K P Mohanan *26 August 2015*

Lecture Notes for Session 3: Discovering Research Questions

1. Components of Research

In session 1, we identified the core components of research as:

1. discovering and formulating worthwhile and feasible ***research questions*** to investigate;
2. identifying ***methodological strategies*** to look for answers;
3. implementing those strategies to find ***answers***;
4. arriving at and formulating ***conclusions*** emerging from those answers;
5. ***justifying*** (proving/defending/arguing for) those conclusions to establish them as knowledge;
6. subjecting to ***critical thinking*** our own and other people's questions, methodological strategies, conclusions, and justification;
7. engaging in ***rational debates*** to choose between competing conclusions;
8. ***communicating*** all of the above clearly and precisely, through written or spoken language;
9. engaging with others through ***reading/listening*** to keep abreast of current knowledge.

(1)-(9) may be viewed as a skeletal syllabus for a course on research (the ordering does not imply sequencing in the steps of research). It may be a good idea now to re-read the Course Outline of "Introduction to Research" to see how these strands connect to it.

2. Research Questions

Broadly speaking, we can identify at least three models of research:

I. research as hypothesis testing/prediction testing/conjecture proving;
II. research as theory building; and
III. research as detailed and comprehensive documenting of information,

The kind of research questions we formulate will depend on which of these models we adopt.

3. Traits and Trait Values

Let us explore these models in scientific research, making occasional connections to mathematical and humanistic forms of research. To understand the nature of the models, we need to first understand the concepts of *trait* and *trait value*.

We will use the term 'trait' (typical of the discourse in biology) for the same concept as denoted by *variable*, *attribute*, *parameter*, *dimension*, *and feature*; and 'trait value' for what is denoted by *value, attribute value, parameter value* and *feature value*. For instance, biological sex is a trait whose values are male and female. Adult human height is a trait whose values range from two feet to nine feet. Human skin colour is a trait whose values may range from dark brown to fair to pale. (Warning: in the literature on research, the term 'trait' is also used for both variables and their values.)

In chemistry, the term 'property' is used for what biologists call trait values. For instance, a reasonable answer to the question: "What are the properties of metals?" would be the following opening from the Wikipedia entry on metals:

> "A metal (from Greek μέταλλον *métallon*, "mine, quarry, metal") is a material (an element, compound, or alloy) that is typically hard, opaque, shiny, and has good electrical and thermal conductivity. Metals are generally malleable — that is, they can be hammered or pressed permanently out of shape without breaking or cracking — as well as fusible (able to be fused or melted) and ductile (able to be drawn out into a thin wire)." https://en.wikipedia.org/wiki/Metal

To see the parallem between ‘properties’ in chemistry and ‘trait values’ in biology, compare the above with the following traits of mammals:

> “Mammals… are… distinguished from reptiles and birds by the possession of hair, three middle ear bones, mammary glands, and a neocortex (a region of the brain).”
>
> (https://en.wikipedia.org/wiki/Mammal)

and in physics, the following account from Wikipedia entry for electrons:

> “The electron is a subatomic particle…with a negative elementary electric charge. …The electron has a mass that is approximately 1/1836 that of the proton. Quantum mechanical properties of the electron include an intrinsic angular momentum (spin) of a half-integer value in units of $\hbar$…. no two electrons can occupy the same quantum state …Like all matter, electrons have properties of both particles and waves, …” (https://en.wikipedia.org/wiki/Electron)

To extend this to literary studies, here is an account of the trait values of the English haiku (as opposed to the Japanese haiku):

> “a three-line format with 17 syllables arranged in a 5–7–5 pattern; or about 10 to 14 syllables, which more nearly approximates the duration of a Japanese haiku with the second line usually the longest. Some poets want their haiku to be expressed in one breath.”
>
> (https://en.wikipedia.org/wiki/Haiku_in_English)

4. Research Questions in Scientific Inquiry

There are four levels of research in scientific inquiry:

A. trait values of a *single member* of a population, stated as an *observational report*;

B. *regularities* governing the co-occurrence of trait values in a population — by generalizing from a sample of the population — stated as *observational generalizations, observational laws, correlational hypotheses,* or *causal hypotheses*;

C. explanation for the observational generalizations resulting from the above, expressed as ***theories***; and

D. theory-internal ***interpretation*** of an observational report.

Research at level (A) yields what is called a ***data point,*** specified as the value of a single trait (in a one-dimensional trait space), the values of a pair of traits (in a two-dimensional trait space), or trait values of a multi-dimensional trait space. If we say “Zen’s height is 170 cm, we are locating Zen (a point) in a *one-dimensional trait space*.

