PAI 705

McPeak

Lecture 1

Epistemology – the science of knowing.

Methodology – a subfield of epistemology, the science of finding out.

How do we know what we know?

How much is direct experience?

How much is passed on?

Do you know the dark side of the moon is cold?

Do you know they speak Portuguese in Guinea Bissau?

Agreement Reality. The things we 'know' as part and parcel of the culture we share with those around us.

In what ways does this help?

In what ways does it hinder?

Role of Tradition in establishing agreement reality; don't necessarily have to experience to have knowledge.

Role of Authority in establishing agreement reality; experts stating assessments about what is a fact.

Agreement reality is not the same as proven reality, which is part of what we try to test in social science inquiry.

Problems in Social Science inquiry.

Inaccurate observation / lack of attention to specific detail. We need to be deliberately paying attention to particular facts and recording them correctly. Problem of recall of specific item if we were not paying attention to this item.

Overgeneralization. We see one case and we think it applies to all cases.

Selective Observation. We see what we expect to see. We dismiss contrary evidence to confirm the version of reality we choose to see rather than allow the contrary evidence to refute our favored interpretation.

Illogical reasoning. The exception proves the rule. Dismiss evidence that refutes one interpretation as exceptions.

The foundation of social science.

Theory, the logical foundation that provides systematic explanations.

Data collection is the observational aspect.

Social theory has to do with what is, not what should be.

Questions about “What is” are distinct from questions about “what should be”

“What should be” draws on your Management classes: what is the objective of our project?

- What are the indicators?
- We will look at some of the issues near the end of the course with monitoring and impact evaluation.
    - We have to first agree on measurable criteria with which we will evaluate.
    - That is the monitoring of indicators approach.
- The logframe / theory of change is a “what should be” statement, and then we use a “what is” research approach to evaluate the evidence of impact.

Looking for patterns of regularity; average overall behavior that may have exceptions, but describes the overall tendency.

- That is our research part of the overall approach.
- Identifying the overall tendency is where you draw on the statistics training.
- Then you take the findings and go back to the management part.
    - How can we take the findings from the research, which are generated using statistical tools, to do a better job in realizing our objectives?

In doing analysis and understanding behavior we need to understand the context.

- What are the formal rules and regulations?
    - Where can you park your car and on which side of the street today?
    - Can you eat food in this classroom?

- What are the informal norms and customs?
    - Which side of the stairs is for going up and which side is for going down?
    - Which side of the sidewalk for east to west crossing of the quad?

- A critique of social science is that when we identify a regularity, that regularity might appear trivial, or confirming the obvious.
    - “Female headed households in rural Senegal are poorer than male headed households”.

- However, what makes things interesting is that which is understood as true might not hold up to investigation.
    - “Actually, they are not statistically poorer due to the role of male labor outmigration and remittances.”

- That which might be seen as obvious or trivial *ex post* was not understood *ex ante*.
    - “Fulani female headed households are poorer, but Wolof are not due to differing land inheritance customs that vary across ethnic groups.”

Figuring out which is the exception, and which is the general pattern, is not always easy, especially if not treated with some systematic design.

- Adopting the formal structure of probability, and the law of large numbers, allows us to investigate data.
    - What is the general tendency and what is an outlier (and what does the outlier tell us).

One approach is to create typologies. Data from the PARIMA study in northern Kenya and southern Ethiopia rangeland areas from 2000 to 2002:

Table. Average herd size and income measures by sub-group

Group	Herd Size TLU	Total Income per capita per day	Cash Income as % of Total Income	Total Income variability (cv)
1) Left out	7.3	$0.20	29%	1.32
2) Moving From	7.2	$0.27	46%	0.90
3) Staying With	23.7	$0.34	21%	0.82
4) Combining	26.0	$0.46	35%	0.63
Significant difference in means by groups, t-statistics	t12 , t13 ***, t14 ***, t23 ***, t24 ***, t34 ,	t12 , t13 , t14 **, t23 , t24 ***, t34 ,	t12 ***, t13 **, t14 *, t23 ***, t24 ***, t34 ***,	t12 **, t13 ***, t14 ***, t23 , t24 ***, t34 ***,

*significant at the 10% level, **significant at the 5% level, ***significant at the 1% level
Panel data, 330 households, observations every 3 months

The rules of statistics allow us to create t-tests to identify when a difference in means is statistically significant and when it is not.

Note that obtaining observations has challenges.

