Conducting the Experiment

Alternatively, this topic might be called experiment pragmatics. We will consider details involved in actually conducting an experiment.

Topics include:

1. Who should be the participants.

2. How independent variables are manipulated and how to check their effectiveness.

3. Choice of dependent variables and checking their effectiveness.

4. Choices for comparisons (control conditions or baselines).

5. Pilot testing to check the design and reveal problems.

I. Research Participants

How do you choose who will be the participants?

The two most common populations for participants are rats and college students. But how representative are they of the larger population to which we would generalize our conclusions?

If we use rats as subjects, can we generalize to other species regarding the mechanisms that control behavior? The basic answer is that we do, but that we continue the studies looking for both similarities and differences so that we know how much generalization is warranted.

How representative are college students of humans? For processes shared by all humans, such as basic sensory capabilities, simple forms of learning and memory or basic language abilities, college students are quite representative. They may also be reasonably similar to other groups in basic social processes (evaluation of self and others), processing of emotional information and personality.

For psychological processes that vary with intelligence, experience, age, and culture, college students may not be representative. For example, imagery is a good memory strategy for college students, but young children and older adults show little benefit from it. There are also differences across cultures in the effectiveness of imagery.

II. The Independent Variable

There are basically two things that you can do: manipulate events or manipulate instructions. In either case, the starting point is to construct an operational definition of the IV based on the ideas that you wish to test.

There are a number of issues that need to be addressed as part of this process:

1. Is the manipulation "transparent"? Do you tell the participants exactly what you want them to do or do you use deception and/or a cover story?

2. How extreme or strong do you make the difference between conditions?

3. What will it cost to do the study and do you and the participants have the resources?

A) Manipulation of the IV

This can be "transparent", meaning that simple alterations in the instructions or materials produce the different conditions.

Alternatively, this can involve deception, staged events, and/or the use of a confederate to make the situation "realistic".

The general principle is that the IV needs to reflect the concepts in the hypothesis in a plausible and realistic way.

1. Straightforward IV Manipulation

As an example of a straightforward or transparent manipulation, consider the "Stroop Effect". The basic question is how well can individuals attend to one set of information and ignore irrelevant information. Here, we are studying *selective attention*.

Construct three lists of items. The first is a series of color words (red, green, blue, yellow) written in black ink. The second is a series of color patches, in different ink colors (red, green, blue, yellow). The third list has color words written in conflicting ink colors (e.g. the word "red" written in blue ink - red).

Two groups of subjects participate. One group is given the list of color words in black ink and asked to **read the words**. They are also give the list of color words in conflicting ink colors and asked to **read the words**. This is the *reading* group. Their IV is the list type (color words in black ink, color words in different ink colors) and the DVs are their speed and accuracy in doing the task. In this case, we find that their speed (and accuracy) for doing the two lists are the same.

The second group is given the list of color patches and asked to **name the ink colors**. They are also given the list of color words in conflicting ink colors and asked to **name the ink colors**. This is the *color naming* group. The IV is again two lists (color patches, color words in different ink colors) and the DVs are speed and accuracy. These participants are MUCH slower for the color words in conflicting ink colors than they are for the color patches.

The materials are illustrated on the next page. Try going down the right hand column and name the ink color for each item. Try going down the center column and name the ink color for each item. Time yourself as you do this for each list.

For reading, try reading each word in the left hand column. Now, do the same (read the word) for each item in the center column. Time yourself. Which of these four conditions was harder (took longer with more errors)?

Sample Stroop Task Lists

Reading Control	Conflict	Naming Control
RED	GREEN	XXXX
BLUE	RED	XXXX
YELLOW	BLUE	XXXX
GREEN	YELLOW	XXXX
YELLOW	BLUE	XXXX
RED	RED	XXXX
GREEN	YELLOW	XXXX
BLUE	GREEN	XXXX
GREEN	BLUE	XXXX
YELLOW	YELLOW	XXXX
BLUE	GREEN	XXXX
RED	RED	XXXX
BLUE	YELLOW	XXXX
GREEN	GREEN	XXXX
RED	RED	XXXX
YELLOW	BLUE	XXXX
GREEN	RED	XXXX
BLUE	GREEN	XXXX
YELLOW	BLUE	XXXX
RED	YELLOW	XXXX

This is a straightforward manipulation. The participants are instructed in the task and the materials. There is no deception and no cover task is necessary.

