

FLORIAN HARTIG, UNIVERSITY OF FREIBURG

RESEARCH SKILLS

AN INTRODUCTION TO THE CRAFTS OF A SCIENTIST

This script is part of the course material for the MSc model Research Skills, University of Freiburg (website [here](#)). First edition 2015, created October 18, 2015.

[Florian Hartig](#)

University of Freiburg
Germany

If you find typos or have suggestions, please report them by creating a ticket [here](#)

Acknowledgments: Severin Hauenstein for contributions to the chapter on scientific writing.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Contents

1	<i>Logic and philosophy of science</i>	4
2	<i>On the shoulders of giants - the scientific literature</i>	12
3	<i>Questions, data and valid conclusions</i>	15
4	<i>Principles and channels of scientific communication</i>	21
5	<i>Scientific writing</i>	25
6	<i>A working scientists</i>	36
7	<i>Bibliography</i>	38

1

Logic and philosophy of science

1.1 *What is science?*

The Oxford dictionary defines science as

Science [mass noun] - the intellectual and practical activity encompassing the systematic study of the structure and behavior of the physical and natural world through observation and experiment.

Or, in the words of Richard Feynman:

Science is a process for learning about nature in which competing ideas about how the world works are measured against observation.

Hence, science is a method to find out how the world works.¹ In fact, "method" is an understatement. In the last centuries, the scientific method has been the reason for an unprecedented increase in knowledge, opportunity and well-being. And at least in the western world, science has become the most widely accepted answer to what one can know and how to find it out (epistemology).

But science is also a human endeavor. "Science" may as well refer to the history of science, the state of scientific knowledge, or science as a social institution. Those latter aspects of science are less formally defined and stable than science as a method - if we look at the history of science, we see individuals that are unified by a common cause, but that are at the same time subject to their national and cultural influences as well as by personal vanity, pride or madness. As a result, the history of science, as well as its present state, is complicated by personal preferences, power games and random events, in the same way as political and economic history is.

In the center of science, however, remains a set of assumptions and methods, a philosophy if you will, that allows us to draw and agree on conclusions from empirical observations. Not every scientist is aware of the "formal rules" of these methods, but that does not mean that (s)he would not apply them. In his book "Induction & Intuition in Scientific Thought", Peter Medawar notes:

¹ Epistemology, from Greek episteme = knowledge and logos = word is the branch of philosophy concerned with what knowledge is, how it can be acquired, and what can be known

Many famous scientists engaged in personal feuds that, looking back, seem to be fueled by irrationality and personal pride. The doubtlessly ingenious Isaac Newton engaged in various spiteful and bitter feuds, including his most famous battle with German mathematician Gottfried Leibniz over the invention of infinitesimal calculus.

The fact that scientists do not consciously practice a formal methodology is very poor evidence that no such methodology exists. It could be said—has been said—that there is a distinctive methodology of science which scientists practice unwittingly, like the chap in Molière who found that all his life, unknowingly, he had been speaking prose.

It may thus be possible to do science without having ever learned the basics of the scientific method. However, that does not mean you should not learn about the methods of science. An amateur can sometimes do great work, from intuition and practice, but without a good education s(he) may fail in other situations. Learning the rules of science will provide you with safeguards against drawing false conclusion, will teach you methods to make discovery and communication more efficient, and will arguable improve your general understanding of the world. Hence, read this manual for science with care, including the rest of this chapter, which is devoted to laying out the assumptions and history of science, and discussing how the "scientific method" works.

1.2 *Philosophical foundations of scientific thinking*

Empiricism and induction

Empiricism is the stance that (scientific) knowledge comes only from observations. For an empiricist, (sensory) experience is more important for the formation of ideas about how the world works than logical reasoning or innate knowledge.

Famous empiricists are, e.g., Hume or John Stuart Mill

Inductive reasoning An important idea in the context of empiricism is the concept of inductive reasoning. An induction is a conclusion from particular instances to the general. Example: all swans I have seen are white, therefore all swans are white. Empiricism does not prescribe the use of inductions for generating knowledge, but it seems obvious that pure empiricism would require to rely on inductive reasoning if any general conclusions are to be generated.

Induction = conclusion from the specific to the general

Empiricism != induction

The problem of induction The problem with inductive reasoning is evident - an inductive conclusion does not need to be correct, even if the assumptions (observations, formal logic) are formally correct. Hence, we cannot be truly be sure if the knowledge that is created by empiricism and inductive reasoning is reliable. This problem has been named the "problem of induction" by David Hume, but it was of course known for a long time. Sextus Empiricus (c. 160-210 AD), the name father of empiricism, observed famously:

Inductions are not always correct

C. D. Broad remarked more shortly: "Induction is the glory of science and the scandal of philosophy"

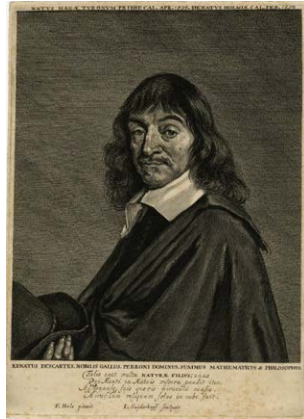


Figure 1.1: Portrait of René Descartes

When they propose to establish the universal from the particulars by means of induction, they will effect this by a review of either all or some of the particulars. But if they review some, the induction will be insecure, since some of the particulars omitted in the induction may contravene the universal; while if they are to review all, they will be toiling at the impossible, since the particulars are infinite and indefinite. Sextus Empiricus (c. 160-210 AD)

Because of the problem of induction, there has always been a stain on inductive reasoning, and philosophers have tried to avoid relying on it for drawing conclusions about the world

Rationalism and deductions

Rationalism, in contrast to empiricism, holds the view that knowledge is created mainly by reasoning. In one of the foundational works of modern science, "Le Discours de la Méthode (1637)" ², René Descartes famously observed:

Good sense is, of all things among men, the most equally distributed; for every one thinks himself so abundantly provided with it, that those even who are the most difficult to satisfy in everything else, do not usually desire a larger measure of this quality than they already possess.³

Pure thinking and logical conclusions are a fine thing, but a problem for a pure rationalist is that s(he) still needs something to think about - from where do the objects and basic assumptions arise to which we can apply our rational thoughts? Rationalists differ though in their view on this matter. Some philosophers have stressed experience as providing these starting points, while others have maintained that there are some innate concepts or truths that we can know without reference to the material world.

² The full title of this famous essay is "Discourse on the Method of Rightly Conducting One's Reason and of Seeking Truth" (Descartes, 1673)

³ These words of one of the founding fathers of what we today consider Science are often read and cited as a sarcastic observation of human vanity. However, if one continues reading, it becomes clear that Descartes may, after all, have been serious about his faith in human rationality. He goes on to state: " And in this it is not likely that all are mistaken the conviction is rather to be held as testifying that the power of judging aright and of distinguishing truth from error, which is properly what is called good sense or reason, is by nature equal in all men; and that the diversity of our opinions, consequently, does not arise from some being endowed with a larger share of reason than others, but solely from this, that we conduct our thoughts along different ways, and do not fix our attention on the same objects. "

Deductive reasoning Wherever the starting points of rationalism come from, the reasoning natural to rationalism is deductive reasoning. A deduction is a conclusion from the general to the specific. Example: all people from Frankfurt like applewine, I am from Frankfurt, therefore I like applewine. An important difference to inductive reasoning is that, if assumptions are correct, conclusions made by deduction are necessarily correct. This is an important motivation for the further development of the scientific method.

