Chapter 7: Within-participants design

A within-participants design is one where instead of randomly assigning participants to different levels of the independent variable, participants each complete both (or all) conditions. This turns out to be an extremely powerful experimental technique that should always be considered in experimental design, although there are many research areas in which it cannot be effectively used.

Using a within-participants design has several important implications for how we approach inferential statistics. The specific tools differ, e.g., a paired-sample t-test instead of independent samples, but this approach often has much greater sensitivity to the reliability of differences between conditions.

## Summary of advantages:

* Highly efficient. Each participant provides data in all conditions so accumulating data collection towards the planned number of participants is more rapid.
* Perfect control of participant variables. Since all participants provided data on all conditions, the conditions are exactly matched for all extraneous variables related to the participants on the task, e.g., motivation, attention, ability.

## Summary of disadvantages:

* Order effects. If conditions have to be given in order, there are many ways in which “history” effects influence the data such that the earlier/later conditions performance on the DV is affected. These can be counterbalanced but never fully controlled.
* Participant demand characteristics. By being exposed to all conditions of the experiment, participants will always be aware of all the levels of the independent variable. This increases the chance that they will understand the experimental hypothesis, which may affect their performance (bias).

# **Within-Participants Experiments**

In a within-participants experiment, each participant is tested under all conditions. Consider an experiment on the effect of a defendant’s physical attractiveness on judgments of his guilt. Again, in a between-participants experiment, one group of participants would be shown an attractive defendant and asked to judge his guilt, and another group of participants would be shown an unattractive defendant and asked to judge his guilt. In a within-participants experiment, however, the same group of participants would judge the guilt of both an attractive and an unattractive defendant.

The primary advantage of this approach is that it provides maximum control of extraneous participant variables. Participants in all conditions have the same mean IQ, same socioeconomic status, same number of siblings, and so on—because they are the very same people. Within-participants experiments also make it possible to use statistical procedures that remove the effect of these extraneous participant variables on the dependent variable and therefore make the data less “noisy” and the effect of the independent variable easier to detect. We will look more closely at this idea later in the book.  However, not all experiments can use a within-participants design nor would it be desirable to do so.

## Carryover Effects and Counterbalancing

The primary disadvantage of within-participants designs is that they can result in order effects. An order effect occurs when participants’ responses in the various conditions are affected by the order of conditions to which they were exposed. One type of order effect is a carryover effect. A carryover effect is an effect of being tested in one condition on participants’ behavior in later conditions. One type of carryover effect is a practice effect, where participants perform a task better in later conditions because they have had a chance to practice it. Another type is a fatigue effect, where participants perform a task worse in later conditions because they become tired or bored. Being tested in one condition can also change how participants perceive stimuli or interpret their task in later conditions. This type of effect is called a context effect (or contrast effect). For example, an average-looking defendant might be judged more harshly when participants have just judged an attractive defendant than when they have just judged an unattractive defendant. Within-participants experiments also make it easier for participants to guess the hypothesis. For example, a participant who is asked to judge the guilt of an attractive defendant and then is asked to judge the guilt of an unattractive defendant is likely to guess that the hypothesis is that defendant attractiveness affects judgments of guilt. This knowledge could lead the participant to judge the unattractive defendant more harshly because he thinks this is what he is expected to do. Or it could make participants judge the two defendants similarly in an effort to be “fair.”

Carryover effects can be interesting in their own right. (Does the attractiveness of one person depend on the attractiveness of other people that we have seen recently?) But when they are not the focus of the research, carryover effects can be problematic. Imagine, for example, that participants judge the guilt of an attractive defendant and then judge the guilt of an unattractive defendant. If they judge the unattractive defendant more harshly, this might be because of his unattractiveness. But it could be instead that they judge him more harshly because they are becoming bored or tired. In other words, the order of the conditions is a confounding variable. The attractive condition is always the first condition and the unattractive condition the second. Thus any difference between the conditions in terms of the dependent variable could be caused by the order of the conditions and not the independent variable itself.

There is a solution to the problem of order effects, however, that can be used in many situations. It is counterbalancing, which means testing different participants in different orders. The best method of counterbalancing is complete counterbalancing in which an equal number of participants complete each possible order of conditions. For example, half of the participants would be tested in the attractive defendant condition followed by the unattractive defendant condition, and others half would be tested in the unattractive condition followed by the attractive condition. With three conditions, there would be six different orders (ABC, ACB, BAC, BCA, CAB, and CBA), so some participants would be tested in each of the six orders. With four conditions, there would be 24 different orders; with five conditions there would be 120 possible orders. With counterbalancing, participants are assigned to orders randomly, using the techniques we have already discussed. Thus, random assignment plays an important role in within-participants designs just as in between-participants designs. Here, instead of randomly assigning to conditions, they are randomly assigned to different orders of conditions. In fact, it can safely be said that if a study does not involve random assignment in one form or another, it is not an experiment.