However, we would delay explaining why we are doing the experiment until after the participant is done. This is to avoid biasing the participant's behavior.

Note that in this example, each participant is responding to two lists. The order in which participants did the two lists would be counterbalanced. If we wanted to make the experiment into a completely within-subjects design, how would we do that?

2. Staged IV Manipulation

Consider the study on bystander assistance by Piliavin et al. (1975). An emergency was faked on a New York City subway car. The dependent measure was whether passengers would come to the aid of the "victim"?

A white male carrying a cane stumbles and falls to the floor. There are two IVs:

Half of the time the victim has a "big, ugly red birthmark" on their face. The other half of the time they do not. The same "victim" is used in both cases and the birthmark is done with makeup.

Half of the time, an observer wearing a white medical jacket is standing nearby. The other half of the time the same observer is present, but without the medical jacket.

This results in four conditions (all possible combinations of the two levels of each of the two IVs): birthmark, no intern; birthmark, intern; no birthmark, no intern; no birthmark, intern. The control or baseline condition here is the no birthmark, no intern condition. The DV is a count of how often someone comes to the aid of the "victim".

The results show that either the presence of an intern or the presence of a birthmark on the victim reduced the percentage of trials on which a bystander came to the aid of the victim. The combination of intern presence and a birthmark on the victim dramatically reduced the percentage of trials with bystander assistance.

3. A comparison of Straightforward and Staged Manipulations – Personal Space

Humans seem to have a conception of a region, around them, which others are not supposed to intrude into (without permission): Personal Space.

The question is how do we measure or evaluate a concept like this which can not be sensed directly? The approach used to evaluate this concept is to invade the personal space of individuals and measure resulting changes in their behavior.

a) Direct assessment. Kinzel (1970) had subjects stand in the center of a 20 x 20 ft room. The experimenter then approached them, from one of eight different directions. The subject was to say "stop" when the experimenter got too close.

The independent variable is the direction, the dependent variable is the distance when the subject said stop.

If there was no "personal-space" around an individual, then the experimenter should have been able to walk right up to each subject. In fact, subjects did say stop before the experimenter got that close.

This is a direct measurement of the space. However, is this measurement merely an example of *demand characteristics* (a type of reactivity)? Is the subject simply doing what he/she thinks the experimenter expects?

b) Indirect assessment. Barefoot et al. (1972) did an indirect test of personal-space by giving subjects the opportunity to invade the experimenter's personal-space. They used *unobtrusive measurement* in a natural setting (a field experiment). The experimenter sat near a public drinking fountain (distances of 1, 5, and 10 feet) and counted the number of people who passed by and the number who paused for a drink.

The independent variable is distance from the fountain. Ten feet is large enough to be outside an individual's personal-space and represents a control condition. The dependent variable is the percentage of individuals passing by who pause for a drink. The percentage is used because there is no guarantee that the same number of individuals will pass by in each condition.

Results:

Distance	Percent who drank from fountain
1	10
5	18
10	22

Using different operational definitions of how to measure personal space, both experiments find support for the concept. Since the second experiment used an unobtrusive measurement in a natural environment, we are fairly certain that there were no demand characteristics (subject reactivity) contaminating this result.

Note – This study involved public behavior in a public setting. All data were recorded anonymously.

B) Strength (Size) of IV Manipulation

Three factors are important:

1. Make the manipulation of the IV large enough to produce an effect.

2. Make the manipulation similar to what happens in the real world.

3. Make the manipulation appropriate for the hypothesis.

When research is in it's early stages on a topic, ensuring that you get an effect is important, so making the strength of the IV manipulation large is appropriate.

C) Pragmatics – Cost of IV Manipulation

Researchers have to live with cost constraints. They have limited money for equipment, personnel, paid participants, etc. If additional resources are necessary to do the research, the investigator tries to interest an outside agency in funding the research.

III. The Dependent Variable

A) Type of Dependent Variable

1. Self Report – Ask the participant to report their feelings, attitudes, thoughts on something. Basic issues on recording responses same as for surveys (e.g. open-ended verses closed-ended questions, etc.).

2. Independent Observer – Direct recording of behavior by observer or machine. This includes what a person does and how long they take to do it.

