

March 24 Surveys

Adler and Clark distinguish between surveys (standard measurements and statistical analysis) and questionnaires and structured interviews (ways of collecting information). But surveys and questionnaires or structured interviews go together so often that people usually don't make this distinction. If a sociologist says "survey research" they probably mean research based on structured interviews. A structured interview is where an interviewer reads a questionnaire to a respondent and writes down the answers. That's the most common form in which serious survey research is done. Sometimes people are given a questionnaire themselves and asked to fill it out. That's common in small-scale survey research, like what we'll do for this class. I won't follow their distinctions in terminology rigorously.

Another term--"poll". Same as a questionnaire or structured interview, but usually used for ones on contemporary political issues, like those reported in newspapers, rather than those intended for academic research. But sometimes social scientists will analyze data from polls, so the distinction is kind of arbitrary.

The survey begins from a simple idea--ask people questions. By doing this, you can find out about what they think, or about what they do. But has a special feature. Basic principle--standardization. Standard format for everyone--if it can be varied, it is only according to explicit rules. E. g., if a person doesn't have children, you would skip questions about children.

This is different from the way that a journalist or sociologist might ask questions in a qualitative interview, where you may adjust your approach from the person. For example, you might change the wording depending on how well informed or interested the person seems to be. If a person doesn't seem to understand, you might explain it further, or give them an example. If the person asks you what you think, you might tell them.

In a survey, interviewers are told not to do any of that. Essentially, the idea is to do it the same way for everyone, and not to add or modify anything on the questionnaire. Also, don't give any extra information. E. g., if a person asks the interviewer what she thinks, the interviewer is supposed to give some polite statement to the effect that I can't say, we're interested in what you think, etc.

Essentially the idea is to make the process completely mechanical--no choice by the interviewer, and no choice by the person being interviewed except what answer to give to each question. Can't achieve this totally, because there is a human element, but that's the idea.

Why? Essentially to eliminate differences caused by the interviewer. You want to know that any differences you find are due to differences in what people think. E. g., if a survey in 2003 finds something, and you compare it to a survey taken in 1983, you want to be sure that differences reflect changes in what

people think, not differences in who the interviewers were or how they did things.

Goal is to make "replication" possible--producing results, doing things again in the same way. Important in science, because it means other people can check your work. If I say I took a survey, and 70% said that they were satisfied with their jobs, then a person who doubts me can go out and check. If he didn't get the same results when following my instructions, that would cast doubt on my claim. On the other hand, if I say I talked to some people, and my judgment is that most people are satisfied with their jobs, can't really be checked. If someone else went out and talk to people and decided that most people weren't satisfied, all you could say is that our judgments were different.

There are some things lost by standardization--a skilled interviewer working with less structure could probably tell you more about what any individual thinks. But then replication wouldn't be possible, and you also wouldn't be able to interview as many people. So surveys are best for comparisons of large groups of people--men vs. women, people in 1983 vs. people in 2003, etc. Not as good for detailed understanding of any one person.

One issue with replication: what is "the same"? There are different languages in the world, and often in the same country. Also, even within the same language, different people may use words differently, and meanings may change over time:

IF THE QUESTION OF NATIONAL PROHIBITION SHOULD COME UP AGAIN,  
WOULD YOU VOTE FOR PROHIBITION OR AGAINST IT?

Responses:

FOR	23%
AGAINST	72
NO OPINION	5

Survey Organization: Gallup Organization  
Population: NATIONAL ADULT  
Population Size: 1357  
Interview method: PERSONAL  
Beginning date: JUN 4, 1950 Ending date: JUN 9,  
1950

R18 If the question of national prohibition should come up  
again, would  
you vote to outlaw the sale of alcoholic beverages, or not?