Deductive reasoning = conclude from the general to the specific.

1.3 *The basic science stance*

In the previous section, we have learned about two fundamentally different thoughts about how knowledge is created (empiricism and rationalism), and we have learned about two types of reasoning (induction and deduction). An attentive reader has probably already drawn a conclusion: in some way, science combines these two views. Before we come to that, however, let me define the basic assumptions that underlie the scientific method:

- There is a material reality, and we can observe it.
- The material reality follows certain general laws. We can discover the nature of these laws (scientific theory), and we can do this objectively, meaning that everyone sharing the same basic principles (the scientific method) will agree eventually
- Discovery happens by (a limited number of) empirical observations (empiricism) and reasoning (rationalism), which are combined in the scientific method.

You may find it trivial to make these statements, but even in academic circles, people often deny these assumptions in particular circumstances. For example: “there are some things that you can simply not measure”, or “not everything can be expressed in a formula”.

The first two statements express what is known in philosophy as positivism. Note that there are parts of social science (but not the natural sciences) that actively deny the truth of these statements. A notorious example are extreme positions of social constructionism and postmodernism, which hold broadly the view that reality and truth is not absolute, but a property that is constructed or negotiated through the interactions in a society.

There have been some fierce debates in the 90's between natural scientists and postmodernist. A very good example is the so-called Sokal affair, in which the physics professor Alan Sokal published a nonsense article full of deliberately incomprehensible jargon in a postmodernist academic journal (Sokal, 2008)

1.4 *The scientific method and its history*

By now, we have learned about rationalism and empiricism, and we have defined the basic assumptions of science, but we are yet to learn how we can bring these parts together to create knowledge - the scientific method.

A wider account of the different views on the scientific method on wikipedia [here](#)

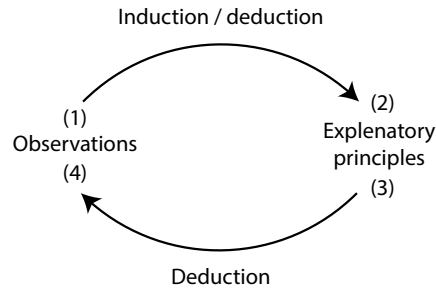


Figure 1.2: Aristotle's inductive-deductive model of the scientific method.

Aristotle's view on the scientific method in his Posterior Analytics

One of the first, certainly one of the best-known explanations of the scientific method is Aristotle's inductive-deductive method (Fig. 1.3). The idea depicted in the figure is the following

1. You make an observation
2. Infer general principles from the observation
3. Infer consequences of these principles
4. Check them against observation
5. Potentially repeat

This basic idea has not changed enormously in the next 2000 years, but it has changed in subtle details. Philosophers and scientists have developed different opinions about how the processes of generating explanatory principles and testing them works exactly. A particularly controversial topic has been the extent to which induction plays a role in this cycle.

Aristotle suggests towards the end of the *Posterior Analytics* that, at first, explanatory principles arise through induction

Thus it is clear that we must get to know the primary premises by induction; for the method by which even sense-perception implants the universal is inductive.

However, he is clearly uncomfortable with the inductive element in this and goes on to say that about the premises created by induction

[...] it follows that there will be no scientific knowledge of the primary premises, and since except intuition nothing can be truer than scientific knowledge, it will be intuition that apprehends the primary premises.

How this intuition works to create our first premises remains then a bit unclear - but you should read Aristotle yourself, maybe you can make more sense of it than me.

Aristotle's six standard works on logic are summarized in the *Organon*. His view on the scientific method are spread around this work, but the main part is in the *"Posterior Analytics"*, (Latin: *Analytica Posteriora*).

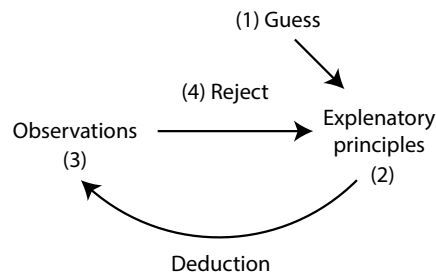


Figure 1.3: Popper's hypothetico-deductive model of the scientific method

Popper and the scientific method

We stay with the problem of induction, but take a 2000-yr leap to the arguably most influential interpreter of the scientific method in the 20th century: Karl Popper. Popper developed the current orthodox interpretation of how science works. You may not like it, but the fact is that many, probably most scientists subscribe to his views as a basic description of the scientific method. Hence, don't act openly against it, and if people ask you how science works, the safe bet is to recite Popper!

Karl Raimund Popper (28 July 1902 – 17 September 1994), born into a Jewish middle-class family in Austria-Hungarian Vienna, is widely regarded as one of the most influential philosophers of the 20th century

Hypothetico-deductive method One of Popper's main contributions to the scientific method was to take a radical step and remove the induction completely from the scientific equation. According to Popper, the essence of science is not to produce new theories, but to have a method to reject them. We don't care how new theories arise: they may be guessed, they may be induced, this is not part of the scientific method. The method starts once we have a theory: we look at the predictions of this theory (deduction), we can test it against observation and see whether the theory is compatible with it (deduction), and if not, we can conclude that the theory is wrong (deduction). Through this construction, science consists only of logically clean deductions. Here a summary of this method - again, if someone asks you what the scientific method is, this is what you should say:

Popper developed the hypothetico-deductive interpretation of the scientific method, the currently accepted orthodox narrative of how science works

1. Form / guess a hypothesis about a phenomenon in nature
2. Derive predictions from hypothesis
3. Test predictions against observation
4. Draw conclusions about the validity of the hypothesis

It follows from Popper that "you cannot prove a theory, you can only reject" - hence, avoid using the word proof in scientific articles - many people care deeply about Popper, and they will react strongly to it!

You cannot prove a theory, you can only reject. Hence, never write: "we proved A" - if you must, write "the data is compatible with A"

Multiple working hypothesis

Popper is the mainstream, but there have been other suggestions about how science works or should work. A counterpoint to Popper's idea of rejection is the idea of multiple working hypothesis. The basic principle is simple - instead of trying to reject an idea, one can work with alternative theories. With time, evidence typically moving to favor one of them. Platt (1964) tracks the history of this idea to an influential essay of the Geologist T. C. Chamberlin, who recommended working with multiple hypothesis to avoid as the best way to avoid "parental affection" (T. C. Chamberlin 1890, Science).