Point	Trait	Trait value
Zen	Height	172 cms

This information can be represented as a number line (dimension/trait), and Zen is located at 170.

If we say, “Zen’s height is 170 cms and his age is 80, we are locating Zen on a *two-dimensional trait space*.

Point	Trait	Trait value
Zen	Height	170 cms
	Age	80 years

This can be represented as a two-dimenstional space, with height along one axis, and age along the other, and with Zen as a point in this space.

If we say “Zen’s height is 170 cms; his age is 80; his weight is 80 kg; he is male; he has brown skin, black eyes, two hands, two legs, two lungs, one heart, red blood; he is an academic; he is married; he is a father; he has ten grand children… ,” we are locating Zen in a *multi-dimensional trait space*.

Point	Trait/Variable	Value
Zen	Height	170 cms
	Age	80 years
	Weight	80 kgs
	Gender	male
	Skin colour	brown
	Eye colour	black
	Number of hands	two
	Number of legs	two
	Number of hearts	one
		etc.

An ***observational report*** locates a member of a population in an n-dimensional trait space. The observation may be

- experimental or non-experimental
- instrumental or non-instrumental
- quantitative (using counting or measurement) or non-quantitative (= qualitative)

Research at level (B) results in an ***observational generalization***, a proper sub-space within a multi-dimensional trait space. If we measure the height of a large random sample of adult humans, we will find that their data points are distributed within the region “more than 60 cms and less than 300 cms”. This is an observational generalization.

If we now include frequency as a parameter, we discover that the highest frequency of data points is located around its center. This is another observational generalization. If represented as a statistical graph, such generalizations yield a curve on the graph. This particular trail (height) is distributed in the population such that it yields a bell curve.

The example given above is that of a single trait/variable generalization. If thegeneralization involves two traits, and the value of one trait can be inferred from that of the other, then we have a ***correlation***. For instance, take the traits eukaryoticity and multi-cellularity. We have the following combinations of trait values in biology:

	eukaryotic	prokaryotic
uni-cellular	√	√
multi-cellular	√	

While uni-cellular eukaryotes, uni-cellular prokaryotes, and multi-cellular eukaryotes exist, multi-cellular prokaryotes don’t. To put it differently, the combination of multi-cellular and prokaryotic is ruled out. This means that if we know that a species is multi-cellular, we can infer that the species is eukaryotic. Likewise, if we know that the species is prokaryotic, we can infer that it is uni-cellular. Thus, there is correlation between the two traits. This can be stated as a correlational law:

<u>Cell-Number and Cell-Nucleus Law</u>

[+multicellular] → [+eukaryotic] i.e., if the value of multicellularity is +,
then the value of eukaryoticity is +.

Even the briefest examination of the possible combinations of trait values suggests a wide range of correlational laws in biology:

if the member of a species has a beak, it has two legs;
if the member of a species has six legs, it has compound eyes;
if the member of a species has lungs, it has a heart;
if the member of a species produces chlorphyl, it does not have a brain;
…

If we use the term ‘quantitative’ to refer to trait values that can be expressed in terms of rational numbers, on which we can perform statistical operations, and reserve the term ‘qualitative’ to categories (discreet) or quatities that can be counted, but cannot be subjected to statistical operations, then the examples of correlational laws given above are qualitative. Correlational laws like Galileo’s law of the simple pendulum (postive correlation between the length of the pendulum and its period) and the law of the falling bodies (correlation between time and distance) are examples of quantitative (gradient) correlational laws.

Similar correlational laws are found in geometry as well: there is a positive correlation between the length of the diameter of a circle and its circumference; there is a positive correlation between the number of lines and the number of angles in a polygon; there is a negative correlation between one angle of a triangle and the sum of the other two angles; and so on.

Once a correlational law is established as true, we can ask why it is true. Such questions fall within the domain of research at level (C), the **theory**. Thus, we can ask why Galileo’s correlational laws of the simple pendulum and falling bodies are true. The answer comes from Newton’s theory of gravity and motion. We can ask why there are periodic correlations between time and temperature — a daily cycle as well as a yearly cycle of time and and temperature. The answer comes from the heliocentric theory of the solar system.

In science, the answers to the questions

“Is this correlation true?”
and, if true
“Why is it true?”

call for two different kinds of research. The first question calls for hypothesis testing research (I). The second calls for theory building research (II). In mathematics, the corresponding questions are

“Is conjecture C true?”
and
“Why is it true?”

These are both answered by the same strategy, namely, that of a proof that shows that the conjecture in question follows logically from the axioms and definitions of the theory. The proof of the Pythagoras Theorem tells us whether the conjecture is true, as well as why it is true.

We have not talked here about Research type III or of Research level D.