Engineering Research Methodology

A Computer Science and Engineering and Information and Communication Technologies Perspective.

Krishnan Nallaperumal

First Edition

1. Basic research methodology

Do we need Research?

"All progress is born of inquiry. Doubt is often better than overconfidence, for it leads to inquiry, and inquiry leads to invention"

- *Hudson Maxim*

Any one can understand the significance of research from the above quote. Humanity gained a lot due to research and many of the comforts we have would not have been realized without research. Increased amounts of research make progress possible for humanity. Research inculcates scientific and inductive thinking and it promotes the development of logical habits of thinking and organization.

What is research? - A search for knowledge - A scientific and systematic search for significant information on a specific topic – A careful investigation through search for new facts in any branch of knowledge – A voyage of discovery – and so on.

When the unknown confront us, our inherited instinct of inquisitiveness (curiosity) makes us probe to attain the full understanding of the unknown. The method we employ to probe for getting the knowledge of unknown can be called research.

Research is defining and redefining problems, formulating hypothesis or suggested solutions; collecting, organizing and evaluating data; making deductions and reaching conclusions; and finally testing the conclusions carefully to determine whether they fit the hypothesis formulated.

Research is the pursuit of truth with the help of study, observation, comparison and experiment; the search for knowledge through objective and systematic method of finding solution to a problem is Research. The systematic approach concerning generalization and the formulation of a theory is also research.

Broadly speaking, research refers to the systematic method consisting of enunciating the problem, formulating a hypothesis, collecting the facts or data, analyzing the facts and reaching certain conclusions either in the form of solution to the concerned problem or in certain generalizations for some theoretical formulation.

This method we employ will vary according to the broader category of knowledge domains: literature, language, arts, social sciences, science and Engineering.

1.1 Objectives and Motivations in Research

The purpose of research is to discover answers to questions through the application of scientific procedures. The main aim of research is to find out the truth which is hidden and which has not been discovered as yet. Research objectives fall into the following broad groupings:

1. To gain familiarity with a phenomenon or to achieve new insights into it (known as exploratory or formulative research studies);
2. To determine the frequency with which something occurs or with which it is associated with something else (known as diagnostic research studies);
3. To test a hypothesis of a causal relationship between variables (known as hypothesis-testing research studies).

Why do people aspire for research? What makes them undertake research? This motivation is of fundamental importance. The possible motives for doing research may be either one or more of the following:

- Aspire to get a research degree along with career benefits;
- Aspire to take up the challenge in solving the unsolved problems;
- Desire to get intellectual joy of doing a creative work;
- Aspire to do research to serve the society;
- Seek to get recognition and respect;
- Many other;

Many more unlisted factors such as directives of government, employment conditions, curiosity about new things, social thinking and awakening, and the like may as well motivate or compel people to do research.

1.2 Distinct Approaches and Significance of Research

There are a few distinct approaches prevalent in doing research which should be known to a researcher in any field of study so that he analyses his problem in proper perspective, understands what methods and tools are needed and decides to choose the appropriate approaches for his research; some among them are worth comparing and contrasting to get a clear distinction of what should be ones approach towards research. The following are discussed with that objective:

1.2.1 Descriptive Research vs. Analytical Research

Descriptive research includes surveys and fact-finding enquiries of different kinds. **The major purpose of descriptive research is description of the state of affairs as it exists at present.** This approach is suitable for social sciences and business and management studies for descriptive research studies. **The main characteristic of this method is that the researcher has no control over the variables;** s/he can only report what has happened or what is happening. Most research projects of this nature are used for descriptive studies in which the researcher seeks to measure factors like frequency of shopping, brand preference of people, most popular media programme etc. Survey methods of all kinds fall under descriptive research, including comparative and correlation techniques. **In analytical research, on the other hand, researcher makes a critical evaluation of the material by analyzing facts and information already available.**

1.2.2 Applied Research vs. Fundamental Research

Research can either be Applied Research or Fundamental Research. The main target of **Applied Research** is **to find a solution for an immediate problem facing a society or an industrial / business organization**, whereas **Fundamental or Pure Research** is mainly **concerned with generalizations and concentrates on the formulation of a theory. "Gathering knowledge for the sake of knowledge"** is termed '**Pure**' or '**Basic**' or **'Fundamental' research.** Examples of fundamental research are: research concerning some natural phenomenon or related to pure mathematics; research studies aimed at studying and making generalizations about human behaviour. While applied research concentrates on discovering a solution for some pressing practical problem, fundamental research is focused towards formulation of theories that may have a broad base of applications either at present or for future which adds more materials to the already existing organized body of scientific knowledge.

1.2.3 Quantitative vs. Qualitative

While **Quantitative research** is applicable to **phenomena** that can be expressed **in terms of quantity**, **Qualitative research** is concerned with **qualitative phenomenon**. For instance, when we are interested in investigating the reasons for human behaviour (i.e., why people think or do certain things), we quite often talk of 'Motivation Research', an important type of qualitative research. Qualitative research is especially important in the behavioural sciences where the aim is to discover the underlying motives of human behaviour.

1.2.4 Conceptual vs. Experimental (or Empirical)

Conceptual research is that **related to some abstract idea(s) or theory**. It is generally used by philosophers and thinkers to develop new concepts or to reinterpret existing ones. On the other hand, **experimental (empirical) research relies on experiment or observation alone**, often without due regard for system and theory. It is data-based research, coming up with conclusions which are capable of being verified by observation or experiment. In such a research it is necessary to get at facts firsthand, at their source, and actively to go about doing certain things to stimulate the production of desired information. **In such a research, the researcher must first provide himself with a working hypothesis** or guess as **to the probable results**. He then **works to get enough facts (data) to prove or disprove his hypothesis**. He then sets up experimental **designs** which he thinks will manipulate the persons or the materials concerned so as to bring forth the desired information leading to the hypothesis. Such research is thus characterized by the experimenter's control over the variables under study and his deliberate manipulation of one of them to study its effects. **Empirical research is appropriate when proof is sought that certain variables affect other variables in some way.** Evidence gathered through experiments or empirical studies is today considered to be the most powerful support possible for a given hypothesis.

1.2.5 How to Approach Research?

From the discussions held above, it is clear that there are two basic approaches to research, viz., **quantitative** approach and the **qualitative** approach. The former **involves the generation of data in quantitative form which can be subjected to rigorous quantitative analysis in a formal and rigorous fashion.** This approach can be **further sub-classified into inferential, experimental and simulation approaches research**. The purpose of **inferential approach to research is to form a data base from which to infer characteristics or relationships of population**. **This usually means survey research** where a sample of population is studied (questioned or observed) to determine its characteristics, and it is then inferred that the population has the same characteristics. **Experimental approach is characterized by much greater control over the research environment** and in this case some variables are manipulated to observe their effect on other variables. **Simulation approach involves the construction of an artificial environment within which relevant information and data can be generated**. This permits an observation of the dynamic behavior of a system (or its sub-system) under controlled conditions. Given the values of initial conditions, parameters and exogenous variables, a simulation is run to represent the behavior of the process over time. Simulation approach can also be useful in building models for understanding future conditions.

Qualitative approach to research is concerned with subjective assessment of attitudes, opinions and behavior. Research in such a situation is a function of researcher's insights and impressions. Such an approach to research generates results either in non-quantitative form or in the form which are not subjected to rigorous quantitative analysis. Generally, the techniques of focus group interviews, projective techniques and depth interviews are used. All these are explained at length in chapters that follow. **Engineering research may not have anything to do with this approach.**

1.2.6 The significance of research

This can be understood keeping in view the following points from different persons' perspective:

- To those students who are to write a master's or Ph.D. thesis, research may mean careerism or a way to attain a high position in the social structure;
- To professionals in research methodology, research may mean a source of livelihood;
- To philosophers and thinkers, research may mean the outlet for new ideas and insights;
- To literary men and women, research may mean the development of new styles and creative work;
- To analysts and intellectuals, research may mean the generalizations of new theories.

Thus, research is the fountain of knowledge for the sake of knowledge and an important source of providing guidelines for solving different business, governmental and social problems. It is a sort of formal training which enables one to understand the new developments in one's field in a better way.

1.3 Research Methodology versus Research Methods

It seems appropriate at this juncture to explain the difference between research methods and research methodology. **Research methods may be understood as all those methods/techniques that are used for conducting research.** Research methods or techniques, thus, refer to the methods the researchers adopt to carry out their research.

Research methods can be put into the following three groups:

1. In the first group we include those **methods** which are concerned with the **collection/ acquisition of data**; these methods will be **used** where the **data already available are not sufficient to arrive at the required solution**;
2. The second group consists of those **mathematical/statistical techniques** which are used **for establishing relationships between the data and the unknowns;**
3. The third group consists of **those methods which are used to evaluate the accuracy of the results** obtained.

Research methods falling in the above stated last two groups are generally taken as the analytical tools of research.

Research methodology is a way to systematically solve the research problem; it may be understood as **a science of studying how research is done scientifically.** In it **we formulate the various steps that are to be adopted by a researcher in studying his research problem along with the logic behind them. It is necessary for the researcher to know not only the research methods/techniques but also the methodology,** Researchers not only need to know how to develop certain indices or tests, how to calculate the mean, the mode, the median, the PSNR, or 'the standard deviation or chi-square, 'how to apply particular research techniques, but they also **need to know which of these research methods or techniques, are relevant and which are not, and what would they mean and indicate and why**. Researchers also need to **understand the assumptions underlying various techniques.** All this means that it is necessary for the **researcher to design his methodology for his problem as the same may differ from problem to problem.** For example, an architect, who designs a building, has to consciously evaluate the basis of his decisions, i.e., he has to evaluate why and on what basis he selects particular size, number and location of doors, windows and ventilators, uses particular materials and not others and the like. Similarly, in research the scientist has to expose the research decisions to evaluation before they are implemented. He has to specify very clearly and precisely what decisions he selects and why he selects them so that they can be evaluated by others also.

It is clear from the above discussions that research methodology has many dimensions and research methods do constitute a part of the research methodology. The scope of research methodology is wider than that of research methods. Thus, when we talk of **research methodology** we not only talk of the **research methods** but also **consider the logic behind the methods we use in the context of our research study and explain why we are using a particular method or technique and why we are not using others so that research results are capable of being evaluated either by the researcher himself or by others.** Why a research study has been undertaken, how-the research problem has been defined, in what way and why the hypothesis has been formulated, what data have been collected and what particular method has been adopted, why particular technique of analyzing data has been used and a host of similar other questions are usually answered when we talk of research methodology concerning a research problem or study.

1.4 Research Process

1.4.1 Finding a research Advisor/Guide

The choice of research supervisor is probably the most important one that a research scholar faces.

1.4.1.1 What to Look for in a Potential Research Advisor/Guide

The ideal advisor might have the following traits:

Has research interests in common with the scholar.

Work is easier when both you and your advisor find the research area fun.

Has a national or international reputation among researchers.

Someday the scholar will finish his/her degree work and be looking for a job. Advisor's reputation and professional colleagues could be key in opening opportunities for the scholar.

More immediately, the advisor will be leading scholar's research, at least at the beginning, and it is important that the advisor knows how to do quality research.

Has grant support for research.

If the scholar is working his/her way through school as a teaching or research assistant, he/she may well want to be supported as a research assistant by the advisor. Even if the scholar has fellowship support or an outside job, grant support is a sign of the advisor's skill as a researcher, although in subjects such as pure mathematics, research assistantships are rare.

Has successfully directed students in the past.

The scholar is new at research; it helps if the advisor has some experience in dealing with research students.

Has a reputation as a fair and reasonable advisor.

No scholar wants to work with an advisor who never shares credit for ideas, who expects every student to spend 7 years as a research scholar, or who doesn't know how to motivate and encourage students.

Has a high probability of staying at the university.

If the advisor moves to another university before the scholar finishes his/her degree, while the scholar is trying to finish his/her works, the scholar will have a severe handicap. The scholar may need to consider moving with the advisor or changing advisors--a difficult situation at best. Even if the advisor goes travelling for a year on sabbatical, communication can be temporarily difficult.

Is someone scholars like and admire.

The scholar will be working closely with this person until his/her graduation, and the relationship will not end even then. The more comfortable the two of them are with the relationship, the fewer distractions from the research at hand.

Has an active research group.

Scholars can learn a tremendous amount from more advanced students, and the opportunity to work in a group of motivated researchers working on similar topics is quite stimulating. Be aware, though, that if the group is too big, the scholar will have little time with the advisor, and may be directed by a post-doctoral student or a more advanced graduate research scholar.

1.4.1.2 How to Find an Advisor/Guide

Before any scholar come to the university, he/she should have made sure that some faculty members were active researchers in areas of interest to him/her. Now is the time to consider each of those candidates as a potential advisor, measuring them up against the criteria in the previous section.

Use of departmental annual reports, *Science Citation Index*, *Math Reviews*, or electronic resources will help the scholar to find recent publications by each advisor. The scholar should read a few of these publications, and try to understand them enough to be able to ask intelligent questions and to see directions for further work.

Scholar should get to know potential advisors by taking courses from them, attending seminar talks given by them, and by seeing them in their offices (by appointment or during office hours) to talk about their research interests. The scholar should ask for relevant papers to read.

Scholar should talk to other students about various candidate advisors.

Scholar should get advice from faculty members that he/she respects or from the graduate office of his/her department.

Good indicators are strong publications in major journals, some grant support, and a good teaching record.

Once the scholar has a good research advisor, he/she should ask that person to be the advisor. Don't be discouraged by a ``no''--try a different advisor. Good advisors are much in demand, and they don't remain good if they stretch themselves too thin.

1.4.1.3 The Advisor-Advisee Relationship

The best analogy for the relationship between an advisor and a student is probably that between a parent and a child.

At the beginning, the child has little independence, and almost every action is directed by the parent. Initially, most students need close supervision, being told what papers to read and what tasks to accomplish.

As the child grows, independence develops. A student begins to ask interesting research questions with minimal prompting and can set the direction of the next week's work. The advisor still plays a crucial role as catalyst and evaluator of ideas.

As adolescence sets in, conflicts arise. The student realizes that all too soon, school days will end, and it will be essential to be able to function on one's own. A student eventually may feel that research would be more easily finished without the advisor's ``interference,'' even though the student may lack the detachment necessary to evaluate the work. Independence is frightening, but dependence is resented, and frustration can run high.

In adulthood, parent and child redefine the relationship. The process of graduate school should transform the advisor's student into the advisor's colleague. The two may or may not continue to collaborate after the student graduates, but future contact is ideally built on mutual respect, gratitude from the student to the advisor for the professional formation, and pride of the advisor in the student's accomplishments.

It is important that the advisor and advisee develop a compatible working style. Some people thrive on regular weekly meetings between the two that force the student to synthesize the week's accomplishments (or to explain the reasons for the lack of progress). Others rely on chance encounters in the hall. Some advisors have weekly group meetings at which each student discusses progress and everyone can comment. Some advisors expect students to attend seminars or journal clubs in order to keep up with recent research results.

If the scholar feels that he/she is floundering (as every student sometimes does), should ask his/her advisor for extra meetings, should send frequent email messages asking for pointers, or discuss his/her work with another trusted faculty member or student.

As in any relationship, conflicts should be faced and discussed. Cultural and generational differences can lead to misunderstandings that are easily resolved once they are recognized. Sometimes a fellow faculty member or graduate student can lend some insight.

In rare cases, the relationship just does not work. In such cases, the student should seek another advisor, leaving the first with as little ill-will as possible.

1.4.2 Finding a Topic and Beginning Research

There are some aspects of graduate school that are more daunting than others, and finding a research topic is perhaps the biggest obstacle for most students. The characteristics of an ideal topic are to some extent incompatible:

> The subject should be timely. Previous groundwork should leave the scholar’s research problem ripe for completion, and it should be in an active area with potential for future work and employment.
>
> On the other hand, if a field is too crowded, and the subject too prominent, then the scholar stand a risk being ``scooped'' by a more experienced researcher who is able to work faster than him/her. In this case, the scholar may be forced to start over again (rather disastrous) or at least publish jointly (possibly a blessing, but surely an inconvenience).
>
> The scholar’s work should lead to a well defined set of results to which he/she can lay claim. In particular, employment prospects will be lessoned if the scholar merely completes a small piece of a very large project or piece of software which is closely identified with his/her advisor, or is published with a long list of collaborators.

On the other hand, it is impossible to work in a vacuum, and the scholar's task can be significantly harder if he/she doesn't have a group of people working on closely related problems with whom he/she can interact and share code.

The best theses show a high level of creativity - and are often somewhat speculative. It is often unclear at first how the ideas will develop.

On the other hand, a multiyear plan of research is a very valuable asset.

The scholar should really enjoy the subject, and wants to spend the next several years with it!

On the other hand, an ideal subject is of no use without a thesis advisor who is willing to direct you in it.

Clearly some compromise is necessary here!

1.4.2.1 Getting Research Ideas

1.4.2.1.1 Becoming an Active Reader and Listener

It is very important to make the transition from the passive mode of learning that traditional lecture courses encourage to an active and critical learning style. Whenever one reads technical material, evaluates a piece of software, or listens to a research talk, should ask him/her these *canonical questions*:

From where did the author seem to draw the ideas?
What exactly was accomplished by this piece of work?
How does it seem to relate to other work in the field?
What would be the reasonable next step to build upon this work?
What ideas from related fields might be brought to bear upon this subject?

One technique that some find helpful is to keep a written log of one's technical reading and listening. Review it periodically to see if some of the ideas begin to fit together.

1.4.2.1.2 Exposing Yourself to Research

The scholar should set aside some time every week for trying to generate research ideas. Some possible catalysts are:

Making a weekly trip to the library to read at least the abstracts from the premier journals in one's field. One should choose an article or two to read in depth and critique.
The scholar should make a weekly investigation to find technical reports in his/her field, using electronic resources or libraries. Read selectively and critique.
The scholar should attend at a research seminar or colloquium series. Listen and critique.

The scholar should maintain a log and add these to the log, and ask the canonical questions. As the scholar reviews the log 6 months from now, he/she may find something that strikes a chord then but is beyond him/her now.

1.4.2.1.3 Directed Study

Which comes first: the thesis advisor or the thesis topic? The answer is, both ways work. If the scholar has identified a compatible advisor, he/she could ask for an independent study course. Both of them together set the focus for the course, with the scholar having more or less input depending upon his/her progress in identifying a subfield of research.

1.4.2.1.4 Formulating the Research Problem: Develop the Nucleus of an Idea

There are two types of research problems, problems which relate to states of nature and those which relate to relationships between variables. In the beginning stage itself the scholar must single out the problem he/she wants to study, i.e., must decide the general area of interest or aspect of a subject-matter that he/she would like to inquire into. Initially the problem may be stated in a broad general way and then the ambiguities, if any, relating to the problem be resolved. Then, the feasibility of a particular solution has to be considered before a working formulation of the problem can be set up. The formulation of a general topic into a specific research problem constitutes the first step in a scientific enquiry. Understanding the problem thoroughly and rephrasing the same into meaningful terms from an analytical point of view are the very important steps in the process of formulating the research problem.

The best way of understanding the problem is to discuss it with the scholar's own colleagues or with those having some expertise in the matter. In an academic institution the scholar can seek the help from an advisor/guide who is usually an experienced person and has several research problems in mind. Often, the advisor/guide puts forth the problem in general terms and it is up to the scholar to narrow it down and phrase the problem in operational terms.

Once the scholar has identified a topic that looks feasible, he/she should make sure that he/she is aware of all of the literature in the area. The scholar should keep reading and listening and keep distinct in his/her mind what is different between his/her work and others. If the scholar does not frequently review the literature he/she had read months ago, he/she may find him/her unconsciously claiming credit for other people's ideas. On the other hand, the scholar should not let other people's frame of mind limit his/her creativity.

The scholar may review two types of literature-the conceptual literature concerning the concepts and theories, and the empirical literature consisting of studies made earlier which are similar to the one proposed. The basic outcome of this review will be the knowledge as to what data and other materials are available for operational purposes which will enable the scholar to specify his own research problem in a meaningful context.

1.4.2.2 Extensive Literature Survey: A Trap to Avoid

It is possible to spend almost all of one's time in literature review and seminars. It is easy to convince oneself that by doing this he/she is working hard and accomplishing something. The truth of the matter is that nothing will come of it unless the scholar is an **active** reader and listener and unless the scholar assigns him/her time to develop his/her own ideas, too. It is impossible to ``finish a literature review and then start research." New literature is always appearing, and as the scholar's depth and breadth increases, he/she will continually see new connections and related areas that must be studied. Active listening and reading must be viewed as ``continuing education" that will involve the scholar for the rest of his/her career. The scholar should not be fooled into thinking it must be finished before he/she can begin research.

Once the problem is formulated, a brief summary of it should be written down. It is compulsory for a scholar writing a thesis for a Ph.D. degree to write a synopsis of the topic and submit it to the Doctoral Committee or the Research Board for approval. At this juncture the scholar should undertake extensive literature survey connected with the problem. For this purpose, the abstracting and indexing journals and published or unpublished bibliographies are the first place to go to. Academic journals, conference proceedings, government reports, books etc., must be tapped depending on the nature of the problem. In this process, it should be remembered that one source will lead to another. The earlier studies, if any, which are similar to the study in hand, should be carefully studied.

1.4.2.3 Choosing an Idea

From reading, interacting with the advisor during independent study, or work on a research assistantship, some possible projects will emerge. The scholar should make a list of open problems and possible projects that are of interest to him/her, and discuss them with potential advisors.

1.4.2.4 Stay Active

Even after the scholar has decided on his/her initial focus, it is important to continue a routine of reading new journals and technical reports and attending seminars. All of these sources can contribute to the development of the scholar's idea.

At this stage the scholar can add one question to the canonical list: How can these ideas help the scholar solve his/her research problem?

The scholar should remember that often the initial idea is quite far from the final thesis topic. If the scholar remains active in reading and listening, it will be much easier to generate alternative topics if the time comes.

1.5 Measure of Good Research

Whatever may be the types of research works and studies, one thing that is important is that they all meet on the common ground of scientific method employed by them. One expects scientific research to satisfy the following criteria: II

1. The purpose of the research should be clearly defined and common concepts be used.
2. The research procedure used should be described in sufficient detail to permit another researcher to repeat the research for further advancement, keeping the continuity of what has already been attained.
3. The procedural design of the research should be carefully planned to yield results that are as objective as possible.

