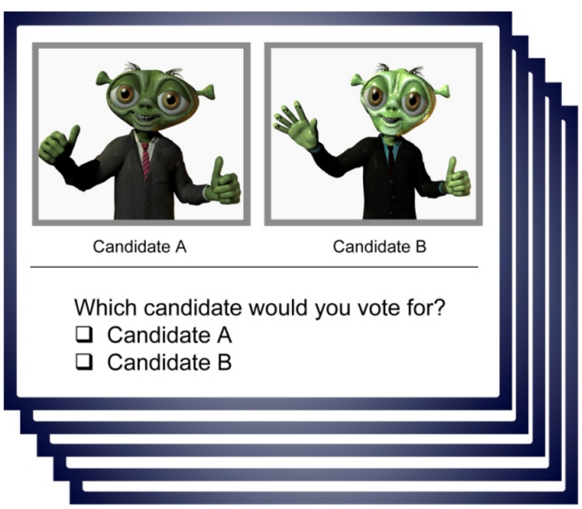
Suppose also that under-the-arm thermometers measure body temperature to the nearest degree – 97, 98, 99, and so on. What if body temperature really is raised when people are in love, but the amount it is raised is about 0.3 of a degree? You could miss the entire effect because the measure you are using is not precise enough for this research. The under-the-arm method of taking body temperatures could be just fine if you were measuring increases in body temperature caused by the flu because they would alert you to a high fever, and measurement to the nearest degree would

I hope you are confident in your understanding about reliable, accurate and sensitive dependent variables because we build on them in the next chapter.

be accurate enough for detecting a high fever.

Chapter 3:
Dependent Variables are Scored Objectively

In thinking about dependent variables, it is also important that they be **scored objectively.**
Here's an example. The Human Beans tend to elect their leaders – weird, I know. Let's say I am a Fuath Scientist who wants to learn about the election process. So, if I wanted to know if voters are more likely to prefer an older candidate, I could put together different combinations of two photos and ask the Bean participants in my study to indicate which candidate they would vote for, if they were both running for the same elected office. There are many ways to design this experiment. One way is to put pictures of two hypothetical candidates on each of 5 pages (10 photos in all). Go to the next page to see an example.



In this experiment the dependent variable is the preference for older candidates versus younger candidates. One way to determine preference for the older candidate is to count the number of times each participant selected the older candidates and compare that number to the number of times the younger candidates were selected. I would add the number of times each participant selected either the older or younger candidates. This count would be the **dependent measure**, which corresponds to the *dependent variable*. Although there is a distinction between the terms "dependent variable" and "dependent measure" it is not important for this training course, so I use the terms interchangeably. Since there are a total of 5 choices that a participant can make, I know that this is the maximum number of times an older or younger candidate can be chosen. So, each participant's score would range between 0 and 5 for either the older or younger candidates.

OK, so, just to make sure you understand the study design. What is the *independent variable*?

The independent variable is the age of the candidate.

What is the *dependent measure*?

The dependent measure is the actual count of the number of times the older or younger candidates were chosen.

So to test the hypothesis that Beans prefer older candidates over younger ones, we'd compare the number of times the older candidates were chosen to the number of times the younger candidates were chosen.

Suppose I collected data from 100 Beans, and now I need help scoring the responses. I ask two assistants, Quark and Blorq, to take half of the booklets and score them. Assuming that these two assistants are pretty careful, should it matter who does the scoring? Is it likely that both scorers would come up with the same score for each participant? The answer is yes because this is an example of **objective scoring**. **Objective scoring occurs when scoring the data doesn't require any judgment of quality on the part of the experimenter.** With objective scoring we should get the same results regardless of who does the scoring if they are being careful. In the candidate example, Quark or Blorq would just count up the number of times the older or younger candidates were selected by each participant. We should get the same results regardless of which of these fine Fuaths does the work or if they decide to split it up and each do part of the work of scoring.

What is likely to happen if, instead of asking each participant to select the preferred candidate, I asked them to write a short essay describing each candidate? I might do something like this if I wanted to know if participants described the older candidates with more favorable terms, such as competent and trustworthy. I would score these essays by counting up the number of favorable statements made about each candidate. Of course, scoring these essays would require some judgment on the part of the scorer because sometimes it would not be clear whether a statement is favorable, neutral, or unfavorable. For example, if someone wrote that one of the candidates "looked ok," it would be hard to know if this was intended as a positive statement or perhaps a neutral one. This sort of scoring is called **subjective scoring. Subjective scoring requires judgment on the part of the scorer, and two people may get**

different scores after reading the same

____
____

The important point to get here is that although objective scoring is desirable, it's not always possible. You shouldn't necessarily question the quality of the study if it has **subjective** scoring. It's a big misconception to think that **objective** scoring is ALWAYS better than subjective scoring. However, you need to make sure that the experimenter took steps to make sure that the subjective scoring wasn't biased in any way and that more than one person agreed on their subjective

essay.

judgements. This is really important to look for when evaluating these kinds of studies.

I should mention that one way to obtain objective scoring is to have an instrument do the recording – like a thermometer, a scale, or a computer. There are a zillion other types of instruments. Contrary to popular belief, instruments are not always perfect. The term **instrumentation** refers to errors that can occur over time as an instrument becomes less accurate. Springs loosen; parts wear down and so on. The term *instrument* in this case can refer to a variety of things, such as machines, chemicals, and even Fuaths or Human Beans!

Consider the following example:

Remember the huge salt scare years back on Thoth when the townspeople of Gorth thought there was salt water getting into the only fresh water lake nearby? Fuaths all over town reported a very salty taste in their water. This obviously developed into mass panic because too much salt dries us out and makes our eyes collapse. So, scientists from the University of Gorthland measured the salt levels in the lake water using their saltometers. They showed that there were in fact significant levels of salt in the fresh water lake. This suggests that the subjective scoring made by the townspeople matched the objective scoring of the saltometers.

As you may recall from the newspaper accounts published at the time, the salt measurements from the saltometers decreased over time, but the townspeople kept reporting saltiness. The scientists kept saying that they were just imagining it. Well, it turned out that the saltometers were no longer providing accurate measures of the salt levels in the water because they were corroding from all the salt. The instrument became less accurate over time while the lake was still being infiltrated by salt from a local salt mine. Instrumentation had struck! Fortunately, we discovered the source of the salt, and the crisis was averted.

Let's conclude with a reminder: objective scoring is fine, but be aware of instrumentation, which makes the scores less accurate with time.. Subjective scoring is fine when raters agree and when biases are absent or taken into account.

Chapter 4: Experimental and Control Groups

Sometimes it is hard to know what is really causing some outcome. Have you ever noticed that the richest Beans are also the healthiest? It's true, but why? Is it because they can afford healthier food or because they have more time for exercise or maybe they are better educated about how to be healthy. To decide which of these explanations is best, we would need to study one of these variables at a time, while controlling the other two. To establish control is to isolate and identify variables that might affect the results. In this chapter, I will talk about one way to establish control which is measuring variables that could affect or otherwise clarify the results. Let's start with an example – I know, UFOs are a great example! UFOs stand for Unidentified Flying Objects. But when the Beans talk about UFOs, they are really talking about beings from another planet who are flying to earth. (Of course the UFOs that the Beans see are not Fuath ships because those silly Beans do not have the technology to detect our transportation devices).

Some Beans have seen UFOs. In one case, 16 people from one town reported seeing a UFO, whereas 25 people from that town did not report seeing the UFO. From the perspective of the Bean scientists, does that mean that UFOs actually exist?

Before you say 'yes,' consider that some Beans who see UFOs often have consumed large quantities of alcohol, a drink that makes the Beans even odder than they already are. Is it possible that alcohol creates the tendency to see UFOs when there are none? If the answer is 'yes' then perhaps UFOs do not exist. If the answer is 'no' then UFOs might exist. A Bean scientist would want to 'control' alcohol in some way to rule out this alternate explanation.