Kuhn and variations of the scientific method

The argument for multiple working hypothesis is a first example that includes social aspect of how science works into the method - the problem that Chamberlin has with the idea of Popper (he formulated his view before Popper of course) is not that Popper's method wouldn't work in principle, but that in practice, researchers may be hesitant to reject ideas, because they like them to be true.

A more modern position that goes in a similar direction are the ideas of Thomas Kuhn: in his seminal work "the structure of scientific revolutions", Kuhn posits that sciences works within certain assumptions. In a normal science phase, we may apply Popper's idea to test and improve these assumptions, but we would not reject the basic premises even if there is evidence against them, unless a better theory emerges (revolutionary phase). The phases of science, according to Kuhn, are thus:

- Pre-paradigm phase – no consensus about a certain phenomenon, incomplete theories
- Normal science – dominant paradigm emerged, progress in small steps within, but no questioning of paradigm, problems accumulate
- Revolutionary science – paradigms redefined

There is probably a good deal of truth in this observation - in the history of science, there are many examples of the fact that an old theory is only rejected if a better theory is on offer.

1.5 Summary: the principles of science

Bringing these different theories together, I think it's important not to interpret Aristotle, Popper, Chamberlin or Kuhn as opposing each

Thomas Samuel Kuhn (July 18, 1922 – June 17, 1996) was an American physicist, historian, and philosopher of science.

other - I would simply say that they highlight different aspects of the scientific enterprise.

Popper offers arguably the most pure account of a logically consistent way to use empirical arguments to deductively reject a scientific explanation. Chamberlin and Kuhn, on the other hand, tell us something about the social rules and necessities that affect scientists and therefore the scientific process. There are many more philosophers worth reading: Francis Bacon, Descartes, Hume, Wittgenstein, Russell, Lakatos - in the larger picture, neither of them is entirely right nor wrong. They simply highlight different aspects of science as a complex mix of philosophy, people and institutions.

In the heart of all that, however, remains the idea of creating “near-certain” knowledge by combining rational thinking with empirical observations.

Repetition of central concepts in this chapter:

- Science is based on empiricism and rationalism
- Philosophers have argued the removal of the inductive element from the scientific method is key, stressing the deductive parts of science
- Popper’s scientific method is the orthodox view, and you should know it by heart, but there are many other views on science. Some of the stress more strongly science as a human endeavor, for example the idea of multiple hypothesis to avoid falling in love with your theory.

2

On the shoulders of giants - the scientific literature

Science today is highly specialized. Most basic questions have long been solved. You may think of the scientific enterprise as a "house of knowledge", where one story is built on the next. This is often expressed by a famous quote by Isaac Newton

If I have seen further it is by standing on the shoulders of Giants.

¹

Regardless of whether the quote was meant as it is interpreted today, the fact remains that scientists today are hardly explorers that go far on unmarked territory. Before you start a new research project, you should therefore find out what has already been done in on the question you are interested in, to understand better what you are doing, but also to avoid doing something that has already been done.

¹ Brian Gottesman notes [here](#): "Sir Isaac's most famous quotation may well have been an exercise in sarcastic, spiteful anger. In February 1676 Newton wrote to Hooke "if I have seen further it is by standing on the shoulders of Giants." Often taken as a sign of Newton's great humility, this famed quote was almost certainly intended as an insult to Hooke, who was hunchbacked and may have suffered from a form of dwarfism."

2.1 Finding relevant literature

How to search for relevant literature depends on the your field, but basically there are three main databases

- Google scholar
- ISI web of knowledge (proprietary, slow, but most "official")
- Scopus (proprietary, competitor to ISI)

All of them work fine, with slightly different interface, search options, advantages and disadvantages. More details in our [lab on literature research](#).

2.2 Organizing your literature

Also the answer to how to organize your literature is pleasingly short - people have created specialized software for this purpose.

These reference managers work great, so use them! They will allow you to search in your papers for keywords, names, or anything else, link your pdfs, and they will also couple with your word processing software (MS Word, Open Office, LaTeX).

See our lab on working with a reference manager [here](#)

The only question is which of the available programs work best for you. A general recommendation is difficult, as the choice depends next to your personal preferences also on your operating system and your word processing software. More details in our [lab on literature research](#). You will also find lots of hints on the internet.

2.3 *Citation rules*

Given the amount of PhD titles that have recently been revoked from prominent public figures for not correctly citing the literature in their PhD theses (a charitable expression for outright plagiarism in some cases), it seems a good idea to make sure that everyone is aware of a basic rule in science:

High-profile cases of "faulty" citations include the former German Minister of Defense Karl-Theodor zu Guttenberg as well as the former German Federal Minister of Education and Research, Annette Schavan

If you use the ideas of another person in your work, you need to cite them!

There are different types of citations:

- Reference: It is like that (Smith, 2001), or „Smith (2001) expresses the opinion that ...“
- Quotation: „It is because it is“ (Smith, 2001, p.23)
- Personal communication “Frogs are green (S. Smith, personal communication)”

Taking larger passages / ideas from a source without making this clear is plagiarism. Changing only some words will make it obvious that deliberate intention was involved. Naming the source alone doesn't make it better when you design the presentation in a way that credit is wrongly attributed to you. By and large, the rules dictated by common sense and honesty apply!

Citation Signals Citations can be modified by citation signals. An example is the signal “but”, as in “It is like it is (but see Adams, 2009)”. These signals are meant to clarify the meaning of a citation, e.g. as supportive, related, or even contradicting the statement made in the main work. Citation signals are therefore crucial, and wrong use can also constitute misconduct! (a few of the high-profile cases with politicians turned around a cf. that was inserted although it should be a direct citation). A list of explanations of the different signals is provided [here](#). Some of the most important are:

You should read the papers you cite A recurrent problem, both in student's assignments and in the academic literature, is that papers are cited too broadly or wrongly. Editorial (2015) gives an account of that. Greenberg (2009) examine citation networks of papers that are miss-cited. The bottomline is: 1) sure you read the papers that you cite. 2) cite specifically, e.g. by noting in a short sentence for what you cite a paper or what they found, instead of writing only (see also Smith, 2001).

Repetition of central concepts in this chapter:

- You find literature through literature databases
- Use a reference manager to organize your literature
- Read the papers you cite, cite them correctly for what they say, and don't forget to mention all sources that you use!

3

Questions, data and valid conclusions

For to be possessed of a vigorous mind is not enough; the prime requisite is rightly to apply it. The greatest minds, as they are capable of the highest excellences, are open likewise to the greatest aberrations; and those who travel very slowly may yet make far greater progress, provided they keep always to the straight road, than those who, while they run, forsake it. (Descartes, *Le Discours de la Méthode*, 1637)

If you start a new research project, the first thing to decide on is your question. From that, you develop a hypotheses and experimental procedures to test it and so on. It cannot be stressed enough that the choice of the question is probably the single most important decision in a scientific project. It determines

- What you do for the next year / 5yrs / decade
- Whether you identify with your work or not
- If you can succeed
- Whether anyone cares or not

3.1 What makes a good scientific questions?

Many essays have been written on what makes a good question. If you listen to advice nobel price winners, they often tell you: don't listen to the opinions of the others, do your own thing. This may have become true for them, but certainly not everyone that avoids the mainstream becomes successful. The sad truth may be: it's hard to know in advance what question will make you famous.