4. The researcher should report with complete frankness, flaws in procedural design and estimate their effects upon the findings.
5. The analysis of data should be sufficiently adequate to reveal its significance and the methods of analysis used should be appropriate. The validity and reliability of the data should be checked carefully.
6. Conclusions should be confined to those justified by the data of the research and limited to those for which the data provide an adequate basis.
7. Greater confidence in research is warranted if the researcher is experienced, has a good reputation in research and is a person of integrity.

In other words, we can state the qualities of a good research is as under:

1. Good research is systematic: It means that research is structured with specified steps to be taken in a specified sequence in accordance with the well defined set of rules. Systematic characteristic of the research does not rule out creative thinking but it certainly does reject the use of guessing and intuition in arriving. at conclusions.

2. Good research is logical: This implies that research is guided by the rules of logical reasoning and the logical process of induction and deduction are of great value in carrying out research. Induction is the process of reasoning from a part to that; whole whereas deduction is the process of reasoning from some premise to a conclusion which follows from that very premise. In fact, logical reasoning makes research more meaningful in the context of decision making.
3. Good research is empirical: It implies that research is related basically to one or more aspects of a real situation and deals with concrete data that provides a basis for external validity to research results.
4. Good research is replicable: This characteristic allows research results to be verified by replicating the study and thereby building a sound basis for decisions.

1.6 Common Problems for Research Scholars in Engineering

Research Scholars, particularly those engaged in engineering research, are facing several problems. Some of the important problems are as follows:

1. The lack of a scientific training in the methodology of research is a great impediment for research scholars across the world. There is paucity of competent research Advisors/Guides in engineering sciences. Many research scholars in engineering sciences take a leap in the dark without knowing formal research methods required for engineering research. Most of the work, which goes in the name of research, is not methodologically sound. Many research scholars do research by cut and paste methods without any insight shed on the collated materials. The consequences are obvious; quite often the research resultsdo not reflect the realities. Thus, a systematic study of research methodology is an urgent necessity. Before undertaking research projects, researchers should be well equipped with all the methodological aspects. As such, efforts should be made to provide short- duration intensive courses for meeting this requirement.

2. There is insufficient interaction between the university research departments on one side and industries and research institutions on the other side. A great deal of primary

data of non-confidential nature remains untouched/ untreated by the researchers for want of proper contacts. Efforts should be made to develop satisfactory liaison among all concerned for better and realistic researches. There is need for developing some mechanisms of a university-industry interaction programmes so that academics can get ideas from practitioners on what needs to be researched and practitioners can apply the research done by the academics.

3. Most of the industrial units do not have the confidence that the material supplied by them to researchers will not be misused and as such they are often reluctant in supplying the needed information to researchers. The concept of secrecy seems to be sacrosanct to business organizations so much so that it proves an impermeable barrier to researchers. Thus, there is the need for generating the confidence that the information/data obtained from a business unit will not be misused.

4. Research studies overlapping one another are undertaken quite often for want of adequate information. This results in duplication and fritters away resources. This problem can be solved by proper compilation and revision, at regular intervals, of a list of subjects on which and the places where the research is going on in engineering sciences. Due attention should be given toward identification of research problems in various disciplines of engineering which are of immediate concern to the industries.

5. There does not exist a code of conduct for researchers and inter-university and inter-departmental rivalries are also quite common, Hence, there is need for developing a code of conduct for researchers which, if adhered sincerely, can win over this problem.

6. There is also the difficulty of timely availability of published research data from various government and other agencies doing this job. Researcher also faces the problem on account of the fact that the published research data vary quite significantly because of differences in coverage by the concerning agencies.

2 OVERVIEW OF THE THEORY OF SCIENCE AND HISTORY OF SCIENTIFIC RESEARCH

INTRODUCTION

In the 21th century we live in a world that is more scientific then ever in history of humanity. The number of active scientists today exceeds considerably the total number of all scientists in the past. Our civilization is permeated by science. Science has become in one way or another a part of every humans life, deeply integrated in our culture.

What is then this thing called science?

We hear that something is scientifically proved. It sounds like its true. What does true mean? What is scientific truth? How is science made? How is it formulated? How is it communicated? What is communication? How do we use different languages?

There is a red thread in this course and that is critical thinking. We are going to use critical thinking as method when approaching science, and we are going to think (critically!) about critical thinking.

Last but not least we take a look at science and scientists from the ethical point of view. Just not to forget that science is made by humans, and indeed for humans.

2.1 WHAT IS SCIENCE?

"Scientists are people of very dissimilar temperaments doing different things in very different ways. Among scientists are collectors, classifiers and compulsive tidiers-up; many are detectives by temperament and many are explorers; some are artists and others artisans. There are poet-scientists and philosopher-scientists and even a few mystics."

-Peter Medawar. Pluto's Republic, Oxford University Press, NY, 1982, p. 116.

Science is an extremely complex phenomenon, and difficult if not entirely impossible to define in a simple way. Here is an attempt to determine science by goal and process as well as by contrast (i.e. by defining what is not science).

2.1.1 Definitions by Goal and Process

Science (Lat. scientia, from scire, “to know”) is wonder about nature. Like philosophy, science poses questions - but also has the means to answer them, as long as they concern the state and behavior of the physical world.

Science is the systematic study of the properties of the physical world, by means of repeatable experiments and measurements, and the development of universal theories that are capable of describing and predicting observations. Statements in science must be precise and meaningful, such that other people can test them (in order to establish “universality”).

Science is

- ✓ the systematic observation of natural events and conditions in order to discover facts about them and to formulate laws and principles based on these facts.
- ✓ the organized body of knowledge that is derived from such observations and that can be verified or tested by further investigation.
- ✓ any specific branch of this general body of knowledge, such as biology, physics, geology, or astronomy.

- Academic Press Dictionary of Science & Technology

Science involves more than the gaining of knowledge. It is the systematic and organized inquiry into the natural world and its phenomena. Science is about gaining a deeper and often useful understanding of the world.

- Multicultural History of Science page at Vanderbilt University

Science alone of all the subjects contains within itself the lesson of the danger of belief in the infallibility of the greatest teachers in the preceding generation…As a matter of fact, I can also define science another way: Science is the belief in the ignorance of experts.

-Richard Feynman, Nobel-prize-winning physicist, in The Pleasure of Finding Things out

2.1.2 Definitions by Contrast

To do science is to search for repeated patterns, not simply to accumulate facts.

- Robert H. MacArthur, Geographical Ecology

Religion is a culture of faith; science is a culture of doubt.

- Richard Feynman

I think that we shall have to get accustomed to the idea that we must not look upon science as a “body of knowledge”, but rather as a system of hypotheses, or as a system of guesses or anticipations that in principle cannot be justified, but with which we work as long as they stand up to tests, and of which we are never justified in saying that we know they are “true”.

- Karl R. Popper, The Logic of Scientific Discovery

2.2 WHAT IS “THE SCIENTIFIC METHOD”?

If there is anything that can characterize science then it must be its method. However, there are many people, both philosophers of science and scientists who would resolutely deny the existence of anything that could be called “the scientific method”. Let us despite everything take the risk and try to define a common methodological picture that can be applied to science in general as it, even if it never can be perfect, can anyway serve as a helpful model.

The scientific method might be seen as the logical scheme used by scientists searching for answers to the questions posed within science, as well to formulate theories as to assure the means for producing them (instruments, tools, algorithms). The simple version might look something like this (see Figure 1):

1 Pose the question in the context of existing knowledge (theory & observations). (It can be a new question that old theories are capable of answering (usually the case), or the question that calls for formulation of a new preliminary theory).

2 Formulate a hypothesis as a tentative answer.

3 Deduce consequences and make predictions.

4 Test the hypothesis in a specific new experiment/theory field. The new hypothesis must prove to fit in the existing world-view (1, ´normal science´, according to Kuhn). In case the hypothesis leads to contradictions and demands a radical change in the existing theoretical background, it has to be tested particularly carefully. The new hypothesis has to prove fruitful and offer considerable advantages, in order to replace the existing scientific paradigm. This is radical redefining of the fundamental assumptions is called "scientific revolution" (Kuhn) and it happens very rarely. As a rule, the loop 2-3-4 is repeated with modifications of the hypothesis until the agreement is obtained, which leads to 5. If major discrepancies are found the process must start from the beginning, 1.

5 When consistency is obtained the hypothesis becomes a theory and provides a coherent set of propositions that define a new class of phenomena or a new theoretical concept. Theory at that stage is subject of process of natural selection among competing theories (6). A theory is then becoming a framework within which observations/theoretical facts are explained and predictions are made. The process can start from the beginning, but the state 1 has changed to include the new theory/improved old theory.

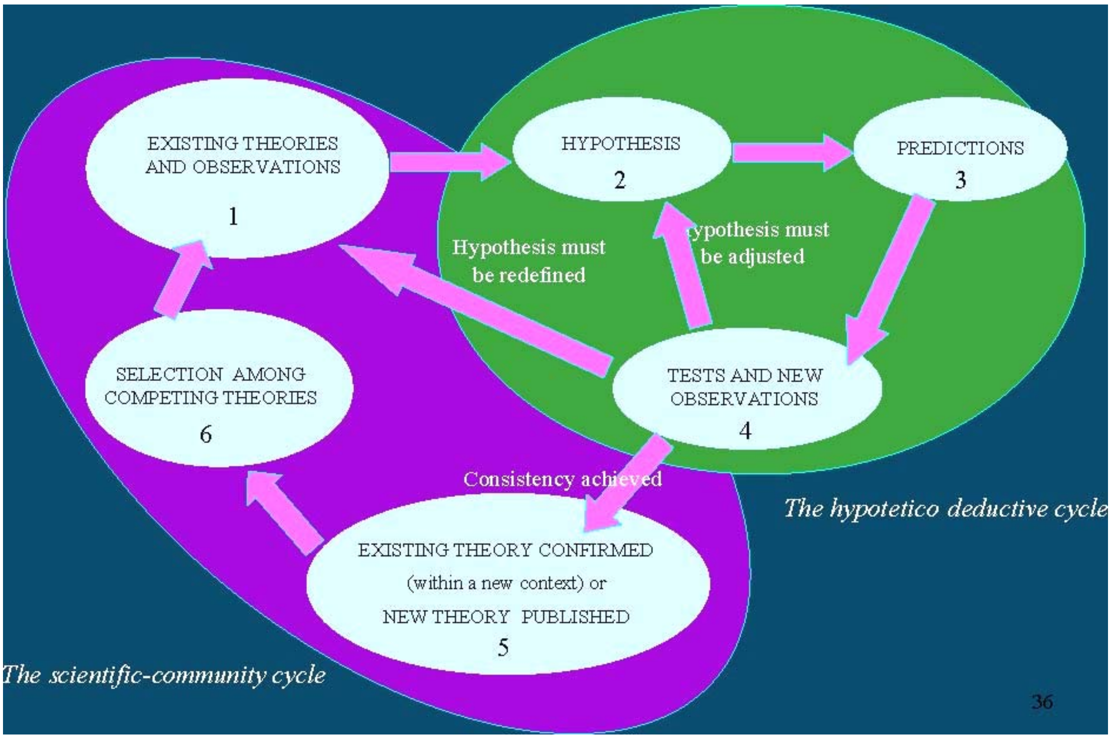


Figure 1 Diagram describing iterative nature of the scientific method (hypothetico-deductive)

Figure 1 describes the logical structure of scientific method in a very general scheme. As the flow diagram suggests, science is in a state of permanent change and development. The one of the most important qualities of science is its provisional character: it is subject to continuous re-examination and self-correction.

For an individual scientist it is of course possible to enter any phase of a scientific project. One can for example start by analyzing some experimental/observational results (4). However, it is crucial to understand that the logic of science is recursive: prior to every observation/experiment/theoretical test it is a hypothesis (2) that has its origins in the pre-existing body of knowledge (1). Every experimental/observational result has a certain world-view built-in. Or, to say it by Feyerabend, every experimental data is “theory-contaminated”.

Here it is also interesting to mention that designing new experimental equipment or procedure match the same scheme. So we proceed as follows: (1) Start from existing theoretical/experimental facts; (2) Formulate the problem; (3) Infer consequences; (4) Test if it works as expected; (5-6) Accept. As soon as a piece of equipment or method is designed and used as a tool for testing new hypotheses it is supposed that it works according to the design specification. The detailed information about its internal structure is therefore hidden.
The scheme of the scientific method in Figure 1 is without a doubt an abstraction and simplification. Critics of the hypothetico-deductive method would argue that there is in fact no such thing as “the scientific method”. By the term “the scientific method” they actually mean the set of rules defining how to proceed in posing new relevant questions and formulating successful hypotheses (1)→(2).

Indeed, the formulation of hypothesis (1)→(2) is a part of scientific work where we can accept Feyerabends formulation “anything goes”. (Educated) guess, chance, intuition, metaphysical concepts…anything can lead to formulating new hypothesis.

Essential however is that the new result/concept/hypothesis have to withstand tests and observations/experiments and moreover it has to survive the "natural selection" among competing theories/concepts/results. The reaction of the scientific community is decisive in this context. New results have to endure cross-examinations and show their correctness, usefulness, power and beauty.

Figure 1 leaves out many significant aspects of scientific work as for example its social nature. Modern science is an activity that often includes big groups of scientists that are communicating all over the world across national borders, sharing scientific results and ideas. The co-operation between scientists is indeed a crucial part of the scientific way of work.

Let us mention only the very evident fact that it is impossible for every scientist to independently do every experiment or derive every theorem to confirm every theory. Because time at our disposal is limited, scientists have to rely on other scientists. That is, the scientific community provides the support for individual scientists. One can also say that the body of scientific knowledge and thought belongs to scientific community and that individual scientists can contribute to the development of their scientific field only through the process of communication with the scientific community.

It is a tradition that scientific methods, results and ideas, within fundamental science at least, are the common heritage of humanity and they are shared without profit. They are general (universal) and unrestrained. In that respect, science belongs to very few fields of human activity where information is free. It is the open information exchange that constitutes the very backbone of modern science.

In the process of development of science, experiments/theories usually do get repeated as part of other experiments/theories. Most scientific papers contain suggestions for other scientists to follow up. Usually the first step in doing new research is to repeat earlier work. So there is a principle that a new theory has to reproduce results of old theory within the domain where the old theory has been demonstrated to work. This is basically the demand for the consistency between the old and the new theories. For example Einstein's special theory of relativity gives the Newtonian theory in the limit $v/c \rightarrow 0$ (i.e. where velocities are neglectable compared to speed of light).

The important advantage of the scientific method is that it is impartial:[2] one does not have to believe a given researcher, one can (in principle) repeat the experiment and determine whether certain results are valid or not. The conclusions will hold irrespective of the state of mind, or the religious persuasion, or the state of consciousness of the investigator. The question of impartiality is closely related to openness and universality of science, which are its fundamental qualities.

A theory is accepted based in the first place on the results obtained through logical reasoning, observations and/or experiments. The results obtained using the scientific method have to be reproducible. If the original claims are not verified, the causes of such discrepancies are exhaustively studied.

It is not always possible to repeat the studied event many times under controlled forms. There are for example only few large laboratories in the world running experiments in particle physics. Very complicated and sophisticated experiments are not so unique. Also, when studying certain astronomical objects like distant stars we cannot perform experiments; all information is thus obtained from observations and measurements. Theories are then devised by extracting some regularity in the observations and formulating this into physical laws (theories).

There is an important characteristic of a scientific theory or hypothesis that differentiates it from, for example, a religious belief: a scientific theory must be "falsifiable". This means that there must be some observation/experiment or another well-approved theory that could disprove the theory in question. For example, Einstein's theory of relativity made predictions about the results of experiments. These experiments could have produced results that contradicted Einstein, so the theory was (and still is) falsifiable.

Observe yet an important distinction (Feyerabend): "No theory ever agree with all facts in its domain, yet it is not always theory that is to blame. Facts are constituted by older ideologies, and a clash between facts and theories may be a proof of progress."

So it would be naïve to believe that single experimental discrepancy would be able to falsify any of well-established theories. As Kuhn describes the process, it is the act of revolution, and it calls for very dramatic changes in science, which means fundamental (and very expensive) re-building that is done only if there is no other solution to the problems but to abandon the established world-view ("paradigm").

A frequent criticism made of the scientific method is that it cannot accommodate anything that has not been proved. The argument then points out that many things thought to be impossible in the past are now everyday realities. This criticism is based on a misinterpretation of the scientific method.

All scientific truths are provisional. But for a hypothesis to get the status of a theory it is necessary to win the confidence of the scientific community. In the fields where there are no commonly accepted theories (as e.g. explanation of the process of creation of the universe-where the "big bang" hypothesis is the most popular one) the number of alternative hypotheses can constitute the body of scientific knowledge.

2.2.1 Socratic Method vs. Scientific Method

As we already pointed out, science is developing during history. It is therefore instructive to compare the classic Socratic Method (based as a rule on "thought experiments") and the modern Scientific Method. A very similar structure is found comparing those two!

Table 1 Socratic Method vs. Scientific Method

Socratic Method	Scientific Method
1. Wonder. Pose a question (of the "What is X ?" form).	1. Wonder. Pose a question. (Formulate a problem).

2. Hypothesis. Suggest a plausible answer (a definition) from which some conceptually testable hypothetical propositions can be deduced.	2. Hypothesis. Suggest a plausible answer (a theory) from which some empirically testable hypothetical propositions can be deduced.
3. Elenchus ; "testing," "refutation," or "cross-examination." Perform a thought experiment by imagining a case which conforms to the definiens but clearly fails to exemplify the definiendum, or vice versa. Such cases, if successful, are called counterexamples. If a counterexample is generated, return to step 2, otherwise go to step 4.	3. Testing. Construct and perform an experiment, which makes it possible to observe whether the consequences specified in one or more of those hypothetical propositions actually follow when the conditions specified in the same proposition(s) pertain. If the test fails, return to step 2, otherwise go to step 4.
4. Accept the hypothesis as provisionally true. Return to step 3 if you can conceive any other case which may show the answer to be defective.	4. Accept the hypothesis as provisionally true. Return to step 3 if there are predictable consequences of the theory which have not been experimentally confirmed.
5. Act accordingly.	5. Act accordingly.

In order to pose the question and formulate hypothesis (Figure 1, (1)→(2)), we use different approaches:

1 Intuition
2 Analogy
3 Paradigm
4 Metaphor

Last but not least: a very important factor leading to many scientific discoveries is chance!

2.2.2 Criteria to Evaluate Theories

When there are several rivaling hypotheses number of criteria can be used for choosing the top candidate for a best theory (Figure 1, (6)). Following can be evaluated:

Theoretical scope
Appropriateness
Heuristic value
Validity
Parsimony (Simplicity, see Ockham's razor in Dictionary)

The theory must necessarily fulfill the following criteria
Logically consistent

Consistent with accepted facts
Testable
Parsimonious
Consistent with related theories
Interpretable: explain and predict
Pleasing to the mind (Esthetic, Beautiful)
Useful (Applicable)

We can conclude that the process of asking a relevant question and formulating a successful hypothesis/theory is essentially not different from any other creative process like writing music, or poetry or performing art. That specific part of the working process within science has lead Feyerabend and other critics to claim the non-existence of a scientific method.

What is crucial for science however is that the hypothesis (theory), must be thoroughly investigated and either fit into the existing framework (normal science) or lead to thorough reformulation of the existing scientific paradigm (scientific revolution).

Let us have a look on the following table illustrating how the general context influences our focus, including the type of questions we ask as well as type of answers we search for.

Table 2 Focus of Classical Antiquity Science vs. Contemporary Science
Dominant Characteristics

Science of Classical Antiquity	Contemporary Science
why?	how?
eternal	present
explain	predict/use
holism	reductionism
individual	groups/institutions
local	global
knowledge for its own sake	knowledge for application
reasoning	experiment

2.3 WHAT IS KNOWLEDGE?

Knowledge is justified, true belief. - Plato in dialogue Meno

The objective of the analysis of knowledge is to state the conditions that are necessary and sufficient for propositional knowledge: knowledge that such-and-such is the case. Propositional knowledge must be distinguished from non-propositional knowledge (tacit knowledge) that is the knowing how to do something.

2.3.1 Abstraction

Our knowledge is built out of concepts or notions, i.e. theoretical representations that are the abstract “image” of the objective world. The first step of abstraction is the immediate processing of sensory impression (visual, auditory, tactile…). The consciousness is a dynamic interplay of these two types of abstractions. Instant sensory representation is also mediated: that is, it is with human, cultured sense organs that we view nature. There is nothing like “direct” (unmediated) observation.

Concepts are abstractions. They are extracted from the concrete world and interconnection with everything else, idealized, concentrated and simplified. No abstraction exists in nature, outside of human consciousness, because everything outside of consciousness moves, changes, is immensely complex, interconnected, and in perpetual transition and movement.

It is however not to say that we cannot know matter. It simply means that our knowledge is not the same thing as matter. Indeed, there is nothing to know of matter in itself.

Our knowledge is given in the form of immediate, essential or theoretical perception, and it is indeed knowledge of the world beyond sensation.

If the conceptual abstractions by means of which we are able to grasp the immediate abstractions of sense perception are not to be found in the material world, but only in our heads, how do we understand the relation between these abstractions and the materiality of the world?

How do we explain the fact that we can build bridges which don't fall down even after thousands of trucks have been driven over them? When we conjecture things about what we can't actually see, is there any basis for doing so? And if X and Y see different things, is it sensible to talk about something that exists independently of X or Y seeing it?

The key idea is information hiding. In order to build a bridge we do not need to have an exact understanding of how the basic physical laws act in a bridge. It is sufficient to start with previous experience in building bridges.

2.3.2 The Origins of Knowledge

Conceptual knowledge dates back to the time when human beings emerged as a species. Pre-humans developed certain practices, ways of living, in which they took from nature what they needed and changed it to meet their needs, initially just by carrying it elsewhere, later by breaking it up and then putting pieces together in new ways. At this point humans are much like any other animal.

2.3.3 The Role of Science in Knowledge Creation

The major purpose of science is to develop laws and theories to understand, explain, predict, and control phenomena.

Science must have a distinct subject matter, a set of phenomena which serves as a focal point for investigation. The discovery of the underlying regularities among these phenomena yields empirical generalizations, laws, principles, and theories. Through this process, science produces knowledge of the world by establishing generalizations governing the behavior of the world. How does this process relate to the "scientific method"?

2.3.4 Knowledge and Objectivity

Knowledge can be stable and it can be in a state of transition. It may be available in the form of public beliefs shared by all, and it may reside in special individuals.

As Kuhn, Popper, and others have noted, observations are always interpreted in the context of an a priori knowledge. The history of science provides numerous examples of the fact that "what a man sees depends both upon what he looks at and also upon what his previous visual-conceptual experience has taught him to see".

