Research Methodology

Part 1: Introduction to Scientific Research

Christian Hilbe, christian.hilbe@it-u.at
Version May 28, 2025

1 Preamble: About this class

Remark 1.1 (**A brief intro session**)

- *Overview.* This course will be taught by two instructors, Christian and Jie, over the next three weeks. This course should give students an overview on the research process, and different methods that are used along this process.
- *Introduction.* Name; Own research topic; Expectations on course; Previous exposure to research methodology

Remark 1.2 (**Motivation for this course**)
This course serves at least three goals

1. *Practical relevance.* To do research effectively, it is just important to have a good grasp of what research is about, what questions will come up along the way, and which options are out there to carry out research. In their scientific lives, students will often be asked specifically to describe their own research methodology. The first such instance is probably when they need to hand in their research proposal at the end of the first year. However, later there will be numerous other occasions, for example when presenting papers or when applying for grants. To this end, two competencies are important: (1) Having the right vocabulary, and a knowledge what is considered to be research methodology, and (2) Ability to express one's thoughts clearly. This course should prepare you for the first point; the 'Scientific Writing' course that is offered in parallel will prepare you for the second point.
2. *Allowing you to be an informed consumer of science.* It may well be that some concepts you will hear about during this course have no immediate relevance for your own research. For example, when we talk about research methods and the different data types they produce, we might briefly encounter *ethnographic research.* Most of you will probably not engage in this kind of research. Yet, as a scientist, and as a consumer of science, it is still important to have an overview on different methods that allow people to generate knowledge. This comment seems to be particularly important at an interdisciplinary university as ours. Knowledge of each others' methods also makes it easier to start collaborations.
3. *Philosophical aspects.* In courses like this, it is useful to stress that the concepts that will be taught are human constructs. In this case, they are even relatively recent constructs. For example, modern

hypothesis testing and many of the associated statistical methods have only been developed a bit more than a hundred years ago. Even the idea that (controlled) experiments are a viable way to generate knowledge about the world is fairly recent (one of the early pioneers here is Robert Boyle who discovered that at a given temperature, the volume of an ideal gas and the respective pressure are inversely related. This was in the 17th century). So this course should also help students understand why smart people at some point found it useful to think of science in this organized way, and why they came up with certain experimental or statistical methods. It should also invite students to think critically about this subject.

Example 1.3 (Appreciating the success of science)
Before starting with this course, it is perhaps worth to appreciate just the enormous success humankind had by employing scientific methods. To this end, note that on a genetic level, there is not much difference between humans and apes (for example, humans and chimps share around 98.8% of their DNA). Yet for some reasons, humans were able to make remarkable discoveries even without modern technologies. For example, Eratosthenes of Cyrene (276 BC – 195/194 BC) was able to approximate the circumference of earth with an error margin of less than 1% (already the realization that the earth is not flat is a non-trivial task). Similarly, already earlier Aristarchus of Samos (310 BC – 230 BC) estimated the size of the moon, relative to the size of earth (he estimated earth's diameter to be the 2.85-fold of the moon; the true number is 3.67). If I asked you to come up with these numbers (without looking it up online), how would you even do it? It requires human ingenuity, sound research methods, and precise measurements!
Discussion. What is your favorite example of human ingenuity? (e.g., Einstein's relativity theory, Gödel's impossibility theorem, Arrow's theorem).

Remark 1.4 (Some general administrative remarks)
There are no prerequisites for this course. Grading will be based on in-class participation, quizzes and minor assignments. For further readings, take a look at the books this course is based on (1, 2). Structure:

- Monday, May 26: Introduction to Scientific Research (Christian)
- Wednesday, May 28: Research designs and methods (Christian)
- Monday, June 2: Quantitative and qualitative methods (Jie)
- Wednesday, June 4: Presenting research findings (Jie)
- Thursday, May 12: Resources, tools, and techniques (Jie)

Group Exercise 1.5 (An ice-breaker)
Before going into the details, let us start with some group activity:

Group 1: *Why do we do research?* (Motivations and external incentives). For example: Curiosity, learning useful knowledge or skills, working with mentors, prestige, getting access to a network, getting paid, graduation requirement, career development.

Group 2: *What are some of the challenges we might face?* (Risks). For example: Time management, lack of support, finding a topic, publication pressure, workload, funding issues, failed experiments, unexpected or insignificant results.