That being said, I think it is quite easy to give a few minimum requirements for a good question.

It has to be new! A basic requirement is that your question is new, or, as we say in science: novel! Note there may be different levels of new

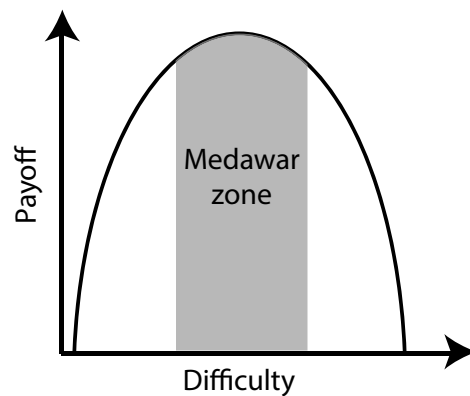


Figure 3.1: The Medawar zone is an illustration of the sweet spot between too trivial and too difficult questions. The term was coined in [Loehle \(1990\)](#), in honor of Sir Peter Medawar, who gave a verbal description of this zone in his book "The Art of the Soluble" ([Medawar, 1967](#)).

- Probably thought of, but not done before (no one bothered)
- Thought of, but couldn't be done (because methods lacking)
- Done, but wrongly (correct previous results)
- Not thought of before (really new)

Obviously, the last item would be the most desirable form of novelty. In any case - check the literature, make sure you understand to which extent your study is different from previous studies

It has to be exciting / relevant! Your question should be relevant to you, to your community, to society.

It needs to be a valid scientific question! We will discuss validity in more detail in the next chapter, but roughly, what is meant by that is: is the question logically consistent, and can we hope to generate a valid answer to it?

Risks and benefits should balance In all that, however, you should also make sure that risks and benefits balance - there may be great questions like "a general cure of cancer" - it's new, exciting and a valid scientific question. However, if you would try to find an answer in a MSc thesis, you would be set up for failure. Try to find questions that

- Don't test things when the answer is clear
- To avoid this, don't pose the question too widely

- Is a negative result nontrivial? If you can't show the thing that you expected, is the result still interesting?

Great questions are also to test a common belief that was not tested before, or question something which was considered established knowledge.

If you want to aim really high, you can read the essay "Ten Simple Rules to Win a Nobel Prize" by (Roberts, 2015)

3.2 *Experimental design and validity*

Assuming we have a question of interest and novelty, that seems to be worth pursuing - how can we make sure that we set up a valid scientific procedure to answer it?

A framework that allows us to classify things that can go wrong when we try to apply the scientific method as defined in chapter 1.4 is the concept of validity (Shadish et al., 2002). The application of these terms is more common in the social science / economics than in natural sciences, but the issues are universal, and for discussing and checking problems I find it useful to have a classification.

Construct validity and bias

The first problem that may appear when we try to test a theory are problems of construct validity. Basically, construct validity means: is what we measure well defined and corresponds to what we mean? Does it correlate with measures it should correlate with (convergent validity)? Is it unrelated to measures it should not correlate with (discriminant validity)?

Construct validity = is what we measure well defined and corresponds to what we mean.

Construct validity of proxy variables In "hard sciences" such as physics this is often not so much a concern, but the more we go to complex or social systems, the more likely it is that we measure variables that are proxies for concepts rather than a physical quantity. Examples are "intelligence", "biodiversity", "competition", "light climate". In all these cases, there is probably room for interpretation how to define and measure the concept exactly. It is very advisable to think about this hard, because it can have a strong effect on results.

Bias A part of construct validity that is important in either case are potential biases. Biases can appear through instrumental errors, but frequently also through observer bias. More on that in the lecture.

Internal validity

Internal validity is concerned with the question whether a causal effect can be demonstrated within the experiment / system under

Internal validity = correlation is not causation

study. The basic problem is that correlation does not prove causation. As an example - if we find that trees have started to grow taller in the last 50 years, and we find that atmospheric carbon dioxide concentration have risen in the same time, we can still not be sure that the CO₂ is the causal factor for tree growth.

Inferences are said to possess internal validity if a causal relation between two variables is properly demonstrated. A more detailed description on this in the [accompanying statistics script](#), section on experimental design.

Inferences are said to possess internal validity if a causal relation between two variables is properly demonstrated.

Statistical validity

Statistical validity is concerned with the question if an effect has been demonstrated with sufficient statistical (probabilistic) certainty.

Threats to statistical validity are wrong statistical assumptions, but also explicit and implicit biases through malpractices that are known as multiple testing, researcher's degrees of freedom, low power, P-value fishing / hacking. A more detailed description on this in the [accompanying statistics script](#).

External validity

External validity asks the question: Does our inference about correlations, or cause-effect relationships hold outside the empirical observations available to use. In other word - can we generalize our study, and if so, to what extent.

External validity = do extrapolations hold

The problem with external validity is that it's hard to demonstrate - if we have measured a plot in a particular location, we cannot be sure a priori how representative it is. In practice, it is a matter of experience and common sense to establish the domain on which your results are valid. Nevertheless, this is a very important task, because it is of course crucial for the results of a study to know how generalization they are.

Checklist validity

Before everything else, make sure the question is well defined. I recommend writing it down to avoid ambiguity. Now, check validity of:

- Construct validity
 - Do the variables that you can measure match your question?
 - What could be sources and extend of errors in your measurement (stochastic, systematic), think also about observer bias

etc.

- Internal validity
 - What are potential confound or interacting variables?
 - How are you planning to control / randomize / measure them?
Remember, if you can control is better than measure.
 - Are you sure about the direction of the causality?
- Statistical validity
 - How will you analyze your data?
 - What is the sample size that you need (power)?
 - Are there problems with the independence of your data? How can you minimize those problems (pseudo-replication, blocked design, ...)
- External validity
 - What generalization are you planning to make from this study?
 - What are threats to these generalizations / what could be different in other sites
 - Is there anything in your design that you can change to allow better generalizations?

3.3 *Planning and analysis of experiments*

Assuming that you have a good question, and that you have checked that you have validity of answering it, how do you perform the measurements and how do you analyze them? A detailed description on all this is in the [statistics script](#) accompanying these lecture notes. I highly recommend reading it. I therefore provide here only bullet points of the things that are absolutely crucial to know.

Planning and experimental design

If you have considered internal validity, you should already know which variables you need to consider. A more detailed discussion on the choice of variables to maintain internal validity (confounding effects) is in the [statistics script](#).

As a next step, you will have to decide how you vary the value of those variables (if you have control), and how many observations (replicates) you need. Again, this is discussed in more detail in the [statistics script](#), section on experimental design.