So, one possibility is that the Beans who claim to see UFOs are actually under the influence of alcohol. To test this possibility, a researcher would want to control for drinking alcohol. There are lots of ways to do this. One way would be to find people who report seeing a UFO and for each of these people, find someone else who was nearby at the time of the UFO spotting. Then ask all of the participants in this research if they consumed alcohol one hour before the sighting. Suppose you found that most of the people who reported seeing a UFO had a specified amount of alcohol just prior to spotting the UFO, and the people who were nearby but did not see a UFO did not drink alcohol at the time. You can think about this research design in the following table.

	Drank Alcohol	Did Not Drink Alcohol
Saw UFO	14	2
Did Not See UFO	3	22

If you got the results shown in this table, then you would conclude that it was probably drinking alcohol that caused people to see UFOs (although we wouldn't know for sure). In this way, you would have **controlled for** drinking alcohol because you have isolated the effects of alcohol on seeing UFOs.

Here is a key idea: when researchers control for something, they eliminate possible confounds. **A confound is something that could really be causing the effect researchers found, but they might not know it.** In this example, people who saw UFOs (one condition) were more likely to have drunk alcohol and those in the "did not see UFO" condition were less likely to have drunk alcohol. Thus, drinking alcohol is confounded with whether or not someone sees a UFO. By controlling for and eliminating confounds, scientists get closer to the truth.

Let's consider a common way of exerting experimental control—the use of control groups.

You have probably heard the term **control group**, but you may not be sure about what it actually means. **A control group is a group of participants (could be people, other animals, plants, or anything else depending on what we are studying) that does not receive the treatment that we are studying**. Control conditions are needed to see if the independent variable causes differences on the dependent variable. (Quick check on your comprehension—Do you know what independent and dependent variables are? If not, go back and review chapter 1; independent and dependent variable are key concepts, so be sure you understand these terms.)

A control group is a critical concept to understand when evaluating experiments. Let's start with an example.

Many Beans, especially those in high school and college, believe that by drinking energy drinks, they have more energy and do better on tests. To test this hypothesis, we could find people who drink the energy drinks, ask them about how energetic they feel and collect information about their grades. Even the dullest Fuath should be able to sense that something is wrong with this approach. If canned drink enthusiasts believe that the drink gives them more energy, then they will rate themselves high on energy, but their energy ratings might reflect their beliefs about the drink and not the actual effects of the drink. Grades would not be useful in determining if the drink affects grades because we would have nothing to compare them to.

So, to develop a sound study that tests the effect of energy drinks on perceived energy and grades, we need at least two groups of participants.

One group is called the **experimental group. An experimental group is a group of participants who receive the intervention or treatment in which we are**

interested. The other group is the control (or comparison group), and this group should be as similar to the experimental group as possible.

In the study we are discussing, consuming energy drinks is the independent variable. We want to know if it affects the dependent variables, which are ratings of energy and grades. To test this hypothesis, we could recruit a group of participants for our study. Because we want to know about students in high school and college, we would use students who are in high school or college as participants. Each participant would be assigned to one of two groups. The two groups would be as similar as possible, except that one group would drink the canned energy drinks and the other group would not.

The group of participants that drinks the energy drink is the *experimental group,* and the group that does not drink the energy drink is the *control group*. To make the control group as similar as possible to the experimental group, they would drink something from a can that looks the same as the one from which the experimental group drinks. No one would know what they are drinking because we want to know if the drink (and not expectations about the drink) is responsible for ratings of energy levels and grades.

Some people believe that the term "control group" means that the participants in this group do not get any intervention at all, but since they drink from the same sort of can as the experimental group, you can see that this is a misunderstanding. They do not get the independent variable of interest, which is the actual drink, but in order to be a good comparison, they usually get some intervention to make the groups as similar as possible. In this case, similarity is achieved by having both groups drink something. Many Bean medical researchers do something similar by having control groups receive a "placebo" (or fake) treatment.

In running the experiment, we would give all participants similar-looking cans to drink from, being sure that they all drink the same amount. Then wait a short time, let's say 15 minutes, and ask them to rate how energetic they feel, perhaps using a scale in which 1 = like a slug and 10 = like an Olympic track star at the start of a race. We would then give them some test, perhaps a test of scientific reasoning, and see if the participants in the group that drank the real stuff that comes in the cans scored better (or worse) on average than those who drank something like flavored water from the other can. We can compare the scores from the control and experimental groups, using appropriate statistical tests, and determine if the ratings and test scores of the two groups can be considered different from one another.

It's pretty important to point out that we controlled for beliefs about the effects of the energy drinks by having participants in both groups drink from similar-looking cans and perhaps by telling everyone that they were drinking the energy drink. Thus, the use of a control group and the ways we made the groups think they were drinking the same thing are both types of experimental control.

Let's stop for a minute (the Earth 60-second kind, not the Fuath minute of 549 Fuath days) and think about the importance of having control groups.

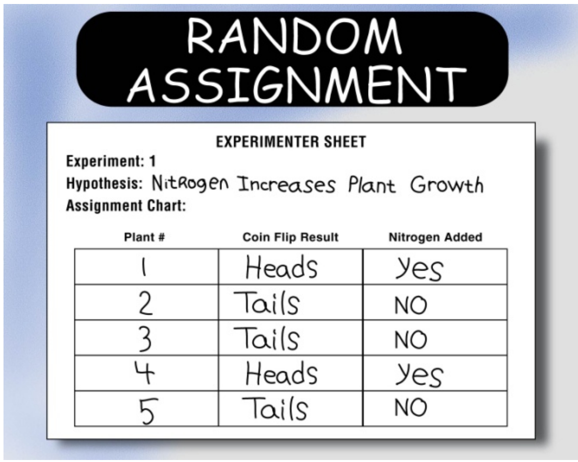
Sometimes, when reading popular Bean magazines, it seems that it will be easy to take over the puny planet Earth with bad science. There are so many times when Beans believe almost anything without even thinking about whether there is a control group to show if it works. People are willing to pay lots of money for products that claim to help them become more popular, earn more money, get better grades, and feel healthier. Sometimes, these claims are supported with information about people who used the products and got the desired results. Consumers of these products and programs must ask for—actually they should demand —evidence from experimental studies that used a control group. Without these experimental controls, the odds are good that the product or program will not do what it promises. So remember, control groups are essential when evaluating any claim.

The Beans conduct research because there is something they want to know. There are many methods for conducting research, and the method researchers use in a study depends on what they want to know. For example, if they want to be able to describe behavior, perhaps they can make careful observations of how people behave when they are a minority in some group. But, most often the Beans want to know if something (some variable) causes some other thing (another variable) to change. Does taking vitamins cause people to have fewer colds? Does sticking tiny pins into people, they call this acupuncture, reduce the pain of a migraine headache? Does keeping up with the reading for a class consistently throughout the semester cause better grades than other methods of studying? Those Beans want to know about all sorts of possible *causal* relationships.

If you want to know if some variable caused a change in another variable, one of the most important things to remember has to do with the way participants are assigned to the different groups – the *experimental* and *control* groups. Let's just do a "quicky review" before we go on to discuss what is important about the way participants are assigned to groups.

When the Beans conduct experiments, they are concerned with two types of variables. The independent variable (Yup--here it is again. It is so important to review this term!) is what the experimenter manipulates or changes to see if it has some effect on the dependent variable – what is being measured (e.g., participants' behaviors or maybe on how they think or feel). If the researchers wanted to know if a new "memory drug" (independent variable) really helped people remember better (dependent variable), they would give some participants the drug and others would get something that looks like it might be a drug, but really isn't--maybe a pill that is made of sugar. (Of course, they would have to be careful that whatever is in the fake pill does not affect memory.)