3. Physiological Measure – A measure of the physiological response of the body, including heart rate, brain activity (the electroencephalogram or EEG), galvanic skin response (GSR), and electromyogram (EMG). These measures have long been known to be related to psychological processes (e.g. anxiety, stress, arousal and attention).

B) DV Sensitivity

In all cases, we want our DV to be sensitive to the range of variation that we expect to find.

The DV should not be so easy that everyone does very well. In this case, we would have a tough time observing the effect of the IV because the DV is so easy everyone is at *ceiling*.

The DV should not be so hard that everyone does poorly. In this case, we would have a tough time observing the effect of the IV because the DV is so hard that everyone's performance is at *floor*.

C) Multiple Measures

Sometimes, no one measure is the "best" for our hypothesis. In this case, we may use multiple (two or more) DVs. We could use GSR, EMG, and heart rate together as physiological measures. To investigate effects on people's health, we might tabulate doctor visits, number of days missed from work, and subjective ratings of health. In studies where speed of response is important, we also measure accuracy.

D) Dependent Variable Pragmatics

Some types of measurement are easier to make than others. Some require special equipment (e.g. GSR, EEG, EMG). Using multiple human observers to code behavior takes more time/people than an automated measurement system, but might be necessary for certain types of behavior.

IV. Control Conditions and Baselines

A) Subject Reactivity

Participants may behave differently when they know that they are being observed. However, they may also form expectations about the experiment and behave according to these expectations. This is called *demand characteristics* (a type of reactivity).

If a participant forms their own hypothesis about what is being studied and attempts to cooperate with the investigator by behaving accordingly, this could be a problem. Orne (1962) has shown that participants are, in general, cooperative, so this could be a big problem.

The issue is not whether participants do what the experimenter has requested, but whether they alter their behavior to conform to an expectation that *they* (participants) have about what should happen in the study.

To avoid this problem, we take a couple of steps:

1. We don't tell our participants why (theory, hypothesis) we are doing the study until after their participation is done.

2. A cover task (deception) could be used to disguise the true nature of the study.

3. We embed our critical measurements of behavior in a larger set of measures. The extra measures are fillers. They are present to disguise what we are doing.

4. We ask participants about their perception of the task. Had they guessed the hypothesis? Were there other features that led everyone to behave in a particular way?

5. *Placebo Control Groups*. To control for participant expectations, we run them through a sham treatment. By disguising which are the real treatments and which are not, we can control our participant's expectations (basically, they all have the same expectation). This is commonly done in drug studies. This approach is known as a single-blind design. Participants are blind as to the condition that they are in.

We can even study the effect of participant expectations. This is a part of the work of Marlatt on behavioral effects of alcohol.

Example - AT&T ran a study at one of its Western Electric plants on worker productivity. Six workers participated in a longitudinal study of factors affecting the assembly of telephone switching relays. The hours of work were lengthened, shortened. Breaks were added, deleted. The workers method of payment of wages was changed (piece work, hourly), and even a light lunch was served.

Most of these changes had the same effect: They *increased* productivity. The usual interpretation is that the workers, *knowing* that they were the subject of a study, *cooperated with the study and worked harder*.

This is known as the **Hawthorne effect**. It is an example of demand characteristics and is a form of subject reactivity.

B) Experimenter Bias

The researcher knows the hypotheses being investigated. This could lead to one (or both) of two problems:

1. The researcher unintentionally treats the participants in different conditions differently.

2. In coding the data, the researcher treats ambiguous cases differently for different conditions.

In both cases, the results of the study would be suspect since we don't know how much of the participants' behavior was due to the IV and how much was due to the researcher's actions. Together, these two possible problems are called *Experimenter Bias*.

Is this a real problem?

3. Clever Hans, the horse who could "count" and do arithmetic.

4. Graduate students who trained rats were told that the rats were maze bright or maze dull. They found differences in the rats' behavior (speed of learning) even though all rats were from the same genetic strain.

5. Clinical psychologists who are told a video is about a patient rate the person's behavior as more disturbed than when told the video is about a job interview.

Clearly, experimenter bias *could* be a big problem.