Apart from that, I think the most important tip for planning an

Simulate your analysis before going into the field!

experiment is: play it though! What I mean by that is: simulate how you take your data (for real with a pilot study, or make up some data), analyze it, and see if the results of this analysis correspond, in principle, to what you want to know. In most cases, you will notice that the way you were taking data is not optimal for the statistical analysis, that the analysis you do with your data does not answer what you want to know, or something alike. If you notice that before your experiment, it's easy to fix. If you notice it afterwards, it isn't.

Analysis

See chapter "inferential statistics" in the [statistics script](#) accompanying these lecture notes.

Visualization

See chapter "descriptive statistics and visualization" [statistics script](#) accompanying these lecture notes.

Repetition of central concepts in this chapter:

- A good scientific question is novel and important, and you should expect to be able to give a valid answer with a acceptable risk of failure
- Validity is a classification scheme of problems that can arise when trying to give an answer to a scientific question. See the validity checklist.
- You should plan your experiment thoroughly. Part of good planning is that you play the complete process through before taking the data, including the analysis.

4

Principles and channels of scientific communication

If you have made a scientific discovery, or at least a little bit of scientific progress, you probably want to let the world know about it. Hence, you need to communicate it.

4.1 Principles of communication

Thinking about communication requires a fundamental change in our attitude - up to now, we have dealt with ourselves, and how to generate knowledge. It was all about finding out the truth. Communication, however, is different. It is not necessarily about saying things right - it is about saying the right things.

What I mean by that is that the end of communication is not to say things as correct or truthful as they are - the thing that counts is that they are received as correct and truthful as possible by your audience. Good communication therefore requires adjusting to your audience. Before starting to communicate anything, you should therefore ask yourself the following basic questions:

1. Rule1: Who is my audience?
2. Rule2: What is my main message?

Based on the answer, adjust your words, level of detail, and jokes. There are a few general hints, however, that are always good advice:

Clarity Science is about getting this right - the main goal of (scientific) communication is therefore clarity and logic.

KISS People in science, however, are also busy - this means you need to communicate efficiently. Both goals are served well by the KISS principle - Keep it short and simple! However, don't forget Einstein: "as simple as possible, but not simpler".¹

The KISS principle - Keep it short and simple!

¹ This is also reflected in the following remark by Niels Bohr "Never express yourself more clearly than you are able to think."

Jargon To reach the goals outlined above, Jargon can be a good and a bad thing. It can make communication more efficient, but you have to be careful - this is only the case if you and the recipient have the same jargon. Jargon used in a wrong way is worse than no jargon.

4.2 *Communication channels*

Books and articles

The most important form of scientific communication is arguably the written word, in form of books and articles. A longer description on how a scientific article is structured is given in the following chapter.

Oral presentations

The probably second most important channel are presentations. That is, in the perception and attention of many people. Especially young scientists are very concerned with presentations, because they are often grade-relevant, or they feel the social pressure to speak in front of a group. But also for older scientists, presentations are crucial, for example in job interviews. It is therefore important to have some skill in this type of communication (see, e.g. the tips of [Kelleher and Wagener, 2011](#)). More material on poster presentations on our [lab on oral presentation](#).

Personally, I think the importance of presentations has decreased. That is not to say that oral presentations are not important - people still give presentations, and you shouldn't suck at it. Often, your job may rely on it. However, it seems to me that, somewhat surprisingly, the age of the internet and new media has moved the scientific debate more strongly back on an old technique - twitter, blogs, forums all work through writing, and these media, as the scientific article, are able to reach a much wider audience than a typical conference talk that is visited by maybe 50 people.

Poster presentation

Poster presentations are offered on many conferences for participants that don't give a talk. Poster presentations are a great opportunity to meet people and start discussing about your work, but I don't think they rank nearly as high in terms of importance than journal articles. More material on poster presentations on our [lab on posters](#).

Blogs, social media, and such

Blogs, social media and similar outlets have become more important in recent years. A collection of things that are worth checking out is provided in our [lab on social media](#).

4.3 Arguments and logical fallacies

In his article “Sounder thinking through clearer writing”, [Woodford \(1967\)](#) cites jargon as one of the key contributors to inefficient communication. He goes further, however, stating that if you communicate without clarity, you will start thinking without clarity yourself, leading to a tendency to draw wrong conclusions. Hence, clear communication is important for you and for others.

A similar case are rhetorical tricks and logical fallacies. There is a long list of archetypical conclusion structures that sound right, but that are typically wrong. Examples of these are

A [long list of logical fallacies](#) is provided on Wikipedia

- False dilemma - if we want everything to be completely correct, we can stop doing science!
- False equivalence - no model is completely perfect, therefore there is no difference between using model A and B
- Strawman - proving an argument wrong that is similar, but not identical to that of the opponent - A: Diversity can increase ecosystem function. B: my opponent suggests that diversity increases human well-being. But this is clearly wrong. Imaging a world where we have no HIV virus. Now, if I introduce the virus, we have more diversity. But human well-being is clearly lower!

Getting to know these fallacies has many advantages: you will train your logical thinking in general. You will be able to spot these fallacies when used by others. And lastly, you will notice if you use them yourself, avoiding wrong conclusions and sloppy arguments. Therefore, I highly recommend looking at this material!

Repetition of central concepts in this chapter:

- Good communication means not saying the right thing, but that your recipient understands what you want him/her to understand - the road to hell is paved with good intentions!
- You should know the different communication channels of science. The most important ones are the oral presentation and the scientific paper (next chapter)

- Be aware of rhetoric structures and logical fallacies. As a good scientist, you must be able to see through them!

5

Scientific writing

As already said in the previous chapter: writing is an essential skill for an academic, as for many other professions. Despite its central role, however, few master it well. Writing is a subtle art. You can read some advice, but to become a skilled and effective writer requires a long process of observation, practice and reflection: read widely, and think about why some writers appeal to you more. Look critically on your own writing, and think about the things that you can improve. The rest of this section is intended to help you start this journey.

The importance of the written word does not seem to change at all in the modern world. The art of writing, similar to public speaking, seems to be one of the stable skills that one can think of.

5.1 Some basic writing advice

Start writing early Writing structures your thinking, and you will find flaws in your logic that you didn't see before. "The time to begin writing an article is when you have finished it to your satisfaction. By that time you begin to clearly and logically perceive what it is that you really want to say." (Mark Twain)

Please, follow this point in your thesis - if you don't it leads foreseeable into disaster!

Know your audience Who is going to read what you write? You need to adjust your text to the knowledge and background of your audience.

[Gopen and Swan \(1990\)](#) state "it matters only whether a large majority of the reading audience accurately perceives what the author had in mind".

Know what you want to say, and make a selection! Sounds trivial, but in my experience it isn't. Before spending hours on writing the wrong text, try to explain a friend in 5-10 min the essential points of piece, and note the structure. If you fail with this, think again about your story. Also remember: Writing a good paper is not about telling everything you know or have done about a topic – it is about telling a story and selecting the relevant details.