- One is the direct logistical challenge. To have observations on 330 households every 3 months over 2 years is expensive and takes a lot of logistical coordination.
- Another is more subtle. The knowledge by people we are observing that they are being observed may lead to changed behavior, making

the observed behavior not the 'natural' behavior but the response to being observed. They might be 'performing' for us, so we are recording their constructed behavior, not their natural behavior.

- In social science research, this is sometimes called the "Hawthorne Effect".
    - Managers were observing workers in a plant in Illinois in the 20s and 30s.
    - They changed working hours, break times, lighting in the plant.
    - Worker productivity went up when each of these changes were made.
    - Eventually they figured out the productivity went up because people were aware they were being observed, not due to the changes.

Hard to observe behavior if behavior changes in response to observation.

Note the challenge of obtaining informed consent. We will talk more about this later when we cover research ethics, but note the requirement that we have to tell people we would like to observe their behavior means we run the risk of recording behavior that reflects they know we are observing and that differs from what they would have been doing if they did not know we were recording behavior. There is an inherent conflict we will return to later.

What you see is also filtered against what you are prepared to see.

Sometimes our approach is grounded in a discipline that might frame how we perceive things:

http://www.jstor.org/stable/665280?seq=1#page_scan_tab_contents

A literary version of this is an interview with William Faulkner about the nature of perceptions and how they can add up to a reality.

Unidentified participant: Mr. Faulkner—

William Faulkner: Yes, sir.

Unidentified participant: In *Absalom, Absalom!* is any one of the people who talks about Sutpen have the right view, or is it more or less a case of thirteen ways of looking at a blackbird with none of them [getting it] right?

William Faulkner: That's it exactly. I think that—that no one individual can—can look at truth. It—it—it blinds you. You look at it, and—and you—you see one phase of it. Someone else looks at it and sees a slightly awry phase of it, but taken all together, the truth is—is in what they saw, though nobody saw the truth intact. So—so these are—are true as far as—as Miss Rosa and as Quentin saw it. Quentin's father saw what—what he believed was truth. That was all he saw. But the old man was—was himself a little too big for—for people no greater in stature than Quentin and Miss Rosa and Mr. Compson to see all at once. It would've taken, probably, a wiser or more tolerant or more sensitive or more thoughtful person to see him as he was. It was, as you say, thirteen ways looking at a blackbird. But the truth, I would like to think, comes out, that when the reader has read all these thirteen different ways of looking at the blackbird, the reader has his own fourteenth image of that blackbird, which I would like to think is the true one.

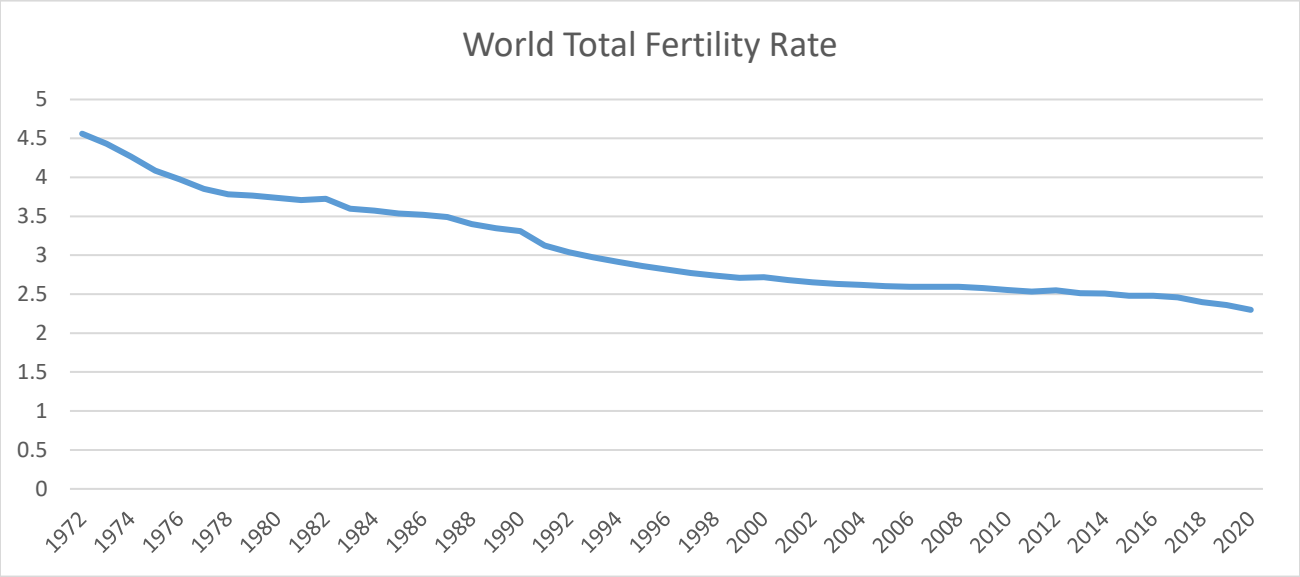
http://faulkner.lib.virginia.edu/display/wfaudio29_1

Social science is study of the aggregate, rather than the individual.