Responses:

Outlaw	20%
Do not outlaw	78
Don't know	3
Refused	*

\* = less than .5 percent

Survey Organization: I.C.R.-International Communications  
Research  
Research Sponsor: Harvard School of Public Health  
Population: National adult  
Population Size: 1001  
Interview method: Telephone  
Beginning date: JUL 13, 2001 Ending date: JUL 17,  
2001

March 26

Methods of administering surveys:

4 basic methods: (1) face-to-face (2) telephone (3) mail (4) internet. The book mentions group-administered questionnaires, like the one described in the rape attitudes article, which was given out in classes. However, this isn't a major method. Only possible if you want to survey people who naturally would be in one place, like students or people who work at a particular company. If you want to survey a cross-section of the public, you have to go to them individually.

Traditionally, face-to-face interviews were most popular. Now telephone surveys have become widely used. In the US, they are the most common--pretty much all commercial polls are by telephone, and most academic surveys too. Mail surveys have never been as popular for major studies, although they are often used for small projects. Commercial polls don't use them because they're slow--many people don't fill them out right away, and may need reminder letters. The most prominent example of a mailed questionnaire is the US Census. Web surveys have started to appear recently. Most useful for special populations, since many or most people (even in America) don't use the internet.

Biggest advantage of telephone surveys is lower cost and greater speed. Biggest advantage of in-person survey is that it's possible to have longer surveys and more complex questions, including questions using visual cues. The major disadvantage is cost--they are much more expensive than the other kinds because you have to add the interviewer's travel time. Advantage of mail questionnaires is probably that they minimize "social desirability bias." People may be more willing to admit to things that might be embarrassing when they don't have to actually speak out. Web survey is similar to mail, with the advantage of greater speed. The major disadvantage is that a lot of people don't have web access, or aren't very comfortable using it. So it's useful for some sorts of people--e. g., professors or students. One additional point: sometimes an in-person survey contains a mail-in supplement. Usually these get a pretty good response rate because people already feel some commitment to you, and possibly some interest in the survey.

Key components of a serious survey--identify population, select people from it (preferably random selection) and make serious effort to get those people to respond. "Surveys" based on asking for volunteers to call in or go to a web site are essentially worthless--they're basically a form of entertainment. Only a small fraction of the population will respond, and they're very likely to be different from the people who don't respond. Also, people may respond multiple times.

A good electronic survey contacts people by phone or mail, and gives them an individual PIN number which they need to access the site, and which can only be used once.

It's also important to make multiple efforts to contact people, or remind them to return a form. There will still be people who you can't reach, but you can reduce the number a lot by repeated efforts. Also, there will be some who simply refuse to respond. You can't prevent them from doing this, but you need to keep count of them. It's also sometimes possible to get at least a little information on them (e. g., in a phone survey, their number gives some information about where they live, and you can guess their sex with high accuracy; in an in-person survey, you know where they live, their sex, and can make pretty good guesses about age and race). The reason to do this is to see if they are different from the people who do respond in any obvious way.

March 29

Designing survey questions:

Closed-ended vs. Open-ended. Closed-ended questions include a list of possible answers—like a multiple choice test. People are asked to choose some number (usually one, but sometimes more). Open-ended just gives a question, doesn't provide possible answers.

Closed-ended is more common. Reasons (1) less work for interviewer and analyst. With open-ended, you have to write down exactly what people say, and then classify ("code") it before you can analyze it. (2) Less need for interpretation by person who classifies. For example, suppose you ask a person how he likes his job, and he tells you that it has some good things—a,b,c—and some bad things—d,e,f. How do you classify him—favorable, unfavorable, somewhere in the middle? If you give him the possible answers—very well, pretty well, not too well, not at all, then he will make the judgment himself. Opposite problem also comes up, and actually seems to be more common—people may not tell you much in an open-ended question. E. g., if you ask someone how well he likes his job, and he says "it's OK". That could mean a lot of things. If you give him the answers I list above, you know that it's a middle response. (3) May be easier for people to answer a closed-ended questions. E. g., suppose you ask people what's the biggest problem facing the US today. There are lots of things that people might be concerned about, but it's hard to bring them all to mind and sort through them. If you give people a list, like economic situation; threat of terrorism; crime; environment; taxes; ..., that may jog their memory.