"If you don't know where you are going, you'll end up someplace else" (Yogi Berra).

Mindset Make your goal, above all, clarity of thought and expression, and impervious logic of your argument. Read [Woodford \(1967\)](#).

Respect conventions Readers, particularly experienced readers, are familiar with writing conventions expect to find arguments, sections and statements in particular positions and structures. If the writer respects these structural standards the readability is greatly improved.

Modesty Think of a scientific text as a piece of craftsmanship, like a chair. What is expected of you is to produce a simple chair that respects the usual standards (e.g. four legs), and that is, above all, solid. You are not expected to add fancy ornamentation that might be used by an ingenious master craftsmen, and produce a entire set of furniture that would suffice to fit out a house. If you could do the later, you would gain admiration and praise, but when trying, you will nearly certainly fail to even produce the very thing you were asked for - a simple, solid chair. Bottomline: stay modest in what you want to achieve, but stay ambitions in how you achieve it!

Style Writing is difficult. Good writing is incredibly hard. If you think writing is no problem for you, you probably haven't even realized the extent of your writing problems. Read the following section twice!

5.2 *Style*

Style is personal, and there are many small aspects of style that we cannot cover in a short introduction. However, there are 4 basic properties of style that I find are nearly always missing in undergraduate writing, and then can improve the readability of your text dramatically.

Sentence Structure

The first is sentence structure. A sentence has two particularly important positions - the beginning and the end. These two positions have distinct functions (Gopen and Swan, 1990; Tischler, 1978).

The topic position The beginning of a sentence provides us with the context. Typically, we also use it to connect to previous information.

The stress position The end of the situation provides us with important new information. If it is not clear from the context that we have a logical closure, it also raises the expectation that this new information will be further developed in what is to come.

Strictly keeping this old-new structure in sentences creates a logical thread, and makes your text easier to read. Reversely, if you bring

For example, the default scientific article is divided into a title, abstract, introduction, materials and methods, results and discussion. Within these structural elements there are further recognizable patterns. Maintaining this structure makes it easier to understanding of the article's messages (Gopen and Swan, 1990; Tischler, 1978)

new information at the beginning of a sentence, it is often hard to understand. An additional bonus of checking your sentence structure is that you are more likely to notice flaws in your logic. I close with a quote from Brian McGill:

The battle for good writing is won sentence by sentence! A good sentence is: short, has the subject and verb together, has an active verb, has the points of emphasis at the beginning and end, and moves the reader along from a familiar launch point at the start to the new information at the end. (Brian McGill, on the blog Dynamic Ecology)

Paragraph structure

A paragraph is supposed to be a logical unit of text. The beginning of a paragraph provides the context. Different to a sentence, the new information should be in the middle of the paragraph. The end of a paragraph is reserved for conclusions (what have we learned in this paragraph). It may also bring up the connection to the next paragraph, e.g. by raising the necessity to look at a particular aspect in more detail. In your writing, make sure your paragraphs have a beginning and a conclusion, and that both are standing out verbally and through simplicity. The latter helps readers that only skimming the text.

Logic

Your text should be logically flawless. Read again about fallacies. Read about the scientific method and induction. Be careful with the use of logical indicators such as "thus", "hence", "however" - is the implication of this indicators really supported by your argument? More hints in the later section on common mistakes.

Clarity

Keep it simple.¹ More hints in the later section on common mistakes.

¹ Mark Twain is famous for his effective writing. A tip from him: "As to the Adjective: when in doubt, strike it out."

5.3 *The structure of a scientific article*

A paper from far above

Structure Seen far from above, a research paper consists of title, abstract and keywords, and four sections: introduction, methods, results and discussion section.

Story A scientific paper, however, is also a story. Many forget that it should be exactly one story. One paper, one story. Be sure you

know what your story is. A paper tells the story several times: once in title, once in the abstract, then in the whole paper, and you give additionally summaries of your story in the introduction, and in the discussion and conclusion.

The hourglass One purpose of telling your story several times is to provide the same information with different levels of details to your reader. The most detailed account is the full paper. Your reader needs to be prepared and led to this level of detail. The idea is that the introduction starts very general, and then becomes more and more specific, until we arrive at the most detailed level at the start of the methods. After the results, we discuss first the immediate results of the paper, and then become wider again, going to implications towards general conclusions. This is known as the hourglass structure. The abstract follows the same structure.

Going closer - structural elements in detail

Title The title is the first time you tell your story. The main purpose of the title is to inform the reader about the content of the paper. A good and informative title is crucial. Remember that most of your readers will only read the title of the paper.

Abstract The abstract is the second time you tell your story. Think of the abstract as a mini paper - it has all elements of a paper, but in 300 words. It starts off by setting the scene with a very short and general introduction. This usually links to the second part: Raise the problem. You can emphasize the research question by indicator words like "however". Third, you briefly delineate your approach. The fourth subsection serves to state the results. Finally, you present your conclusions and discuss the broader significance.

I provide an example later. It may be helpful to copy this as a template into your manuscript when you write your abstract.

Keywords and other info Articles usually also have some a list of keywords, usually below the abstract. These terms allow used to be important for indexing purposes in libraries or in the ISI web of knowledge database. Yet, since Google scans whole documents, their importance has decreased. Nevertheless, a good strategy is to choose keywords complementary to title and abstract, e.g. for important technical terms that you did not want to include in the abstract nor title.

Introduction The first main section of the article is the introduction. The introduction goes from the general to the specific, but not back

The introduction goes from the general to the specific, but not back to the general.

to the general. It should consist of four parts, each of which typically has 1-2 paragraphs

1. Introduce the topic. Lay out relevance and what has been done before. Here you usually cite a lot of literature showing that you know the current state of knowledge concerning the topic of your study.
2. Explain the problem, i.e. the gap of knowledge. You should demonstrate novelty as well as relevance.
3. Explain your approach: What are your methods; what are the data you used?
4. List your specific research questions. This subsection does not need to be clearly separated from the approach subsection. It should however, provide a link to the result sections by outlining the work flow of the analysis.

To distinguish these parts, you can make use of indicator words: however; problem/ challenge; here we used; we applied; we asked; we tested.

Methods The methods section explains the methods. It needs to be short, because paper space is limited. Technical details (Software version, etc.) are nowadays often moved into an online supplementary. Nevertheless, the methods should provide sufficient information to repeat the study in principle.. The method section is often divided in subsections such as study area, data, statistical analysis/ model, analysis, etc. Another helpful tip is to add figures. One further tips: problems with your methods are not discussed here, but in the conclusions.

Keyword is reproducibility.

Results The results present the results of the analysis, as clear as possible. They state only facts, but do not interpret them (unless the interpretation is obvious). The latter is the job of the discussion. The presentation of the results should follow the questions in the introduction. If you have three questions, you would usually have three subsections.

Discussion/ Conclusion The discussion summarizes the findings of the paper, and discusses it's implications and broader relevance. It moves from the specifics of your particular study to back to the general level.

The discussion moves from the the specific to the general.