An efficient but rarely-used way of counterbalancing is through a Latin square design which randomizes through having equal rows and columns. For example, if you have four treatments, you must have four versions. Like a Sudoku puzzle, no treatment can repeat in a row or column. For four versions of four treatments, the Latin square design would look like:

|  |  |  |  |
| --- | --- | --- | --- |
| A | B | C | D |
| B | C | D | A |
| C | D | A | B |
| D | A | B | C |

You can see in the table above that the square has been constructed to ensure that each condition appears at each ordinal position (A appears first once, second once, third once, and fourth once) and each condition precedes and follows each other condition one time. A Latin square for an experiment with 6 conditions would by 6 x 6 in dimension, one for an experiment with 8 conditions would be 8 x 8 in dimension, and so on. So while complete counterbalancing of 6 conditions would require 720 orders, a Latin square would only require 6 orders.

More commonly, when the number of potential orders is large, experiments simply use random counterbalancing in which the order of the conditions is randomly determined for each participant. Using this technique every possible order of conditions is determined and then one of these orders is randomly selected for each participant. This is not as powerful a technique as complete counterbalancing or partial counterbalancing using a Latin squares design. Use of random counterbalancing will result in more random error, but if order effects are likely to be small and the number of conditions is large, this is an option available to researchers.

There are two ways to think about what counterbalancing accomplishes. One is that it controls the order of conditions so that it is no longer a confounding variable. Instead of the attractive condition always being first and the unattractive condition always being second, the attractive condition comes first for some participants and second for others. Likewise, the unattractive condition comes first for some participants and second for others. Thus, any overall difference in the dependent variable between the two conditions cannot have been caused by the order of conditions. A second way to think about what counterbalancing accomplishes is that if there are carryover effects, it makes it possible to detect them. One can analyze the data separately for each order to see whether it had an effect.

## Simultaneous Within-Participants Designs

So far, we have discussed an approach to within-participants designs in which participants are tested in one condition at a time. There is another approach, however, that is often used when data is collected across multiple trials (events). For example, if participants were asked to judge the guilt of 10 attractive defendants and 10 unattractive defendants, these could be presented in an intermixed order instead of having people make judgments about all 10 defendants of one type followed by all 10 defendants of the other type. The researcher could then compute each participant’s mean rating for each type of defendant. Or imagine an experiment designed to see whether people with social anxiety disorder remember negative adjectives (e.g., “stupid,” “incompetent”) better than positive ones (e.g., “happy,” “productive”). The researcher could have participants study a single list that includes both kinds of words and then have them try to recall as many words as possible. The researcher could then count the number of each type of word that was recalled. This approach removes some of the concerns of order and carryover effects across conditions by having them presented roughly at the same time. Some attention still has to be to the specific order of stimuli in these kinds of approaches. If the stimulus order is constructed using a purely random mechanism, the resulting order can end up not looking very “random” from a human perspective (humans are poor at recognizing truly random sequences) because there can be long subsequences of the same condition across several trials. Typically, the stimulus order in this kind of design is constructed using a “pseudo-random” sequence that limits the number consecutive stimuli in a row of the same condition. Note that if this is not done, the apparent randomness of the sequence ends up being an extraneous variable that will add noise to the dependent variable, weakening the reliability of the results, but not confounding or weaking the validity of the study.

When this approach can be used, it is often highly effective in that it removes many of the concerns about history effects across experimental conditions. However, it should also be noted that it is still vulnerable to concerns about participants potentially being aware of the experimental hypothesis. Since they are participating in both conditions, they are no longer blind to the experimental design and the possibility that this may bias their responses needs to be considered.

## **Between-Participants or Within-Participants?**

Almost every experiment can be conducted using either a between-participants design or a within-participants design. This possibility means that researchers must choose between the two approaches based on their relative merits for the particular situation.

Between-participants experiments have the advantage of being conceptually simpler and requiring less testing time per participant since only one condition is being completed, in addition to avoiding any possible carryover effects. Shorter tests reduce the risk of fatigue and also interact with other details of the procedure for data collection. For in-person data collection where participants come to a laboratory to complete an experiment under controlled conditions, a significant component of time and effort is scheduling and traveling to the lab. Shortening a protocol from 50 minutes to 25 minutes is of limited value in this case. However, for online data collection, shorter protocols may have better engagement with participants and reduce fatigue and the probability that the participants drop out of the study before completion. Practical questions of recruiting and carrying out data collection may be important for planning whether a between or within participant design is preferred.