2 Definitions and concepts

Remark 2.1 (Overview)
Before looking at the scientific process itself, we should first have a common understanding of some of the key concepts. For example, what is science, and how does it differ from pseudoscience or ordinary common sense? What are theories, laws, hypotheses, and models? What are properties of a good theory? We will discuss and define these concepts in the following.

Definition 2.2 (Science)
The term *science* is used in several contexts.

1. *Science as a process.* Science consists of the systematic observation of natural events and phenomena, to discover facts about them, and to formulate laws and theories based on these facts.
2. *Science as a body of knowledge.* Science can refer to the body of knowledge that is derived from such observations.
3. *A subfield of research.* The term 'science' can also refer to specific branches of this body of knowledge, e.g., physics, biology, etc.

The US Academy of Sciences defines science as *the use of evidence to construct testable explanations and predictions of natural phenomena, as well as the knowledge generated through this process* (3).

Some key features of scientific knowledge are that it is *objective* (reproducible), *self-corrective*, *cumulative*, *tentative*, and *parsimonious* (Occam's razor).

Remark 2.3 (Academic disciplines)
In academia, it is common to organize the different disciplines into different categories. For example, natural sciences are concerned with the description, understanding, and prediction of natural phenomena (e.g., physics, biology, chemistry, earth science, astronomy). Formal sciences explore abstract structures and their relationships (e.g., logic, mathematics, computer science). The social sciences explore human societies and the relationships among the members of these societies (e.g., economics, sociology, psychology, anthropology, political science). Finally, the humanities are academic disciplines that are concerned with human society and culture (e.g., classics, religion, history, language, law, performing arts). These various branches of academic disciplines differ in their methods, in the types of data they use, in the kind of knowledge generated, etc. In particular, some of the methods that we will learn later on will be important in one branch but not in another. In fact, for many of the methods discussed later, we will implicitly assume that the method of inquiry is based on empirical data of some kind.

Remark 2.4 (How to distinguish science from pseudoscience or common sense? – See also Ref. (1))

1. **Science vs Pseudoscience.** There are plenty of pseudosciences out there, with astrology or psychoanalysis sometimes mentioned as examples. Pseudoscience literally means 'false science'. The distinction between pseudoscience and science can be blurry. Both aim to provide some support for an idea, but the methods of pseudoscience do not have the same standards of rigor. What are some characteristics of pseudoscience that allow us to make a distinction compared to science?

- Proponents would use situation-specific hypotheses to explain away falsification of a pseudoscientific claim
- There is no mechanism for self-correction, and hence the field stagnates
- Shifting the burden of proof to skeptics
- Relying on non-scientific anecdotal evidence
- Failing to build on an existing base of scientific knowledge
- Failing to specify conditions under which ideas or claims would not hold true

2. **Science vs common sense.** Everyday experiences can serve as another source of explanations for a phenomenon. These explanations can be termed 'common sense'. Sometimes political parties diminish science and prefer focussing on common sense; hence it becomes important to compare scientific and common sense explanations. Both types of explanation start by observing the real world. However, compared to scientific explanations, common sense explanations again lack the necessary rigor. Such explanations are not based on careful, systematic observation; these explanations are usually not compared to other equally plausible alternative explanations; and usually there is no attempt made to see how well the explanation accounts for similar behavior in other circumstances. However, everyday experience can often serve as a motivation for a scientific research program.

Remark 2.5 (Theories / hypotheses / laws / models – See also Ref. (1))
There are a few more terms mentioned in scientific discussions that require careful definitions.

- **Theory.** Everyday language uses the term 'theory' rather loosely. Meanings range from well-tested explanations to simple guesses that happen to be consistent with the observations we have. Unfortunately, this loose understanding can lead to fundamental misunderstandings about the merit of scientific theories. For example, critics of Darwin's theory of evolution may argue that this is just one theory among several others (e.g., intelligent design). However, in science the term *theory* has a rather specific meaning. For example, according to the US academy of sciences, a theory is a *plausible or scientifically acceptable well-substantiated explanation of some aspect of the natural world; an organized system of accepted knowledge that applies in a variety of circumstances to explain a specific set of phenomena and predict the characteristics of as yet unobserved phenomena* (3). It is in this sense that Darwin's theory of evolution is a theory.

- **Hypothesis.** Unlike theories, hypotheses are claims that are not yet well substantiated. Moreover, hypotheses are often more limited in scope, and they do not involve a complex set of assumptions and logical relationships as theories usually do. According to the US academy of sciences, *A tentative explanation for an observation, phenomenon, or scientific problem that can be tested by further investigation. Scientific hypotheses must be posed in a form that allows them to be rejected* (3).