1. Questions and main findings in short
2. Comparison to other findings
3. Limitations and further research
4. Conclusions, and broader impacts for science and society

The first part has the function of providing a lazy reader with the question and directly the findings of the study. The second part will discuss how these findings fit to previous research, if knowledge needs to be updated and so on. You will then typically proceed to limitations of your approach. The goal is to establish the domain under which your results are valid; do not question your approach in general, or list everything that could have been done as well. The final paragraph is very important - it provides your conclusions and outlook on the subject.

Only discuss the limitations that question/limit the credibility of your results.

5.4 *Showcase: A commented paper*

Fancy analysis supports hypothesis X as the main cause high species diversity in tropical forests

Florian Hartig and Severin Hauenstein

*Department of Biometry and Environmental System Analysis, University of
Freiburg, 79106 Freiburg, Germany*

October 18, 2015

Abstract

Tropical forests are some of the most species-rich ecosystems of the world. The reason for this, **however**, is still widely debated. Hypotheses range from processes related to productivity over environmental stability to the historical changes in geography. **Here**, we contrasted these different hypotheses by using data from ... together with ... (fancy new method). **We find that** hypothesis X seems to be significantly better supported by our data than all alternatives we tested. Specifically ... **In conclusion**, our study supports the hypothesis that species diversity in the tropics is mainly driven by higher productivity. These results challenge some long held ideas about geographical stability being the main reason for global diversity patterns. They also have important practical applications for mitigation of climate change, as ...

Keywords: Biodiversity – Primary production – Floristic structure ...

Try to capture the questions, methods, and main results of your study in the title, if possible

Remember: the abstract is a mini-paper

First words set the scene

Raise the problem. Indicator word: however

Introduce your approach. Indicator word: here

State your results. Indicator word: We find, our results are ...

Give your conclusions and discuss wider significance. Indicator word: in conclusion, ...

Keywords for indexing purposes.

Introduction

Tropical forests are some of the most species-rich ecosystems. Ac-

Introduce the topic. Explain relevance. Remember intro should mirror the abstract, one sentence in the abstract = one paragraph intro

According to XY (19XX) tropical forests make up $x\%$ of the planet's biodiversity.

...

The reason for high diversity in the tropics is still widely debated. Historically, it was thought [...]. More recently, however, the view has emerged that [...] (e.g. ABC, 19XX; D&E, 20XX; FG, 20XX). Up to now, none of these theories has been tested to an extent that it could be excluded as a potential explanation for the high tropical biodiversity

Introduce the problem (1-2 paragraphs)

...

While the hypotheses described earlier cannot be tested directly, a possible route would be to conduct an indirect test via the method of ABC (19XX). This method has originally been developed for X, but it should be able to distinguish among Y and Z. To our knowledge, no such test has been performed yet.

Introduce your approach ... could also be, as here, by explaining a new option to address the problem

...

We therefore applied X to compare the support for the main hypotheses published by ABC (19XX), D&E (20XX), FG (20XX). For the latter we used X data provided by XYZ (2014). We first test X in order to find the difference in Y. As an alternative approach we choose to look at Z for the purpose of XY.

Demonstrate novelty

Approach, now more specific

...

We ask the following questions:

Questions at the end of the intro should mirror sections in the results and discussion in the conclusion.

1. Which of the three hypotheses receives most support from the data
2. Question 2
3. Question 3

Methods

This section is typically divided into

1. Study area
2. Experimental procedure
3. Statistical analysis

Results

The presentation of the results should follow the questions at the end of the introduction. If you had three questions, you would typically make three subsections.

Discussion

In this study, we used method X to test whether prevalence of A is associated to factor B. Our main result is that there is a significant correlation of B with A. This correlation, however, was only found when factor C was also present.

Summarize main conclusions

Our findings support earlier findings of (X, 2005) and (Y, 2002) who also found a positive correlation between A and B. The fact that this correlation was only present in samples under conditions C may also explain previous opposite results such as (ZZ, 2006,2008), as neither of these previous studies controlled for C. The reason for this positive association and the fact that it is affected by C is still unknown. We speculate that mechanism R could be a reason for this. This idea is given further support by the observation that the correlation between A and B is affected by C. Such an effect would be expected if R is really the cause of this Correlation, as R is dependent on C (XAY, 1968).

Discuss relation to other studies and scientific implications

A limitation of our study is the geographical limitation of our data. As we could only retrieve data from the neotropics, but not from Africa or Asia, we can currently not say if our results will hold globally.

Discuss limitations and scope of the conclusions (see external validity). Limitations is not everything that could also have been done.

A further issue is the possibility that niche conservatism contributes to the observed signal. Currently, we lack sufficient genetic data to obtain high-resolution phylogenies that could be used to test for this effect. However, as these data become increasingly available, future studies should include phylogenetic information to exclude this possibility.

Limitations are often connected to ideas for further reasearch

In conclusion, the results of our study supports the idea that X is the main mechanism for the high diversity in the tropics. This has important implications for global biodiversity conservation [...], as well as for our understanding of the maintenance of diversity.

End with conclusions and wider importance for science and society

Some people end with further research here. I find this a bit weak and prefer to deal with this topic in the earlier paragraphs.

5.5 *A list of common writing problems*

(A) Strategic writing problems

- (a) Do not make statements that are unnecessary for your argument, but could be questioned. Example: "The two most classic games to simulate the behavior of animals are the Hawk-Dove game and the Prisoner's Dilemma." → Others might disagree. Is it really important to be so definite here? Why not say: "Two classic games..."?
- (b) Passive/ Active: See the comments in the lecture notes on scientific writing. In general you should write in an active voice, unless 1) Sentence structure C₃ takes precedence, or 2) the thing you talk about is impersonal, was done by another person. More details in the lecture notes. In particular, opinions of your own SHOULD be active.
- (c) Be definite - remember Yoda: There is no try, do it. Phrases like: "we attempted to find whether there is" or "this study

aims at" sound weak and defensive. Write "we studied whether ...". Also: "to the best of our knowledge ... " This is a borderline case. You will see this expression in scientific articles, and it's OK to use it. However, it does seem insecure. Better to indicate that you made some effort to find out together with such an expression.

(B) Logical problems

- (a) The stated implication does not follow the argument logically. Example: "Because both models follow different pay-off schemes, it is important to know how a simulation model reacts to different settings." → Unclear why it is important to know how simulation reacts to different settings if two models are different.
- (b) "Micrologic": Logical indicators such as "thus", "hence", "however", "nevertheless" have a clear and distinct meaning. "thus", "hence" indicate an implication. "however", "nevertheless" indicate contradicting statements or evidence to the previous proposition.