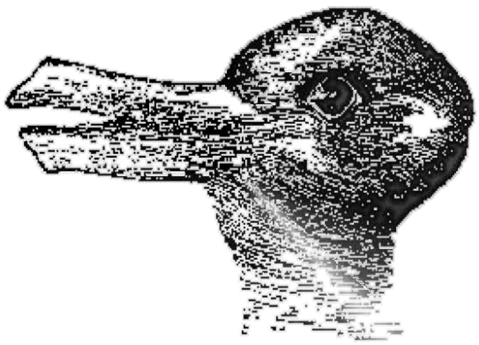
All observation is potentially ”contaminated”, whether by our theories, our worldview or our past experiences. Nevertheless, it does not mean that science cannot ”objectively”[6] choose from among rival theories on the basis of empirical testing. Obviously, if objectivity requires certainty (no possibility of error), then science is not objective. In science, all knowledge claims are tentative, subject to revision on the basis of new evidence. Although science cannot provide one with hundred percent certainty, yet it is the most, if not the only, objective mode of pursuing knowledge.

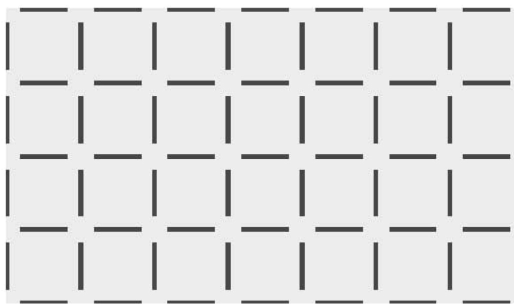
[6] What is here meant by ”objectively” does not refer to ”object” i.e. things as they are, but instead points out non-subjective nature that is the fact that it must be agreed upon (within scientific group or community).

2.3.5 Perception and "Direct Observation"

A common idea about the scientific method is that one makes observations, and then systematizes facts about the objects under investigation (see also Appendix 5). Here are some comments on "direct observations" and why it is impossible to perform experiments independent of theory (Mach, often cited by A Einstain at that point).

Conceptually and logically we always start from existing ideas (1) in Figure 1, when making experiments/observations.





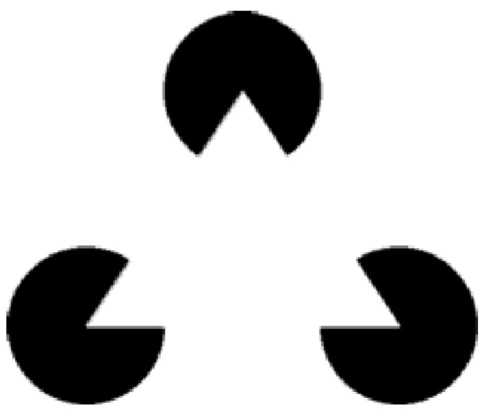


Figure 2 What do you see here?(A rabbit or a duck? Circles? A black vase or white profiles? Triangle?) Observing implies a variety of choices and characterizations. The description and the explanation depend on what aspect of a phenomenon we are focused on.

Take another (non-visual) example. Onomatopeic words used in different languages to describe the sounds made by familiar animals. What french, english, german or japanese hear that cat says is very different (Miaow? Mjau? …). What is actually the true thing cat says?

Some more "I see!" examples…

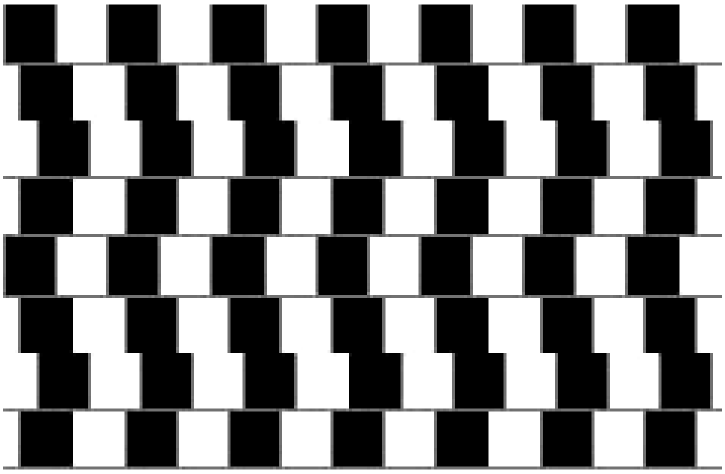


Figure 3 Parallel lines!

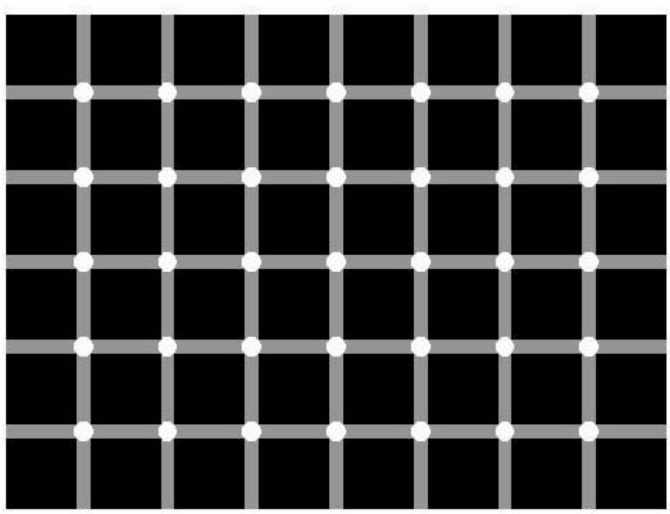
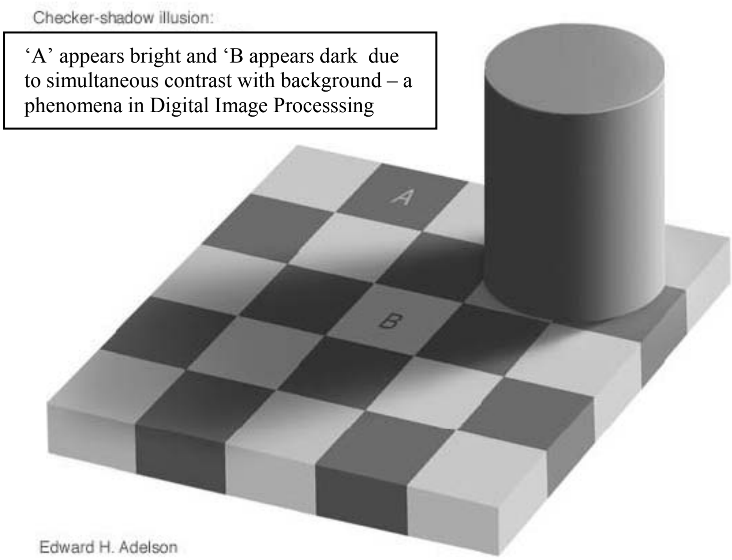


Figure 4 Count black dots. Look sharp! Here is yet another example:



So it is not obvious that our observations, not even the most direct ones are reliable…

Not to mention problems coupled with the complexity of experimental apparatus (do we really measure what we intend to?), measurement uncertainties (how good/reliable/precise are our results?), and alike. There is a specific scientific discipline called theory of measurement…so it is not an entirely trivial problem.

2.4 SCIENCE AND TRUTH

There are two conceptions of science that embody two different views of scientific life and of the purpose of scientific inquiry.

2.4.1 Science as Consensus

According to this approach, scientific knowledge is the product of a collective human enterprise to which scientists make individual contributions that are purified and extended by mutual criticism and intellectual cooperation. According to this theory the goal of science is a consensus of rational opinion over the widest possible field.

Scientific knowledge is distinguished from other intellectual artifacts of human society by the fact that its contents are consensible. This implies that scientific communication should be so transparent and unambiguous that the recipient is able either to give it approval or to offer well-founded objections. The goal of science, moreover, is to achieve the maximum consensus. Ideally the general body of scientific knowledge should consist of facts and principles that are firmly established and accepted without serious doubt, by a majority of competent, well-informed scientists.

The origin of the consensus in science may be found in the scientist's adherence to the logic of scientific inference. Sociologists argued that science exhibited so high a degree of agreement because scientists shared a set of norms or standards, which governed the professional life of the scientific community. Based upon the consensual view of science, science was thought to be strictly cumulative and converging.

2.4.2 Science as Controversy

Kuhn has, however, noted that in the consensual approach the emergence of new scientific ideas “requires a decision process which permits rational men to disagree, and such disagreement would generally be barred by the shared algorithm which philosophers have generally sought. If it were at hand, all conforming scientists would make the same decision at the same time.” Kuhn maintains that it is only the existence of different preferences and values among scientists that results in new theories. What makes the broad degree of agreement in science even more perplexing is the fact that the theories around which consensus forms do themselves rapidly come and go.

The rival theories are often radically incommensurable. The impossibility of full translation between rival paradigms is further enhanced by the fact that the advocates of different paradigms often subscribe to different methodological standards and have non-identical sets of cognitive values.

Feyerabend has argued that many highly successful scientists have repeatedly violated the norms or canons usually called scientific. Specifically, Feyerabend believed that it is undesirable for scientists to ever reach consensus about anything. His ideal of science is the sort of endless questioning of fundamentals which one associates with pre-Socratic natural philosophy: nothing is taken as given, everything can reasonably be denied or affirmed.

For the supporters of this doctrine, scientific debate and disagreement is far more likely the “natural” state of science than consensus is. In this view, science is diverging.

2.4.3 Critical Thinking: Logical Argument

Reserve your right to think, for even to think wrongly is better than not to think at all.

- Hypatia[7], natural philosopher and matematician

What is an argument? An argument is a statement logically inferred from premises. Neither an opinion nor a belief can qualify as an argument.

How do we analyze the soundness of an argument? It is a good practice to begin a critical analysis of an argument by isolating the conclusion. By examining the conclusion we find the point of the argument. The next step is examining of premises that the argument rests on.

It is important to note that some premises can be implied and not stated explicitly within the argument. We might call these premises assumed premises or underlying assumptions. Here is yet another reason to reflect critically upon ideas. Many arguments would never be made if the arguer were forced to make explicit her assumptions; i.e. the underlying assumptions can be extremely erratic!

Not all the discussions contain arguments consisting of a supporting premises and a conclusion. Discussions often contain statements that are not part of the argument. Such statements can be included in order to give necessary background information in support of anargument.

Typically, arguments are understood as either being deductive or inductive. As long as the structure corresponds to the given forms, the argument will attain validity. Let us define validity as follows:

A valid argument is an argument that is built according to the demonstrated rules of construction.

Note that validity does not guarantee truth of an argument. For instance, even with false premises we can still have a valid argument.

What does it mean now that the conclusion necessarily follows from the premises? In the given examples, the premises, if true, force us to accept the conclusion. In a deductive argument the conclusion is irrefutable. If an argument is valid but the conclusion is wrong, we may describe the argument as possessing validity but not possessing soundness.

A sound argument is an argument that is valid in its construction and whose conclusion is true. A cogent argument is an argument that possesses validity, soundness, and is convincing.

2.4.4 Deduction

Typically, a deductive argument is defined as:

- constructed according to valid rules of inference
- the conclusion necessarily follows from the premises.

First, what is meant by "constructed according to valid rules of inference"? There are valid ways for constructing arguments. A well-constructed argument will support its conclusions and endure against criticism. What follows below are four of the more common models for constructing arguments:

MODUS PONENS

$$\begin{array}{l} P \supset Q \\ \underline{P} \\ \therefore Q \end{array}$$

Let P represent "All humans" and let Q represent "are mortal." The horseshoe symbol can be read as "then." The three dots symbol can be read as "thus." A Modus Ponens argument would be formulated as follows:

All humans are mortal. (premise)
Kevin is human. (premise)
Thus, Kevin is mortal. (conclusion)

MODUS TOLLENS

$$\begin{array}{l} P \supset Q \\ \underline{\sim Q} \\ \therefore \sim P \end{array}$$

Let P represent "All birds" and let Q represent "have wings." A Modus Tollens argument would be formulated as follows:

All birds have wings. (premise) Kevin has no wings. (premise)
Kevin is not a bird. (conclusion)

DISJUNCTIVE SYLLOGISM

$$\begin{array}{l} P \vee Q \\ \underline{\sim Q} \\ \therefore P \end{array}$$

HYPOTHETICAL SYLLOGISM

$$\begin{array}{l} P \supset Q \\ \underline{Q \supset R} \\ \therefore P \supset R \end{array}$$

Let P represent "The baby is a boy" and Q "The baby is a girl."A Disjunctive Syllogism argument would be formulated as follows:

The baby can either be a boy or a girl. (premise)
The baby is not a girl. (premise)
The baby is a boy. (conclusion)

Let P represent " If Karro is a terrier, Karro is a dog " and let Q represent " If Karro is a dog, Karro is a mammal." A Hypothetical Syllogism argument would be formulated as follows:

If Karro is a terrier, Karro is a dog. (premise) If Karro is a dog, Karro is a mammal. (premise) If Karro is a terrier, Karro is a mammal. (conclusion)

In order to build a valid deductive argument, we must follow the above patterns. There are of course many more rules, but we have listed only the most common ones.

2.4.5 Empirical Induction

In a simplified scheme, deductive inferences can be said to move from general statements to particular conclusions. Inductive inferences move from particular assertions to general conclusions.

Here is the generic form of an inductive argument:

Every A we have observed is a B.
Therefore, every A is a B.

Many of scientific hypotheses are formulated via induction. Consider the following:

Every instance of water (at sea level)[8] we have observed has boiled at 100° Celsius.
Therefore, all water (at sea level) boils at 100° Celsius.

It is important to note that even the best inductive argument will never offer 100% probability in support of its position. An inductive argument poses general statement in a way that it assumes a general domain for phenomena that are tested in some part of that domain.

The classical example is that by induction we know that sun rises every morning in the east. Our conclusion is based on experience that so was the case at least as long as people inhabit earth. But we are not aware of any law that would imply the necessity[9] of being so.

In other words, an inductive argument have no way to logically (with certainty) prove that:

the phenomenon studied do exist in general domain, and
that it continues to behave according to the same pattern.

Therefore we can say that the induction hypothesis is a tentative conclusion. Hypothesis supports working theories or models that are based on the collected evidence. These models, which are induction-based, are not described as certainty as they will no longerbe accepted as soon as one certain relevant counter-instance is found.

Counter-example 1

Perhaps the most well known counter-example was the discovery of black swans in Australia. Prior to the point, it was assumed that all swans were white. With the discovery of the counter-example, the induction concerning the color of swans had to be re-modeled.

[8] Note the text ”at sea level”.

A typical problem with inductive argument is that it is formulated generally, while the observations are made under specific conditions. We could add here ”in an open vessel” as well.

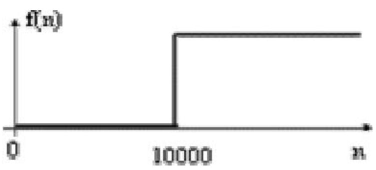
[9] Law of gravitation tells us how the celestial bodies (sun, earth) are placed and how they move, but nothing says that gravitation constant for example, can’t change!

Counter-example 2

Suppose the following are the results of measurements (observations): f(0) = 0, f(1) = 0, f(2) = 0, …f(10) = 0,…f(100)= 0, ….f(1000) = 0….f(5000) = 0…

Here we maybe conclude that the result is zero for every n! It is however possible that the law that f follows is of the form:

f(n) = n/10000 (integer division)



which agrees with our measurements, but contradicts our conclusion! (f(10000) = 1).

The question is: how much problem this can cause? For the most practical purposes this discrepancy is of no importance whatsoever as the parameter range of interest is far below, i.e. we focus on investigating parameters in the relevant range.

Talking about scientific method we pointed out the importance of thorough examination of the whole range of parameters of interest. It is clear that we never can examine all parameters in all possible ranges. But as far as we are sure that (in this case) range of $n \geq 10000$ is of no interest for us, we can accept the inductive conclusion, with no severe consequences.

In short, although induction cannot guarantee certainty, it can be a very useful method for constructing working models and theories. As long as we understand the limits of induction we may use induction productively.

Actually the induction is the only method we have to explore enlarged domains for phenomena under investigation.

3. OVERVIEW OF RESEARCH METHODOLOGY FOR ENGINEERING RESEARCH

3.1 WHAT IS AFTER ALL THIS THING CALLED SCIENCE?

The whole is more than the sum of its parts. - Aristotle, Metaphysica

Up to now an attempt has been made to define science. The definitions have been defined by goal and process and the definitions by contrast (i.e. what is not science). The scientific method was described and for searching the truth, the ancient Greek Socratic Method was compared with pure logical reasoning. The ideas of knowledge, truth, and objectivity have been problematized. In a short historic part the evolution of scientific thought was studied.

Let us now apply that given theoretical framework and try to describe sciences as specific branches of this general knowledge, and look for some regularities and patterns that can help us to get a general view of relationships between them.

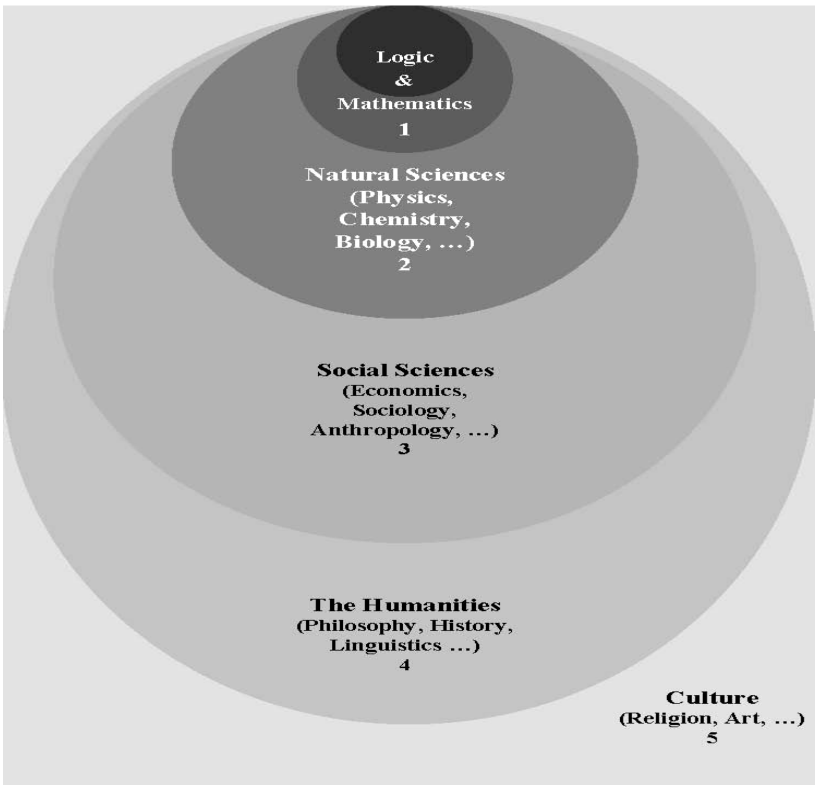


Table 3 Sciences, Objects and Methods

SCIENCE	OBJECTS	DOMINATING METHOD
	Simple	Reductionism (analysis)
Logic & Mathematics	Abstract objects: propositions, numbers, ...	Deduction
Natural Sciences	Natural objects: physical bodies, fields and interactions, living organisms ...	Hypothetico-deductive method
Social Sciences	Social objects: human individuals, groups, society, ..	Hypothetico-deductive method + Hermeneutics
Humanities	Cultural objects: human ideas, actions and relationships, language, artefacts…	Hermeneutics
	Complex	Holism (synthesis)

One more time we have to point out that the definition of science is not simple and unambiguous. For example, history and linguistics are often but not always catalogued as sciences.

The next thing we can realize from the scheme of Figure ? is that sciences have specific areas of validity. The logic and mathematics (the most abstract and at the same time the most exact sciences) are more or less important part of every other science. They are very essential for physics, less important for chemistry and even less for biology, and their significance continues to decrease towards the outer regions. The logical reasoning as a basis of all human reasoning is of course present in every kind of science as well as in philosophy.

We can imagine this as analogy of looking into a microscope. With the highest resolution we can reach the innermost regions. Inside those regions logic is not only the tool used to make conclusions. It is at the same time (together with mathematics) the object of investigation. Even though big parts of mathematics can be reduced to logical reasoning (Frege, Rusell and Whitehead) it is impossible to reduce the whole of mathematics to pure logic. On every next level of lower magnification the inner regions are given as prerequisites for the outer ones. Physics is using mathematics and logic as tools, without questioning their internal structure. In that way information about the deeper structure of mathematics and logic is hidden looking from the outside. In much the same way, physics is a prerequisite for chemistry, that is a hidden level inside biology etc.
The basic idea of Figure ? is to show in a schematic way the relation between the three main groups of sciences (Logic & Mathematics, Natural Sciences and Social Sciences) as well as the connections to thought systems represented by the Humanities. Finally the whole body of human knowledge, scientific and speculative is immersed in and impregnated by the cultural environment.

The transition between Social science and Humanities is not a sharp one. Moreover there are many connections between inner and outer regions of Figure ?, and the most important fact is that our cultural framework constitutes the basis that everything is built upon. For example metaphysics that is a philosophical discipline has a very strong coupling to physics.

In our scheme, the innermost sciences, Logic and Mathematics are the most fundamental ones and the ones with the highest degree of certainty. They have the most abstract and the simplest objects of investigation. Their language is the most formal one. They rely predominantly on the deductive method. From the existing axioms one is deducing theorems. It is however important to notice that the basic elements in both logic and mathematics have been extracted from our real-life language and purified into set of well defined formulae/symbolic expressions via an essentially inductive process. For instance, in order to define a triangle, one have to idealize (generalize, abstract) geometric shapes that can be found in nature, which is made by observing big enough number of such objects. Natural numbers have been defined similarly. The first step on the way to Peano axioms was to count concrete physical objects, as our fingers (probably the first thing used for counting), people, animals, etc.

The next region, Natural Sciences, is not an axiomatized theory as the previous one. Physics, which is the ideal of science for many philosophers of science (Popper, Carnap, Kuhn, Chalmers) contains both theoretical parts with pure mathematical formulations derived from "first principles" and parts that are empirical i.e. shortcut expressions for observed facts that are built-in into system as they proved useful. Far away from all theoretical physics can be axiomatized. Even less so is the case for experimental physics, for quite obvious reasons.