In this example, the experimental group would get the memory drug, and the control group would get the fake pill. The dependent variable is the measure of memory the experimenter uses. In this example, the participants might take their pill, wait an amount of time equal to how long it would take for the pill to "work," then the experimenter would read them a long list of words. The dependent variable in this example would be the number of words that are recalled. On average, if the memory drug really improved memory, we would expect that the people in the group that got the drug would recall more words than those in the group that got the fake pill. So far, all of this is review, so it should be fairly easy to understand. But even if the experimenter conducted the research this way, she could not conclude that the memory drug caused better memory. She needs to do one more thing before she can make a causal claim: She needs to assign participants to either the control or experimental group at **random**, which is referred to as **random assignment**.



Random assignment is a procedure for assigning participants to the experimental and control group so that each participant has an equal chance of being in the different groups. **Random assignment is one important way that experimenters can ensure that the participants in the experimental and control groups are similar in all respects except for the manipulation of the independent variable** (whether or not they get the memory drug). Let's think carefully about what this means.

INSERT A VIDEO OF ZLOTSKY RANDOMLY ASSIGNING FUATHS

If you were a Bean, we would say it is time to put on your thinking cap.

Those Beans are so silly. Do you think they really believe that putting a cap on their head will help them to think better? Probably not--it is just a colorful expression that they use, probably because of their boring lives, but then they must feel so bad about being mere Beans and not superior like us Fuaths. Sigh. Yes, I feel sorry for them too.

Let's suppose that you are planning an experiment to test a memory drug. You need to find participants for your study. Let's suppose that the teachers at a local high school agreed to participate in your research. After all, who would be more interested in finding out if a memory drug really worked than teachers? You also found a second group of willing participants--students at a local college that teaches students how to curl hair and apply that waxy stuff those Beans call make-up. Schools like that are called Beauty Colleges because the students study how to make Beans look more beautiful (as if there's any hope for that!). It would be easiest if you put all of the high school teachers in one group and all of the beauty school students in another group and then performed the experiment the way we just discussed.

Can you think of a BIG, HUGE, HUMUNGOUS problem with this decision? Type the problem in the box provided:

The problem with putting all of the teachers in one group and all of the beauty school students in another group is that on average, these two groups may not be the same in how well they remember information. It seems that the high school teachers would be better because they might have more practice memorizing things like lectures and names of students. But the opposite could also be true. The best way to reduce possible differences between the control and experimental group that exist before you start your experiment is to **assign participants at random to each group – this is random assignment.** With a two-group design like the one we are planning in this study, you could flip a coin for every participant to decide which group the participant will join. With this method of random assignment, the people in the groups will differ in many ways--some will be high school teachers, some will be beauty college students, some will be tall, some will be rich, and so on. But, the two groups will not differ in a systematic way that would bias the results.

When we use random assignment of participants to experimental and control groups, we can make causal claims, such as the memory drug caused better remembering, if we found that the group that took the memory drug performed better than the group that took the fake pills.

Of course, we would need to use statistical techniques to decide if the differences in performance were due to chance variations between the two groups. If our statistics support the conclusion that the two groups performed differently on the memory test and the experimental group performed better, then we could conclude that the memory drug caused people to have better memories. You will not learn about statistics in this book but we recommend that you read “Introduction to Statistics” in your spare time. (Really, it’s a great book.)

As the Beans say: Here is the bottom line (weird expression they like to use when drawing a conclusion—bottom line of what?!)--If you want to be able to make a causal claim, the study must use random assignment of participants (or animals, plants, etc.) to groups.

Chapter 6: Experimenter Bias

There was once a popular television show about a smart talking horse named "Mr. Ed". You might think the Beans actually believed there really were talking horses. But no, even the Beans knew way back then that horses do not talk.

Nonetheless, over 100 years ago in human time (approximately one heckle-era in Fuath time) there were a lot of Beans who believed that there was a horse that could answer questions. The horse could even spell words and do arithmetic! This is the true story about a horse known as Clever Hans. The horse lived with his owner in Germany; Hans is a common name (even for a horse) in Germany. The "clever" part of his name came from his high intelligence.

Here is the story: Around the turn of the 20th century, a German school teacher named Mr. von Osten showed off his very smart horse. Of course, not even this smart horse could talk, so he communicated by tapping his hoofs. One tap was either the number one or the letter A; two taps were either the number two or letter B, and so on. Amazingly, this horse could even calculate square roots! That must have been some horse!

Two psychologists were skeptical about this amazing horse. Being skeptical is a good thing. Skeptical means keeping an open mind, while also looking for evidence that something that seems too good to be true really is true. At first, it seemed that Hans the Horse was really as clever as his owner said he was. The horse could even answer questions that were asked in languages that he had never heard before!

How would you go about testing whether Hans really could answer questions about math, geography, and other subjects?

Take a minute or two to think about this question and type your answer in the box below. (Try your best, but know that this is a really hard one.)

Let's see what the psychologists did to study the amazing thinking horse, Clever Hans.

They did several different experiments, each time trying to figure out how Hans knew the answer to difficult questions. It seems that Hans was generally correct except in two different circumstances.

Hans could not answer questions when he could not see the person who asked the questions or when the person asking the questions did not know the correct answer. Can you use these facts to figure out how Hans was able to perform so well when he could see the person asking the question and when the person who asked the question knew the correct answer?

It seems that when the person asking the question knew the answer, he or she would make some slight movement that Hans could use when answering. Suppose, for example, that you asked Hans an easy question, like "What is the square root of 9?" Hans would begin tapping his foot. At the third tap, you might smile or tilt your head or raise your eyebrows in a way that indicated to Hans that this was the correct answer and he should stop tapping his foot. These slight movements were not done intentionally, but instead were part of the usual way we communicate nonverbally with each other, and apparently with horses as well. Now, Hans really was pretty clever because he learned to look for small nonverbal gestures that signaled to him to stop tapping, but he was not really able to calculate square roots, spell, or do other cognitive tasks. The term **experimenter bias is used to describe the ways an experimenter can unknowingly alter the response from any participant in a study.** A bias is a predilection or slant in a certain direction. In the case of Clever Hans, the person asking the question was biased in that she or he made certain gestures that communicated the correct or expected response. When the person asking the questions was behind a wall so that Hans could not see these small gestures, the horse could not pick up on the biases and he could not answer questions correctly.

____

One way to prevent experimenter bias from influencing the outcome of any event is to use **blind** or **double blind** procedures. A blind study means that participants do not know what group they are in (Experimental vs. control) whereas double blind means that both the researcher and the participants do not know what group the participants are in during the study. Examples of these two concepts will help you understand: Earlier you learned about "placebos," which are substances or procedures that are not expected to have an effect on an outcome. If you had a cold and you wanted to know if chicken soup really helps people recover from colds sooner than not taking chicken soup, you could design an experiment in which some people with colds eat chicken soup and other people with colds take some placebo treatment--perhaps they drink water that is the same color as chicken soup. If the participants are unaware of their group membership (real chicken soup or the placebo), the research design is called "blind" because the participants are blind to (do not know) whether they received the real chicken soup treatment or the placebo.

When neither the participant nor the experimenter knows which group the participant is in, it is called "double-blind" because both of them are "blind" to the condition. This sort of design is one way around the problem of experimenter bias. When the experimenter is unaware of the participant's group membership, she or he cannot use subtle cues that might suggest to people with colds whether they felt better sooner in one condition or the other.

Let's try one more example of experimenter bias. Keep in mind that with experimenter bias, the experimenter is NOT deliberately giving cues about how to respond. The small gestures and movements are done without the experimenter being aware that they are happening.

Let's consider an example from chemistry and medicine. Homeopathy is a theory of treating illnesses in which a person who has some illness is best treated with very, very, very dilute medicines, usually poisonous substances. The idea is that a chemical can work to cure a disease even when it is given in very tiny amounts. How tiny? By one account, you would take one drop of medicine (often a poison) and dilute it with 100 drops of water. But with some poisons that would still be too concentrated and cause harm, so dilute that solution again with another 100 drops and so on. How long would the dilution go on? The most common solution is diluted 30 times. According to one source, that would be equal to the number 1 divided by 1 followed by 60 zeros (1/1000000000000000000000000000000000000000000000000000000000000)!