Within-participants experiments have the advantage of controlling extraneous participant variables, which generally reduces noise in the data and makes it easier to detect any effect of the independent variable upon the dependent variable. Within-participants experiments also require fewer participants than between-participants experiments to detect an effect of the same size.

A good rule of thumb, then, is that if it is possible to conduct a within-participants experiment (with proper counterbalancing) in the time that is available per participant—and you have no serious concerns about carryover effects—this design is probably the best option. If a within-participants design would be difficult or impossible to carry out, then you should consider a between-participants design instead. For example, if you were testing participants in a doctor’s waiting room or shoppers in line at a grocery store, you might not have enough time to test each participant in all conditions and therefore would opt for a between-participants design. Or imagine you were trying to reduce people’s level of prejudice by having them interact with someone of another race. A within-participants design with counterbalancing would require testing some participants in the treatment condition first and then in a control condition. But if the treatment works and reduces people’s level of prejudice, then they would no longer be suitable for testing in the control condition. This difficulty is true for many designs that involve a treatment meant to produce long-term change in participants’ behavior (e.g., studies testing the effectiveness of psychotherapy). Clearly, a between-participants design would be necessary here.

Remember also that using one type of design does not preclude using the other type in a different study. There is no reason that a researcher could not use both a between-participants design and a within-participants design to answer the same research question. In fact, professional researchers often take exactly this type of mixed methods approach.

## Pre/post designs

In some cases, our hypothesis about our dependent variable can be described as a change score, such as a change in mood after some intervention (e.g., music, performance feedback, etc.). These will generally be implemented in a design where there is a measurement before the independent variable is manipulated (between-participants), a pretest, and another measurement after the manipulation, a posttest. For the analysis of these data, there is a choice to be made about how to operationally define the dependent variable. In many cases, with relatively simple designs, it is permissible to define the DV as the subtraction between the pre and post measurements and then do the statistical analysis treating that score as the DV. This results in a design that appears to have a within-participants aspect but is effectively a between-participants design.

Many designs that use this approach are more complex in nature and rather than rely on calculated subtraction score, both the pre and post measures will be incorporated into a more complex analysis approach. This type of design is very common in research that examines the effect of different kinds of interventions. Intervention research is very common in applied areas within psychology as well as public health, economics, and public policy. We will briefly review some of the common methodologies employed in invention-based research in Chapter 14.

# Within-participants Statistics

Within-participants designs are much more statistically powerful than between participants designs for the same number of participants. We can describe this as an increase in “sensitivity” to differences between conditions (statistical sensitivity and effect sizes will be discussed in Chapters 11/18). The actual execution of the statistical routine for calculating a within-participants effect is generally similar to between-participants tests with a few notable exceptions.

First, in a within-participants experiment, we will tend to organize the data somewhat differently. In a data spreadsheet, we will continue to generally organize the data with one participant per row in our data matrix. However, now we will have multiple columns of data for scores of the dependent variable across the conditions tested. This allows us to visually inspect the data and often makes condition differences quite easy to see. If in each (or many) of the participants one of the conditions is consistently producing a larger score on the DV, we are likely to be observing a reliable difference. In fact, it is often useful to calculate a subtraction score between the conditions for each participant. An alternative approach to running a paired sample t-test is to do a one-sample t-test on the subtraction scores (these are formally identical procedures with respect to the math).

However, running a paired-sample t-test using R software does require one extra step for reorganizing the data. The input to analysis in R requires only one data point per row in the data matrix, so the data must be reorganized and recoded so that the two conditions are listed as the same participants on two rows. In addition, a participant number is now required in order to connect the two measures. The specific process for carrying out the data reorganization and the analysis will be reviewed in Chapter 11 in consideration of factorial designs with within-participants factors (often called **repeated measures** designs).

To this point, we have only considered simple designs with a single independent variable that we will very either between or within participants. We will shortly extend the general model of experimental design to more complex **factorial designs** with multiple independent variables (factors). An issue to be aware of with the types of designs discussed here that are intended to be simple within-participants studies is that they can accidentally become more complex designs due to order effects. As noted above, when a design has to be administered in order, e.g., condition A then condition B, we should always counterbalance the order of conditions to ensure that order is not confounded with condition. In this case, we should also always check to see if the first measurement and the second measurement differ from each other regardless of the A/B conditions. This could happen due to a fatigue effect (second score is always lower) or a learning effect (second score is always higher). If this occurs, the design should no longer be analyzed as a simple paired-sample t-test but requires using a factorial analysis approach that simultaneously considers both condition and order effects (technically this is called a mixed-model factorial ANOVA). Needing to do this does not make the analysis intractable in general, but the risk of discovering the experimental design requires more complex analysis is one to be aware of when planning a within-participants design.