Open-ended questions are used mainly with things where there are a limited number of possible answers—e. g., what's your major; in what state were you born. Sometimes used to supplement closed ended questions. For example, you might follow up a closed ended question by asking "why do you feel that way."

Good questions. Goals (1) clear; mean the same thing to everyone (2) balanced—don't trick people or suggest that they ought to answer in a certain way.

There are no absolute rules for designing questions. Basically, the right question is one that "works" in practice, and you can't know that until you try. So when in doubt, it's a good idea to pre-test questions.

Pre-test—give a trial version of survey. Then ask people if they were sure about everything, if there were any questions where they weren't offered an answer that really fits with what they think. Also, you might ask people to explain what they had in mind when they answered some questions.

For understanding, try to use everyday language. Avoid ambiguous terms, or terms that aren't widely understood. E. g., in controversy over President Clinton a few years ago. "Impeach"—brought to trial before the Senate; "Convict"—removed from office by the Senate. But a lot of people may not understand this distinction, and may think "impeach" means "remove from office."

So when polls asked about this, they didn't say "should President Clinton be impeached?" Rather, they normally asked something like: "do you think President Clinton should be removed from office?" Sometimes they would include some explanation of terms:

#### **QUESTION:**

As you may know, the House of Representatives is expected to vote soon on whether or not to impeach Bill **Clinton**. If the House impeaches him, the Senate will hold a trial to decide whether or not **Clinton should be removed from office**. Based on what you know, do you think the House should or should not impeach **Clinton**?

#### **RESULTS:**

Should impeach	- 40%
Should not	- 58
No opinion	- 2

Some questions are short and simple. For our survey, we should try to keep to that sort of question. But sometimes it's a good idea to include some background on your topic. There are some issues that aren't all that familiar to most people. For example, if you're talking about some proposed bill before Congress, many people might not have heard of it at all, others will have heard of it but won't know the details. So if just named the bill, most people might not be able to offer any opinion. However, if you describe the basic features, people may have something to say. And sometimes you might want to ask about hypothetical ideas that are pretty remote from everyday experience. In this case, you will definitely need to give some background.

Examples :

**QUESTION:**

The so-called '**flat tax**' would generally reduce tax rates but it would also eliminate many of the current deductions. How do you feel about the '**flat tax**'?

I'm for it	-	31%
I'm against it	-	15
Don't know	-	54

[mail survey 1985]

**QUESTION:**

There has been a lot of talk about changing the income tax laws to make the system fairer to everyone as well as make the forms easier to fill out. One proposal is to do away with all exemptions and deductions and impose on everyone a **flat tax** of 14% on all income above \$8,000 a year. Would you be in favor of a flat income tax of 14% a year or do you prefer the income tax system we now have?

Flat 14% tax	-	39%
System we now have	-	42
* Don't know	-	20

[Phone 1982]

**QUESTION:**

Thinking about the federal tax system--which of the following would you prefer?...The current system in which people with large incomes pay higher tax rates than people with smaller incomes--and taxpayers can take deductions for charitable contributions, interest paid on home mortgages, and other items, or a different system in which all people would pay a **flat tax** rate of 17% regardless of income and no deductions would be allowed. Families with four or more people and incomes of less than \$36,000 per year would pay no income tax.

**RESULTS:**

Current system	-	46%
Different system	-	50
No opinion	-	4

[1996, telephone]

March 31

Fairness: in general, should ask questions in balanced form—"do you agree or disagree" rather than "do you agree." On controversial issues, sometimes start with a "some people think this, other people think that" introduction. Try to avoid using loaded or emotional terms—e. g., if you're asking about people who work for the government, call them people who work for the government rather than "government bureaucrats." If you give information or argument on one side, give it on the other too. E. g., question on welfare system

### QUESTION:

On another subject, Congress has passed a bill that makes basic changes in the **welfare** system, ending the federal guarantee of public assistance for the poor, requiring able-bodied recipients to work after two years, cutting off benefits after five years and cutting back on food stamps. Supporters say this will end **welfare dependency** and save more than 50 billion dollars over six years. Opponents say it will let children go hungry and won't help poor people find the jobs they need. What's your view--do you support or oppose this bill?