How can such a dilute solution of anything still work? According to homeopathy, water has memory, so it can remember that it started with one drop. Therefore, it doesn't matter that the substance has been diluted. It should still work.

So, imagine that you were ill, and an experimenter who believes in homeopathy gave you this solution and asked you to record how you felt every day for a week. Do you think that the experimenter would verbally or nonverbally communicate their beliefs to you? Probably so. Do you think these communications would affect how you rated your day-to-day feelings? You bet it would!

By now, you should be able to think about ways to get around this bias to create an unbiased test of whether the solution worked. One way is to have the experimenter not know the solution - homeopathic medicine or just plain tap water – that is given to each participant. If the experimenter is blind to the condition, then he or she cannot bias the participants and affect the results.

In summary, to prevent experimenter bias from contaminating the results, the experimenter should either not know the hypothesis or be blind to the conditions.

By the way, there have been many tests of homeopathy, and when they are done correctly, there is no evidence that they work. James Randi, a show person who loved to debunk myths offered a one million dollar prize for anyone who could prove that homeopathic drugs worked beyond placebo effects. No one has yet claimed that prize.

References:

Description about diluting homeopathic medicines was retrieved from Digital Bits Skeptic on April 8, 2009 from http://www.dbskeptic.com/2008/05/12/homeopathy-diluted-and-deluded/

CH7: Participant or Subject bias

Every year at about the same time, we Fuaths line up to take our annual exams. We heard that on Earth, exams are only taken by people who are in school or who want to go to school. Frankly, I don't understand why only some Earthlings take exams. All Fuaths at every age take exams every year. Even Granddaddy Sheldon Fuath takes exams, and he is older than Earth-dirt. This is just the way things are on Thoth. But, I have wondered if Fuaths might learn better if they did not take exams every year. This is what the Beans call an empirical question, which is a fancy way of saying that we could find the answer by conducting an experiment. But, what sort of experiment would it be?

One possible experiment designed to answer the question of whether we learn better when we have to take exams or when we don't would be to get a group of Fuaths, teach them all the same thing, but test some and not others. We could then determine which group learned better. Does this sound good to you? Hmmm, I think there is a problem, but I can't put my arm-extension digit on it.

Oh, I see the problem now. How will I know which group learned better if I only test one group? Now that is a problem. But, there must be a solution.

Suppose I change my research design just a bit and give a written exam to one group of the students and not to the other group. Then, about a week later I phone students from both groups and ask them some questions about what they learned. Would that work? Would this design answer the question of whether taking an exam or not taking an exam produces better learning?

- Yes, it sure will
- No, it won't

I agree with you if you said “yes, it sure will.” The idea of giving some students an exam, others no exam, and then calling them at home a week later will help us answer the questions about whether taking an exam or not taking an exam helps students learn better. So far, so good. Let's plan the study.

I hate to take exams, so I would certainly sign up for the group that doesn't have to take exams. My oh-so-nerdy friend Anna Fuath will sign up to be in the group of Fuaths who take an exam. She is like that--always gets good grades and always showing them off. The way she is going, she will probably end up as president of Lemon, that new computer company that makes computers that are so small they will fit in your wallet.

My research plan is complete. I will ask Fuaths to sign up for a class in astral-physics-and-rap-studies. They can decide if they want to take an exam or not take an exam when the class is over. When the class is over and the students in the exam group have taken their exam, I can call each Fuath at their domicile and ask them questions about astral-physics-and-rap-studies. Then, by comparing the number of correct responses in each group, I can find out if taking an exam or not taking an exam is good for learning.

Here is a diagram of this research design:

Group 1: Takes an	Group 2: Does not

exam	take an exam			

What is the independent variable in this design and what is the dependent variable? Write in your responses in the box below:

The independent variable is   The dependent variable is

Let's see how well you did. You wrote:

The independent variable is

The dependent variable is

The correct answers are
The independent variable is the exam group--there are two levels (or conditions) for this independent variable--(1) take an exam or (2) do NOT take an exam.

The dependent variable is the number of questions answered correctly when Fuaths are phoned at their domicile after the course. If we asked them 20 questions, then the value of the dependent variable could range between 0 (no questions answered correctly) and 20 (every question answered correctly). We could compute the mean (the word "mean" is the same as "average"--those research types need special words for everything) number correct for each group, use statistics to determine if the mean difference between groups is statistically significant and then, Voila! (French makes me sound so-o-o sophisticated), we can answer our research question.

So, are we ready to start our experiment?

You probably guessed that this was too easy to be true. Have you already thought about a problem with this study? Here is a big hint--it is the problem we talked about in the last chapter. To be specific, this study would not involve *random assignment* when placing the participants into groups. Instead, they would get to choose which group to join.

This is a poor study design because any significant results could be due to **participant or subject bias.** (Earth scientists often call the participants in an experiment 'subjects'.) **Participant (or subject) bias means that rather than the experimental manipulation strictly causing the outcome in the study, the participants themselves also affected the differences found between the groups.** So, subject bias occurs when the subjects bias the results. Subject bias can only occur when the experiment is testing human behavior, not when they are animals or plants because plants and (nonhuman) animals can't guess the researcher's hypothesis or act in ways that will make them look good.

There are different ways in which subjects can bias results. One occurs when the participants (the subjects) get to choose which group to join. This is called **self-selection**. Self-selection is a type of selection bias. Selection bias violates random assignment to groups, which is SO important! The problem with self-selection is that the participants (e.g., Beans, Fuaths, monkeys, etc.) who join one group might be very different than those who join another group. Consequently, the researcher would not be able to know if the differences on the dependent variable occurred because of the independent variable or because of the different participants. There is a lack of control.

Another type of subject bias occurs when participants act differently because they KNOW that they are being studied. I mean, wouldn't you? You might like the experimenter and think you know the purpose of the experiment, and so you might try to do things that you think are expected of you. Or perhaps you are feeling nasty, and perform in a way that you think would mess up the study. Participants can figure out the purpose of a study by paying attention to cues in the environment. These cues are called **demand characteristics** because they can place a demand on what types of behaviors are expected from the participants, and they are something that Bean psychologists and sociologists have to consider in their studies.

First, let’s take a closer look at self-selection.

Would you guess that Fuaths who sign up to take an exam are different in some unknown ways from those who do not choose to take an exam? They probably are, even if we cannot specify the differences in advance of running the experiment. Maybe Fuaths who want to take an exam tend to study harder or get better grades than those who would avoid an exam if they could. Don't you think so? This is an example where the participants in a study (remember that sometimes we call them subjects--who knows why--maybe because they are subject to what we tell them to do) will probably differ between groups in systematic ways that are unrelated to whether or not they actually took an exam.

Let's try another example, just to be sure that you have the main idea. Suppose some doctors wanted to know if exercise helped older Fuaths with the creaky joint problems they get as they age. To test this hypothesis (more exercise causes fewer creaky joint problems), they offer an exercise program to the old Fuaths. Some agree to participate and some don't. After a few months of exercise, the researchers compare the amount of joint-creakiness for those who exercised to those who did not agree to exercise.
Sounds good--right?

Ok, guess not. Older Fuaths who joined the exercise program are probably different from those who did not join. Maybe the ones with really creaky-joints did not want to do exercise, so their joints were creakier than those who exercised before the exercise program even began. You can probably think of other ways that the exercise group differed from the no exercise group before the experiment even started.

Both of these examples provide evidence for the dangers of allowing subjects to choose their group membership rather than using random assignment

---

Ok, now let’s talk about an example of *demand characteristics*.

You may recall the transportation scare that occurred during the later part of the Klorth eon. Many Fuaths began reporting that their pod-mobiles seemed to have trouble moving as quickly as they had in the past. There was concern that Earth-like oxygen was getting into our atmosphere each time a Fuath rocket came back from Earth, and the oxygen was clogging our machinery. That’s when the researchers stepped in.