- **Law.** A law is an empirically verified (or mathematically derived) quantitative relationship between two or more variables. A mathematical form is not necessarily required for a relationship to be a law (e.g., law of supply and demand in economics).

- **Model.** Usually, a model is a specific formal implementation of a more general theoretical view (e.g., Lotka-Volterra model of predator-prey interactions). Mathematical or computational models can help to identify hidden assumptions in a theory, to eliminate ambiguity, and to make predictions.

Group Exercise 2.6 (What makes a good theory? – See also Ref. (1))
What are theories even good for – couldn't we just run a series of experiments and human knowledge would be the sum of these experiments? Also, what are the characteristics of good theories?
One answer: Theories should help us understand and predict a certain aspect of the world, and it should help us organize and interpret research results. Ideally, such theories can also generate future research projects. The quality of a theory can be evaluated according to the following criteria: **(1)** Ability to account for data; **(2)** Explanatory relevance (the explanation provided by the theory needs to offer grounds for believing that the phenomenon would occur under the specified conditions); **(3)** Testability / Falsifiability; **(4)** Prediction of novel events; **(5)** Parsimony.

3 Steps in the research process

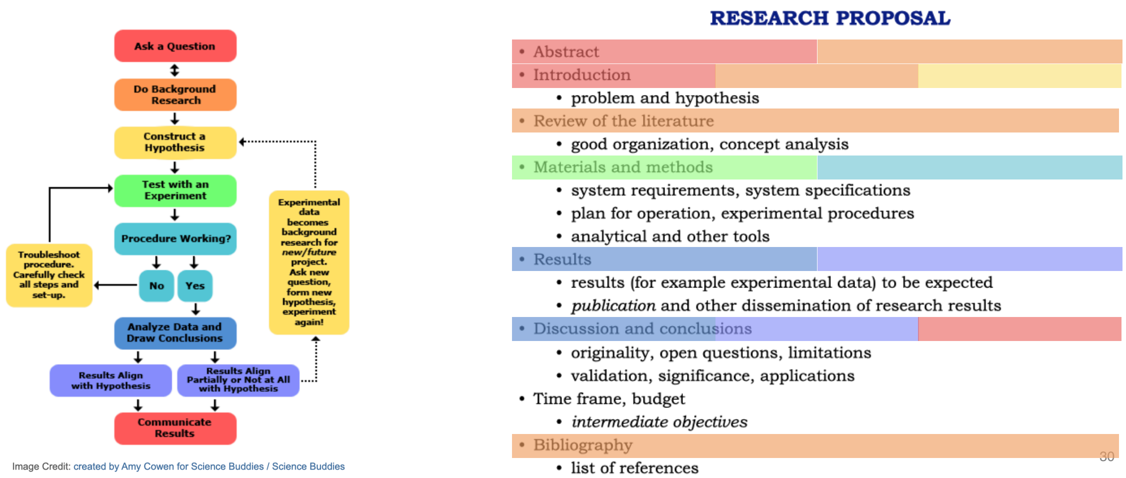
Remark 3.1 (Motivation)
The previous section provided some perspective on the big topics: what is science, what are scientific theories, etc. However, in their daily lives, researchers usually do not work on grand theories, but rather on some smaller and well-defined research projects. The aim of this next section is to discuss the stylized steps of such a usual research process.

Group Exercise 3.2 (Steps of a research process)
Describe the typical steps of a research process in your field, beginning from the conception of an idea to the publication of the results. What are the challenges in each step?

Remark 3.3 (The scientific method)
The scientific method is a scheme of how research should be done based on testable hypotheses. It follows the following steps:

1. Define a general question
2. Gather information and resources (e.g., literature review, theory building)
3. Form a hypothesis
4. Perform an appropriate experiment and collect data
5. Analyze the data to derive the results of your study
6. Interpret the results and draw conclusions (this may serve as a starting point for a new hypothesis)
7. Publish results

Remark 3.4 (Communicating the steps of a research process in a research proposal)
When writing a research proposal, it is often a good idea to write a document that follow the steps of the scientific method (see **Fig. 1**). In this document (much similar to a grant application), you describe how you will follow through your research plan for the given project. Ideally, when reading this document, reviewers should get the feeling that after a formal 'ok', you would be able to start conducting the research



Mämmelä, A. (2006). HOW TO GET A PH.D.: Methods and Practical Hints.