### RESULTS:

Support the bill	- 74%
Oppose the bill	- 23
Don't know/No opinion	- 3

However, difficult to get agreement on what's fair. Some people may think a term is loaded. People may disagree with the definitions or the factual information you give. If you give arguments on both sides, they may think the argument on their side isn't as strong as it ought to be.

In cases like this, usually a good idea to ask a number of different questions, changing the wording, or adding different information, or considering different possibilities. If you get a similar distribution no matter how you ask it, you have a clear conclusion. If not, you can look more closely at the questions and think about why you are getting different answers.

Information you get from surveys doesn't "speak for itself." It has to be interpreted. In particular, you need to have some idea about how people understood a particular question. Example:

Question:

R47 When a person has a disease that cannot be cured, do you think doctors should be allowed by law to end the patient's life by some painless means if the patient and his family request it?

Responses:

Yes	72%
No	26
No opinion	2

Survey Organization:	Gallup Organization
Population:	National adult
Population Size:	1012
Interview method:	Telephone
Beginning date:	MAY 6, 2002

If you think about the words used, they can all be interpreted in different ways. E. g., "disease that can't be cured." Some people suffer from diseases that can't be cured, and reduce their quality of life, but aren't fatal. In this case, I'm confident this is not what people had in mind--they meant the last stages of a fatal illness. And what does "family" mean? Aunts, uncles, cousins? Minor children? Again, probably not what people had in mind here. I assume they mean spouse, adult children, parents, maybe brothers and sisters. I would say that these assumptions are "common sense"--ie they're things I know from living in this society,

But its often useful to check the "common sense", especially when interpreting what questions mean to people. This is the idea of the "random probe" discussed in the article. Simply ask people to briefly explain why they answered the way they did. Article on random probes--usually reasons people gave made sense. Not always articulate, but usually fit with how they answered. But some cases in which they often didn't, especially rating as liberal or conservative. Suggests that many people don't understand these terms, or understand them in idiosyncratic ways. Also, sometimes made references to things that might help to explain why they answered that way--e. g., people who said the taxes they have to pay are too high usually mentioned unfairness rather than the direct financial burden.

Sometimes a random probe will reveal that some people understand the question differently than the interviewer meant it. Example: survey of black people in 1968:

"Do you feel Negroes have more, less, or the same duty as whites to obey the law?"

More	5%
Same	90%
Less	5%

Assumption was that most people would say same, but that some would say "less," on the grounds that they would think that the political process and laws were biased. Didn't really expect any significant number to say more, but just included it for the sake of completeness. Surprised that about as many said more as said less. But they'd done a random probe, and found that most of the "mores" said things like "I do because they get picked up more at night. Also, they can't go downtown at night without being

searched." That made it clear that people who said "more" didn't think "duty" meant moral obligation, but thought of it as a practical requirement.

Figuring out reasons. Most sociological research is devoted to explanation. E. g., why people do or do not smoke, why they are liberal or conservative, why they do well or badly in school. How do you find out about explanation from surveys?

One possibility is to just ask people. On some issues, this is reasonable. You figure that they'll know the answer and be willing to tell you. But on some questions, people may not know the answers, or not want to admit them. On others, they may give valid answers, but not at the level you're interested in. What counts as a good answer to "why" depends on your point of view. When people are asked about why, they probably focus on causes that are immediately visible to them, not on larger social forces. E. g., if you asked people who made a lot of money how they did it, they would probably tell you a lot about their hard work and say less about the general economic trends that might have helped them. Their answers aren't necessarily wrong, just incomplete.