A study was conducted comparing the speed of pod-mobiles in our normal atmosphere and those in a large underground Earth-oxygen-rich container. Although it was found that the pod-mobiles in the underground container moved slower than those in our normal atmosphere, there was concern that the drivers in the underground group became aware of the purpose of the study because they noticed an “anti-Earth travel” sign on the entrance to the underground container, and they were motivated to purposely drive slower. In fact, several drivers reported later that they didn’t approve of the Fuath interest in taking over the Earthlings, and they thought trips to Earth should end. This suggested that the differences in speed that were found might have been due to *demand characteristics*. The underground group knew the purpose of the study; therefore, they behaved in ways that ensured support for the hypothesis. In other words, they purposely drove slower to show that the oxygen really was slowing the cars in the hopes that this would end the trips to Earth. Needless to say, a follow-up study showed that it was the percentage of Thoth-worms in our mud that was slowing down the pod-mobiles and not the Earth-oxygen. What a relief – I love my pod-mobile!

As you can see, researchers need to be aware of both demand characteristics and self-selection when they set out to design a well- controlled study. And, when you read about an experiment, you need to look for the possibility that subject bias is responsible for the results. If you find that it is, then you cannot trust the findings from the study. Here is something else to add to your list of "bad science detectors."

Ch8: Sample size, attrition, and mortality

Research is about finding answers to questions. Of course, there are lots of ways to find answers to questions. When Beans follow the method described in this book--the scientific method--they can have more confidence in the answers they come up with than if they use other methods. Some methods that are NOT scientific are guessing, flipping a coin, or just making up an answer that sounds good.

Here's a great question: Did you know that some Beans carry around the foot of a rabbit because they believe that it will make them lucky? Sometimes, the foot is not from a real rabbit, but they believe it will bring them good luck anyway. Can you guess what they call it? Okay, they call it a "Lucky Rabbit's Foot." I wonder, "Are Beans who carry this odd furry thing really luckier than those who don't?" Think about how you would go about finding an answer to this question.

After you have thought about how you might answer this question, then click here □

----Did you really think about this question?----

One way to find out if people who carry a Lucky Rabbit's Foot are luckier than those who don't might be to find people who carry one and ask them if they are lucky people.

That might work, but there are problems with this plan.

If people who carry a Lucky Rabbit's Foot say they are "pretty lucky" that would be interesting, but we still would not know if they are luckier than people who don't carry a Lucky Rabbit's Foot, so we would need to find a group of people who don't carry one and ask them the same question.

In this example, we probably can't just ask people if they are lucky or unlucky, or even ask them to rate how lucky they are, because people who carry a Lucky Rabbit's Foot may *think* they are luckier than those who don't, but they really may not be.

We could try a different way of determining who is or isn't lucky. For example, we could have people in each group buy a lottery ticket and then we could see if the people in the Lucky Rabbit's Foot group won more money than those in the group that doesn't carry a Lucky Rabbit's Foot.

Regardless of how we decide to test for luckiness, we know for sure that we will need to find people who carry a Lucky Rabbit's Foot and those who don't in order to get an answer.

This brings us to an important issue. Our question in this example is about all of the Beans who carry a Lucky Rabbit's Foot and all of the Beans who don't. There are a lot of Beans on Earth. We could never find all of them to find out how lucky they are. The group that we want to know about is called a **population**. In this example, there are two populations: the population of all Beans who carry a Lucky Rabbit's Foot and the population of all Beans who don't. So, as you can see, that pretty much includes everyone who lives on Earth. Even though we want to know about everyone in each of these two groups, we can't find them all and ask them what we want to know. Because we cannot use everyone in our populations of interest, we need to select a subset of people from each population. The group that we select for our study is called a **sample**. The process of selecting people who will be in a sample is called **sampling**.

The process of sampling is very important, so let's consider another example.

Any time we think about sampling, we need to decide how many participants

we will need in our study. The number of participants in a study is the sample size. Any discussion of sample sizes always takes me back to the interesting findings reported by Dr. Oshlosher (best known as Dr. O) from the Gothland Research Institute a number of years ago. As you may recall, Dr. O was fascinated by the differences in the height of a blue-leafed native plant called the Blupblop. The Blupblops were found in both the northern and southern Thoth hemispheres. It seems he was convinced that a recent change in the air quality on Thoth (particularly in the southern hemisphere) was the reason that these plants landed on the endangered species list at that time. Now we know that he was correct, but back then it stirred quite a controversy.

Dr. O suspected that no one was capable of collecting information on the entire populations of northern and southern Blupblops. It just seemed like too large of a task. As it turned out, the populations were so small that Dr. O was able to assess each Blupblop in each population. Below are the data from his original investigation:

Height (in Earthling inches) of all northern blue-leafed Blupblops

6	5	8	5	6	7
7	6	4	7	6	6
5	7	4	5	6	8
6	6	5	8	7	4
8	6	6	5	4	7

Height (in Earthling inches) of all southern blue-leafed Blupblops

5	8	19	2	4	7
10	2	4	1	3	2
8	6	3	5	13	1
1	9	7	15	4	7
14	2	5	9	1	3

As you can see from these numbers, there is variability among the heights of the Blupblops. **Variability** refers to the amount of difference between the participants, subjects, or observations in samples or in populations. Right away, you probably noticed that the northern plants range in height from 4 to 8 inches, but those in the southern hemisphere range in height from 1 to 19 inches. Therefore, in both populations, Blupblops are variable. You probably noticed that the populations differ in their amount of variability. Which of the two populations appears to be more variable (shows greater differences among the heights) than the other?

Type answer here:

The correct answer is that the southern blue-leafed Blupblops have more variability in their heights than those found in the northern hemisphere. What's especially interesting about this is that Dr. O also learned that the average heights of the northern and southern Blupblops were identical. Both populations had an average height of 6 inches! This may seem odd considering the difference in their variability, which is what led Dr. O to consider these findings further. He decided to take a variety of samples from the two populations beginning with 1 plant per sample and gradually increasing the sample sizes. (He tells us that he also picked each plant in a sample randomly.) He did this procedure in order to look at how the sample averages changed between the two groups of Blupblops as the sample sizes increased.

Here is what he found:

Sample Size	Average height for a **sample** of northern Blupblops	Average height for a **sample** of southern Blupblops	Average **population** heights for Blupblops northern and southern	
1	7	10	6	6
2	4	4	6	6
3	6	8	6	6
4	5	13	6	6
5	8	6	6	6
6	6	3	6	6
7	6	9	6	6
8	6	7	6	6
9	6	6	6	6
10	6	6	6	6

Do you notice anything interesting about the changes in the averages for the samples of northern and southern Blupblops?

Dr. O's findings presented in the table above show two important things. The first key thing to notice here is that for both populations, the *sample averages were more likely to match the population averages as the sample sizes increased.* The second thing to pay attention to in the table is *that the variability within a population influences the degree that samples of different sizes will have averages that match the population average*. You should recall that the southern Blupblop population has more height variability than was found in the northern population. Dr. O's data show that this more highly variable population required larger sample sizes in order to match the population average. The key idea is this: The need for a larger sample size is especially great when the population is highly variable. So this suggests that highly variable populations are best studied using larger samples, and less variable populations can be safely studied using smaller samples. Although this guideline is helpful, we still don't know the ideal sample size to use for a given study – because we usually do not have all of the scores of a population. Dr. O was lucky because he did.

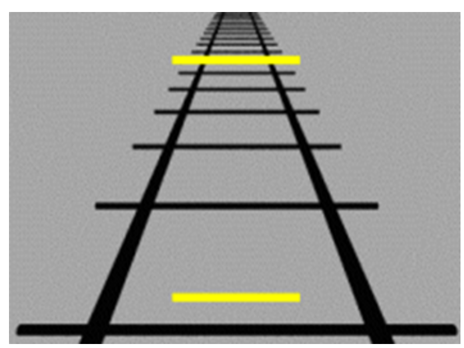
It isn't necessary for you to worry about the specific statistical tests needed for determining ideal sample sizes. This will be covered in the *Fuath Book of Statistics*. But, you can keep some guidelines in mind. For example, many researchers follow the rule that at least 20 to 25 participants are needed in each group in most studies that involve Fuaths or Human Beans.