We are looking for collective regularities.

Individual decisions, but in aggregate, they create aggregate facts.

For example, Total Fertility Rate (TFR) is the average number of children a woman will give birth to in her reproductive years conditional upon her birth cohort.



There are changes over time but is not a collective decision, but a characteristic of individual decisions aggregated to the collective level. Not an outcome of social planning (though may reflect social programs and policies).

The goal is not to understand an individual person, but ‘to understand the systems in which people operate, the systems that explain why people do what they do’ (p. 14).

Elements of Analysis.

A variable. Sets of attributes that can have different values.

Attributes. Characteristics of peoples or things.

Variable	Attributes
Age	Number of years since birth
Gender	Female, Male, Non-Binary
Occupation	Professor, Lawyer, Candymaker
Social Class	Upper, Middle, Lower, Lumpenproletariat
Height	Value in centimeters

Attributes are agreed upon concepts that sort complicated reality into conceptual 'bins'.

The bins are constructs, choices.

- We have to figure out what to do with a transgender lawyer who makes candy on the weekends from a wealthy family who self identifies as working class.

Above, we used constructed groupings:

Group
1) Left out
2) Moving From
3) Staying With
4) Combining

These were defined by two different measures. Herd size per capita and total income per capita.

- Using above and below the median for each of these two measures created 4 types.

So sorting cases into bins is part of the art, or our first step from complicated reality to our artificial constructed version of reality used for analysis.

When looking at two variables, we might see some kind of pattern in correlation. The variables move together in some sense.

Rho=Expected value [((x – mean x)*(y-mean y))]/(standard deviation x*standard deviation y)

That computation is mechanical based on statistical theory.

The theory that lies behind explaining it is based in some kind of social science research approach.

- Do things move together positively or negatively?
    - Total Fertility rate and wealth?
    - Height and lifetime expected income?
    - Household self-assessment for resilience on a 1-5 scale and responses to how many months out of a year are the able to provide sufficient grains for the family rho=.52.

To put structure on the correlation, we can reach for the idea of causation. That moves beyond statistical manipulation and forces us to reach for a behavioral theory.

Much of what we do in social science research is a search through correlations to try to identify causation.

    - Why does Total Fertility Rate decrease as expected income increases?
    - Why do taller people have higher expected lifetime earnings?
    - Why is resilience positively correlated to food security?

We are trying to identify independent variables, the things on the right hand side, the predictors, from the dependent variables, the things on the left hand side, the outcomes.

In classic y=f(x) notation, the y is the outcome, the dependent variable, the x is the independent variable(s).

We spend time arguing about what is the outcome and what is the input. What is the dependent variable? What is jointly determined?

- Higher income countries have better governance; countries with better governance have higher incomes.
- Higher wages go to people who are less likely to be absent for health problems; People with higher incomes eat more healthy diets.
- Q=286-20*p is the demand; Q = 88+40*p is the supply

We also have to worry about the outcome being related to the relationship in question.

- Students are highly satisfied with the IR curriculum based on our survey.
    - Students who have higher satisfaction are the ones that complete the survey.
- Results indicate students are not satisfied with the IR curriculum.
    - Students who are unhappy are more likely to fill out the survey.
- Our survey indicates 95% of people who went through our treatment program are employed and no longer homeless.
    - We did not find 40% of the sample who went through treatment and are still homeless. B
    - ecause we could not contact them we have 95% of 60% we could locate so .95*.60+0*.40 =0.57 is a worst case possibility.

Purposes for conducting Social science research:

1) **Exploratory**. What do people around here mean by the word 'household'? What is a working definition of this concept that we can use to pose questions. Herd owning, common cooking pot, joint labor in fields, wives and co-wives, labor migration concepts….
2) **Descriptive**. How many households in this area are female headed? What is the average number of children? What is their experience with formal education? What is the mid upper arm circumference for the kids?
3) **Explanatory**. Do the kids residing in female headed households have higher or lower educational attainment than kids from non-female headed households? How about nutrition as measured by MUAC? How can we explain differences in childrens' nutrition with differences in the share of household income controlled by women?

