

8 Control Problems

Ya know, someday these scientists are gonna invent something that will outsmart a rabbit —Bugs Bunny

Hewwo! Acme Pest Contwol? Weww I have a pest I want contwolled. —Elmer Fudd

Elmer Fudd is always being outsmarted by Bugs Bunny. If Elmer Fudd was a psychologist, he might say “Be vewy vewy quiet, I’m conducting an expewiment”. But, we all know how the experiment would turn out. Bugs Bunny would find a way to foil Elmer Fudd’s plans and ruin his experiment. Bugs Bunny is a metaphor for all of the control problems that can plague experiments. In this chapter we discuss many common problems that are well known, how they confound the interpretation of experimental results, and practical solutions for solving the problems. Bugs Bunny wasn’t wrong: Scientists have learned how to outsmart rabbits.

Ceteris Paribus

Ceteris paribus is the Latin phrase for “All other things being equal”, or “All other conditions remaining the same”. Ceteris paribus is a fundamental logical condition for making inferences about cause and effect when conducting experiments. Experiments are a method for making inferences about causal forces, and the effects they have on observable outcomes in the world. Potential causes are identified by manipulating an independent variable, and measuring potential effects on a dependent variable. If, the only change between conditions involved change in the independent variable, then we can safely infer it caused a change in the dependent variable. Critically, we can make this inference when, ceteris paribus, or all other things (besides the independent variable) are equal or remain the same between the conditions of the manipulation. However, in practice, ceteris paribus is almost never guaranteed, and there are many possible confounding variables that vary between conditions of the independent variable. These confounding variables place constraints on the experimenters ability to make strong inferences about causality. The solution in practice is to identify and eliminate or reduce the influence of all of these confounding variables when conducting experiments.

All the other things

In psychology experiments there are numerous other things besides the independent variable that can cause change in the measured dependent variable. In chapter’s two to five we discussed at length how *chance* and measurement variability (or error) can produce the appearance of differences. We also discussed choices that experimenters can make when designing their experiments to reduce the influence of chance, such as increasing the number of observations, increasing the number of subjects, and improving measurement precision.

In this chapter we will discuss a broad range of *confounding variables* (also termed extraneous variables) that can influence results, as well as how to construct designs to guard against these confounding influences. For fun, you might think of confounding variables as Bugs Bunny showing up in one of your conditions and screwing around with the data. When this happens, we know the difference we observed was not due to our manipulation, but to a wascally wabbit.

Ligation: a tale of an unexpected confound

Prior to the 1950s, ligation was a common surgical procedure for heart disease. The procedure involves cutting a patient open to expose the internal mammary artery, and then ligate the vein. Ligating the vein involves tying it off, so it no longer flows. This procedure was thought to lessen heart pain caused by angina.

In the 1950s some doctors became skeptical of the ligation procedure. There were no very good medical reasons why the procedure should work, and there was increased awareness of the possibility of placebo effects. Could it be that patients feel better simply because they knew they were getting surgery to fix a problem, and not because the surgery actually fixed the problem?

Two separate studies tested this possibility in an ethically questionable experiment (Cobb et al. 1959). Each experiment involved groups of patients who were scheduled for ligation surgery. Secretly, and unbeknownst to the patients, the surgeon randomly chose whether each patient would either, 1) receive the full ligation surgery, or 2) follow all of the surgery steps (being anesthetized, cut-open, sown back up, etc.), except for the ligation (the vein was never tied off). This was an experiment with two levels, one group got the treatment (ligation), and the other group was treated as similarly as possible (*ceteris paribus*), but did not the the ligation. The outcome shocked the medical community. Both groups of patients reported the same levels of post-surgery decreases in heart pain. The inference was that the ligation manipulation was not causally responsible for decreasing heart pain; instead, the decreases were the result of a confounding variable. In this case, the confound appeared to be the process of having a surgery in general. As a result, the medical community abandoned the ligation of the internal mammary artery as a medical practice.

Third-variable problems

In correlational research a goal is to determine relationships between measured variables. For example, does the cost of your shoes predict your current level of happiness? If people who wore more expensive shoes were happier than people who wore less expensive shoes, there would be a positive correlation. However, anyone could easily object that the cost of shoes may not be the causal variable that influences happiness. Instead, any number of other variables, such as socio-economic status, age, geographic location, etc. could both cause changes in happiness

and the kinds of shoes that people by. All of these other possible variables that could have causal influences are called third variables. Third variables can also influence the results of experiments, especially when they are confounded with the experimental manipulation.