But, sample size is not the only thing we need to think about when thinking about sampling—we also need to consider attrition and mortality.

Attrition and **mortality** sound like scary or sad concepts. And, in general, they are. But the terms have a different meaning when we talk about research than they do in other contexts. For researchers, attrition means that they lost some of their participants (or animals, plants, etc.) during the experiment. When a researcher loses participants, it means that they can no longer use their data because the participants did not complete the study. There are lots of ways to lose participants. Let's consider some of them.

Sometimes even our superior Fuath eyes can deceive us. Beans love to study visual illusions - things that are not what they look like. Researcher Beans study what people think they see when they are shown pictures that can be deceiving.

For example, here is an illusion called the Ponzo illusion. Who know why it is called this? Probably because somebody named Ponzo was the first to come up with it.



Which line is longer—top, bottom, or are the same? We could ask participants to answer this question. The independent variable is the figure as it is drawn; the dependent variable is the answer each participant gives—in this case there are three choices.

Now, who should be the participants?

We could use students at Greater Gothland University - they are a handy group. We could get a group of them together and ask them each to tell us which line is longer. So far, so good.

But, what if some students who start this study find it boring and they drop out of our study ?

In the formal language of research design, this loss of participants is called attrition. The participants who dropped out and the data that they provided would be gone. Is this a big deal? Yes, it might very well be! Do you think the students who think this is boring might give different answers from those who finished the study? It's hard to know for sure, but probably so. The loss of participants for reasons related to the study is a bad thing. Did you recognize the loss of participants who finds this boring and refuse to participate another type of subject bias? In an earlier chapter, we talked about the way subjects sign up for research as a type of bias. When we lose subjects in ways that are related to what we are studying, this is another example of bias. Again, it isn't a serious problem if only a few participants are lost to attrition, but if *many drop out especially for systematic reasons, then we have a problem*. Another example should help with this concept.

This is sort of yucky, but did you know that lots of Beans stick rolled up bundles of leaves in their mouth, then set them on fire, and breathe in the smoke? Those Bean habits always amaze me. They call this "smoking cigarettes." I have often wondered if the ones who do this strange ritual are, well, not as smart, as those who don't. Maybe smoking leaves that are on fire makes them dumber. Suppose we wanted to study this question in old Beans.

If we wanted to find out if smoking made Beans dumber, we would have to find a group of Beans and make some of them breathe in the smoke of leaves that are on fire (they call this inhaling) and make some Beans never smoke at all. Surely you realize that if we want to know about CAUSE (does smoking cigarettes cause people to become dumber), we need to assign people at random to different conditions. We couldn’t do this because it wouldn’t be ethical to make people inhale smoke that could make them dumb. But suppose we really could do this experiment – you know, determine who would smoke or not smoke for their life time. What would we need to do? What might happen?

You are probably thinking (I hope, I hope) that we would first need to provide an operational definition for smarter and dumber. Beans have all kinds of tests to decide who is smarter and who is dumber. We could use one of their intelligence tests. We could measure the intelligence of the Beans in the smoking group (let's say they had to smoke 20 cigarettes a day from age 18 to age 78) and the group that never, ever smoked. In this example, the IV is whether the Beans smoked or did not smoke cigarettes. And the DV would be performance on an intelligence test - Beans sometimes call them IQ tests.

Because we want the smoking group to smoke cigarettes for 60 years, we would have to measure their intelligence at age 78 if they started smoking at 18. But wait an Earthling minute! What if more ple in the group that smoked cigarettes for 60 years died before they were 78 than the people in the group that never smoked cigarettes? In this example, we would have attrition that is related to the variable we are interested in studying. In this case, the attrition is caused by increased mortality in the smoking group. **Mortality occurs when subjects die during the course of an experiment**. So, in this example, it appears that the reason for the mortality was related to smoking - the variable we wanted to study. If many of the participants did not complete the study because they died, it would be very hard to determine if our hypothesis was supported.

Ch 9: Sample selection--Representative of the Population

Although most Beans tend to look pretty much alike, after you live with them for a while (and get used to their humany smell), Fuaths like us begin to notice that Beans come in different sizes and shapes. I consulted the historical records that we have on Beans. It seems that over the last several hundred years, they have grown taller. I wonder if being a tall Bean provides an advantage. Maybe tall Beans earn more money or are more successful at work using other measures of work success.

It is important to understand that experiments are usually about AVERAGE differences between groups. Even if we found out that on average, tall Beans make more money or are more successful at work than short Beans, there might still be many poor and unsuccessful tall Beans and many rich and successful short Beans.

How could we find out if tall Beans earn more money than short Beans?

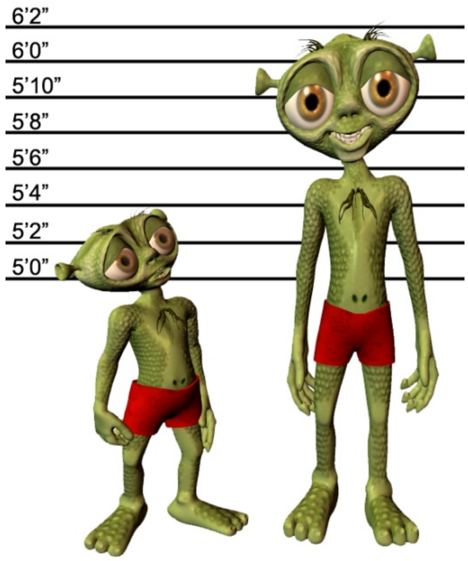
In thinking about this question, what are the populations we want to know about?

Type your answer here:

In the example about tall and short Beans, the populations we are interested in are all tall Beans on Earth and all short Beans on Earth.

No one can find all the tall Beans and all the short Beans on Earth and find out how much money each one earns, so we will have to *sample* from each of these two populations.

First, we need to *operationally define* "tall Bean" and "short Bean" which means we define it in a way that we can measure it. Suppose for this example, we decide that any Bean taller than 6 feet is tall, and any Bean shorter than 5.4 feet is short. In this example, we are leaving out middle-height Beans.



Even though researchers might be interested in all of the Beans on Earth, they really cannot sample from this population in a way that is **representative** of the population. **A representative sample has characteristics that accurately reflect characteristics of the population of interest.** In this case, the population is all Beans on Earth. A representative sample is needed in order to be able to adequately generalize research findings from the sample to the larger population. Because of this, it is sometimes more realistic to sample from a smaller population and accept a more limited generalization of findings. So, to make things easier in our example, suppose we decide that we really want to know about tall and short Beans just in North America. (The "population" here would be the Beans in North America.) We would need to sample just from North America, and we would generalize our results just to those Beans living in North America.

Next, we need a plan for sampling. Our contacts tell us that there are three countries in North America--Canada, the United States, and Mexico. Let's say that I decide to sample Beans under 5.4 feet only in Mexico and Beans over 6 feet only in Canada and measure their incomes. Does this sound like a reasonable plan? I hope you said "no," because it is possible that tall people in Canada are not representative of all tall people in North America. Likewise, short people in Mexico may not be representative of all short people in North America. In fact, there might be several problems. First, besides the height of a Bean, the amount of money a Bean has surely depends on the country in which he or she lives. Second, countries also might differ on how much they pay short and tall beans ('height prejudice'). Lastly, a "short" Bean in one country might be considered "tall" in another. What a mess!

There needs to be a way to select samples which are representative of the entire population of North America. In fact, there are many ways to sample. One of the best ways to sample from a population is to do it at **random**. **Random sampling means that every person in the population has the same chance (theoretically) of being selected for the sample.** It is as though we had a list of all tall people in North America and then drew their names out of a very large hat. If we decided to sample 500 tall people, we would draw 500 names at random from all tall people in North America and use them as our sample. We would use the same method for sampling short people.