Natural sciences are dominated by an empirical method which Popper calls hypotetico-deductive method that we have already studied in the chapter about scientific method. Physics is a basis for chemistry, which is the basis for biology. It means that through information hiding mechanism physical levels are taken for granted in chemistry, while chemical levels (including physical levels (including mathematical (logical levels))) are taken for granted in biology. It is however instructive to notice that neither the whole of physics nor chemistry or biology can be derived from "first principles". So in biology for example we do not start from atoms to build up living organisms, but instead go in the opposite direction. Biologist starts from the macroscopic facts and proceeds towards more and more detailed microscopic levels.

Social Sciences include sociology, pedagogic, anthropology, economics etc. The objects studied are humans as social beings, alone or in a group. Social sciences rely primarily on the qualitative methods. The aim is to understand (hermeneutics) and describe phenomena. The quantitative aspects of their methodology are related to statistical methods.

The Humanities (The Liberal Arts) include philosophy, history, linguistics and similar. The difference between Humanities and Social Science is not a very sharp one, but we can say that Humanities predominantly have a qualitative approach, and very rarely depend on any statistical methods.

It is worth to point out that moving from the central towards the outer regions implies more complex objects as well as more holistic (synthetic) concepts. Language used in wider regions is more and more free, with all the ambiguities it implies.

Figure 10 represents a dynamic scheme seen in a specific moment. For example a corresponding scheme for the medieval age (compare to Table 4) would be very different. We can also remark that the culture is like a flow that all sciences follow. Albeit very slow, that flow steadily changes the framework for all the sciences.[16]

3.1.1 Sciences belonging to several different regions

Our scheme represents the classical groups of sciences. It says nothing about a modern type of science such as psychology. It gives us no clue whether to categorize psychoanalysis as "cultural heritage", some sort of fine art or a science. The answer is not unambiguous!

The development of human thought parallel to the development of human society has led to a emergence of sciences that do not belong to any of the classic types we have described earlier, but rather share common parts with several of these.

Computer science for example includes the field of artificial intelligence that has its roots in mathematical logic and mathematics but uses physics, chemistry and biology and even has parts where medicine and psychology are very important.

Moreover, we seem to be witnessing an exciting paradigm shift [21]:

"We should, by the way, be prepared for some radical, and perhaps surprising, transformations of the disciplinary structure of science (technology included) as information processing pervades it. In particular, as we become more aware of the detailed information processes that go on in doing science, the sciences will find themselves increasingly taking a meta position, in which doing science (observing, experimenting, theorizing, testing, archiving,) will involve understanding these information processes, and building systems that do the object-level science. Then the boundaries between the enterprise of science as a whole (the acquisition and organization of knowledge of the world) and AI (the understanding of how knowledge is acquired and organized) will become increasingly fuzzy."

Many of the modern sciences are of interdisciplinary, "eclectic" type. It is a trend for new sciences to search their methods and even questions in very broad areas. It can be seen as a result of the fact that the communications across the borders of different scientific fields is nowadays much easier and more intense than before. Here we can find a potential of the new synthetic (holistic) worldview that is about to emerge in the future.

The opposite is of course also true: there is a two-way communication between culture and science.

3.2 FUSION OF SCIENCE, RESEARCH, TECHNOLOGY

3.2.1 Aristotle's Distinctions between Science and Technology

In his famous reflections on science and technology, Aristotle has identified four key distinctions that are frequently quoted until today, giving them a kind of common sense status. It is therefore instructive to try to analyze modern science and technology in the light of the original formulations of Aristotle.

3.2.2 Objects: Unchangeable vs. Changeable

According to Aristotle's first distinction, science (episteme) is about the unchangeable, while technology (techne) is about the changeable.

In order to illustrate one possible interpretation, we take chemistry as an example. The change of chemical substances by chemical reaction is the very essence of chemistry. So, chemistry would be a kind of technology according to this criterion. But the same would be for high-energy physics, cosmology, biology, geology, and so on. Nearly all of modern sciences are about changeable things.

In fact, the same is true for Aristotle's own science of nature (episteme) that is also about changeable things, in contrast to mathematics and his 'theology', which are about the unchangeable.

This apparent contradiction may be solved only by distinguishing between two types of objects, the concrete empirical objects changeable by nature, and their abstractions including principles and patterns of change that remain constant.

We can also say that science is studying the behavior of substances that are unchangeable, in our time-perspective[17] at least, such as elements, atoms, elementary particles, living species, etc. Objects of technology are steadily changing like computers (including their constituent hardware and software) in computer technology.

3.2.3 Principles of Change: Inherent vs. Imposed

Aristotle made a distinction between theoretical science (theoretike) and practical science (poietike) with regard to the different principles of motion of their objects'. The objects of Aristotle's physics (as a theoretical science) bear their own principle of motion, i.e., they are moved or generated by their own inherent forces.

The objects of techne or practical science, on the other hand, have their principle of motion outside of the object, i.e., they are moved or generated by the technician's activity, according to the aims that may change from day to day.

Experimental chemistry has accordingly the status of a practical science, since most of its objects motions are generated by chemists. Chemists lose their status as scientists of nature by the intended intervention, by disturbing the inherent principle of motion of their objects.

An objection can be made here that even elementary forces of nature are (very slowly, but anyway) changing according to modern physical theory. However, the metaphysical distinction based on the concept of principle of motion has been dropped since the dawn of mechanistic philosophy. Physicists do not care, whether their objects are of artificial or natural origin as long as these are describable in terms of physical laws. While motions may be explained in terms of causes, Aristotle's principle of motion has lost any meaning in science. Secondly, and more important, generating phenomena in artificial contexts is the very essence of the experimental method.

Aristotle's concept of theoretical science is the 'spectator notion of knowledge' (Dewey). This has been discarded in most sciences for centuries - in chemistry/alchemy, for millennia.

Hence, since according to this Aristotle's criterion all experimental science would be technology, we should better abandon it today.

3.2.4 Ends: General vs. Specific

Starting from sensations of concrete things, science finally aims at general knowledge, whereas technology applies general knowledge back to concrete things.

The latter seems to be again a distinguished characteristic of preparative chemistry. Chemists apply general chemical knowledge to change concrete material samples. But our modern concept of science is methodologically different to Aristotle's concept.

Since Bacon Aristotelian inductivism has lost its prestige. General knowledge always remains preliminary knowledge, subject to empirical falsification. As soon as a general law is born, we try to apply it to new experimental arrangements, i.e., to concrete things, in order to test or to exhaust the law.

Moreover, once a general law has been established, it is used for new instrumental skills. It is actually hard to imagine modern science simply stopping at a certain general law, without trying to apply it.

3.2.5 Goals: End in Itself vs. End in Something Else

Scientists look for theoretical knowledge (theoria), that is an activity having an end in itself (and as such being a candidate for the highest form of happiness!). Technicians, on the other hand, produce new things, and such an activity has always an end in something else.

In other words: the purpose of scientific activity is just that activity itself, whereas technical activity is always good for something else.

Producing new chemical substances is a typical technical activity. Moreover, the production of new substances is usually intended to be good for something else. Consequently, preparative chemistry would be a technology par excellance. However, the aim of producing new substances is, for the most part, to improve the abilities of the field. Hence, though the individual practical activity has not an end in itself, it has an end in its own field: supporting the scientific community. Modern ('big') science is a complex network of cooperative research based on the division of labor that was quite unknown to Aristotle. If we do not consider that, substantial parts of modern science would again be characterized as technology.

It should be emphasized, that ordinary distinctions between different 'qualities' of aims, say, between 'practical use' and 'understanding the world', do not lead back to Aristotle's point. As soon as a certain activity is instrumentalized, which is necessarily the case in cooperative research of 'big science', it definitely loses its character of having an end in itself (and misses Aristotle's ideal of happiness).

In the light of modern “big science”, we might be willing to reformulate Aristotle's distinction: if the activity aims to contribute to its own field, it is science; if the end is outside of the field, it is technology. But even that can run into trouble, since, first, we have interdisciplinary research undermining the field distinction. And, secondly, scientific fields are usually heterogeneous and intricate with regard to ends; some, for sure, always get outside of the field.

So we may conclude that Aristotle's four distinctions between science and technology, though frequently repeated in various combinations until today, fail, because the structure of science has basically changed since the ancient Greek. Today's science is:

- (1) about changeable things (revealing general, unchangeable laws, however)
- (2) very often experimental science, or based on experimental science
- (3) following different methodology, and
- (4) 'big science' in the sense of complex research cooperation based on the division of labor.

3.2.6 Recent Methodological Distinctions

More recent philosophers have claimed some further methodological differences between science and technology.

3.2.6.1 Abstracting vs. Modeling Complex Systems

Scientists, it is said, use abstracting methods to discover general laws. Engineers, on the other hand, apply multi-factorial models to approach concrete and complex systems.

Yet, that sounds rather like a distinction between physics and chemistry! In fact, chemists are concerned with developing multi-factorial models to approach the concrete and complex systems of chemical substances. Every substance is different, and it is just the difference that chemists are interested in. But the same kind of approach can be found in biology, mineralogy, geology, meteorology, and the rest.

What is wrong with this distinction is that it mixes up characteristic of theoretical physics with science in general.

3.2.6.2 Conceptualizing vs. Optimizing Processes

It is sometimes stated that scientist try to understand a process in principle, while engineers try to optimize the process for certain aims.

Even that distinction runs into trouble when applied to chemistry. In fact, chemists are very busy to optimize the yield of their preparation according to theoretical standards. The higher the yield, the better the scientific result. The reason behind that is quite simple: Modern science has a cooperative structure. The outcome of a certain research serves as a means for later research and is, consequently, subject to optimizing.

Optimizing strategies are not restricted to chemistry. You can find the same even in physics, especially in solid-state physics. For instance in the field of superconductivity the physics Nobel Prize was awarded for materials exhibiting the superconductivity at until then highest temperature.

So again this distinction fails because of a one-sided conception of science.

3.2.6.3 Meta-methodological Distinction: Discovery vs. Invention

An old dispute among philosophers is whether scientific innovations should be named 'discovery' or 'invention'. According to the discovery-paradigm, chemists discover new changeability, e.g., that substance A changes to B under certain conditions. According to the invention paradigm, chemists invent new procedure, e.g., producing substance B from A under those conditions.

Similar approach is used to distinguish between science and technology: science discovers, technology invents.

Accordingly, there are two different types of results: science produces law-like statements, while technology produces rule-like statements.

Even this criterion applied in practice works poorly. It is related to two opposite philosophical assumptions:

- empiricists prefer the discovery paradigm, so they reconstruct any activity as science
- constructivist prefer the invention paradigm, so they see the same activity as technology.

Searching for empirical evidence, we can analyze the linguistic form of scientific reports. There are indeed some perceptive linguistic studies on chemistry papers and textbooks giving the impression that the use of rule-like or law-like statements is just a temporarily fashion.

Rule-like and law-like statements seem to be pragmatically (though not syntactical) equivalent, in the sense that every rule-like statement is easily translated into a law-like statement, and vice versa, without losing information.

Standard sharp binary distinctions between science and technology seem to fail, because the underlying concepts of science are out-dated, one-sided or arbitrary. Today's science is much more complex and heterogeneous than science of the Aristotle's time, the fact that even many modern philosophers have difficulty to admit.

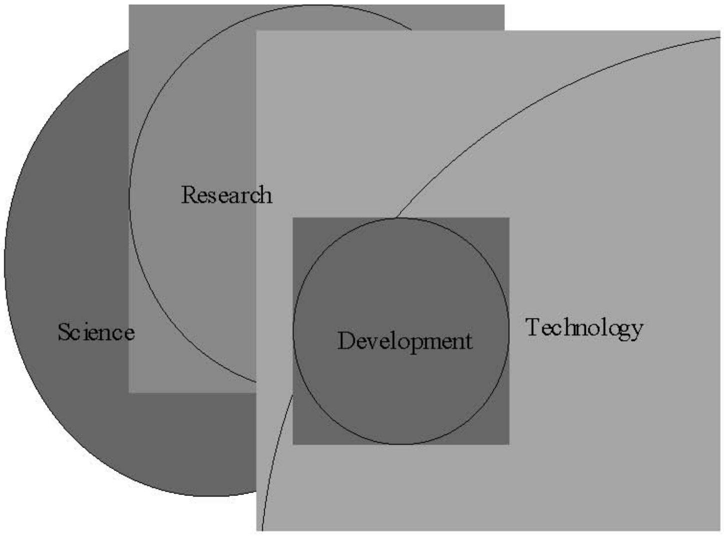
Philosophers of science have nearly neglected what Kuhn called the 'Baconian Sciences' covering more than 90% of today's science, i.e. experimental solid state physics and the whole of material sciences, molecular biology, genetics and genetical engineering, pharmacy and biomedical sciences, mineralogy, petrology and the major part of earth sciences, and so on. All these fields are more closely related to chemistry than to (the idealized philosophical image of) physics.

That is why philosophy of science is in vital need of a deeper, more realistic understanding of contemporary sciences. The time is ripe for paradigm change in philosophy of science!

Table 4 Standard Distinctions between Science and Technology

	Science	Technology
Object	unchangeable	changeable
Principle of motion	inside	outside

End	knowing the general	knowing the concrete
Activity	theoria: end in itself	poiesis: end in something else
Method	abstraction	modeling concrete (complex)
Process	conceptualizing	optimizing
Innovation form	discovery	invention
Type of result	law-like statements	rule-like statements
Time perspective	long-term	short-term



Mutual relations between science, research, development and technology

3.3 PSEUDOSCIENCE

A pseudoscience is a set of theories presented as scientific, while they are not. A theory about empirical facts is scientific if and only if it explains a range of observed phenomena and can be empirically tested in some meaningful way. Scientific testing involves deducing empirical predictions from the theory. To be meaningful, such predictions must, at least in theory, be falsifiable (Popper). A pseudoscientific theory claims to be scientific, but either the theory is not really falsifiable, or it has been falsified but its adherents refuse to accept that fact.

Pseudoscientists claim to base their theories on empirical evidence, and they may even use some elements of the scientific method, though often their understanding of a controlled experiment is inadequate. For example, "the truth of the hypothesis that plague is due to evil spirits is not established by the correctness of the deduction that you can avoid the disease by keeping out of the reach of the evil spirits."

There are several characteristics of pseudoscience that help us to identify them:

1 The tendency to propose theories that are put forth as scientific, but which cannot be empirically tested in any meaningful way. That is, the theory is consistent with every conceivable empirical event and no deduced prediction from it could ever falsify it.

2 The dogmatic refusal to give up an idea in the face of overwhelming evidence that the

idea is false, and the use of ad hoc hypotheses to try to explain away contrary evidence.

3 The selective use of data: the tendency to take into account only confirming instances and to ignore disconfirming instances.

4 The use of personal anecdotes as evidence.

5 The lack of concern over the absence of evidence in support of one's theory.

6 The use of myths or ancient mysteries to support theories which are then used to explain the myths or mysteries.

7 Naivety, especially about paranormal, supernatural or extraterrestrial claims.

3.4 SCIENCE AND ETHICS

3.4.1 The Conventionally Accepted Ethical Norms

R.K. Merton first suggested in 1942 that the behavior of scientists could be exemplified by a coherent set of norms.

Communalism requires that scientific knowledge should be public knowledge; that the results of research should be published; that there should be freedom of exchange of scientific information between scientists everywhere, and that scientist should be responsible to the scientific community for the trustworthiness of their published work.

Universalism requires that science be independent of race, color, or creed and that it should be essentially international.

Disinterestedness requires that the results of bona fide scientific research should not be manipulated to serve considerations such as personal profit, ideology, or expediency, in other words they should be honest and objective; it does not mean that research should not be competitive.

Organized skepticism requires that statements should not be accepted on the word of authority, but that scientists should be free to question them and that the truth of any statement should finally rest on a comparison with observed fact.".

C – communalism U – universalism D – disinterestedness O – originality OS - organized skepticism

3.4.2 Precautionary Principle

An ethical principle for guiding human activities, to prevent harm to the environment and to human health, has been emerging during the past ten years. It is called the "principle of precautionary action" or the "precautionary principle" for short.

An international group of scientists, government officials, lawyers, and grass-roots environmentalist met 1998 at Wingspread, Wisconsin to define and discuss the precautionary principle. The group issued the following consensus Wingspread Statement on the

Precautionary Principle:

"The release and use of toxic substances, the exploitation of resources, and physical alterations of the environment have had substantial unintended consequences affecting human health and the environment. Some of these concerns are high rates of learning deficiencies, asthma, cancer, birth defects and species extinctions, along with global climate change, stratospheric ozone depletion and worldwide contamination with toxic substances and nuclear materials.

We believe existing environmental regulations and other decisions, particularly those based on risk assessment, have failed to protect adequately human health and the environment - the larger system of which humans are but a part.

We believe there is compelling evidence that damage to humans and the worldwide environment is of such magnitude and seriousness that new principles for conducting human activities are necessary.

While we realize that human activities may involve hazards, people must proceed more carefully than has been the case in recent history. Corporations, government entities, organizations, communities, scientists and other individuals must adopt a precautionary approach to all human endeavors. Therefore, it is necessary to implement the Precautionary Principle:

"When an activity raises threats of harm to human health or the environment, precautionary measures should be taken even if some cause and effect relationships are not fully established scientifically. In this context the proponent of an activity, rather than the public, should bear the burden of proof. "

The process of applying the Precautionary Principle must be open, informed and democratic and must include potentially affected parties. It must also involve an examination of the full range of alternatives, including no action."

Thus, as formulated here, the principle of precautionary action has four parts:

1 People have a duty to take anticipatory action to prevent harm.
2 The burden of proof of harmlessness of a new technology, process, activity, or chemical lies with the proponents, not with the general public.
3 Before using a new technology, process, or chemical, or starting a new activity, people have an obligation to examine "a full range of alternatives" including the alternative of doing nothing.
4 Decisions applying the precautionary principle must be open, informed, and democratic and must include affected parties.

The precautionary principle is not really new. The essence of the principle is captured in common-sense aphorisms such as "An ounce of prevention is worth a pound of cure," "Better safe than sorry," and "Look before you leap."

However, environmental policy in the Asia, U.S. and Europe for the past 70 years has been guided by entirely different principles perhaps best reflected in the aphorisms, "Nothing ventured, nothing gained". The ethics comes into the picture when we understand that risks are as a rule taken (decided upon) by one group, and hazards are taken by some other group of people.

3.5 Science versus Engineering

3.5.1 Importance of Knowing What Engineering is all about

Engineering is the application of scientific and technical knowledge to solve human problems. Engineers are people who solve problems and focus on making things work more efficiently and effectively. Engineers apply the theories and principles of science and mathematics to research and develop economical solutions to technical problems. Engineers use imagination, judgment and reasoning to apply science, technology, mathematics, and practical experience. The result is the design, production, and operation of useful objects or processes. Their work is the link between perceived social needs and commercial applications.

The crucial and unique task of the engineer is to identify, understand, and integrate the constraints on a design in order to produce a successful result. It is usually not enough to build a technically successful product; it must also meet further requirements. Constraints may include available resources, physical or technical limitations, flexibility for future modifications and additions, and other factors, such as requirements for cost, marketability, producibility, and serviceability. By understanding the constraints, engineers derive specifications for the limits within which a viable object or system may be produced and operated.
Engineers design products, machinery to build those products, plants in which those products are made, and the systems that ensure the quality of the products and the efficiency of the workforce and manufacturing process. Engineers design, plan, and supervise the construction of buildings, highways, and transit systems. They develop and implement improved ways to extract, process, and use raw materials, such as petroleum and natural gas. They develop new materials that both improve the performance of products and take advantage of advances in technology. They harness the power of the sun, the Earth, atoms, and electricity for use in supplying the Nation's power needs, and create millions of products using power. They analyze the impact of the products they develop or the systems they design on the environment and on people using them. Engineering knowledge is applied to improving many things, including the quality of healthcare, the safety of food products, and the operation of financial systems.

3.5.2 Problem solving– Engineers! Know Yourself!!

“A scientist builds in order to learn,
but an engineer learns in order to build”
- *Anonymous*

Engineers use their knowledge of science, mathematics, and appropriate experience to find suitable solutions to a problem. Creating an appropriate mathematical model of a problem allows them to analyze it (sometimes definitively), and to test potential solutions. Usually multiple reasonable solutions exist, so engineers must evaluate the different design choices on their merits and choose the solution that best meets their requirements. Genrich Altshuller, after gathering statistics on a large number of patents, suggested that compromises are at the heart of "low-level" engineering designs, while at a higher level the best design is one which eliminates the core contradiction causing the problem.

Engineers typically attempt to predict how well their designs will perform to their specifications prior to full-scale production. They use, among other things: prototypes, scale models, simulations, destructive tests, nondestructive tests, and stress tests. Testing ensures that products will perform as expected. Engineers as professionals take seriously their responsibility to produce designs that will perform as expected and will not cause unintended harm to the public at large. Engineers typically include a factor of safety in their designs to reduce the risk of unexpected failure. However, the greater the safety factor, the less efficient the design may be.

3.5.3 Scientific Research vs. Engineering Research

> *You see things; and you say "Why?" But I dream things that never were; and I say "Why not?" —George Bernard Shaw*

Engineering is concerned with the design of a solution to a practical problem. A scientist may ask **why** a problem arises, and proceed to research the answer to the question or actually solve the problem in his first try, perhaps creating a mathematical model of his observations. By contrast, engineers want to know **how** to solve a problem, and **how** to implement that solution. In other words, **scientists attempt to explain phenomena, whereas engineers use any available knowledge, including that produced by science, to construct solutions to problems. This is no contradiction.**

There is an overlap between science (fundamental and applied) and engineering. It is not uncommon for scientists to become involved in the practical application of their discoveries; thereby becoming, for the moment, engineers. Scientists may also have to complete engineering tasks, such as designing experimental apparatus or building prototypes. Conversely, in the process of developing technology engineers sometimes find themselves exploring new phenomena, thus becoming, for the moment, scientists.

However, engineering research has a character different from that of scientific research. First, it often deals with areas in which the basic physics and/or chemistry are well understood, but the problems themselves are too complex to solve in an exact manner. **The purpose of engineering research is then to find approximations to the problem that can be solved.** Examples are the use of numerical approximations to the Navier-Stokes equations to solve aerodynamic flow over an aircraft, or the use of Miner's rule to calculate fatigue damage to an engineering structure. Second, engineering research employs many semi-empirical methods that are foreign to pure scientific research, one example being the method of parameter variation.