(C) Style

- (a) Drop redundant words: Any of the following **words** can be erased: "**a total of** n subjects", "the simulations were run with the **exact** same parameter set as before"
- (b) Pompous/ pontificating style: If something can be said in simple words, it should be said in simple words, i.e. if the only purpose of a sentence is to "sound scientific", it should not be said at all. Examples: The data **featured/comprised** a pixel size of 10 (had). We utilized (use).² We used the model in an integrated framework to facilitate a holistic forecast" (predicting from the model). Tip: Read [Woodford \(1967\)](#).
- (c) Topic position/ Stress position not properly occupied: The first part of a sentence should give the context. The last part of the sentence should provide the crucial new information, and should relate to what is further discussed. Tip: Read [Gopen and Swan \(1990\)](#).
- (d) A paragraph structures the text into topical units. Use them! Each paragraph should introduce its topic in the beginning, and ideally lead to some conclusion or summary at the end.
- (e) Unnecessary adjectives. Mark Twain: "As to the adjective: when in doubt, strike it out." Especially adjectives such as "very", "particularly" are often not necessary. Applies also to larger constructs such as "Evolutionary Game Theory is an

² Styleguides advice against complicated words such as utilize if there are simpler equivalent alternatives, although others have argued that there is a different between use and utilize - but could you name it? See discussion here [here](#).

important tool to investigate(s) ..." – if you erase the underlined text, the sentence remains fine.

- (f) Unclear/ ambiguous expressions: Write as clear and as informative as possible. Avoid ambiguous words like "tends to", "mostly" etc. if you can make a definite statement. Example: "the frequency of cooperators always tends to be zero" → What does that mean? 1) It was always zero or 2) It was zero in 95 % of the cases or 3) something else.

(D) Grammar and spelling

- (a) English commas have different rules than in German and other languages <http://www.grammarbook.com/punctuation/commas.asp>
- (b) Wrong use of **which** and **that** Let's say that we have 10 samples. Compare the two sentences: 1) the samples, which were expose to radiation the day before, were analysed. 2) The samples that were exposed to radiation the day before were analyzed. The important difference is that 1) says all samples were exposed to radiation and then analysed, while 2) implies that not all samples were exposed to radiation, and only those that were were analysed. 1) is called a non-restrictive clause, and 2) is called a restrictive clause. "that" always implies a restrictive meaning. Additional confusion, however, is created by the fact that "which" can in fact be used for both, with a minor modification: if you want to use which restrictive, you have to write 3) The samples which were exposed to radiation the day before were analyzed. Notice that there is a tiny difference to 1) - there is no comma before which. This is the only indicator that allows us to know whether the restrictive or the non-restrictive meaning is implied. Many people are not aware of this difference though. In prose, the distinction may matter little, but for science, my opinion is that clarity goes before style, and I therefore recommend to strictly stay with 1) which for non-restrictive and 2) that for the restrictive meaning.

5.6 Examples

In the following, I highlighted redundant words and other problems

- A subsequent ANOVA analysis **enabled a quantification of** the impacts of the varied factors (quantified)
- There are a number of records in the literature focusing on comparisons between sets of modeling approaches while predicting biomass at

plot scale (A number of previous studies has compared modeling approaches to predict biomass at the plot scale)

- As such, the explicit findings of our two experiments showed that a total of 9 samples ...

Repetition of central concepts in this chapter:

- A sentence starts with the context and ends with the new info
- A paragraph starts with the topic, discusses the topics, and ends with a summary and connection to the next point
- Write with clarity and logic
- Follow the standard structure. No experiments!
- Read through the list of common writing problems!

6

A working scientists

We have covered by now the primary scientific process from deciding on a scientific question, generating valid data and conclusions, and communicating those in writing or through other channels. What remains are the aspects that relate to science is also a human and professional endeavor. The purpose of this chapter is to give a very fast overview on questions such as professional norms and ethics, as well as career options and the working life of scientist.

6.1 Professional ethics and good scientific practice

The term "good scientific practice" encompass the accepted rules and norms of the field. They structure around subjects such as

- Honesty in data generation and the presentation of results (don't falsify data)
- Honesty in attribution (authorship, plagiarism, etc)
- Standards of professional work (data storage, lab notebook, ...)
- Contribution to the scientific society (share data, contribute to common goods such as reviewing)

Many universities, institutes, and funding organizations have explicit rules of good scientific practice [Forschungsgemeinschaft](#) (e.g. 2013). The following paragraphs give a fast overview about the most important topics:

The rules of the German Science Foundation (DFG) are available [here](#). Other guidelines for good scientific practice are available [here](#)

Fabrication and falsification You must not do this!

Misrepresentation of results You must not do this either!

Misrepresentation would be knowingly using wrong statistics, or claiming something you haven't shown

Documentation and organization There is no clear legal rule about how long data, documentation and samples have to be stored, but most funding organizations recommend a 10yr period. Depending on the discipline, not having a lab notebook or storing your data would be considered something between a sloppy behavior and a severe and highly suspicious misconduct (many ecological researchers, for example, have very sloppy or no lab/field notebooks. In medical research, lab notebooks are taken extremely serious.)

Attribution and authorship A common problem is also the attrition of work, i.e. authorship and plagiarism. You know you should not copy text from others (plagiarism). You also should not copy text from yourself and publish it again to appear to be more productive (self-plagiarism). Authorship is a difficult issue. More on that in the lecture.

Simple things like having a good folder structure for each project help a lot. Have a look at our lab on [project organization](#)

When working with code, specially collaboratively, there is no way around using a version control system. See our lap on version control [here](#)

Repetition of central concepts in this chapter:

- Read and follow the rules of good scientific practice

7

Bibliography

Descartes, R., 1673. Discourse on the Method of Rightly Conducting One's Reason and of Seeking Truth in the Sciences.

Editorial, 2015. Cause, correlation, conjecture. *Nat Genet* 47 (4), 305–305.

Forschungsgemeinschaft, D., 2013. Rules of Good Scientific Practice, 2nd Edition. WILEY-VCH.

Gopen, G. D., Swan, J. A., 1990. The Science Of Scientific Writing. *Am. Sci.* 78 (6), 550–558.

Greenberg, S. A., 2009. How citation distortions create unfounded authority: analysis of a citation network. *BMJ* 339.

doi: 10.1136/bmj.b2680

Kelleher, C., Wagener, T., 2011. Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling & Software* 26 (6), 822–827.

Loehle, C., 1990. A guide to increased creativity in research: Inspiration or perspiration? *Bioscience* 40 (2), 123–129.

Medawar, P. B., 1967. The art of the soluble.

Platt, J. R., 1964. Strong Inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science* 146 (3642), 347–353.

doi: 10.1126/science.146.3642.347

Roberts, R. J., 2015. Ten Simple Rules to Win a Nobel Prize. *PLoS Comput Biol* 11 (4), e1004084–.

doi: 10.1371/journal.pcbi.1004084

- Shadish, W. R., Cook, T. D., Campbell, D. T., 2002. Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin Boston.
- Sokal, A., 2008. Beyond the hoax: Science, philosophy and culture. Oxford University Press.
- Tischler, M. E., 1978. Scientific writing booklet. University of Arizona.
- Woodford, F., 1967. Sounder thinking through clearer writing. Science 156 (3776), 743.