Consider the famous experiment showing that unconsciously priming concepts of being elderly can influence how quickly people walk down a hallway (Bargh, Chen, and Burrows 1996). In this experiment, subjects completed anagram word puzzles. Some subjects received words that primed the concept of being elderly (e.g., bingo, florida, old, etc.), and other subject received neutral words. After completing the word puzzles, subjects left the laboratory and walked down the hallway to an elevator. A confederate standing in the hallway used a stopwatch to time how long each person took to walk from the door to the elevator. The results showed that subject primed with the elderly concept walked more slowly down the hallway than subjects who received neutral primes. Neat!

Interestingly, a recent replication study showed that the original result could have been due to a confounding variable (Doyen et al. 2012). One concern with the original study was that the confederate in the hallway may not have been blind to the experimental design. That is, the person taking the measurements with the stopwatch may have known which subjects were primed and which subjects were not primed. As a result, it is possible that the confederate may have expected the primed subjects to walk more slowly than the unprimed subjects. If so, these expectations could plausibly influence how the confederate used the stopwatch. For example, the confederate might accidentally record longer times for the primed than unprimed subjects. If this occurred, then the priming manipulation was confounded with the expectations of the confederate taking the measurement.

The replication study eliminated the possibility of confounds due to the confederate by using an objective measure of walking time. In one study they installed infrared sensors to measure walking speed, rather than a confederate with a stopwatch. They found no differences in walking speed between priming conditions. In another study they manipulated whether the experimenters were blind to the priming conditions of each subject. They showed no differences in walking speed when the experimenters were blind to the priming condition. However, when the experimenters were told which conditions the participants were supposedly in, those participants did show differences in walking speed, even with the object timing measure using infrared sensors. This suggests that the experimenters were biasing how the subjects walked, perhaps by interacting more slowly or more quickly with different subjects as they left the room.

Controlling the problems

Researchers have developed numerous strategies to control for confounding variables in experimental designs. We have already briefly discussed some of these strategies, including random assignment and counterbalancing in Chapter 4. Here, we again review and dig a bit deeper into these concepts.

Remember, the ideal experiment manipulates only the independent variable (s) of interest, and holds all other possible variables constant. Psychology experiments are never ideal. And, many other variables can not be held constant. For example, in a between-subjects design there are different subjects in each condition. So, the different subjects are inextricably confounded with the manipulation. In a within-subjects design, the same person performs in both conditions, but those conditions happen at different points in time. So, time is inextricably confounded with the manipulation. The solution to these confounds is to approximate the ideal experiment. This is accomplished by attempting to equate the confounds across conditions so that their influences are the same in both conditions. If the confounds are the same in each condition, then they cannot produce differences between conditions.

Random Assignment

The process of random assignment is a tried and true method for equating confounds between conditions. Before we look at this in the context of a psychology experiment, let's take a historical look at how random assignment was used in an agricultural example.

Imagine you are a corn farmer, and you want to know which fertilizer to use to make your corn grow the best. What do you do? You could run an experiment. For example, you could by Fertilizer A and Fertilizer B, plant corn in your field, and then apply Fertilizer A to half of your field and Fertilizer B to the other half. At the end of the growing season you simply measure the corn yield from both halves of the field. If the section with Fertilizer A grows more corn, then you might use that Fertilizer A next year because it works better than Fertilizer B. Case closed?

Maybe not. What kind of confounding variables could have been responsible for the result? Maybe the soil on one side of the farm happened to have more nutrients than the other side, maybe the sun is better on one side than the other, maybe one side got more or less rain than the other, etc. There are lots of factors that can influence corn growth, and you could be very unlucky if all of the "good stuff" happened to be on one side of the field, and all of the "bad stuff" happened to be on the other side of the field. In this case, the fertilizer manipulation would be completely confounded with the confounding variables.

Real-world farmers do not have the luxury of creating ideal fields that are completely identical all respects when performing their fertilizer experiments. Instead, they have to deal with the fact there are numerous confounding influences all across the field. However, the influence of these confounds can be substantially reduced by the process of careful random assignment.

The above corn experiment is a simple example of a split-plot design. The field is a plot of land, and it is split into two parts. Each part receives a different manipulation (Fertilizer A or B). We have already discussed how this kind of experiment could be vulnerable to confounding influences. How could the split design be improved?