3.6 Distinct Perspective of Goals

3.6.1 Importance of Knowing How Engineering Research is done

The study of engineering research methodology gives the student the necessary training in gathering material and arranging them, participation in the field work when required, and also training in techniques for the collection of data appropriate to particular problems, in the use of statistics, questionnaires and controlled experimentation and in recording evidence, sorting it out and interpreting it. In fact, importance of knowing the methodology of research or how research is done stems from the following considerations:

(i) For one who is preparing himself for a career of carrying out research, the importance of knowing research methodology and research techniques is obvious since the same constitute the tools of his trade. The knowledge of methodology provides good training especially to the new research worker and enables him to do better research. It helps him to develop disciplined thinking or a 'bent of mind' to observe the field objectively. Hence, those aspiring for career in research must develop the skill of using research techniques and must thoroughly understand the logic behind them.

(ii) Knowledge of how to do research will inculcate the ability to evaluate and use research results with reasonable confidence. In other words, we can state that the knowledge of research methodology is helpful in various fields such as engineering, technology, applied science, government or business administration where persons are increasingly called upon to evaluate and use research results for action.

(iii) When one knows how research is done, then one may have the satisfaction of acquiring a new intellectual tool which can become a way of looking at the world and of judging every day experience. Accordingly, it enables us to make intelligent decisions concerning problems facing us in practical life at different points of time. Thus, the knowledge of research methodology provides tools to take at things in life objectively.

(iv) In this scientific and information knowledge age, all of us are in many ways consumers of research results and we can use them intelligently provided we are able to judge the adequacy of the methods by which they have been obtained. The knowledge of methodology helps the consumer of research results to evaluate them and enables him to take rational decisions.

3.6.2 Stages in Research Execution Process

A brief overview of the stages in research execution process is presented here as a precursor for general research methodology and techniques which are essential to effectively carry out research.

The research process consists of a number of closely related activities, which may overlap continuously rather than following a strictly prescribed sequence. At times, the first step determines the nature of the last step to be undertaken. If subsequent procedures have not been taken into account in the early stages, serious difficulties may arise which may even prevent the completion of the study. One should remember that the various steps involved in a research process are not mutually exclusive; nor they are separate and distinct; they do not necessarily follow each other in any specific order and the researcher has to be constantly anticipating at each step in the research process the requirements of the subsequent sets. However, the following order concerning various steps provides a useful procedural guideline regarding the research execution process:

1) **formulating the research problem**
2) **extensive literature survey**
3) **developing the hypothesis**
4) **preparing the research design**
5) **determining sample design**
6) **collecting the data**
7) **execution of the project**
8) **analysis of data**
9) **hypothesis testing**
10) **generalizations and interpretation**
11) **preparation of the research report**

Formulating the research problem and extensive literature survey has been discussed in chapter 1. Other topics are discussed briefly.

3. Development of working hypotheses: After extensive literature survey, researcher should state in clear terms the working hypothesis or hypotheses. Working hypothesis is tentative assumption made in order to draw out and test its logical or empirical consequences. As such the manner in which research hypotheses are developed is particularly important since they provide the focal point for research. The role of the hypothesis is to guide the researcher by delimiting the area of research and to keep him on the right track. It sharpens his thinking and focuses attention on the more important facets of the problem. It also indicates the type of data required and the type of methods of data analysis to be used.

How does one go about developing working hypotheses? The answer is by using the following approach:

(a) Discussions with colleagues and experts about. the problem, its origin and the objectives in seeking a solution;
(b) Examination of data and records, if available, concerning the problem for possible trends, peculiarities and other clues;
(c) Review of similar studies in the area or of the studies on similar problems; and
(d) Exploratory personal investigation which involves original field interviews on a limited scale with interested parties and individuals with a view to secure greater insight into the practical aspects of the problem.

4. Preparing the research design: The research problem having been formulated in clear cut terms, the researcher will be required to prepare a research design, i.e., he will have to state the conceptual structure within which research would be conducted. The preparation of such a design facilitates research to be as efficient as possible yielding maximal information. Research purposes may be grouped into four categories, viz., (i) Exploration, (ii) Description, (iii) Diagnosis, and (iv) Experimentation. A flexible research design which provides opportunity for considering many different aspects of a problem is considered appropriate if the purpose of the research study is that of exploration.

The preparation of the research design, appropriate for a particular research problem, involves usually the consideration of the following:

(i) the means of obtaining the information;
(ii) the availability and skills of the researcher and his staff (if any);
(iii) explanation of the way in which selected means of obtaining information will be organized and the reasoning leading to the selection;
(iv) the time available for research; and
(v) the cost factor relating to research, i.e., the finance available for the purpose.

5. Determining sample design:

The researcher must decide the way of selecting a sample or what is popularly known as the sample design. In other words, a sample design is a definite plan determined before any data are actually collected for obtaining a sample from a given population. Thus, the plan to select 12 of a city's 200 drugstores in a certain way constitutes a sample design. Samples can be either probability, samples or non-probability samples. With probability samples each element has a known probability of being included in the sample but the non-probability samples do not allow the researcher to determine this probability. Probability samples are those based on simple random sampling, systematic sampling, stratified sampling, cluster/area sampling whereas non-probability samples are those based on convenience sampling, judgement sampling and quota sampling techniques. A brief mention of the important sample designs is as follows:

(i) Deliberate sampling: Deliberate sampling is also known as purposive or non-probability sampling. This sampling method involves purposive or deliberate selection of particular units of the universe for constituting a sample which represents the universe. When population elements are selected for inclusion in the sample based on the ease of access, it can be called convenience sampling.

(ii) Simple random sampling: This type of sampling is also known as chance sampling or' probability sampling where each and every item in the population has an equal chance of inclusion in the sample and each one of the possible samples, in case of finite universe, has the same probability of being selected. For example, if we have to select a sample of 300 items from a universe of 15,000 items, then we can put the names or numbers of all the 15,000 items on slips of paper and conduct a lottery.

(iii) Systematic sampling: In some instances the most practical way of sampling is to select every 15th name on a list, every 10th house on one side of a street and so on. Sampling of this type is known as systematic sampling. An element of randomness is usually introduced into this kind of sampling by using random numbers to pick up the unit with which to start.

(iv) Stratified sampling: If the population from which a sample is to be drawn does not constitute a homogeneous group, then stratified sampling technique is applied so as to obtain a representative sample. In this technique, the population is stratified into a number of non- overlapping subpopulations or strata and sample items are selected from each stratum.

(v) Quota sampling: In stratified sampling the cost of taking random samples from individual strata is often so expensive that interviewers are simply given quota to be filled from different strata, the actual selection of items for sample being left to the interviewer's judgement. This is called quota sampling.

(vi) Cluster sampling and area sampling: Cluster sampling involves grouping the population and then selecting the groups or the clusters rather than individual elements for inclusion in the sample. Suppose some departmental store wishes to sample its credit card holders. It has issued its cards to 15,000 customers. The sample size is to be kept say 450. For cluster sampling this list of 15,000 card holders could be formed into 100 clusters of 150 card holders each. Three clusters might then be selected for the sample randomly. The sample size must often be larger than the simple random sample to ensure the same level of accuracy because is cluster sampling procedural potential for order bias and other source of error is usually accentuated. The clustering approach can, however, make the sampling procedure relatively easier and increase the efficiency of field work, specially in the case of personal interviews.

(vii) Multi-stage sampling: This is a further development of the idea of cluster sampling. This technique is meant for big inquiries extending to a considerably large geographical area like an entire country. Under multi-stage sampling the first stage may be to select large primary sampling units such as states, then districts, then towns and finally certain families within towns. If the technique of random-sampling is applied at all stages, the sampling procedure is described as multi-stage random sampling.

(viii) Sequential sampling: This is somewhat a complex sample design where the ultimate size of the sample is not fixed in advance but is determined according to mathematical decisions on the basis of information yielded as survey progresses. This design is usually adopted under acceptance sampling plan in the context of statistical quality control.

6. Collecting the data: In dealing with any real life problem it is often found that data at hand are inadequate, and hence, it becomes necessary to collect data that are appropriate. There are several ways of collecting the appropriate data which differ considerably in context of money costs, time and other resources at the disposal of the researcher.

Primary data can be collected either through experiment or through survey. If the researcher conducts an experiment, he observes some quantitative meas\1rements, or the data, with the help of which he examines the truth contained in his hypothesis.

But in the case of a **survey,** data can be collected by anyone or more of the following ways:

(i) By observation: This method implies the collection of information by way of investigator's own observation, without interviewing the respondents. The information obtained relates to what is currently happening and is not complicated by either the past behaviour or future intentions or attitudes of respondents.

(ii) Through personal interview: The investigator follows a rigid procedure and seeks answers to a set of pre-conceived questions through personal interviews. This method of collecting data is usually carried out in a structured way where output depends upon the ability of the interviewer to a large extent.

(iii) Through telephone interviews: This method of collecting information involves contacting I the respondents on telephone itself. This is not a very widely used method but it plays an important role in industrial surveys in 4eveloped regions, particularly, when the survey has to be accomplished in a very limited time.

(iv) By mailing of questionnaires: The researcher and the ,respondents do come in contact with each other if this method of survey is adopted. Questionnaires are mailed to the respondents with a request to return after completing the same. It is the most extensively used method in various economic and business surveys.

(v) Through schedules: Under this method the enumerators ~e appointed and given training. They are provided with schedules containing relevant questions. These enumerators go to respondents with these schedules. Data are collected by filling up the schedules by enumerators on the basis of replies given by respondents. Much depends upon the capability of enumerators so far as this method is concerned. Some occasional field checks on the work of the enumerators may ensure sincere work.

The researcher should select one of these methods of collecting the data taking into consideration the nature of investigation, objective and scope of the inquiry, financial resources, available time and the desired degree of accuracy..

7. Execution of the project: Execution of the project is a very important step in the research process. If the execution of the project proceeds on correct lines, the data to be collected would be adequate and dependable. The researcher should see that the project is executed in a systematic manner and in time.

8. Analysis of data: After the data have been collected, the researcher turns to the task of analyzing them. The analysis of data requires a number of closely related operations such as establishment of categories, the application of these categories to raw data through coding, tabulation and then drawing statistical inferences. The unwieldy data should necessarily be condensed into a few manageable groups and tables for further analysis, Thus, researcher should classify the raw data into some purposeful and usable categories. Coding operation is usually done at this stage through which the categories of data are transformed into symbols that may be tabulated and counted. Editing is the procedure that improves the quality of the data for coding. With coding the stage is ready for tabulation. Tabulation is a part of the technical procedure wherein the classified data are put in the form of tables. The mechanical devices can be made use of at this juncture.

Analysis work after tabulation is generally based on the computation of various percentages, coefficients, etc., by applying various well defined statistical formulae. In the process of analysis, relationships or differences supporting or conflicting with original or new hypotheses should be subjected to tests of significance to determine with what validity data can be said to indicate any conclusion(s).

9. Hypothesis-testing: After analyzing the data as stated above, the researcher is in a position to test the hypotheses, if any, he had formulated earlier. Do the facts support the hypotheses or they happen to be contrary? This is the usual question which should be answered while testing hypotheses. Various tests, such as Chi square test, t-test, F-test, have been developed by statisticians for the purpose. The hypotheses may be tested through the use of one or more of such tests, depending upon the nature and object of research inquiry. Hypothesis-testing will result in either accepting the hypothesis or in rejecting it.

10. Generalizations and interpretation: If a hypothesis is tested and upheld several times, it may be possible for the researcher to arrive at generalization, i.e., to build a theory. As a matter of fact, the real value of research lies in its ability to arrive at certain generalisations.1f the researcher had no hypothesis to start with, he might seek to explain his findings on the basis of some theory. It is known as **interpretation.** The process of interpretation may quite often trigger off new questions which in turn may lead to further researches.

11. Preparation of the report or the thesis: Finally, the researcher has to prepare the report of what has been done by him. Writing of report must be done with great care keeping in view the following: .

1. The layout of the report should be as' follows: (i) the preliminary pages; (ii) the main text, and (iii) the end matter.

   In its preliminary pages the report should carry title and date followed by acknowledgements and foreword: Then there should be a table of contents followed by a list of tables and list of graphs and charts, if any, given in the report.

The main text of the report should have the following parts:

(a) Introduction: It should contain a clear statement of the objective of the research and an explanation of ~e methodology adopted in accomplishing the research. The scope of the study along with various limitations should as well be stated in this part.
(b) Summary of findings: After introduction there would appear a statement of findings and recommendations in non-technical language. If the findings are extensive, they should be summarized.

(c) Main report: The main body of the report should be presented in logical sequence and broken-down into readily identifiable sections.
(d) Conclusion: Towards the end of the main text, researcher should again put down the results of his research clearly and precisely. In fact, it is the final summing up.

At the end of the report, appendices should be enlisted in respect of all technical data. Bibliography, i.e., list of books, journals, reports, etc., consulted, should also be given in the end. Index should also be given specially in a published research report.

2. Report should be written in a concise and objective style in simple language avoiding vague expressions such as 'it seems,' 'there may be', and the like.

3. Charts and illustrations in the main report should be used only if they present the information more clearly and forcibly.

4. Calculated' confidence limits' must be mentioned and the various constraints experienced in conducting research operations may as well be stated.

4. RESEARCH METHODOLOGY FOR CIRCUIT BRANCHES

4.1 Formulating the Research Problem

In research process, the first and foremost step happens to be that of selecting and properly defining a research problem. * A researcher must find the problem and formulate it so that it becomes susceptible to research. Like a medical doctor, a researcher must examine all the symptoms (presented to him or observed by him) concerning a problem before he can diagnose correctly; To define a problem correctly, a researcher must know what a problem is?

4.1.1 Understanding a Research Problem

A research problem, in general, refers to some difficulty which a researcher experiences in the context of either a theoretical or practical situation and wants to obtain a solution for the same. Usually we say that a research problem does exist if the following conditions are met with:

(i) There must be an individual (or a group or an organisation), let us call it *'I,'* to whom the problem can be ~attributed. The individual or the organization, as the case may be, occupies an environment, say' *N'*, which is defined by values of the uncontrolled variables, *Y.*

(ii) There must be at least two courses of action, say C1 and C2' to be pursued. A course of action is defined by one or more values of the controlled variables. For example, the number of items purchased at a specified time is said to be one course of action.

(iii) There must be at least two possible outcomes, say *O1.* and *02,* of the course of action, of which one should be preferable to the other. In other words, this means that there must be at least one outcome that the researcher wants, i.e., an objective.

(iv) The courses of action available must provide some chance of obtaining the objective, but they cannot provide the same chance, otherwise the choice would not matter.

Over and above these conditions, the individual or the organization can be said to have the problem only if *'r* does not know what course of action is best, must be in doubt about the solution. Thus, an individual or a group of persons can be said to have a problem which can be technically described as a research problem, if they (individual or the group), having one or more desired outcomes, are confronted with two or more courses of action that have some but not equal efficiency for the desired objective(s) and are in doubt about which course of action is best.

We can, thus, state the components of a research problem as under:

(i) There must be an individual or a group which has some difficulty or the problem.

(ii) There must be some objective(s) to be attained at. If one wants nothing, one cannot have a problem.

(iii) There must be alternative means (or the courses of action) tor obtaining the objective(s) one wishes to attain. This means that there must be *at least two means* available to a researcher for if he has no choice of means, he cannot have a problem.

(iv) There must remain some doubt in the mind of a researcher with regard to the selection of alternatives. This means th~t research must answer the question concerning the relative efficiency of the possible alternatives.

(v) There must be some environment(s) to which the difficulty pertains.

Thus, a research problem is one which requires a researcher to find out the best solution for the given problem, i.e., to find out by which course of action the objective can be attained optimally in the context of a given environment.

4.1.2 Issues with Choosing the Research Problem

The research problem undertaken for study must be carefully selected. The task is difficult one, although it may not appear to be so . Help may be taken from a research guide in this connection. A problem must spring from the researcher's mind like a plant springing from its own seed. If our eyes need glasses, it is not the optician alone who decides about the number of the lens we require. We have to see ourselves and enable him to prescribe for us the right number by cooperating with him. Thus, a research guide can at the most only help a researcher choose a subject. However, the following points may be observed by a researcher in selecting a research problem or a subject for research:

(i) Subject which is overdone should not be normally chosen, for it will be a difficult task to throw any new light in such a case.

(ii) Controversial subject should not become the choice of an average researcher.

(iii) Too narrow or too vague problems should be avoided.

(iv) The subject selected for research should be familiar and feasible so that the related research material or sources of research are witn\n one's reach. Even th~n it is quite difficult to. supply definitive ideas concerning how a researcher should obtain ideas for his research. For this purpose, a researcher should contact an expert or a professor in the University .who is already engaged in research. He may as well read articles published in current literature available on the subject and may think how the techniques and ideas discussed therein might be applied to the solution of other problems. He may discuss with others what he has in mind concerning a problem. In this way he should make all possible efforts in selecting a problem.

(v) The importance of the subject, the qualifications and the training of a researcher, the costs involved, the time factor are few other criteria that must also be considered in selecting a pro!>lem. In other words, before the final selection of a problem is done, a researcher must ask himself the following questions:
  (a) Whether he is well equipped in terms of his background to carry out the research?
  (b) Whether the study falls within the budget he can afford?
  (c) Whether the necessary cooperation can be obtained from those who must participate in research as subjects?

If the answers to all these questions are in the affirmative, one may become sure so far as the practicability of the study is concerned.

(vi) The selection of a problem must be preceded by a preliminary study. This may not be necessary when the problem requires the conduct of a research closely similar to one that has already been done. But when the field of inquiry is relatively new and does not have available a set of well developed techniques, a brief feasibility study must always bc undertaken.

If the subject for research is selected properly by observing the above mentioned points, the research will not be a boring drudgery, rather it will be love's labour.

4.1.3 Need for Defining the Research Problem

Quite often we all hear that a problem clearly stated is a problem half solved. This statement signifies the need for defining a research problem. The problem to be investigated must be defined unambiguously for that will help to discriminate relevant data from the irrelevant ones. A proper definition of research problem will enable the researcher to be on the track whereas an ill-defined problem may create hurdles. Questions like: What data are to be collected? What characteristics of data are relevant and need to be studied? What relations are to be explored. What techniques are to be used for the purpose? and similar other questions crop up in the mind of the researcher who can well plan his strategy and find answers to all such questions only when the research problem has been well defined. Thus, defining a research problem properly is a prerequisite for any study and is a step of the highest importance. In fact, formulation of a problem is often more essential than its solution. It is only on careful detailing the research problem that we can work out the research design and can smoothly carry on all the consequential steps involved while doing research.

4.1.4 Various steps involved in Defining the Research Problem

Let us start with the question: What does one mean when he/she wants to define a research problem? The answer may be that one wants to state the problem along with the bounds within which it is to be studied. In other words, defining a problem involves the task of laying down boundaries within which a researcher shall study the problem with a pre-determined objective in view.

How to define a research problem is undoubtedly a herculean task, However, it is a task that must be tackled intelligently to avoid the perplexity encountered in a research operation. The usual approach is that the researcher should himself pose a question (or in case someone else wants the researcher to carry on research, the concerned individual, organisation or an authority should pose the question to the researcher) and set-up techniques and procedures for throwing light on the question concerned for formulating or defining the research problem. But such an approach generally does not produce definitive results because the question phrased in such a fashion is usually in broad general terms and as such may not be in a form suitable for testing.

Defining a research problem properly and clearly is a crucial part of a research study and must I': in no case be accomplished hurriedly. However, in practice this a frequently overlooked which causes a lot of problems later on. Hence, the research problem should be defined in a systematic manner, giving due weightage to all relating points. The technique for the purpose involves the undertaking of the following steps generally one after the other: (i) statement of the problem in a general way; (ii) *i* understanding the nature of the problem; (iii) surveying the available literature (iv) developing the ideas through discussions; and (v) rephrasing the research problem into a working proposition.

A brief, description of all these points will be helpful.

(i) Statement of the problem in a general way: First of all the problem should be stated in a broad general way, keeping in view either some practical concern or some scientific or intellectual interest. For this purpose, the researcher must immerse himself thoroughly in the subject matter concerning which he wishes to pose a problem. The problem stated in a broad general way may contain various ambiguities which must be resolved by cool thinking and rethinking over the problem. At the same time the feasibility of a particular solution has to be considered and the same should be kept in view while stating the problem.

(ii) Understanding the nature of the problem: The next step in defining the problem is to understand its origin and nature clearly. The best way of understanding the problem is to discuss it with those who first raised it in order to find out how the problem originally came about and with what objectives in view. If the researcher has stated the problem himself, he should consider once again all those points that induced him to make a general statement concerning the problem. For a better understanding of the nature of the problem involved, he can enter into discussion with those who have a good knowledge of the problem concerned or similar other problems. The researcher should also keep in view the environment within which the-problem is to be studied and understood.

iii) Surveying the available literature: All available literature concerning the problem at hand must necessarily be surveyed and examined before a definition of the research problem is given. This means that the researcher must be well-conversant with relevant-theories in the field, reports and records as also all other relevant literature. He must devote sufficient time in reviewing of research already undertaken on related problems. This is done to find out what data and other materials, if any, are available for operational purposes. "Knowing what data are available often serves to narrow the problem itself as well as the technique that might be used". This would also help a researcher to know if there are certain gaps in the theories, or whether the existing theories applicable to the problem under study are inconsistent with each other, or whether the findings of the different studies do not follow a pattern consistent with the- theoretical expectations and so on. All his will enable a researcher to take new strides in the field for furtherance of knowledge i.e., he can move up starting from the existing premise.

IV) Developing the ideas through discussions: Discussion concerning a problem often produces useful information. Various new ideas can be developed through such an exercise. Hence, a researcher must discuss his problem with his colleagues and others who have enough experience in the same area or in working on similar problems. This is quite often known as an *experience survey.* People with rich experience are in a position to enlighten the researcher on different aspects of his proposed study and their advice and comments are usually invaluable to the researcher. They help him sharpen his focus of attention on specific aspects within the field. Discussions with such persons should not only be confined to the formulation of the specific problem at hand, but should also be concerned with the general approach to the given problem, techniques that might be used, possible solutions, etc.

v) Rephrasing the research problem: Finally, the researcher must sit to rephrase the research problem into a working proposition. Once the nature of the problem has been clearly understood, the environment (within which the problem has got to be studied) has been defined, discussions over the problem have taken place and the available literature has been surveyed and examined, rephrasing the problem into analytical or operational terms is not a difficult task. Through rephrasing, the researcher puts the research problem in as specific terms as possible so that it may become operationally viable and may help in the development of working hypotheses. *

In addition to what has been stated above, the following points must also be observed while defining a research problem:

(a) Technical terms and words or phrases, with special meanings used in the statement of the problem, should be clearly defined.
(b) Basic assumptions or postulates (if any) relating to the research problem should be clearly stated.
(c) A straight forward statement of the value of the investigation (i.e., the criteria for the selection of the problem) should be provided.
(d) The suitability of the time-period and the sources of data available must also be considered by the researcher in defining the problem.
(e) The scope of the investigation or the limits within which the problem is to be studied must be mentioned explicitly in defining a research problem.