Different methodologies and skills are called for when trying to address each purpose.

In research teams, different people tend to have comparative advantages in the different purposes, but we need to do all kinds at some point.

Kinds of explanations.

Ideographic explanations. The full listing of all the individual influences that lead to one specific unique outcome. Explain by exhaustive listing the multiple reasons a given unique outcome was realized. Fully explaining the single case / event in question.

- What are all the factors that led to the demise of the Iranian nuclear arms deal?
- What are all the factors that explain the Brexit vote?

Nomothetic explanations. Identify a few causal factors that lead to general classes of outcomes. A general explanation of what leads to a class of conditions and outcomes.

- What variables lead to success or failure when analyzing a sample of 30 nuclear arms deals over time?
- What economic, political, and historical factors have influenced referendum votes on political and economic integration in the post war era?
- What are the most important variables in the characteristics of students in Maxwell that predict they will be enrolled in this class?

Two analytical approaches in social science.

Induction. Moving from the particular details in the observations to the general interpretation and theory formation.
Take observations of individual cases and make statements about general patterns.
Begins with whether (I wonder whether this factor matters for that outcome) and moves to why (Hmm, it seems to matter, I am going to try to explain why it matters). The observations come first, and the theory is generated from the information in the observations.

Deduction. Moving from the general to the specific. Begins with why and moves to whether. We begin by formulating the hypotheses based on theory and then collect and analyze data to see if the hypotheses are supported or not by the evidence.
Figure 1-3 from the textbook.

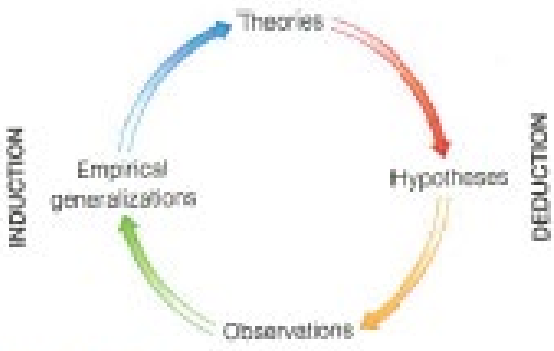
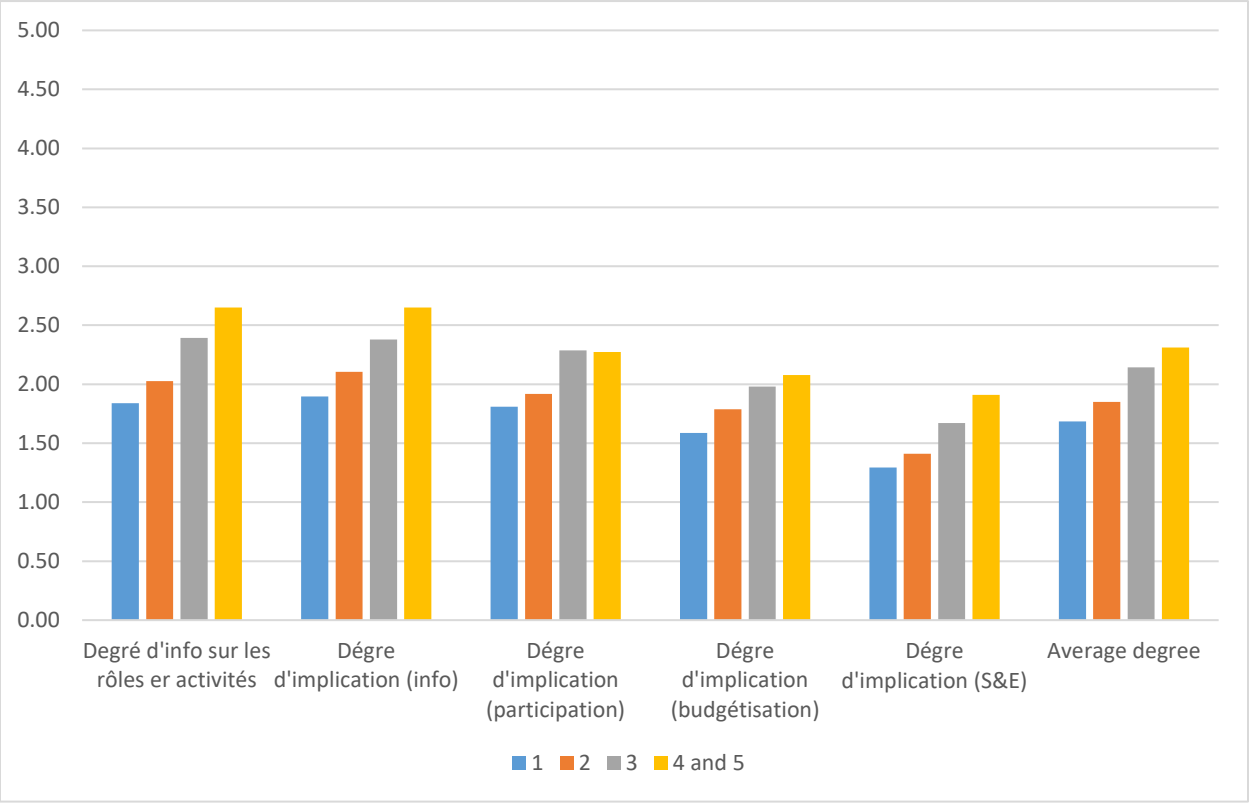