With random sampling, the people in the sample should be representative of the population from which they were drawn. That is, when everyone in a population has an equal chance of being in a sample, it is likely that all of the population's characteristics have an equal chance of being exhibited in the sample. **We have selection bias when we sample in ways that are not representative of the population.** This is really, really important because we are going to use the information we collect from our sample to make inferences about the population, so the sample has to have the same characteristics as the population. If the sample has characteristics that are different from the population, it could be that those different characteristics can account for the results of the study rather than the manipulation of the *independent variable*.

Let's try another example.

Beans are strange in many ways. Lots of them keep dogs in their house and love them a lot. Many of the dog-loving Beans are convinced that the breed of dog that they own is the very smartest. I wonder if some breeds of dogs are smarter than others. How would we answer this question if I asked specifically about Malamutes, Great Danes, and Beagles?

In this example, Beans are not the subjects or participants in our study because our question is about different breeds of dogs. As you already learned, the Beans or dogs or other animals or even organizations that participate in a study are called the *subjects* or, some people prefer, *participants*. Often the terms "subjects" and *"participants"* mean the same thing--the sample that is used in an experiment.

What are the three populations we are interested in for this study?

Type your answer here

The population would be all Malamutes, Great Danes, and Beagles on Earth. Of course, we could never find them all, so like most studies we will sample from each group. Since we really can't select dogs from all over the world to be in our study, let's make our population smaller. Suppose that you are a Bean who lives in Europe. You could then sample from these three populations of dogs in Europe. It is really hard to sample randomly from all of the dogs in any population, so most often researchers use **convenient samples**--samples they can more easily use. If you lived in France, for example, you might find all of the owners of Malamutes, Great Danes, and Beagles in France and select 100 from each group. You might then travel to the home of each dog and measure the dog's intelligence. Ok--so measuring dog intelligence seems strange, but there really are people who have done just that.

Finding a valid and accurate way to measure the intelligence of dogs, our dependent measure in this example, is difficult, but suppose for this example that you will teach each dog a new trick and time how long it takes the dog to learn that trick. The sample that learns the new trick in the shortest amount of time will be considered the smartest.

Sampling is critical to the quality of the results. Your sample of Malamutes would have to be representative of all Malamutes, your sample of Beagles would have to be representative of all Beagles, and your sample of Great Danes would have to be representative of all Great Danes. If you sample *at random* from each of the populations, you would have some old dogs, some young dogs, dogs from rich homes, and dogs from poor homes in each group. The dogs within each sample would differ from each other in many ways, but on average, the three samples would have a similar mix of old and young, and rich and poor dogs.

You would use the average time it took the dogs in each group to learn a new trick to make your conclusion.

You would need to use *statistical techniques* to decide if there were differences among the groups. If your statistical analysis shows that the groups probably are different, then you would make the inference that dogs in the group that learned the new trick in the shortest amount of time are (on average) smarter than dogs in the other two groups. Remember that we are making conclusions about groups of dogs in this example, even though there may be some not-so-smart dogs in the group that we found to be the smartest on average. There also will be some really smart dogs in the groups that we determine are not as smart as the smartest breed.

If one group is statistically faster at learning the trick than the other two groups, we can conclude that the faster group is smarter, and we can generalize our findings such that we infer that the faster group's population is also smarter.

So, when thinking about samples, it is important to consider if the sample is representative of the population it came from. If not, then we have a selection bias. **If we have selection bias, then we cannot make inferences about the population based on information from the sample.**

Chapter 10: Confusing Correlation with Cause

This one has to beat anything you have ever learned about the Nose Breathers. Can you even guess what they do with one-third of their measly life times? They SLEEP. Yup, they really do. They climb into their cozy beds, close their beady little eyes, and do nothing else--every single day of their life. (Actually every single night, but why quibble?) The little ones sleep even more! And, while they sleep, they sometimes make gurgly sounds that almost sound like our beautiful Fuath language, ZZZyAYYZZ.

And, can you even guess which group of Beans sleeps less than all of the others? It seems to be that college students who have tests the next day "pull an all-nighter" which in Bean-speak means that they skip sleep for one whole night. The reason they don't sleep one whole night is so they can study and get high grades in college. Now that seems like a good idea to me. After all, we Fuaths don't really understand this sleep business, but guess what? One researcher studied the sleep habits of 111 students to see if there was a relationship between getting regular sleep and student grades.[1] She found that students who did things like skip a night of sleep to study for a test tended to get lower grades in college than students who had more regular sleep schedules. I guess for the Beans who are college students this conclusion means that skipping sleep one night will cause them to get bad grades. Does that sound right to you?

Although it may seem that the conclusion from this study is that "pulling an all-nighter" will cause student Beans to get low grades, in fact we cannot make this conclusion. Do you know what is wrong with this conclusion?

Think back to earlier chapters about causal relationships. We discussed this idea when we talked about experimental control. If the researcher wants to be able to say that not getting enough sleep *causes* students to get lower grades, she would need to assign students at random to different sleep conditions (the independent variable) and then see if, on average, the students who got less sleep also got lower **grades**. But that is not what the researcher did; the research design used in the sleep study was **correlational**. It tells us that getting less sleep is ***associated*** with getting lower grades, but it does not tell us that getting less sleep ***causes*** lower grades.

Correlations can only lead to conclusions about the association or a relationship between variables. Correlations cannot tell us whether the variables are causally connected, or if they are truly causally connected, the direction of the causal connection (what caused what). For example, it could be that sleep allows Bean brains time to store information that the students learn, so that students who sleep less can't store as much information resulting in lower grades. However, it could also be that students with lower grades worry about their grades, and worrying is what is causing them to get less sleep. Or maybe there is an **extraneous variable,** or third variable that we do not know about, that is causing both lower grades and less sleep. Maybe students who party a lot (by drinking the Nose breather version of Wvvathum juice) tend to get less sleep and lower grades because they spend so much time partying that they don't sleep or study much.

When participants are NOT assigned at random to different groups, we cannot know what caused what to occur. We can only know that getting little sleep is correlated (or associated) with getting lower grades. In this example, the conclusion that getting little sleep *causes* students to get lower grades is inappropriate based on the research that was conducted.

This sleep study is a great example of meaningful results that when interpreted appropriately, can be very informative. Although we can't be sure that less sleep leads to lower grades for Beans, we do know that there is a relationship between sleep and grades, and that's important.

Let's look at another example of a study in which causal claims should be avoided:

Have you ever wondered if good looking people are happier than people whose looks are just average or below average? Now try to think about this question like a Nose Breather. Who knows what they think is attractive? Suppose we find a group of people who model clothes and other things in the Bean magazines. They are supposed to be attractive. We can ask them to fill out a happiness scale. We can then find a group of average or less than average looking Beans and ask them to fill out the same scale. Now keep on supposing that the attractive models are found to be happier than their less attractive Bean colleagues based on the happiness scale. I guess we could say that being attractive makes Beans happy, right?

The answer is nope. In this example, we took groups of people as they naturally exist and did not assign them at random to different groups of attractiveness. Now, we can never take naturally attractive people and put them in an "ugly group" or vice versa, at least without doing plastic surgery. So, we cannot conclude that being attractive makes Beans happy. It could be that being happy makes people attractive or some unidentified third variable is causing people to be happy and attractive or less happy and less attractive. Perhaps being rich or not rich would do both. We just don’t know.

One common misconception is that any study that involves an existing group (e.g., females and males, attractive and less attractive Beans) must always be correlational. This is not always the case. When we design research that uses existing groups of people, we call this a **quasi-experimental design**. A quasi-experiment is like an experiment because we compare different groups of people. But, **a quasi-experiment is not a true experiment because we did not assign people at random to the groups,** and so we cannot conclude that changes in one variable (being attractive or not attractive) caused changes in another variable (being happy or less happy). So, for example, a medical researcher could create a quasi-experimental design to test the effectiveness of a new acne drug by comparing different dosages (control, once a day, twice a day) on male and female participants. The researcher may find differences in the effectiveness of the drug for males and females, but could not conclude that gender caused these differences. Causal claims are not appropriate when conducting quasi experiments because group assignment is not random, yet the results can be highly informative and relevant.