The answer is more splitting and random assignment. For example, consider applying a checkerboard pattern to the field. There would be lots of white squares spread across the whole field, and lots of black squares spread across the whole field. The experimenting farmer could then apply fertilizer A to all of the white squares, and fertilizer B to all of the black squares. If this checkerboard experiment showed that corn grew better when treated with Fertilizer A than B, would you be more confident that the result was real, and that Fertilizer A is better than Fertilizer B? Being able to answer this question is a litmus test for your understanding of the virtues of random assignment.

I would be more confident of the results of the checkerboard experiment. The reason is that assigning the manipulation to lots of little plots all around the field spreads out the influence of confounding variables on corn growth, making it increasingly unlikely that those variables could explain any differences due to the fertilizer manipulation. For example, if the entire left side of the field happened to get better sun than the right half, then a simple split-plot design with two halves (one of the left and one on the right) would be confounded with sun quality. However, the checkerboard experiment is not confounded by sun quality, because there would be just as many white and black squares (containing fertilizer A and B) on the left side of the field as the right the side of the field. So, the effect of sun quality would contribute equally to the corn in the white and black squares. Specifically, the corn on the left side of the field might grow better, but it would grow better both for the white and black squares on the left side. The corn on the right side might grow worse, but would grow equally worse for the white and black squares on the right side. As a result, when you average all of the white squares together, the effect of sun quality averages out; and ditto for the black squares.

The potential confounding effect of sun quality is just one example of a confounding variable, and of course there are many others. However, the checkerboard assignment process accomplishes the result of spreading out the influences of these other confounding variables, so they wash out in the average. The idea is that, the white and black squares are small enough, and spread around enough, that all of the confounding variables occur just as often for the white squares as they do for the black squares. As a result, any difference between white and black squares must be due to the manipulation, because the confounding variables were roughly equal between conditions.

This section is about random assignment, so you might be wondering whether the checkerboard assignment strategy is an example of random assignment. After all, a checkerboard is not very random at all. It is a predefined and highly predictable pattern. On the one hand, you might imagine a seemingly more random way of conducting a similar design. For example you could split the field into a grid of small squares, and the use a coin flip for each square to determine whether it is a white (Condition 1) or a black (Condition 2) square. On the other hand, you might consider that the pure checkerboard pattern and the randomized grid offer fundamentally similar solutions. They both do a good job of spreading out the influence of confounding variables, such that each confounding variable contributes roughly equally to each of the manipulated conditions.

The fundamental property of random assignment that allows it, on the average, to equalize the influence of confounding variables is not that the assignment process itself is random per se, but that the distribution of the resulting conditions are random with respect to the distribution of confounding variables. For example, in the farming example, although the checkerboard pattern is not very random, the distribution of the checkerboard pattern is likely close to random with respect to the distribution of possible confounding variables spread across the plot of land. The actual distribution of confounding variables across the field is an unknown, but there is no good reason to think they would be distributed like a checkerboard pattern or any other specific kind of pattern. We would expect the distribution of confounding variables to be clumpy, some here some there, some everywhere. In principle, any pattern for assigning conditions that was random with respect to the unknown distribution of confounding variables would be a fine way to assign conditions. In practice, these patterns are easily found by randomly assignment.

Random assignment and sample size

Importantly, random assignment does not guarantee that confounding variables with contribute equally in each manipulated condition of a design for any given experiment. By chance alone, the random assignment process could accidentally lump some confounding variables into one condition more than another. The best way to guard against this problem is to increase sample-size. Increasing sample-size increases the number of random assignments that are made, and smooths out the lumpiness of chance.

As we discussed in Chapter 4, random-assignment is often used in between-subjects designs. Between-subjects design are always confounded by the *non-equivalent groups* problem. The subjects in each condition are different people, so they are inherently not equivalent. Or, in other words, all other things besides the manipulation between conditions are not the same—the subjects in each condition are different! The random assignment process is used to create increase the equivalence between the groups.

The results of between-subjects experiments can always be criticized because of the non-equivalent groups problem. Sometimes it is worth being skeptical of the results from these designs. For example, imagine a between-subjects experiment looking at taking notes on a laptop or with pen and paper that found better test performance for the group who used pen and paper. Anyone could suggest they don't trust the result because it is possible that the researchers accidentally put more "smarter" students in the pen and paper group than the laptop group. Although, this kind of concern can be warranted, especially when samples-sizes are small, the concern is not as valid if the sample-size is larger, and good random assignment procedures were used. In this latter case, it would be possible to estimate the chances that random assignment alone would cause unequal distribution of "smarter" students between conditions, and as sample-size increases, would show increasingly lower likelihoods of that uneven assignment occurring. Furthermore, if the result is reproduced over and over again

in many samples and across different research groups, then result deserves to be treated with higher confidence.