FIGURE 1-3
The Wheel of Science. The theory and research cycle for the social sciences can be compared to a relay race; although all participants do not necessarily start or stop at the same point, they share a common goal—to examine all levels of social life.
Source: Adapted from Walter Wallace, The Logic of Science in Sociology (New York: Aldine deGruyter, 1971). Copyright © 1971 by Walter L. Wallace. Used by permission.

Induction: Gather data to formulate a hypothesis.

Degree of involvement in community development activities (y axis) grouped by self-scored resilience on 1-5 scale (x axis)



There is a positive correlation with the household's self-assessed relative resilience (the 1,2,3,4,5) and involvement in community decisions.

We did follow up qualitative field work to explore the nature of this relationship with a focus on females.

That is currently under review at a journal.

DEDUCTIVE

McPeak and Doss (2006) https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1467-8276.2006.00877.x

Husbands solve the following problem:

$$(9) \qquad \max_{d} \ln(\alpha \cdot (m - s^* + \theta \cdot s^* \cdot \eta)) - \omega_2 \cdot d - \frac{\omega_3}{d}$$

while wives are faced with the problem

$$(10) \qquad \max_{s} \ln((1-\alpha) \cdot (m - s + \theta \cdot s)) - \omega_1 \cdot s \cdot d.$$

Table 2. Summary of Model Predictions

	Cooperative/ Individual	Traditional	Contested
Distance variable	Decreasing in s	Not a function of s	Increasing in s^*
Milk sales variable	Decreasing in d	Decreasing in d	Decreasing in d

Table 3. SFIML Simultaneous Tobit Estimation of Distance from Town and Milk Sales

	Dukana		Chalbi	
	Distance	Milk Sales ($\times 10^{-2}$)	Distance	Milk Sales ($\times 10^{-3}$)
Milk sales	3.16611***		3.70025***	
	(1.02678)		(0.25908)	
Distance		−0.658005		−0.686938***
		(0.597104)		(0.250100)

Determinism versus agency; y=f(x) or y=f(x)+u.

By setting up structure of average behavior we have not described individual decision making or motivation, but a description of the aggregate.

Quantitative data.

Putting numbers on things.

- What year were you born?
    - But why years? Months? Days?
- Resilience 1,2, 3, 4, 5
- Degree of implication in selection of the community investment 1,2,3,4,5

Qualitative data.

- Do you consider yourself young or old?
    - What is the threshold that divides?
- Explain to me what has made your household more or less resilient than other households in this community?
- What has been your degree of involvement in the selection of a community investment?
    - How is that related to your age and gender?

We also have quantification of qualitative concepts if we agree to measures and constructs.

Qualitative tends to work better for ideographic and quantitative for nomothetic explanations though this does not have to be the case.

Qualitative looks at the multiple causes of a case, quantitative looks at the overall patterns in multiple cases leading to certain categories of outcomes.

At the end of this class you will be presenting a research proposal. Here is an overview to give you a sense of what this means.

Overview of Research Proposals:

- Abstract / Summary
- Introduction.
- Review of the Literature.
    - Specify the Problem / Question / Topic noting explicitly why this study is needed and what is being added to knowledge.
- Research Design.
    - Data gathering methods
    - Sampling frame and population to whom this research is applicable
    - Research Ethics
- Plan for analysis of data, noting methods and standards of evidence
- Discussion and conclusion; what will we know if you do this research that we don't know now (reiterate now that it is all detailed as described above).

Tell them what you are going to tell them, tell them, then tell them what you told them. Aristotle.