4.1.5 A Case Study

4.2 Designing a Conceptual Structure for Carrying out Research

4.2.1 What is a Conceptual Structure for Research?

The formidable problem that follows the task of defining the research problem is the preparation of the design of the research project, popularly known as the "research design. Decisions regarding what, where, when, how much, by what means concerning an inquiry or a research study constitute a research design. "A research design is the arrangement of conditions for collection and analysis of data in a manner that aims to combine relevance to the research purpose~ with economy in procedure. In fact, the research design is the conceptual structure within which research is conducted; it constitutes the blueprint for the collection, measurement and analysis of data. As such the design includes an outline of what the researcher will do from writing the hypothesis and its operational implications to . the final analysis of data. More explicitly, the design decisions happen to be in respect of:

(i) What is the study about?
(ii) Why is the study being made?
(ill) Where will the study be carried out?
(iv) What type of data is required?
(v) Where can the required data be found?
(vi) What periods of time will the study include?
(vii) What will be the sample design?
(viii) What techniques of data collection will be used?
(ix) How will the data be analysed?
(x) In what style will the report be prepared?

Keeping in view the above stated design decisions, one may split the overall research design into
the following parts:

(a) *the sampling design* which deals with the method of selecting items to be observed for the given study;

(b) *the observational design* which relates to the conditions under which the observations are to be made;

(c) *the statistical design* which concerns with the question of how many items are to be observed and how the information and data gathered are to be analyzed and

(d) *the operational design* which deals with the techniques by which the procedures specified in the sampling, statistical and observational designs can be carried out.

From what has been stated above, we can state the important features of a research design as under:

(i) It is a plan that specifies the sources and types of information relevant to the research problem.

(ii) It is a strategy specifying which approach will be used for gathering and analysing the data, (ill) It also includes the time and cost budgets since most studies are done under these two
constraints.

In brief, research design must, at least, contain-(a) a clear statement of the research problem; (b) procedures and techniques to be used for gathering information; (c) the population to be studied; and (d) methods to be used in processing and analysing data.

4.2.2 Need for the Conceptual Structure for Research

Research design is needed because it facilitates the smooth sailing of the various research operations, thereby making research as efficient as possible yielding ma.xima1 information with minimal expenditure of effort, time and money. Just as for better, economical and attractive construction of a house, we need a blueprint (or what is commonly called the map of the house) well thought out and prepared by art expert architect, similarly we need a research design or a plan in advance of data collection and analysis for our research project. Research design stands for advance planning of the methods to be adopted for collecting the relevant data and the techniques to be used in their analysis, keeping in view the objective of the research and the availability of "Staff, time and money.

Even then the need for a well thought out research design is at times not realised by many. The importance which this problem deserves is not given to it. As a result many researches do not serve the purpose for which they are undertaken. In fact, they may even give misleading conclusions. Thoughtlessness in designing the research project may result in rendering the research exercise futile. It is, therefore, imperative that an efficient and appropriate design must be prepared before starting research operations. The design helps the researcher to organize his ideas in a form whereby it will be possible for him to look for flaws and inadequacies.

4.2.3 Components of a Good Conceptual Research Structure

A good design is often characterized by adjectives like flexible, appropriate, efficient, and economical and so on. Generally, the design which minimizes bias and maximizes the reliability of the data collected and analyzed is considered a good design. The design which gives the smallest experimental error is supposed to be the best design in many investigations. Similarly, a design which yields maximal information and provides an opportunity for considering many different aspects of a problem is considered most appropriate and efficient design in respect of many research problems. Thus, the question of good design is related to the purpose or objective of the research problem and also with the nature of the problem to be studied.
A research design appropriate for a particular research problem, usually involves the consideration of the following factors:

(i) the means of obtaining information;
(ii) the availability and skills of the researcher and his *staff*, if any;
(iii) the objective of the problem to be studied;
(iv) the nature of the problem to be studied; and
(v) the availability of time and money for the research work.

If the research study happens to be an exploratory or a formulative one, wherein the major emphasis is on discovery of ideas and insights, the research design most appropriate must be flexible enough to permit the consideration of many different aspects of a phenomenon. But when the purpose of a study is accurate description of a situation or of an association between variables (or in what are called the descriptive studies), accuracy becomes a major consideration and a research design which minimises bias and maximises the reliability of the evidence collected is considered a good design.

4.2.4 Key Parameters of a Conceptual Research Structure

Before describing the different research designs, it will be appropriate to explain the various concepts relating to designs so that these may be better-and easily understood.

1. Dependent and independent variables: A concept which can take on different quantitative values 'is called a variable. As such the concepts like weight, height, income are all examples of variables. Qualitative phenomena (or the attributes) are also quantified on the basis of the presence or absence of the concerning attribute(s). Phenomena which can take on quantitatively different values even in decimal points are called 'continuous variables'.* But all variables are not continuous. j If they can only be expressed in integer values, they are non-continuous variables or in statistical language 'discrete variables' .** Age is an example of continuous variable, but the number of children is an example of non-continuous variable.

2. Extraneous variable: Independent variables that are not related to the purpose of the study, but may affect the dependent variable are termed as extraneous variables. Suppose the researcher wants to test the hypothesis that there is a relationship between children's gains in social studies achievement and their self-concepts. In this case self-concept is an independent variable and social studies achievement is a dependent variable..

3. Control: One important characteristic of a good research design is to minimize the influence or effect of extraneous variable(s). The technical term 'control' is used when we design the study minimizing the effects of extraneous independent variables. In experimental researches, the term 'control' is used to refer to restrain experimental conditions.

4. Confounded relationship: When the dependent variable is not free from the influence of extraneous variable(s), the relationship between the dependent and independent variables is said to be confounded by an extraneous variable(s).

5. Research hypothesis: When a prediction or a hypothesized relationship is to be tested by scientific methods, it is termed as research hypothesis. The research hypothesis is a predictive statement that relates an independent variable to a dependent variable.

6. Experimental and non-experimental hypothesis-testing research: When the purpose of research is to test a research hypothesis, it is termed as hypothesis-testing research. It can be of the experimental design or of the non-experimental design. Research in which the independent variable is manipulated is termed 'experimental hypothesis-testing research' and a research in which an

independent variable is not manipulated is called 'non-experimental hypothesis-testing research'.

7. Experimental and control groups: In an experimental hypothesis-testing research when a group is exposed to usual conditions, it is termed a 'control group', but when the group is exposed to some novel or special condition, it is termed an 'experimental group'
8. Treatments: The different conditions under which experimental and control groups are put are usually referred to as 'treatments'. In the illustration taken above, the two treatments are the usual studies programme and the special studies programme. Similarly, if we want to determine through an experiment the comparative impact of three varieties of fertilizers on the yield of wheat, in that case the three varieties of fertilizers will be treated as three treatments.

9. Experiment: The process of examining the truth of a statistical hypothesis, relating to some , research problem, is known as. an experiment. or example, we can conduct an experiment to examine the usefulness of a certain newly developed drug. Experiments can be of two types viz., absolute experiment and comparative experiment. If we want to determine the impact of a fertilizer on the yield of a crop, it is a case of absolute experiment; but if we want to determine the impact of one fertilizer as compared to the impact of some other fertilizer, our experiment then will be termed as a comparative experiment. Often, we undertake comparative experiments when we talk of designs of experiments. .

10. Experimental unit(s): The pre-determined plots or the blocks, where different treatments are used, are known as experimental units. Such experimental units must be selected (defined) very carefully.

4.2.5 Types of Research Structures

Different research designs can be conveniently described if we categorize them as: 1) research design in case of exploratory research studies; (2) research design in case of descriptive and diagnostic research studies, and (3) research design in case of hypothesis-testing research studies.

We take up each category separately.

1. Research design in case of exploratory research studies: Exploratory research studies are also termed as formulative research studies. The main purpose of such studies is that of formulating a problem for more precise investigation or of

developing the working hypotheses from an operational point of view. The major emphasis in such studies is on the discovery of ideas and insights. Generally, the following three methods in the context of research design for such studies are talked about: (a) the survey of concerning literature; (b) the experience survey and (c) the analysis of 'insight-stimulating' examples.

The survey of concerning literature! happens to be the most simple and fruitful method of formulating precisely the research problem or developing hypothesis. Hypotheses stated by earlier workers may be reviewed and their usefulness be evaluated as a basis for further research. It may also be considered whether the already stated hypotheses suggest new hypothesis. In this way the researcher should review and build upon the work already done by others, but in cases where hypotheses have not yet been formulated, his task is to review the available material for deriving the relevant hypotheses from it.

Experience survey means the survey of people who have had practical experience with the problem to be studied. The object of such a survey is to obtain insight into the relationships between variables and new ideas relating to the research problem. For such a survey people who are competent and can contribute new ideas may be carefully selected as respondents to ensure a representation of different types of experience. The respondents so selected may then be interviewed by the investigator. The researcher must prepare an interview schedule for the systematic questioning of informants. But the interview must ensure flexibility in the sense that the respondents should be allowed to raise issues and questions which the investigator has not previously considered.

Analysis of ~insight-stimulating' examples is also a fruitful method for suggesting hypotheses for research. It is particularly suitable in areas where there is little experience to serve as a guide. This method consists of the intensive study of selected instances of the phenomenon in which one is interested.

Thus, in an exploratory of formulative research study which merely leads to insights or hypotheses, whatever method or research design outlined above is adopted, the only thing essential is that it must continue to remain flexible so that many different facets of a problem may be considered as and when they arise and come to the notice of the researcher.

2. Research design in case of descriptive **and** diagnostic research studies: Descriptive research studies are those studies which are concerned with describing the characteristics of a particular individual, or of a group, whereas

diagnostic research studies determine the frequency with which something occurs or its association with something else. The studies concerning whether certain variables are associated are examples of diagnostic research studies. As against this, studies concerned with specific predictions, with narration of facts and characteristics concerning individual, group or situation are all examples of descriptive research studies. Most of the social research comes under this category. From the point of view of the research design, the descriptive as well as diagnostic studies share common requirements and as such we may group together these two types of research studies. In descriptive as well as in diagnostic studies, the researcher must be able to define clearly, what he wants to measure and must find adequate methods for measuring it along with a clear cut definition of 'population' he wants to study. Since the aim is to obtain complete and accurate information in the said studies, the procedure to be used must be carefully planned. The research design must make enough provision for protection against bias and must maximise reliability, with due concern for the economical completion of the research study. The design in such studies must be rigid and not flexible and must focus attention on the following:

(a) Formulating the objective of the ~tudy (what the study is about and why is it being made?) (b) Designing the methods of data collection (what techniques of gathering data will be adopted?) (c) Selecting the sample (how much material will be needed?)
(d) Collecting the data (where can the required data be found and with what time period should the data be related?)
(e) Processing and analysing the data.
(f) Reporting the findings.

In a descriptive/diagnostic study the first step is to specify the objectives with sufficient precision to ensure that the data collected are relevant. If this is not done carefully, the study may not provide the desired information.

The data collected must be processed and analysed. This includes steps like coding the interview replies, observations, etc.; tabulating the data; and performing several statistical computations.. To the extent possible, the processing and analysing procedure should be planned in detail before actual work is started.

Thus, the research design in case of descriptive/diagnostic studies is a comparative design throwing light on all points narrated above and mustb~ prepared keeping in view the objective(s) of the study and the resources available. However, it must ensure the minimisation of bias and maximisation of reliability of the evidence collected. The said design can be appropriately referred to as a *survey design* since it takes into account all the steps involved in a survey concerning a phenomenon to be studied.

3. Research design in case of hypothesis-testing research studies: Hypothesis-testing research studies (generally known as experimental studies) are those where the researcher tests the hypotheses of causal relationships between variables. Such studies require procedures that will not only reduce bias and increase reliability, but will permit drawing inferences about causality. Usually experiments meet this requirement. Hence, when we talk of research design in such studies, we often mean the design of experiments. .

Professor R.A. Fisher's name is associated with experimental designs. Beginning of such designs was made by him when he was working at Rothamsted Experimental Station (Centre for Agricultural Research in England). As such the study of experimental designs has its origin in agricultural research. Professor Fisher found that by dividing agricultural fields or plots into different blocks and then by conducting experiments in each of these blocks, whatever information is collected and inferences drawn from them, happens to be more reliable. This fact inspired him to develop certain experimental designs for testing hypotheses concerning scientific investigations.

4.2.6 How to develop an Execution Plan for Research?

Basic Principles of Experimental Designs

Professor Fisher has enumerated three principles of experimental designs:
(1) the Principle of Replication;
(2) the Principle of Randomization;
(3) Principle of Local Control.

According to the *Principle of Replication,* the experiment should be repeated more than once. Thus, each treatment is applied in many experimental units instead of one. By doing so the statistical accuracy of the experiments is increased. For example, suppose we are to examine the effect of two varieties of rice. For this purpose we may divide the field into two parts and grow one variety in one part and the other variety in the other part. We can then compare the yield of the two parts and draw conclusion on that basis. But if we are to apply the principle of replication to this experiment, then we first divide the field into several parts, grow one variety in half of these parts and the other variety in he remaining parts. We can then collect the data of yield of the two varieties and draw conclusion by comparing the same. The result so obtained will be more reliable in comparison to the conclusion we draw without applying the principle of replication. However, it should be remembered that replication is introduced in order to increase the precision of a study; that is to say, to increase the accuracy with which the main effects and interactions can be estimated.

The *Principle of Randomization* provides protection, when we conduct an experiment, against the effect of extraneous factors by randomization. In other words, this principle indicates that we should design or plan the experiment in such a way that the variations caused by extraneous factors can all be combined under the general heading of "chance." For instance, if we grow one variety of rice, say, in the first half of the parts of a field and the other variety is grown in the other half, then it is just possible that the soil fertility may be different in the first half in comparison to the other half. If his is so, our results would not be realistic..

The *Principle of Local Control* is another important principle of experimental designs. Under it the extraneous factor, the known source of variability, is made to vary deliberately over as wide a range as necessary and this needs to be done in such a way til1 at the variability it causes can be measured and hence eliminated from the experimental error.

Important Experimental Designs

Experimental design refers to the framework or structure of an experiment and as such there are several experimental designs. We can classify experimental designs into two broad categories, viz., informal experimental designs and formal experimental designs. Informal experimental designs are those designs that normally use a less sophisticated form of analysis based on differences in magnitudes, whereas formal experimental designs offer relatively more control and use precise statistical procedures for analysis. Important experiment designs are as follows:

(a) Informal experimental designs:

(i) Before-and-after without control design.

(ii) After-only with control design.
(ill) Before-and-after with control design.

(b) Formal experimental designs:

(i) Completely randomized design (C.R. Design). / "
(ii) Randomized block design (R.B. Design).
(ill) Latin square design (L.S. Design).
(iv) Factorial designs.

4. Completely randomized design (C.R. design): Involves only two principles viz., the principle of replication and the principle of randomization of experimental designs. It is the simplest possible design and its procedure of analysis is also easier. The essential characteristic of the design is that subjects are randomly assigned to experimental treatments (or vice-versa). For instance, if we have 10 subjects and if we wish to test 5 under treatment A and 5 under treatment B, the randomization process gives every possible group of 5 subjects selected from a set of l0 an equal opportunity of being assigned to treatment A and treatment B.

5. Randomized block design (R.B. design) is an improvement over the C.R. design. In the R.B. design the principle of local control can be applied along with the other *two* principles of experimental designs. In the R.B. design, subjects are first divided into groups, known as blocks, such that within each group the subjects are relatively homogeneous in respect to some selected variable.

.

6. Latin square design (L.S. design) is an experimental design very frequently used in agricultural research. The conditions under which agricultural investigations are carried out are different from those in other studies for nature plays an important role in agriculture. For instance, an experiment has to be made through which the effects of five different varieties of fertilizers on the yield of a certain crop, say wheat, it to be judged. In such a case the varying fertility of the soil in different blocks in which the experiment has to be performed must be taken into consideration; otherwise the results obtained may not be very dependable because the output happens to be the effect not only of fertilizers, but it may also be the effect of fertility of soil.

7. Factorial designs: Factorial designs are used in experiments where the effects of varying more than one factor are to be determined. They are specially important in several economic and social phenomena where usually a large

number of factors affect a particular problem. Factorial designs can be of two types: (i) simple factorial designs and (ii) complex factorial designs. We take them separately

CONCLUSION

There are several research designs and the researcher must decide in advance of collection and analysis of data as to which design would prove to be more appropriate for his research project. He must give due weight to various points such as the type of universe and its nature, the objective of his study, the resource list or the sampling frame, desired standard of accuracy and the like when taking a decision in respect of the design for his research project.

5. RESEARCH METHODS FOR ENGINEERING RESEARCH

5.1 History of Ideas in computing

5.1.1 Introduction

Computer science includes a variety of topics relating to computers, which range from the abstract analysis of algorithms, formal grammars, etc. to more concrete subjects like programming languages, software, and computer hardware.

As a scientific discipline, it differs significantly from and is often confused with mathematics, programming, software engineering, and computer engineering, although there is some degree of overlap with these and other fields.

The Church-Turing thesis states that all known kinds of reasonable paradigms of computation are essentially equivalent in what they can do, although they vary in time and space efficiency. The thesis is not a mathematical theorem that can be proven, but an empirical observation that all known computational schemes have the same computational power.

Now the problem is what the “reasonable paradigm of computation” is – we will have reason to come back to this point later on in this chapter.

Computer

Most research in computer science has been related to von Neumann computers or Turing machines i.e., computers that do one small, deterministic task at a time. These models resemble most real computers in use today. Computer scientists also study other kinds of machines, some practical (like parallel machines) and some theoretical (like probabilistic, oracle, and quantum machines).

Computer Science

Computer scientists study what programs can and cannot do (computability), how programs should efficiently perform specific tasks (algorithms), how programs should store and retrieve specific kinds of information (data structures and data bases), how programs might behave intelligently (artificial intelligence), and how programs and people should communicate with each other (human-computer interaction and user interfaces).

5.1.2 Computer Science Body of Knowledge

The body of knowledge forming the field Computer Science broadly constitutes the following components:

– Discrete Structures
– Programming Fundamentals
– Algorithms and Complexity
– Programming Languages
– Architecture and Organization
– Operating Systems
– Net-Centric Computing
– Human-Computer Interaction
– Graphics and Visual Computing
– Intelligent Systems
– Information Management
– Software Engineering
– Social and Professional Issues
– Computational Science and Numerical Methods
– ...

Related Fields

Computer science is closely related to a number of fields. These fields overlap considerably, though important differences exist

Information science is the study of data and information, including how to interpret, analyze, store, and retrieve it. Information science started as the foundation of scientific analysis of communication and databases.

Computer programming or software development is the act of writing program code.

Software engineering emphasizes analysis, design, construction, and testing of useful software. Software engineering includes development methodologies (such as the waterfall model and extreme programming) and project management.

Information systems (IS) is the application of computing to support the operations of an organization: operating, installing, and maintaining the computers, software, and data.

Computer engineering is the analysis, design, and construction of computer hardware

Mathematics shares many techniques and topics with computer science, but is more general. In some sense, Computer Science is the mathematics of computing.

Logic is a formal system of reasoning, and studies principles that lay at the very basis of computing/reasoning machines, whether it be the hardware (digital logic) or software (verification, AI etc.) levels. The subfield of logic called computability logic provides a systematic answer to the fundamental questions about what can be computed and how. .

Lexicography and specialized lexicography focus on the study of lexicographic reference works and include the study of electronic and Internet-based dictionaries.

Linguistics is the study of languages, converging with computer science in such areas as programming language design and natural language processing.

5.1.3 Debate over Name of the Field

There is some debate over whether the name of the field should be computer science or computation science or computing. The first name is the original, traditional name; however it implies that CS studies computers. The second name is more recent, and it implies that CS studies what we do with computers. The third is the most general term including not only what we can do with present-day computers but any computing process that any physical system can perform.

5.1.4 Major Fields of Importance for Computer Science

The major fields of importance for Computer Science are Mathematical foundations, Theoretical Computer Science, Hardware, Software, Data and Information Systems, Computing Methodologies, Computer Systems Organization, Computer applications and Computing Milieux ....

Mathematical Foundations consists of

Discrete mathematics (Boolean algebra, Graph theory, Domain theory ..)
Mathematical logic
Probability and Statistics
Information theory
...

Theoretical Computer Science addresses

Algorithmic information theory
Computability theory
Cryptography
Formal semantics
Theory of computation (or theoretical computer science)
– analysis of algorithms and problem complexity
– logics and meanings of programs
– Mathematical logic and Formal languages
Type theory
...

Hardware discusses

Control structures and Microprogramming
Arithmetic and Logic structures
Memory structures
Input/output and Data communications
Logic Design
• Integrated circuits
– VLSI design
Performance and reliability

...

Software consists of

Computer programming (Programming techniques, Program specification, Program verification)
Software engineering
–Optimization
– Software metrics
–Software Configuration Management (SCM)
Structured programming

Computer Systems Organization consists of

Computer architecture
Computer networks
Distributed computing
Performance of systems
Computer system implementation
...

Data and Information Systems consists of

Data structures
Data storage representations
Data encryption
Data compression
Data recovery
Coding and Information theory
Files
– File formats
Information systems
– Databases
– Information Storage and retrieval
– Information Interfaces and Presentation
Data recovery
.....

Computing Methodologies consists of

Symbolic and Algebraic manipulation
Artificial intelligence
Computer graphics
Image processing and computer vision
Pattern recognition
Simulation and Modeling
Document and text processing
Digital signal processing
...

Computer Applications consists of

Administrative data processing
Mathematical software (Numerical analysis, Automated theorem proving, Computer algebra systems)
Physical sciences and Engineering (Computational chemistry, Computational physics)
Life and medical sciences (Bioinformatics, Computational Biology, Medical informatics)
Social and behavioral sciences
Arts and Humanities
Computer-aided engineering
Human-computer interaction (Speech synthesis, Usability engineering)
Robotics
....

Computing Aspects (Milieus) consists of

Computer industry
Computers and education
Computers and society
Legal aspects of computing
Management of computing and Information systems
Personal computing
Computer and information security

What do we infer about Computing?

CS != Programming
CS >> Programming
Computing != CS
Computing >> CS

5.1.5 Big ideas in Computer Science and Engineering

Prof. Gerry Sussman [MIT] said we could write down all the ideas in computer science on 4 pages! CS has added valuable knowledge to our understanding of the world.