C) Dealing with Experimenter Bias

1. Train researchers. This includes practice at dealing with participants consistently.

2. Automate data collection and scoring.

3. Run conditions simultaneously so that role of researcher and her/his behavior is the same for all participants.

4. Keep research assistant who deals with participants blind as to the hypothesis and the condition that each subject is participating in. When this is done along with a placebo control condition, this is known as a *double-blind* study.

In studies of the efficacy of a treatment (drug, behavioral), the double-blind study is the gold standard (best way to do the study).

V. The Pilot Study

This section deals with the process of "trying things out" to see how well they work.

A) Write a Research Proposal.

The process of writing out the details of the study and why they are done a particular way helps to highlight problem areas.

This will also be required for human subjects approval.

B) Run a Pilot Study.

A Pilot Study is a small scale version of the experiment. It is typically done with only a few participants. It is used to check that the instructions are clear, that participants can do the task, to see if the study runs in the time allotted.

Debriefing afterward can include asking questions about how the participants saw the task, their expectations and strategies.

Pilot studies also allow researchers to practice their role.

C) Manipulation Checks.

Did the manipulation of the IV succeed? This is the issue of construct validity that was raised earlier about questionnaires.

If you manipulated anxiety, were the participants in the high anxiety group more anxious?

If you manipulated physical attractiveness of people in photographs, were the photographs perceived as intended?

If you presented words in various backgrounds of noise to vary difficulty in a listening task, did the noise produce the intended differences in difficulty?

Note – we are not asking here if the IV influenced the DV. Rather, were the IV differences were "real".

1. In a pilot study, you could include questions, physiological measures, or a behavioral measure to directly check that the manipulation of the IV was successful.

2. Why not do this in the main experiment?
The manipulation check may make the experiment too long. It may increase subject reactivity. It may be too costly to do in the full experiment.

3. Advantages:
First, if we find that the IV manipulation has failed, we can re-design the study.
Second, if the manipulation works, but in the experiment we find no effect of the IV on the DV, at least we know that the lack of any effect is not because the IV manipulation failed.

VI. Debriefing

A) This is often a requirement, and it can serve an educational component.

It is used to dispel any misconceptions about the research and deal with any use of deception.

When the research involves students as participants, we try to use the debriefing as an educational tool. This is also often done with medical and clinical research with all populations of participants.

B) It is an opportunity to learn from the participants. They can be asked about their expectations, strategies and perceptions of the task that they participated in.

C) Participants may be asked not to discuss the study with "other potential participants", such as classmates. This request is designed to keep other potential participants naïve as to the purpose of the study before they participate. This request is generally used only in studies involving deception or a cover story to conceal the nature of the study.

Answers to Chapter 8 Sample Multiple Choice

1) – a; 2) – d; 3) – d; 4) – c; 5) – c

Chapter 8 Conceptual Review

1. Dependent Variables: Accuracy (words correctly recognized) on the recognition tests.

2. Independent Variables: *One* is the type of question presented with each word. There were two alternatives: focused on word meaning or focused on word sound. *The second* independent variable is the type of recognition test (standard or rhyme).

3. The type of recognition test (standard or rhyme) is between subjects. The description explicitly says that one half of the participants were given one type of test and the other half were given the other. The other variable is not between in that the description says that all participants answered the questions.

4. There are no participant (subject) variables. The study does not describe any characteristics of the participants that were measured. This does not mean that all participants were identical, but for a variable to be a participant (subject) variable, the researcher has to actually measure it.

Chapter 9 Sample Questions

1) In a single blind design: a) the subject is blind to the condition that they are in b) the person testing the subjects is blind to the hypothesis being tested c) the person testing the subjects is blind to the condition the subject is assigned to d) b & c above

2) In an experiment, mice are randomly assigned to two groups. The person who runs the experiment is told that one groups consists of "maze bright" mice and the other of "maze dull" mice and he/she finds a difference in their performance in the experiment. This is an example of: a) subject reactivity b) differential assignment of subjects and the formation of non-equivalent groups c) experimenter bias   d) all of the above

3) The important characteristic of a control condition is that it provides a base line against which the manipulated variable(s) is(are) compared. Therefore, the base line: a) should never have any treatment b) may require some treatment to control subject reactivity c) often involves a placebo when the experimental manipulation involves pharmaceuticals d) b & c above

Answers to Chapter 9 Sample Questions

1) – a; 2) – c; 3) – d;