It is important to be able to identify when causal claims are being made in research and when the claims are legitimate (use of random assignment of participants to groups) and when they are inappropriate (use of correlational methods or quasi-experimental designs).

Got the idea?

When reading about the conclusions from any study, be sure to check for causal statements, and if they are made, check for random assignment of participants to groups. Without random assignment, any other design is either correlational or quasi-experimental, which cannot be used by researchers or anyone else to make causal claims.

Reference: SLU profs study shows all-nighters equal lower grades. Retrieved June 13, 2009 from http://www.stlawu.edu/news/thachersleepstudy.html.

Chapter 11: Premature Generalizations

Sometimes, a good way to learn about the lives of Beans is to just spend some time observing them. If you ever watched them carefully, you will notice an annoying habit: they constantly flick one or more of their opposable thumbs over a small communicator device. They call this texting. The Bean professors think it is annoying when their students text. I guess they are annoyed because the students are texting to friends instead of listening to their professors' well-prepared lectures, like they should.

Lots of Beans gaze into the small flat surfaces of their communicators while they hit their thumbs on the strange characters. I have been told by reliable sources that they are working on their Facebook. A report in the Bean news just found that Beans in college who spend lots of time doing this Facebook thing also get lower grades and spend less time studying than those who spend little or no time on Facebook.[1]

This is sort of interesting. But, what is even more interesting is how the Bean world reacted to this study. Here is a quote from a Bean newspaper, the "Sunday Times of London". You can probably guess about one big mistake the newspaper reporters made:

> Research finds the website [Facebook] is damaging students' academic performance. … Facebook users … are more likely to perform poorly in exams, according to new research. … The majority of students who use Facebook every day are underachieving by as much as an entire grade compared with those who shun the site.

Can you spot some problems with this conclusion?

There are lots of problems with the conclusion from that Bean newspaper. If you read the last chapter carefully, you already know that this was a *correlational* study, so the authors cannot say that using Facebook a lot **causes** students to get lower grades. They can only say that using Facebook is related to getting lower grades. I hope you caught that BIG error. The authors of the study just recorded how often students used Facebook, how much they studied, and their grades. No one was assigned to use or not to use Facebook, so because they took people just the way they were, this is a correlational study.

But, there are additional problems with the way the Bean newspaper reported the conclusions from this study. Some newspapers went on to say that Facebook is hurting businesses because people are spending so much time on Facebook, they are not doing their work. But wait a minute! Can this conclusion be validly made from the original study? I don't think so. The study used students as participants, not people in business. To know what they can safely conclude about this study, we need to know more about the participants.

The participants in the study were college students in Ohio. (Ohio is a region of the land mass between the two blue blobs on most maps of Bean World USA). To know how well designed the study was and consequently how much we can trust the findings, we would need to know if there were enough participants and who they were.

There were over 100 students in the study about Facebook, grades, and studying. The *sample size* was large enough to make a conclusion from the data collected. (You should know that Fuath and Bean scientists have all kinds of fancy formulas for deciding on the right number of participants in a study, but we don't need to go into all of that.) But, besides sample size, the other important question about samples has to do with the way the participants were selected to participate in the study.

Remember that when participants are selected for a study, they are taken from some larger group called the *population*. Researchers use the participants in their sample to make inferences about the rest of the population. So, in the case of the Facebook study, the population was college students. If the sample was selected at random from the population, then we could conclude something about the population – in this case, college students. Were students in the Facebook study randomly selected from *all* college students? No. Those participants were selected at random from all college students at *one* university in Ohio. So, if we are feeling generous, we could say that the participants were a **representative sample** of college students at universities similar to that one in Ohio. But, we cannot say that they are similar to students at other types of universities or to people who work in business or other fields.

Here is the important point: when a sample is representative of the population from which it was drawn, then the results of the study can **generalize** to all of the people in the population (with some caution or with what the scientific Beans call a "margin of error.")

But, we cannot generalize from a study of college students to a population of full-time employees or workers. Therefore, we cannot conclude from the student study that workers are wasting time on Facebook when they should be working. After all, workers are probably older than the college students and may differ from them in other ways. College students usually have lots of computers around, and older people who work may not have any. Workers might also have what are called “bosses” making sure that they do not waste time on Facebook. I’m sure that you can think of other ways in which students and workers differ. So we cannot take the conclusion from a study that used college students as participants and apply it to another population, such as people who work.

Let's try another example of making appropriate generalizations from research.

When Beans have troubles in their personal lives, sometimes they visit a special type of doctor called a “psychologist” who helps them with their problems. Well, one psychologist Bean decided to study the marriage problems that various Beans talked about with her. She kept careful records and found that most of the married Beans she saw in her practice had disagreements about money--how to spend it, how much to save, and so on. So, she wrote a research article in which she told couples to learn how to handle issues about money if they want to have a happy marriage. What do you think about her conclusion?

Actually, it is not necessarily a bad idea for everyone to learn how to handle money, but it is not a valid conclusion from her study. Think about the participants in her (informal) study. Let's say that she has a large number of participants, so don't worry about the size of her sample. What was the population they represented?

Pick one of the following

- married couples
- married couples in therapy
- non-married couples
- assorted odd people

I hope you picked "married couples in therapy" as the population from which they came. We cannot make conclusions from married couples in therapy and apply them to all married couples. Maybe those couples who are in therapy are somehow different from those who are not. Maybe couples in therapy have poorer communication skills than couples who are not in therapy. If that is true, just having a psychologist to talk with will make these couples happier. That is, it has nothing to do with money. For all we know, maybe all Bean couples fight about money, and it has nothing to do with the quality of their marriage. We just don't know, and even though it might be a good idea to learn how to handle money, we can*not* conclude from this study that if couples learn about handling money, they will have happier marriages.

So, what is the take home message here? It is that in order **for a study to generalize to the population of interest, the researchers must use a sample that is representative of that population. In order to get a representative sample, they need to use random selection from the population of interest.** For example, to generalize to a population of Bean adults, the sample should be made up of all Bean adults. To generalize to a population of yellow-bellied chickles, the sample must be made up of all yellow-bellied chickles. To generalize to a population of steel-belted tires, the sample of the study should be made up of all steel-belted tires. You get the idea, don't you?

____

____

_____

Well, we have reached the end of this science manual. I have some closing thoughts for you.

Here are some heavy questions that puzzle even the philosopher Beans: How can we know what to believe and what not to believe? Whom can we trust for good information? Can we every really know about the world?
There are lots of ways of knowing. There is the Fuath Way, which even the tiniest baby Fuaths understand--You tuck your head under your left arm and think deep thoughts until the meaning of Fuath life is understood. It always works for me. But, what if you are a measly Bean who cannot see truth this way?

As you read through this Fuath Guide to the Bean's World of Science, you surely noticed that there is no absolutely certain way that Bean's have for knowing about the world. But, some ways are better than others. Some Beans use their Faith to understand matters that relate to Faith, but when it comes to other more Beanly matters such as whether or not a drug is effective or where to invest their money or whether smoking is good or bad for their health, they need to have other good ways of making decisions. The Beans use the best method they have for making these decisions about their complicated world, and that is the scientific method.

Want to learn more? Try out this fun demonstration on-line:

https://app.skoonline.org/aries/text/ListOfCases.html

References

Grading Facebook's grade link. (2009, April 21). The Wall Street Journal. Retrieved from http://online.wsj.com/article/SB124034834334240435.html on June 18, 2009.

Quote from Holland, E. (2009, May 8). Facebook and procrastination. Columbia Journalism Review, Retrieved from http://www.cjr.org/the_observatory/facebook_and_procrastination.php on June 18, 2009.