April 2

Explanation in surveys usually doesn't involve asking people directly, but looking at relations between variables. Does one thing predict another thing? If so, can you tell a plausible story about how they are connected.

Example: why might people cheat on tests or assignments? There are lots of ideas. Pressure, low moral standards, ignorance of rules, low perceived risk... These have different implications. If it's pressure, we should find that students in tough majors and students with high career aspirations cheat more. If it's generally low moral standards, you wouldn't get that difference among majors. But you would expect to find that people who showed signs of willingness to break rules elsewhere (e. g., traffic violations, unauthorized music downloads, lying to protect friend) would have higher rates of cheating.

Hypothesis—statement about the values of variables, usually the relationship between variables. The simplest hypotheses, which is what we'll look at, involve just two variables. E. g., "On the average, women feel less safe than men." Says value of the variable measuring feeling of safety is related to the value of sex (male vs. female).

A hypothesis may mention other qualities that aren't variables. E. g., "among UConn students, women feel less safe than men." UConn students isn't a variable, it's a specific group--to be a variable, something has to involve differences among groups. The point of mentioning UConn students is to specify where the hypothesis applies. A lot of time, people wouldn't mention it explicitly, but sometimes they would.

## April 6 Hypotheses/Survey research designs

### Points about hypotheses

1. Hypothesis has to involve only facts--to be, in principle, objectively right or wrong. It can't be a matter of taste or values. "Students should meet with their advisors at least once a semester" isn't a hypothesis. You could make arguments for or against this idea, but you can't show it's factually true or false. However, "*most students think* that students should meet with their advisors at least once a semester" is a hypothesis. To find out if that's true, you'd just need to ask a representative sample of students if they think that students should meet with their advisors at least once a semester. If most of them say yes, the hypothesis is supported; if most say no, it's refuted.

2. Not a definition--if I say that prejudice is an irrational dislike or disapproval of some group, that may or may not be a good definition, but it's not something that can be proven true or false.

Often a hypothesis is understood to involve a cover a class of events, not a single event. The reason that a hypothesis is usually understood to be about a class of events is that unless you have multiple events, you can't compare them. And finding out whether a hypothesis is true is basically a matter of comparing--e. g., if you think that men earn more than women, you compare men to women.

Important thing about almost all hypotheses in the social sciences. Probabilistic rather than deterministic--that is "most of the time" rather than "all the time." Sometimes stated explicitly, but often left implicit.

Example: "men are taller than women". If you understand it as deterministic, it would mean any man is taller than any woman--that is, the shortest man is taller than the tallest woman. Clearly false. But if you understand it as probabilistic, it's true--average man is taller than average woman.

Unless someone says that their hypothesis is deterministic, you should understand it as probabilistic. E. g., suppose you have a hypothesis that children with a disorganized home environment are more likely to commit crimes--not all such children grow up to commit crimes. Also, some children have a fine home environment and grow up to be criminals anyway. So if you understand it as a deterministic statement, it's pretty obviously false--all you need to do is come up with one example against it. But if you understand it as probabilistic, it may be true--would have to investigate. Very rare that a social scientist would propose a deterministic hypothesis: in fact, if someone claims that they've discovered a rule that always works is a pretty good sign that they shouldn't be taken seriously.

### Research designs

I'll refer to surveys, since the issue comes up particularly often in survey research, but the basic distinctions among research designs apply to all kinds of information gathering.

1. Cross-section: a snapshot--get information at one point in time. Focus is on the present--what people think or do now: you might ask people about past behavior, but in practice can only ask about recent events or major things in their life, like when you married, or finished school, etc. Asking an older person how they voted in the election of 1956, or how they liked the job they had then, would be pointless, because you couldn't count on their memory being accurate after a long time has elapsed. Even asking a senior in college about how he/she felt freshman year would be questionable, since experiences since then could color memories.