Matching

Matching is a common alternative to random assignment. The aim of matching is the same, to create equivalent groups of subjects in each experimental condition. Rather than leaving group assignment to chance, the researcher hand picks subjects with matching qualities and assigns them to different groups. For example, in the note taking experiment, a research might ensure that all subjects are matched on several variables, such as having the same age, having the same gender, having the same socio-economic-status, having the same level of general intelligence etc.

Matching is often attempted in research with special populations. For example, research comparing the effect of brain damage to a specific region of the brain on some ability often first measures performance among a group of subjects with a specific pattern of brain damage, and then finds matched control subjects who do not have brain damage, but are as similar as possible on other dimensions as each of the impaired subjects.

Matching presents several practical difficulties. For example, it can be difficult to find matching subjects, and it can be difficult to evaluate the goodness of particular matches even if they can be found.

Counterbalancing

You should first refresh your memory of counterbalancing from the section on counterbalancing in Chapter 4. Remember, that an issue with within-subjects designs is that subjects usually contribute data to each condition at different times or orders in the experiment. As a result, the order of the manipulation, rather than the manipulation itself, can sometimes cause differences. Counterbalancing is a technique used to systematically control the order in which subjects receive the manipulation.

For example in simple experiment with two conditions, one group of subjects might receive condition 1 then condition 2, and the other group would receive condition 2 then condition 2. In this example, there are two independent variables. The primary independent variable is manipulated within-subjects (condition 1 vs condition 2), and the counterbalancing order is manipulated between-subjects. With this design, a researcher can determine whether there is an effect of the primary manipulation that is independent of the order. Or, they may also find that the order has an effect, but not the manipulation. There can also be interactions between the two.

Item Effects

Counterbalancing can become complicated, especially when there are multiple variables to counterbalance. Importantly, counterbalancing can be use for variables other than order. For example, counterbalancing can be applied to the items used across trials in a particular task. Consider a memory experiment where the researcher is interested in whether memory performance improves or declines depending on heart rate. The primary manipulation could involve having subjects run on a treadmill to increase heart rate, or simply relax in a chair to reduce heart rate. In each heart rate condition, the subject would be presented with a list of words to memorize, and then subsequently recall to measure their memory for the word lists. In this case, the researcher would counterbalance both the order of the heart rate manipulation, as well as the assignment of the word lists to each condition. How many total conditions does this produce?

There are two heart rate condition (running vs. sitting), and two orders (running then sitting, and sitting then running). So far there are a total of four conditions. One group of subjects will receive the words in list 1 always in the running condition, and list 2 always in the sitting condition. Here, there are still only four conditions, but you should spot a problem. There could be item effects. What if the words in list 1 are easier to remember than the words in list 2. In the above design, this would produce a result showing better memory in the running than sitting conditions. However, we know that result would be the result of an item-level confound. To counterbalance for the item-level confound, we need to run additional groups of subjects who always receive list 1 when they are sitting and list 2 when they are running. This produces a total of eight conditions, shown in the table.

| Order | Running | Sitting |
|---------------|---------|---------|
| Running First | List 1 | List 2 |
| Sitting First | List 1 | List 2 |
| Running First | List 2 | List 1 |
| Sitting First | List 2 | List 1 |

Notice that the process of counterbalancing is accomplishing the same fundamental goal accomplished by random assignment. The goal is to spread out and equalize the influence of possible confounding variables. The two possible confounding variables here are order and list. Notice that all orders and list types and combinations of orders and list types occur equally for both the running and sitting conditions.

References

Bargh, John A, Mark Chen, and Lara Burrows. 1996. "Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action." *Journal of Personality and Social Psychology* 71 (2): 230.

- Cobb, Leonard A, George I Thomas, David H Dillard, K Alvin Merendino, and Robert A Bruce. 1959. “An Evaluation of Internal-Mammary-Artery Ligation by a Double-Blind Technic.” *New England Journal of Medicine* 260 (22): 1115–18.
- Doyen, Stéphane, Olivier Klein, Cora-Lise Pichon, and Axel Cleeremans. 2012. “Behavioral Priming: It’s All in the Mind, but Whose Mind?” *PloS One* 7 (1): e29081.