Figure 1: Mapping the steps of the scientific method to the elements of a research proposal.

the very next day. That requires the research proposal to be written in sufficient detail, and that you address the most important decisions that you will have to face along the way (e.g., what research method you implement, how you sample subjects, how you operationalize your research question, whether there are any foreseeable ethical concerns, what kinds of statistical tests you are going to run, and what other quantities you are going to derive, e.g., benchmarking).

We will talk about the individual steps of the research process for the remainder of this course. In the following, we briefly give an overview on each of the steps, and the decisions that are involved.

Remark 3.5 (Research question; descriptive vs analytical research)
The research question defines the area of interest, but it does not have to be as declarative as a hypothesis. These research questions can be of different types. For example, in *descriptive research*, your aim is to describe or identify the current state of affairs. Here, the researcher aims to observe and document various aspects of the phenomenon, without manipulating any variables to explore cause-and-effect relationships. This is often a useful approach if not much is known about the phenomenon to begin with, and no clear patterns are yet established (i.e., as *exploratory research*). For example, one possible descriptive research question is: *What is the workload of a typical PhD student at IT:U?* On the other hand, *analytical research questions* look for associations and for possible cause and effect. A typical question would be *How does social media influence teenage mental health?* The type of research question determines which kind of hypotheses are made later on (if any), and what appropriate research methods would be (e.g., a survey vs a controlled experiment). Another interesting problem is: What is a *good* research question? We will address this problem in one of the next sessions.

Remark 3.6 (Gather background information)
As a researcher, it is also your task to survey similar and relevant work in the field. On the one hand, this

helps you to determine the state of the art and allows you to reflect on the novelty of your own question (i.e., it allows you to respond to the question: what distinguishes your contribution from previous work?) On the other hand, it also allows you to get a sense of the possible impact your research may have. *Group exercise: What are possible sources of background information?* Your supervisor, your group members, your scientific network; books, reviews, research articles. To get a first overview, also encyclopedias, blogs, or popular science books can be helpful (among these sources, there are tradeoffs with respect to reliability and recency; e.g., in a quickly moving field, books may contain outdated information).

Remark 3.7 (Formulating hypotheses and operationalization)
A scientific hypothesis states the predicted (educated guess) relationship amongst variables. It brings clarity, specificity and focus to a research problem. To be useful, a hypothesis needs to make statements for variables that can be measured. This might require you to operationalize your research question. For example, instead of asking *how do natural distractions (conversations, podcasts) affect human driving abilities*, you may ask *how do two to three different types of podcast affect an individual's ability to keep the track (measured by distance to the roadside)*.

Remark 3.8 (Design of experiments and collecting data)
Together with the specification of your research hypothesis, you also need to design the respective data collection process. This process needs to be planned and documented such that the results and methods are reproducible. Here, many aspects need to be determined; the exact kind of experiment to be performed; the location where data collection is performed (e.g., lab vs field), determining the necessary number of participants and their recruitment; ethics approval; testing of the experimental procedures and pilot studies; preregistration; data management plan. The logistics of these tasks can become considerable; it can be useful to have a checklist with milestones.

Remark 3.9 (Deriving results and interpreting them)
After having the data collected, again a number of important decisions need to be made: What kind of analysis or statistics is appropriate (e.g., *descriptive* or *inferential*; based on p-values or Bayesian statistics)? How do you deal with issues of multiple testing? How do you deal with dropouts or selective attrition? How to interpret the results of a study (e.g., the p-value and the *effect size*).

Group Exercise 3.10 (The publication process)
The way how research is summarized and published and presented can vary considerably between fields. In computer science it is common to publish findings in conference proceedings; in the natural sciences it is common to publish in journals; and in the humanities it is common to publish books. So, how does the publishing process in your field look like? What is the typical length of a research output? How would the review process look like, and how would you address the comments of reviewers? How are different publishing venues ranked?

Remark 3.11 (A summary)
In this first class, students should have learned

- how to define some basic concepts (e.g., science, theories, etc), and to distinguish between science and pseudoscience
- how to conceptualize the research process, and the many decisions that need to be made along the way.

References

[1] Bordens, K. S. & Abbott, B. B. *Research Design and Methods: A process approach* (McGraw-Hill, New York, 2022).

[2] Thomas, C. G. *Research Methodology and Scientific Writing* (Springer, Cham, Switzerland, 2021).

[3] US National Academy of Sciences. Definitions of evolutionary terms (2013). URL `https://www.nationalacademies.org/evolution/definitions`.

[4] Hauser, O., Hilbe, C., Chatterjee, K. & Nowak, M. Social dilemmas among unequals. *Nature* **572**, 524—527 (2019).