2. Panel (longitudinal): follow the same people over a period of time, ask them the same questions at various times. This needs to be planned in advance, because you need to make sure that you keep in touch with people. Usually when you did a survey, you wouldn't preserve the names of the people involved, and therefore couldn't look them up again later. So if you wanted to do a panel, you'd have to get people's names, and also make arrangements for maintaining contact with them.

3. Trend study (repeated cross section): ask the same questions to different samples from the same population at different times. E. g., one important source of information is the Census long form, which is given to a random sample of the American people every 10 years. Each year, the sample is chosen separately, so it's a different group of people each time. But each time it can be taken as representative of the American people at that time. Interesting point about a trend study: you don't have to start out intending to do one--a later investigator can make a trend study simply by repeating questions someone else used before. In fact, I repeated some questions on our class survey, so we will have a trend study for those questions, although it will involve a short period of time.

3a. Cohort study: ask same questions to different samples from the same "cohort" at different times. A cohort is a group of people who started something at the same time. Most common form is a "birth cohort"--people born at a particular time. But you could define cohorts by other experiences. For example, you could study people who started college in 1995. Most of those people would have been born at the same time, but some would be older or younger.

April 7

Cohort study and trend studies are related, because both involve taking cross-sections at different times. But the population from which you take the sample is different.

Limitations of cross-section: (1) doesn't tell you about change, which is often interesting (e. g., is there more or less racial prejudice than there used to be)? (2) Doesn't let you figure out causal order except in special cases. We've mentioned causality before, and will talk about it more again. Causality means that one thing comes first and another follows as a result. With a cross-section, you measure everything at the same time, so you don't know which came first. Sometimes it's obvious (e. g., sex can be regarded as a cause of occupation; occupation obviously

isn't a cause of sex). Or sometimes you can ask about the past and be confident of getting accurate answers--e. g., if you're interested in the relationship between education and having a child. When did you complete your education? When did you have your first child? People are generally going to remember these facts. But there are questions, e. g. opinions about things and what kind of job a person has, where either order might make sense. (E. g., compare people in different jobs and their views on crime and law enforcement). Usually can't rely on people's reports of what they thought in the past.

Trend study tells you about change. It's particularly useful when you're interested in some larger forces that affect all kinds of people. E. g., national trends in things like drug use, racial prejudice, etc. Trend study tells you about change in a whole population, while a cohort study tells you about change in a cohort.

For example, if racial prejudice declines over time, that might be because of changes in the population. The opinions of the population can change even if nobody actually changes their mind (older cohorts dying and being replaced by younger cohorts, as well as by immigrants). Or it might be because individual people become less prejudiced. If you do a cohort study, you can see if there's change within a specific group of people.

Panel helps you figure out causal order; although it's not perfect. The reason is that you can see what came first. E. g., if you could measure someone's opinions when they were still in school, see if they affected the kind of work they went into. And you could measure opinions after someone had been working for a while, see if they were different from when they started work.

Cross-sections are the most common type of survey, despite their limitations. The reason is simply that they're cheaper and easier. (It's good to see if someone has asked about similar issues before, so you can make it a repeated cross-section). The reason that panel studies aren't used more often is the cost and time commitment. Keeping track of people requires a lot of work, especially if the panel is a long one and people widely scattered. Also the issue of "subject fatigue"; and the possibility that participation in the study affects them in some way, e. g., by making them think more about the issue. There are some panels that use national representative samples, but they usually just cover a few years. The few long panel studies are almost all based on local, non-representative samples, because special groups are easier to keep track of.

April 9 Experiments

Sometimes considered the ideal form of scientific study. Experiment is what you'd do if you could, anything else is just a second best when experiments aren't possible in practice. Whether or not this is right, the use of experiments in the social sciences is pretty limited. Rare in economics, political science, anthropology. Common in some areas of sociology, but not most. Psychologists are the only social scientists who use them a lot.

However, experiments need to be considered. They are particularly useful when you want to know whether an association is the result of x causing y, y causing x, or spurious correlation.