CS discipline offers some important concepts which it is useful for everyone to understand. Just as there is a utility for everyone to understand a certain amount of math and science, there is a good reason for people to understand a certain amount of computer science.

Hilbert's program: Mechanical procedures exist for finding the solutions to problems. That is, for many questions/problems, we can write down a series of steps and simple predicates which define precisely how to find the correct solution. This process is completely mechanical, not requiring any "human" judgment to complete.

We can build physical machines which implement these procedures and perform the calculations.

There are simple, universal models of computing which capture the basic capabilities of these machines (e.g. automata, pushdown automata, Turing Machines).

The Turing Machine model is "robust" in the sense that "reasonable" additions to it, or alternative formulations of computing models have the same asymptotic power of computability (Church's thesis).

“Reasonable" meaning, they, for the most part, corresponds to things we imagine a real machine could support. In particular, there are stronger models when the machine is allowed to do "unreasonable" things like consult an oracle.

Deterministic/guaranteed procedures do not exist for all problems (Halting Problem, uncomputability). An important component of CS theory is to classify problems as computable or uncomputable.

Of the problems which are computable, tasks have different computational difficulty (complexity). An important component of CS theory allows us to analyze algorithms and assess their complexity. (complexity classes [P,NP,PSPACE, IP, ...], order analysis [O(), Omega(), Theta()])

For example, Common idioms/solution techniques,

– divide-and-conquer
– linear programming
– dynamic programming
– graph algorithms

There are alternatives to directly solving hard problems optimally. CS theory also tells us what we can give up in the guarantee of solution quality to reduce computational complexity.

– approximation algorithms – online algorithms
– polynomial heuristic solutions
– randomized algorithms

The desired computation can be captured precisely and unambiguously. Computer science deals with how we construct languages to describe computations, and how we make them convenient for human use.

languages
syntax (grammars)
semantics (denotational, operational)

We do not have to emulate the user's description of a computation to implement it correctly. We simply need to implement a computation which gives the same visible result (has the same meaning) as the user's computation (compilation, CAD) which means semantic transformations.
The representation used for data in a computation can have a big effect on efficiency of operation and ease of human comprehension

effects on computational and storage efficiency (e.g. arrays and fixed structures vs. tagged lists, red-black trees; sparse vs. dense representations of data)
easing human comprehension (e.g. rich data structures)

Our physical world allows us to build very large computer systems. The practical limit to the useful size of a computer system (or at least, the size of the function efficiently supported by a computer system) is almost always human comprehension, not the physical capacity required. Consequently, a major concern of computer science is techniques to manage and reduce complexity (abstractions/information hiding, modularity, problem decomposition, hierarchy, component isolation, invariants, common idioms/paradigms for organization (e.g. procedures, frames, streams, objects, APIs, servers)

A computing machine can be implemented out of X.
X=mechanical interactions, relays, tubes, transistors, DNA, molecules...
common/useful abstractions (e.g. digital abstraction, flops, memory, communication channels)
disciplines achieving correctness in the face of noise/uncertainty (e.g. voltage levels, timing models and disciplines)

We can extend our notion of abstraction/information hiding to machine design. In particular, the machine code and operating environment for a machine represents the abstract interface it provides to the outside world. Any implementation which provides the same semantics to the machine code is viable.

Consequently, we have the notion of ISAs or architecture families which all run the same machine code but which admit to a variety of implementations (e.g. IBM 360, VAX, MIPS, SPARC, x86).

Machine code is just another language specifying precisely the computation to be performed.

- a computational engine need only provide the intended semantics, leaving it plenty of freedom as to how it implements the semantics.
- like any other language, it can be translated from the input format to another native format (perhaps another machine's native machine format) as long as it provides the original semantics (e.g. softPC, binary translation)

The engineering side of computer science is about: How do we minimize the resources we use in order to perform a computation (set of computations). Physical machines have finite/limited real resources so time, energy, area (hardware: memory, wires)… must be minimized.

We can provide the abstraction of more physical resources by virtualizing the physical resources. That is, sharing the physical resource among multiple uses over time. To accomplish this, we store the state of a particular usage of the physical resources in cheaper storage, e.g. virtual memory, virtual channels, multitasking, time-sharing

Computations occur at different time scales (rates). To minimize work, when possible, hoist a computation out of a high rate region into a slower region. A trivial example: loop invariants/hoisting.

Feedback is the key to diagnosing discrepancies between one's model of the world and reality. This is really just the heart of the scientific method. It should be used by developers to improve programs (debug functional problems, identify and improve performance problems). Moreover, it can be embedded in programs so that they adapt to their data and environment.

A data structure or computation can either be dynamic or static.

Static structures and computations can be very efficient when the size and shape of the computation is constant or has little variance.

Dynamic structures/computations are necessary when the size or scope of the problem is unbounded. They cost more per element or item, but they only have to be as large (as complex) as a particular problem instance.

There are many big ideas in engineering, e.g.

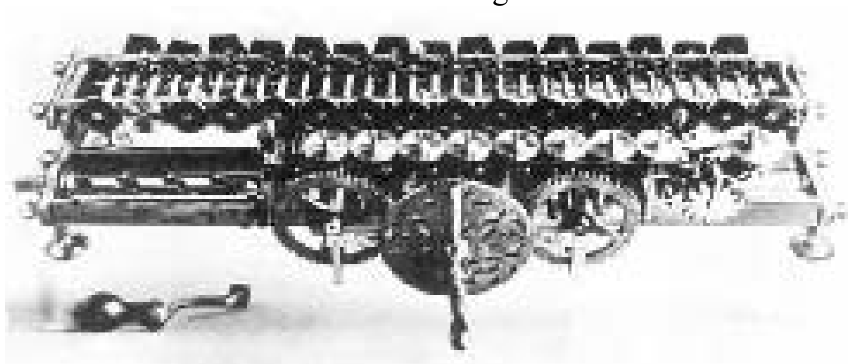
iterative design
real-world constraints
tradeoffs
feedback
complexity management techniques

that are important for understanding not only classical engineered systems but also for understanding social systems and the natural world.

5.1.6 The Great Scientists and their Contributions to the World of Computing

LEIBNIZ: LOGICAL CALCULUS
Name: Gottfried Wilhelm von Leibniz
Born: 1 July 1646 in Leipzig, Saxony (now Germany)
Died: 14 Nov 1716 in Hannover, Hanover (now Germany)

Leibniz´s Calculating Machine



"For it is unworthy of excellent men to lose hours like slaves in the labor of calculation which could safely be relegated to anyone else if the machine were used."

Leibniz´s Logical Calculus

DEFINITION 3. A is in L, or L contains A, is the same as to say that L can be made to coincide with a plurality of terms taken together of which A is one. $B \oplus N = L$ signifies that B is in L and that B and N together compose or constitute L. The same thing holds for larger number of terms.

AXIOM 1. $B \oplus N = N \oplus B$.

POSTULATE. Any plurality of terms, as A and B, can be added to compose $A \oplus B$.

AXIOM 2. $A \oplus A = A$.

PROPOSITION 5. If A is in B and A = C, then C is in B.

PROPOSITION 6. If C is in B and A = B, then C is in A.

PROPOSITION 7. A is in A.

(For A is in $A \oplus A$ (by Definition 3). Therefore (by Proposition 6) A is in A.)

....

PROPOSITION 20. If A is in M and B is in N, then $A \oplus B$ is in $M \oplus N$.

BOOLE: LOGIC AS ALGEBRA
Name: George Boole
Born: 2 Nov 1815 in Lincoln, Lincolnshire, England

Died: 8 Dec 1864 in Ballintemple, County Cork, Ireland
George Boole is famous because he showed that rules used in the algebra of numbers could also be applied to logic.
This logic algebra, called Boolean algebra, has many properties which are similar to "regular" algebra. These rules can help us to reduce an expression to an equivalent expression that has fewer operators.

FREGE: MATEMATICS AS LOGIC
Name: Friedrich Ludwig Gottlob Frege
Born: 8 Nov 1848 in Wismar, Mecklenburg-Schwerin (now Germany)
Died: 26 July 1925 in Bad Kleinen, Germany

The Predicate Calculus (1)

In an attempt to realize Leibniz's ideas for a language of thought and a rational calculus, Frege developed a formal notation (Begriffsschrift).
He has developed the first predicate calculus: a formal system with two components: a formal language and a logic.

The Predicate Calculus (2)

The formal language Frege designed was capable of expressing:

- predicational statements of the form
  'x falls under the concept F' and 'x bears relation R to y', etc.,
- complex statements such as
  'it is not the case that ...' and 'if ... then ...', and
- 'quantified' statements of the form
  'Some x is such that ...x...' and 'Every x is such that ...x...'.

The Analysis of Atomic Sentences and Quantifier Phrases

Fred loves Annie.
Therefore, some x is such that x loves Annie.

Fred loves Annie.
Therefore, some x is such that Fred loves x.

Both inferences are instances of a single valid inference rule.

Proof

As part of his predicate calculus, Frege developed a strict definition of a **'Proof'.**
In essence, he defined a **proof** to be any finite sequence of well-formed ***statements*** such that each statement in the sequence either is an ***axiom*** or follows from previous members by a valid ***rule of inference.***

CANTOR: INFINITY

Name: Georg Ferdinand Ludwig Philipp Cantor
Born: 3 March 1845 in St Petersburg, Russia
Died: 6 Jan 1918 in Halle, Germany

Contribution: Infinities

Set of integers has an equal number of members as the set of even numbers, squares, cubes, and roots to equations!

The number of points in a line segment is equal to the number of points in an infinite line, a plane and all mathematical space!

The number of transcendental numbers, values such as π and e that can never be the solution to any algebraic equation, were much larger than the number of integers.

Hilbert described Cantor's work as: - "...the finest product of mathematical genius and one of the supreme achievements of purely intellectual human activity."

HILBERT: PROGRAM FOR MATHEMATICS

Name: David Hilbert
Born: 23 Jan 1862 in Königsberg, Prussia (now Kaliningrad, Russia)
Died: 14 Feb 1943 in Göttingen, Germany

Hilbert's program

Provide a single formal system of computation capable of generating all of the true assertions of mathematics from "first principles" (first order logic and elementary set theory).
Prove mathematically that this system is consistent, that is, that it contains no contradiction. This is essentially a proof of correctness.
If successful, all mathematical questions could be established by mechanical computation!

GÖDEL: END OF HILBERTS PROGRAM

Name: Kurt Gödel
Born: 28 April 1906 in Brünn, Austria-Hungary (now Brno, Czech Republic)

Died: 14 Jan 1978 in Princeton, New Jersey, USA

Incompleteness Theorems

In any axiomatic mathematical system there are propositions that cannot be proved or disproved within the axioms of the system.
In particular the consistency of the axioms cannot be proved.

TURING: UNIVERSAL AUTOMATON

Name: Alan Mathison Turing
Born: 23 June 1912 in London, England
Died: 7 June 1954 in Wilmslow, Cheshire, England

When war was declared in 1939 Turing moved to work full-time at the Government Code and Cypher School at Bletchley Park.

Together with another mathematician W G Welchman, Turing developed the Bombe, a machine based on earlier work by Polish mathematicians, which from late 1940 was decoding all messages sent by the Enigma machines of the Luftwaffe.

At the end of the war Turing was invited by the National Physical Laboratory in London to design a computer.

His report proposing the Automatic Computing Engine (ACE) was submitted in March 1946.

Turing returned to Cambridge for the academic year 1947- 48 where his interests ranged over topics far removed from computers or mathematics, in particular he studied neurology and physiology.

1948 Newman (professor of mathematics at the University of Manchester) offered Turing a readership there.

Work was beginning on the construction of a computing machine by F C Williams and T Kilburn. The expectation was that Turing would lead the mathematical side of the work, and for a few years he continued to work, first on the design of the subroutines out of which the larger programs for such a machine are built, and then, as this kind of work became standardized, on more general problems of numerical analysis.

1950 Turing published Computing machinery and intelligence in Mind

1951 elected a Fellow of the Royal Society of London mainly for his work on Turing machines

1951 was working on the application of mathematical theory to biological forms.

1952 he published the first part of his theoretical study of morphogenesis, the development of pattern and form in living organisms.

VON NEUMAN: COMPUTER

Name: John von Neumann
Born: 28 Dec 1903 in Budapest, Hungary
Died: 8 Feb 1957 in Washington D.C., USA

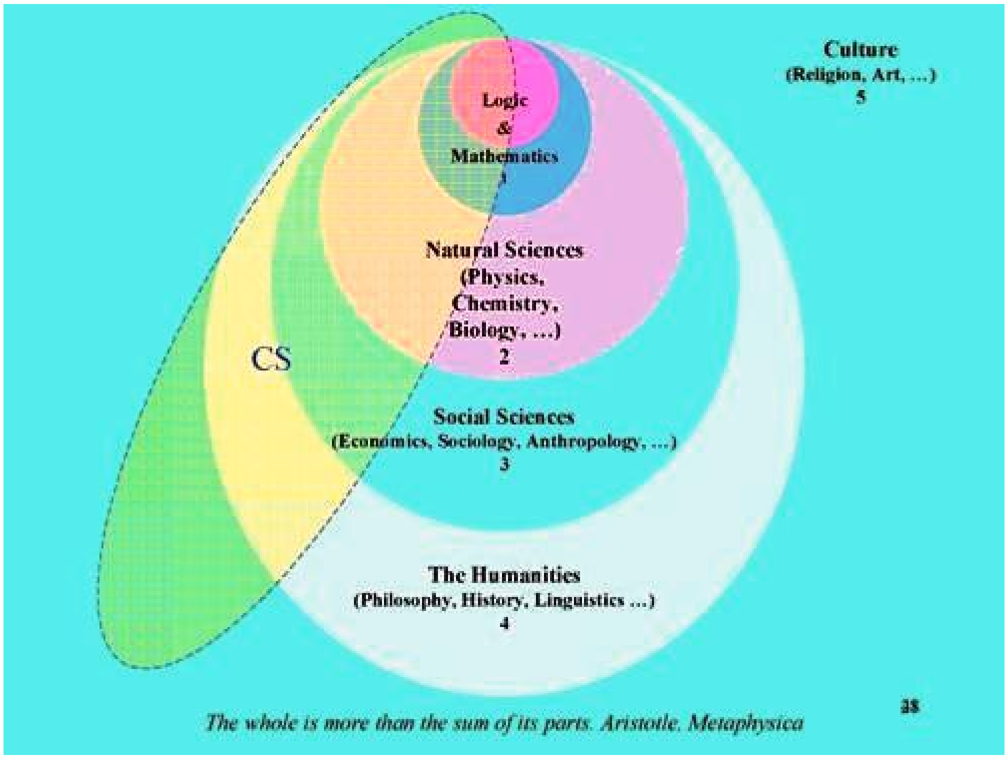
In the middle 30's, Neumann was fascinated by the problem of hydro dynamical turbulence. The phenomena described by non-linear differential equations are baffling analytically and defy even qualitative insight by present methods.

Numerical work seemed to him the most promising way to obtain a feeling for the behaviour of such systems. This impelled him to study new possibilities of computation on electronic machines ...

Von Neumann was one of the pioneers of computer science making significant contributions to the development of logical design. Working in automata theory was a synthesis of his early interest in logic and proof theory and his later work, during World War II and after, on large scale electronic computers.

Involving a mixture of pure and applied mathematics as well as other sciences, automata theory was an ideal field for von Neumann's wide-ranging intellect. He brought to it many new insights and opened up at least two new directions of research.

He advanced the theory of cellular automata, advocated the adoption of the bit as a measurement of computer memory, and solved problems in obtaining reliable answers from unreliable computer components.



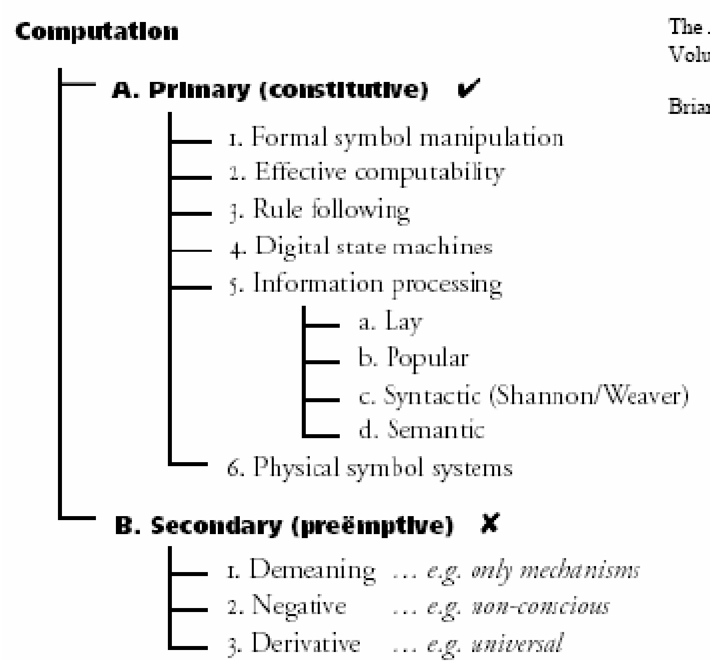
5.1.7 Philosophical Problems of Computing

The computer presents itself as a culturally defining technology and has become a symbol of the new millennium, playing a cultural role far more influential than the mills in the Middle Ages, mechanical clocks in the seventeenth century, or the steam engine in the age of the industrial revolution. (Bolter 1984)

"The important difference is that the computer (the physical object that is directly related to the theory) is not a focus of investigation (not even it he sense of being the cause of certain algorithm proceeding in certain way) but it is rather theory materialized, a tool always capable of changing in order to accommodate even more powerful theoretical concepts. "

Philosophy of Computing

What is Computation?



Construal of Computing

1. Formal symbol manipulation (FSM): the idea, derivative from a century's work in formal logic and meta mathematics, of a machine manipulating (at least potentially) symbolic or meaningful expressions without regard to their interpretation or semantic content;

2. Calculation of a function (FUN): behavior that, when given as input an argument to a (typically mathematical) function, produces as output the value of that function on that argument.

3. Effective computability (EC): what can be done—and how hard it is to do it—mechanically, as it were, by, an abstract analogue of a "mere machine";

4. Rule-following or algorithm execution (RF): what is involved, and what behaviour is thereby produced, in following a set of rules or instructions, such as when cooking dessert;

5. Digital state machines (DSM): the idea of an automaton with a finite, disjoint set of internally homogeneous states—as parodied in the "clunk, clunk, clunk" gait of a 1950's cartoon robot;

6. Information processing (IP): what is involved in storing, manipulating, displaying, and otherwise trafficking in "information," whatever information might be; and

7. Physical symbol systems (PSS): the idea, made famous by Newell and Simon, that, somehow or other, computers interact with and perhaps are also made of symbols in a way that depends on their mutual physical embodiment.

Church-Turing Thesis [ ]

A Turing machine is an abstract representation of a computing device. It is more like a computer program (software) than a computer (hardware). LCMs [logical computing machines: Turing's expression for Turing machines] were first proposed by Alan Turing, in an attempt to give a mathematically precise definition of "algorithm" or "mechanical procedure".

Effective Computability

The Church-Turing thesis concerns an effective or mechanical method in logic and mathematics.
A method M is called 'effective' or 'mechanical' iff:

1. M is set out in terms of a finite number of exact instructions (each instruction being expressed by means of a finite number of symbols);
2. M will, if carried out without error, always produce the desired result in a finite number of steps;
3. M can (in principle) be carried out by a human being unaided by any machinery save paper and pencil;
4. M demands no insight or ingenuity on the part of the human being carrying it out.

Misunderstandings of the Turing Thesis

Turing did not show that his machines can solve any problem that can be solved "by instructions, explicitly stated rules, or procedures" and nor did he prove that a universal Turing machine "can compute any function that any computer, with any architecture, can compute".

Turing proved that his universal machine can compute any function that any Turing machine can compute; and he put forward, and advanced philosophical arguments in support of, the thesis here called Turing's thesis.

Turing introduces his machines as an idealized description of a certain human activity, the tedious one of numerical computation, which until the advent of automatic computing machines was the occupation of many thousands of people in commerce, government, and research establishments.

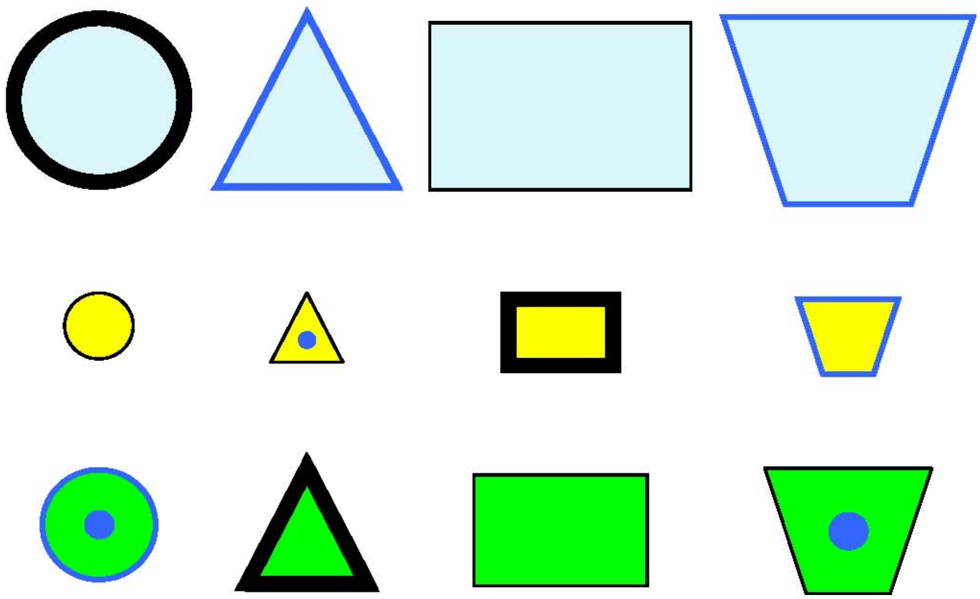
Turing's "Machines". These machines are humans who calculate. (Wittgenstein) A man provided with paper, pencil, and rubber, and subject to strict discipline, is in effect a universal-machine.(Turing)

A thesis concerning the extent of effective methods procedures that a human being unaided by machinery is capable of carrying out -has no implication concerning the extent of the procedures that machines are capable of carrying out, even machines acting in accordance with 'explicitly stated rules'.

Among a machine's repertoire of atomic operations there may be those that no human being unaided by machinery can perform.