Experiment: a planned "intervention" and observation of its result. Contrast with observational studies--just see what happens naturally. Some gray areas: for example, in surveys you're taking an action--asking people questions, not just listening to their conversation and seeing what comes up. But usually wouldn't be considered an experiment, on the grounds that it's a minor intervention. It would be considered an experiment if you exposed people to some experience like seeing a movie or doing some task and then took a survey afterwards to see how their attitudes were affected.

Basic components of social science experiments: At least 2 groups. Measure whatever you're interested in (pretest). Then subject the groups to different experiences. Often you would have a "control" group to which you do nothing, or subject to some "neutral" experience. Then measure again (posttest). Pretest is usually desirable, but not essential. The other parts are essential.

Why use experiments, rather than surveys, observation, etc.? Identification of causality. With a properly designed experiment, you can eliminate the possibility that Y may cause X or that both depending on common or correlated causes. No other method lets you do this with as much certainty.

Example: Suppose you're interested in seeing if students benefit from going to faculty or TA office hours.

	Average course grade
Visited	2.5
Didn't	3.0

Does that prove that the going to office hours was harmful? No-- maybe the people who visited tended to be people who weren't doing well and thought they needed extra help. Without it, maybe they would have done worse than they did.

or suppose it's the other way around:

	Average grade
Visited	3.0
Didn't	2.5

Does that prove that it helped? No--maybe the people who went to office hours were just more motivated or interested, and that was what made the difference in their grades. That is, maybe they would have done just as well (or conceivably even better) even if they hadn't gone to office hours.

Both of these hypothetical cases involve spurious correlation--"something else" explains the differences in group performance. You can control for it to some extent with statistical methods, but to do that you need to measure the "something else". In some cases this is pretty straightforward (family income), but in others it's hard (motivation). And maybe there's some factor you haven't even thought of that makes a difference.

Experiment:

1. Take group of students.
2. Pick half to go to office hours, half not to. Important--need to make sure those two groups of people are the same (or at least similar) in all "important" ways. Sometimes experiments explicitly match people, but usually they just check that they're the same on the average.
3. Measure performance. Would be best to do it at start as well as finish. E. g., give a test of general knowledge of the subject at the start. If your assignment to groups has worked properly, the average should be about the same in both groups. Definitely would need to measure performance in the course.

Problem with this research--will people who are forced to go, get the same benefit as someone who goes voluntarily? And how can you forbid people to go to office hours? Seems to be an obvious ethical problem there, especially if going to office hours really has some benefits. So designing an experiment that is practical and ethical often requires some ingenuity.

## April 12 Experiments

Causal (internal) validity--validity within the experiment. Basically amounts to asking if you really eliminated the possibility of spurious correlation--difference between the groups being explained by something other than the difference in treatment.

External validity (generalizability)--do the conclusions apply in other settings, like other times and places?

Threats to internal validity:

Are groups really (almost) identical to start with? Two ways to make groups identical: matching or random assignment. You can deliberately try to make the groups similar by matching them as closely as possible. But you can only do this with characteristics that you can observe, like age, sex, etc. Hard to do it on more subtle things, like motivation. Random assignment: assign based on a random process like flipping a coin. This usually won't make the groups exactly the same. But they will be similar, and will get more similar as the number of people participating gets larger. Important thing--they'll be similar in all respects, including ones that you haven't measured and maybe even haven't thought of.

Random assignment vs. random sampling: similar idea, but not the same thing. For practical reasons, most experiments aren't based on random samples; usually basically convenience samples.

Random sampling--makes sample similar to population

Random assignment--makes treatment groups similar to each other

Random assignment is important in experiments because experiments are based on comparing groups.

1. May be departures from random assignment. This is likely to happen when experiments are in a natural setting, like a school. Political or social pressures make it difficult to keep up random assignment.

2. Attrition--people drop out. People who drop out may be different from those who stay in. And the people who drop out of different groups may be different. So even if the groups that you start with are the same, the groups that you end up observing may be different. Also more of a problem in natural settings, especially where experiments cover a substantial period of time (e. g., experiment in the effect of reducing class sizes in school).

3. People are affected by knowledge that they're in an experiment (Hawthorne effect) This can also happen in other kinds of research--for example, people taking a survey could be self-conscious rather than answering in a "natural" way. However, it's particularly important with experiments. Various ways in which knowledge can make a difference: people might have an idea about the purpose of the experiment, and try to prove something one way or another. Might be influenced by what the investigators think will happen. Might develop a rivalry with other group. Might feel proud or resentful about being picked for one group, and act differently because of that.

General way to deal with this is to try to control what the participants know. Often investigators try to mislead or confuse people about the purpose of the experiment. Also, usually try to keep groups apart--ideally ignorant of each other. In that case, rivalry can't develop, and people are less likely to figure out what the goal of the experiment is. Separation of groups is pretty easy to do in a laboratory setting; harder in a natural setting (e. g., different classes in a school will know about each other. You could do an experiment based on comparing classes in different schools, but that would make it larger and more complex). If you can't keep them apart, at least try to check for signs of the "Hawthorne effect." After the experiment, ask people what they were thinking about, whether they had a sense of what the experiment was about.

External validity: how widely do the results apply? If they just apply to the particular people and the specific conditions you studied, they're not of much interest. How far beyond that do they go?

Population--what population do your conclusions apply to? It's difficult to use random samples for experiments. Hence, replication with other samples is important. If you just have a convenience sample, you can't say anything for sure about generalizability. But with replication, you can gradually build up some knowledge, although you can't establish it with certainty, the way you could with random sampling.

E. g., one researcher finds something when doing a study based on a sample of undergraduate volunteers at UConn, another might check it with undergraduates at some other college, or a sample of volunteers from a particular community, or people in some other kind of institution (say a retirement home). The more settings that something works in, the more confident you can feel in generalizing. If things work in some cases but not in others, you have another potentially interesting problem--explaining why you get the differences among different groups.

Setting. Under what conditions do your results apply? This is especially important when the experiment is done in a laboratory rather than a natural setting. People may be self-conscious, or otherwise experience things differently. E. g., watching a movie as part of an experiment vs. going to a movie theater with friends or renting it from a video store and watching at home. The results may be valid within the special setting of the experiment, but not apply in the "real world." (Or at least not very widely).

A related issue is the degree to which the proposed cause is manipulated or the duration of the experiment. E. g., studies of violence in the media and attitudes toward violence. Is looking at a few hours of violent TV going to have similar effects to watching many hours over a period of many years? Are the effects you get shortly after the experiment going to continue for a period of time, or will they wear off quickly?

Usually in the social sciences you can only do experiments using small "doses" and short-term effects. For ethical and practical reasons, you usually can't do something that would have a major and lasting effect on people's lives. Example of Stanford mock "prison" experiment, where people were assigned to play "prisoner" or "guard" roles for a period of several weeks. Unlikely that it would be permitted today. But even that couldn't reproduce one of the key aspects of a real prison, which is that you can't have yourself let out.

Interpreting results of experiments: Even though experiments can eliminate the possibility of spurious correlation, there's often some room for argument about how you should interpret the results of an experiment. That is, you may be able to establish that something which went on in the experiment made a difference to the outcome, but you may not be sure what it was and why it mattered. E. g., famous experiment done at Yale by Stanley Milgram on 'obedience to authority'. What was it that made people obey? Would people just go along with anyone who told them to do something? Or was it the prestige of science, or of affiliation with a university? Or was it trust in the laws regulating research ('they wouldn't allow something like this if it really weren't safe')? Or did people treat the whole thing as a game rather than a serious situation?

Some things that you can do to check interpretations:

1. changing conditions of the experiment (as Milgram did by removing signs of Yale affiliation).
2. asking participants what they thought (did you think the man was really being shocked?); looking for other evidence that indicates how they felt (signs of discomfort).
3. Trying to reconcile the experimental results with other information (obedience to "authority" in real life).