Two Most Fundamental Functions
SEARCH (identify objects = divide universe in parts)
SORT (organize: what is the same, what is different)

Repetition, Similarity



As repetition is based upon similarity, it must be relative. Two things that are similar are always similar in certain respects.

Searching for similarity and differences leads to classifications i.e. the division of objects or events in different groups/classes.

The simplest tool by for classification is the binary opposition or dichotomy (dualism). When we use dichotomy, we only decide if an object is of kind A or of kind ~A. Examples of frequent dichotomies are yes/no, true/false, before/after, more/less, above/below, etc.

Identity

The basic feature of experimental method is its reproducibility: It must be possible to establish essentially the same experimental situation in order to obtain the same results. This means that the experimental arrangement can be made with essentially equivalent parts.

What we call "essentially equivalent" (or we can call it "essentially the same") depends on situation. Even here the principle of information hiding helps us to get a practical "level of resolution" which means information hiding for all objects below that level.

Declaring two systems/particles/states as identical is entirely the matter of focus.

For example if we focus on question of how many people in this country are vegetarians, we just treat all people as equal units. If on the other hand we want to know how many women in this country are vegetarian, we discriminate between men and women in our analysis of people, so they are no longer identical.

Declaring two systems/particles/states as identical is entirely the matter of focus.

We can e.g. assume that bacteria of particular sort are interchangeable (indistinguishable) in certain context. That enables us to make repeated experiments with different agents and to treat all bacteria of the same type as equal. It does not mean that they are identical in the absolute sense. It only means that for our purpose the existing difference does not have any significance.

Example of ancient atomic theory. The problem of showing that one single physical body-say piece of iron is composed of atoms is at least as difficult as of showing that all swans are white. Our assertions go in both cases beyond all observational experience.

The difficulty with these structural theories is not only to establish the universality of the law from repeated instances as to establish that the law holds even for one single instance.

A singular statement like "This swan here is white" may be said to be based on observation. Yet it goes beyond experience-not only because of the word "white", but because of the word "swan".

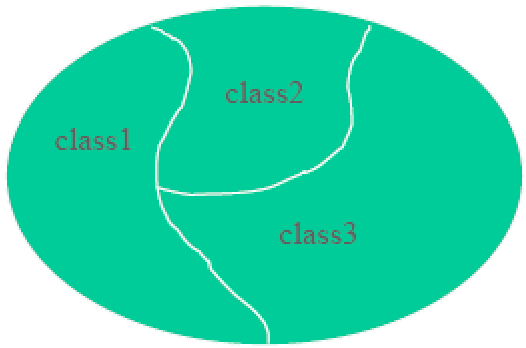
For by calling something a "swan", we attribute to it properties which go far beyond mere observation. So even the most ordinary singular statements are always the interpretations of the facts in the light of theories!

Classification (1)

- A relation is an equivalence relation if it is reflexive, symmetric and transitive.
- An example of such is equality on a set.

Classification (2)

CLASSES SHOULD BE DISJUNCT ...



class 1:
positive effect

class 2:
negative effect

class 3: no effect

Universe here is a group of patients who test a new medicine.

... AND CHOSEN ACCORDING TO SAME CRITERIA ...

Classification (3)

Jorge Luis Borges,
"The Analytical Language of John Wilkins“
(Incongruity)

Borges's fictive encyclopedia divides animals into:

(a) those that belong to the Emperor,
(b) embalmed ones,
(c) those that are trained,
(d) suckling pigs,
(e) mermaids,
(f) fabulous ones,
(g) stray dogs,
(h) those that tremble as if they were mad,
(i) those that resemble flies from a distance
(j) those drawn with a very fine camel's hair brush,
(k) innumerable ones,
(l) others,
(m) those that have just broken a flower vase

5.1.8 Philosophy of Science – A New Paradigm due to the Field of Computing

It is time again to direct gaze upward and see Nature as a whole. What have we learned last couple of centuries of scientific specialisation and division to ever smaller fields?

Field of computing seems to be the best candidate to replace physics as paradigmatic mirror of Philosophy of Science used to reflect the Universe that matters.

This time Universe has to include human, and not only as a machine!

Unifying concept of information has even the potential to bring together Sciences and Humanities. A New Natural Philosophy indeed!

5.2 Evolution of Computing Research

5.2.1 Research in Computing Science

The expanding scope of `computing science' makes it difficult to sustain traditional scientific and engineering models of research. In particular, recent work in formal methods has abandoned the traditional empirical methods. Similarly, research in requirements engineering and human computer interaction has challenged the proponents of formal methods. These tensions stem from the fact that `Computing Science' is a misnomer. Topics that are currently considered part of the discipline of computing science are technology driven rather than theory driven. This creates problems if academic departments are to impose scientific criteria during the assessment of PhDs. It is, therefore, important that people ask themselves `What is Research in Computing Science' and understands its evolution to pursue research in computing with right perspective.

5.2.2 The Dialectic of Research

The highest level of logical argument can be seen in the structure of debate within a particular field. Each contribution to that debate falls into one of three categories:

thesis
This presents the original statement of an idea. However, very few research contributions can claim total originality. Most borrow ideas from previous work, even if that research has been conducted in another discipline.

antithesis
This presents an argument to challenge a previous thesis. Typically, this argument may draw upon new sources of evidence and is typically of progress within a field.

synthesis
This seeks to form a new argument from existing sources. Typically, a synthesis might resolve the apparent contradiction between a thesis and an antithesis.

A good example of this form of dialectic is provided by the debate over prototyping. For example, some authors have argued that prototypes provide a useful means of generating and evaluating new designs early in the development process (thesis), (Fuchs, 1992). Others have presented evidence against this hypothesis by suggesting that clients often choose features of the prototyping environment without considering possible alternatives (antithesis) (Hayes and Jones, 1989). A third group of researchers have, therefore, developed techniques that are intended to reduce bias towards features of prototyping environments (synthesis) (Gravell and Henderson, 1996). Research in a field progresses through the application of methods to prove, refute and reassess arguments in this manner.

5.2.3 Models of Argument

A more detailed level of logical argument can be seen in the structures of discourse that are used to support individual works of thesis, antithesis or synthesis.

5.2.3.1 Proof by Demonstration

Perhaps the most intuitively persuasive model for research is to build something and then let that artefact stand as an example for a more general class of solutions. There are numerous examples of this approach being taken within the field of computing science. It is possible to argue that the problems of implementing multi-user operating systems were solved more through the implementation and growth of UNIX than through a more measured process of scientific enquiry.

However, there are many reasons why this approach is an unsatisfactory model for research. The main objection is that it carries high risks. For example, the artefact may fail long before we learn anything about the conclusion that we are seeking to support. Indeed, it is often the case that this approach ignores the formation of any clear hypothesis or conclusion until after the artefact is built. This may lead the artefact to become more important to the researcher than the ideas that it is intended to establish.

The lack of a clear hypothesis need not be the barrier that it might seem. The proof by demonstration approach has much in common with current engineering practice. Iterative refinement can be used to move an implementation gradually towards some desired solution. The evidence elicited during previous failed attempts can be used to better define the goal of the research as the work progresses. The key problem here is that the iterative development of an artefact, in turn, requires a method or structure. Engineers need to carefully plan ways in which the faults found in one iteration can be fed back into subsequent development. This is, typically, done through testing techniques that are based upon other models of scientific argument. This close relationship between engineering and scientific method should not be surprising:

> **engineering** n. an application of science to the design, building and use of machines, construction etc. (The Oxford Concise Dictionary).

5.2.3.2 Proof by Empirical Method

The Western empirical tradition can be seen as an attempt to avoid the undirected interpretation of artefacts. It has produced the most dominant research model since the seventeenth century. It can be summarized by the following stages:

Hypothesis generation
This explicitly identifies the ideas that are to be tested by the research.

Method identification
This explicitly identifies the techniques that will be used in order to establish the hypothesis. This is critical because it must be possible for one's peers to review

and criticise the appropriateness of the methods that you have chosen. The ability to *repeat* an experiment is a key feature of strong empirical research.

Result compilation
This presents and compiles the results that have been gathered from following the method. An important concept here is that of statistical significance; whether or not the observed results could be due to chance rather than an observable effect.

Conclusion
Finally, the conclusions are stated either as supporting the hypothesis or rejecting it. In the case that results do not support a hypothesis, it is important always to remember that this may be due to a weakness in the method. Conversely, successful results might be based upon incorrect assumptions. Hence, it is vital that all details of a method are made available to peer review.

This approach has been used to support many different aspects of research within Computing Science. For example, Boehm, Gray and Seewaldt (1984) used it to compare the effectiveness of specification and prototyping techniques for software engineering. Others have used it to compare the efficiency of searching and sorting algorithms. Researchers in Information Retrieval have even developed standard methods which include well known test sets to establish performance gains from new search engines.

There are many problems with the standard approach to scientific empiricism when applied to computing science. The principle objection is that many aspects of computing defy the use of probabilistic measures when analysing the results of empirical tests. For example, many statistical measures rely upon independence between each test of a hypothesis. Such techniques clearly cannot be used when attempting to measure the performance of any system that attempts to optimise its performance over time; this rules out load balancing algorithms etc. Secondly, it can be difficult to impose standard experimental conditions upon the products of computer science. For example, if a program behaves in one way under one set of operating conditions then there is no guarantee that it will behave in the same way under another set of conditions. These conditions might go down to the level of alpha particles hitting memory chips. Thirdly, it can be difficult to generalise the results of tightly controlled empirical experiments. For example, just because a user finds a system easy to use in a lab-based evaluation, there is no guarantee that another user will be able to use that product amidst the distractions of their everyday working environment. Finally, it is difficult to determine when a sufficient number of trials have been conducted to support many hypotheses. For example, any attempt to prove that a program always satisfies some property will be almost certainly doomed to failure using standard experimental techniques, The number of potential execution paths through even simple code makes it impossible to test properties against every possible execution path.

5.2.3.3 Mathematical Proof

The dissatisfaction with empirical testing techniques has led many in the computing science research community to investigate other means of structuring arguments in support of particular conclusions. In the United Kingdom, much of this work has focused upon argumentation techniques that were originally developed to model human discourse and thought within the field of philosophy. For example, Burrows, Abadi and Needham (1990) adopted this approach to reason about the correctness of network authentication protocols. The central idea in this work is that mathematics can be used to set up a system of rules about valid and invalid inferences. These rules can then be applied to work out whether or not a conclusion is a valid inference given some initial statements about a program or some hardware.

The field of mathematical reasoning is a research area in its own right. It is, however, possible to identify two different approaches to the use of formal proof as a research technique in computing science:

the argument of verification.
This attempts to establish that some good property will hold in a given system. The classical approach is to allow a human to interactively guide a theorem proving system towards some sequence of proof steps that support the conclusion. The problem here is that if the human cannot construct a proof, this does not imply that the conclusion is invalid. Simply that they have failed to prove it. Another person might be capable of constructing the necessary mathematical argument.

the argument of refutation.
Rather than attempting to prove the correctness of an argument, this approach attempts to refute it. Typically, this is done by setting up a description of the intended system behaviour. Model checking tools then automatically explore the state space of the proposed application in an attempt to find a situation in which the desired conclusion does not hold.

The attractions of mathematical proof techniques are very strong. They provide a coherent framework for analysing research questions in computing science. They also explicitly state the criteria for valid inferences, as well as the environmental conditions, that restrict the scope and applicability of the reasoning process. There are, however, many problems that limit the utility of this approach as a general research tool.

The first is that incredible care needs to be made over the interpretation of results from mathematical proof. Formal methods are nothing more than a system of argumentation and mistakes are to be expected. Problems arise because mistakes can be very difficult to detect given the complex nature of the mathematics that are often used. Recall that a central feature of the empirical approach was that open peer review should be used to check that your method is correct.

The second problem with formal reasoning is that their scope is limited. Interactive and time critical systems pose specially challenges for the application of mathematics. These issues are being addressed but many problems remain.

The third problem relates to the cost of applying formal techniques. It takes a long time to acquire the necessary skills. Similarly, it can take several months to conduct relatively simple proofs for medium to large scale applications.

Finally, it can be argued that there is inadequate discussion about the failures of formal methods. Again, it is important to recall that a failure to prove a hypothesis was a valuable result for empirical techniques. Exaggerated claims have been made for formal reasoning, typically not by the researchers themselves, and many of these claims have been falsified. As a result errors in the application of mathematical reasoning can be seen as a source of shame rather than a learning opportunity for one's colleagues and peers.

1. Proof by Hermeneutics

Formal proof techniques rely upon the development of a mathematical model of the artifact that is being created. This raises important questions about the relationship between that model and the reality which it is intended to represent. For example, if a model omits some critical aspect of a program's environment then it may be proven to be safe but may well fail when implemented. The distance between mathematical models and reality is, popularly, known as the formality gap. Hermeneutics provide an alternative that addresses this problem. Hermeneutic research methods have been pioneered within the field of sociology. The term itself means:

> `adj. concerning interpretation, esp. of Scripture or literary texts'. (The Oxford Concise Dictionary).

In practice, these approaches force researchers to observe the operation and use of an artifact within its intended working environment. The basic premise is that abstract models provide no substitute for real application. Similarly, the results of controlled experiments fail to provide generic results that can be accurately used to assess performance outside of those controlled settings. In particular, the Hawthorne effect suggests that people and, indeed systems, will perform very differently when they are placed within an empirical setting. Repair and maintenance activities are very different for equipment that is provided in a laboratory setting. Individuals react differently when they know that they are being observed. Hermeneutic research, therefore, relies upon the interpretation of signs and observations in the working context rather than on explicitly asking people about the performance of their systems. Hermeneutics techniques urge researchers to enter into the workplace. Taken to an extreme, the performance of an algorithm could only be assessed in field trials with real sets of data on existing architectures with `real' levels of loading from other applications. This stress upon the analysis of a final implementation closely resembles proof by demonstration. The major difference, however, is that the researcher approaches the context of work with an open mind and without any set hypothesis to prove or disprove (Suchman, 1987). This raises problems for the conduct of directed research because users may not use programs in the manner that was intended. For example, it can be difficult to demonstrate that one search engine is faster than another if users continually abandon their requests after one or two items are returned or if they only use those search engines once or twice in their working day.

5.2.4 To Leap Forward...

Computing science is an emerging discipline and hence yet to evolve its own theories, rules and regulations. Vast resources have also been poured into the subject in a relatively short period of time. This has brought startling advances in both hardware and software engineering. Unfortunately the development of computing technology has not been matched by a similar development in academic research techniques. Computing science, as a separate identity is yet to blossom and hence relies on other sciences for help. This has to continue till it reaches its perfection. All the same one cannot wait for that; nor it is a defect. It is a problem to be faces and we have to make use of any tool relevant and handy. In the pursuit of technological goals, researchers have borrowed models of argument and discourse from disciplines as varied as philosophy, sociology and the natural sciences. This lack of any agreed research framework reflects the strength and vitality of computing science. An optimist might argue that we have learnt greatly from the introduction of hermenuetics into the field of requirements analysis. Similarly, we have profited from the introduction of mathematical models of argument to specify and verify of complex systems. A key point is, however, to encourage people to think about the costs that have also been incurred by the heterogeneous nature of research in our discipline.

I do not argue that we must develop a single research model for Computing Science. I do, however, argue that researchers must actively think about the strengths and weaknesses of the research tradition that they adopt. Too often, PhD theses slavishly follow empirical or formal proof techniques without questioning the suitability of those approaches. For example, the hermeneutic tradition has delivered results that ignore the constraints of time and money on commercial system development. Formal methods research has produced results that abstract so far away from the problem domain that they cannot be applied or validated. The tragedy is that unless we begin to recognize these failures then we will continue to borrow flawed research methods from other disciplines. Our goal should be to identify the right approaches, right methods and tools for the various varieties of research sub-domains of computing science backed by sound methodology.

5.3 Overview of Research Methods

5.3.1 Introduction to Scientific Methods

Theoretical methods: Theoretical methods Create formal models (mathematics, logic). They define concepts within these models; prove properties of the concepts; Abstraction, hide details to make the whole more understandable (and to make it possible to prove properties of it). Theoretical methods provide proofs of properties by deductive methods.

Empirical methods: We perform experiments using empirical methods and see how it turned out to draw conclusions.

Simulation: Here, we start with a formal model at some "easy-to-understand" level; Make "artificial experiments" in a computer; Collect statistics and draw conclusions.

In physics: We make hypotheses about the surrounding world (theory), observe it (experiment); Relate the result of experiment to theory; Adjust the theory if it doesn't predict the reality well enough; Theory is used to predict the future (e.g., if a bridge will hold for a certain load, or an asteroid fall down on our heads)

Common pattern in Computing Science:

The system is constructed to behave according to some theoretical model. Deviations are seen as construction errors rather than deficiencies in the theory (hardware error, bug in OS, ...). In both cases the theory helps us understand and predict, but in different ways!

5.3.2 Theoretical vs. empirical methods in Computing Science

Computing Science really has a "spectrum", from "extreme constructivism" to a use of theory close the one in physics as explained below:

"Extreme constructivism": (ideal) Programming language design:

Formal semantics for the language, pure construction of model defining the mathematical meaning of each program.

Abstraction of details to make the meaning of the language simpler (for instance, assume that data structures can grow arbitrarily big).

Implement the language according to the semantics.

One can prove formally within the model that a program is correct – valuable!

But the model does not cover all kinds of errors. E.g., hardware errors, or stack overflow (or an asteroid falling down on the computer)

Extreme "physics" approach: Performance modeling of complex computer-and communication systems

Extremely hard to make analytical calculations

Simplified performance models, tested against experiments(e.g., long suites of benchmarks)

Discrepancy leads to a modified theory, as in physics

Often we do simulation (desire to evaluate systems before building them)

In-between: Algorithm analysis

Build on some form of formal model for how the algorithm executed (meta language with formal semantics), and some performance model (how long does a step in the algorithm take, how much memory is needed to store an entity)

Performance model often of type "one arithmetic operation =one time unit"

Given that the performance model is correct, one proves mathematically that the algorithm needs certain resources (time, memory) to be carried out

But the performance model is often very approximate

Sometimes possible to refine the performance model, but this can make it impossible to calculate the resource needs of the algorithm

5.3.3. Theoretical Models in Computing Science

Discrete mathematics: basic set theory, relations, functions, graphs, algebra, combinatorics, category theory, etc.

The *science* logic: different logical systems, how to make "proofs about proofs"

Theory for complete partial orders (formal semantics)

Topology (mathematics with notions of distance and convergence)

Probability theory

Traditional analysis

5.3.4. Theoretical Problems in Computing Science

What do we want to prove theoretically within Computing Science?
For instance, we want to prove properties of programs, systems, algorithms, and problems.

Some examples:

"FFT uses $O(n \log n)$ operations"

"With 99% confidence the program p runs faster than 1.3 ms on machine m"

"The program p terminates for all in-data"

"If the method M says that a program terminates then this is true"

"There is no method that can decide, for any program, whether it terminates or not"

"For each CREW PRAM-algorithm there is an EREW PRAM-algorithm that can simulate it with a certain slowdown"

$P=NP$ (or $P=NP$)

"The two semantics S_1 och S_2 agree for each program in the programming language P "

5.6. Deductive Methods in Computing Science & engineering

5.6.1. "Ordinary" mathematical proofs

Often *finite entities*: defined *recursively*, properties proved with *induction*

But also reasoning about limits ("go to the limit"), when infinite behaviours are modeled

Encodings and translations – common in Complexity Theory

Sometimes also more conventional mathematical techniques

5.6.2. Inference systems and their applications

Direct modeling with logical inference systems

Common in semantics of programming languages (operational semantics)

Proof methods from logic (proofs about proofs), again induction!

Let's see some examples...

Example 1: Probabilistic Cache Analysis

Purpose: estimate number of cache misses in loops

Probabilistic cache analysis can be used to direct optimizing program transformations, or to estimate the upper bounds on the running time of a program.

Idea: a memory reference can be a *hit* only if the contents of the address have been brought to the cache by a previous access, *and* no other access (taking place between these) has caused the address to be evicted from the cache.

If it is not surely a hit, then we assume it is a miss.

We must model the cache mathematically, as well as the possible program flows and the memory references in the loop.

The goal is to set up equations that tell for each execution of a reference whether is surely a hit, and then count the number of executions where the equations hold.

Dividing this with the total number of reference executions gives an upper bound on the cache miss ratio.

Modeling the cache:

Caches consist of *cache lines* that hold segments of the address space.

We can divide the address space into *memory lines*: mapping M from addresses to memory lines.

In the simplest case the cache is *direct-mapped*: then there is a fixed mapping C from memory lines to cache lines.

Thus, each address *a* will always be stored in cache line C (M (*a*)).

(Caches with a higher degree of associativity are a little more complex to handle, but the principle is the same).

Loops and references

Consider a for-loop with two references:

```
for i = 1 to n
..... r(i) ....
   ...r'(i) ....
```

If, for two possible values i_1 and i_2 of i in the range 1,...,n, it holds that $i_1 \leq i_2$ and M(r(i_1)) = M(r′(i_2)) then there is a possible *reuse* from r(i_1) to r′(i_2) .

If for *no* other reference r″(i_3), executed between r(i_1)and r′(i_2), holds that C(M(r″(i_3))) = C(M(r(i_1))) but M(r″(i_3)) ≠ M(r(i_1)), then we know r′(i_2)is *surely a hit* (since then the memory line was brought into the cache when accessing r(i_1), and has not been thrown out by some other access using the same cache line since).

The above equations can be formalized and in principle applied to all i_2 *{1… n}* . Counting the number of i_2's where the equations hold gives a lower bound to the number of cache hits for r′(i)in the loop, and thus an upper bound for the number of cache misses.

The above is easily extended to nested loops.

How solve the equations?

Testing for all possible values of indices in the loop can be very costly (similar to simulation). For certain kinds of references (Fortran arrays with subscripts that are linear functions of loop indices) the number of solutions can be calculated analytically.

This is also, however, very costly in practice!

Another approach is to *sample* the *iteration space* (set of values for loop indices).

The equations are then tested only in the sampled points. If the sampling is done with an independent distribution, then the result will still say something about the number of cache misses *with some level of confidence.*

This confidence level can be determined by statistical methods (This method has proved to be very efficient in practice).

What kind of theory did we use?

Modeling of cache typically uses integer mathematics (integer division, remainder,...).

Reasoning about (array) references in loops is often done with linear algebra and vector spaces.

Ensuring properties of the sampling method requires probability theory.

Note that there is a lot of hidden modeling of the program itself: how a loop works, how arrays are laid out in memory, etc.

Example 2: Algorithm Analysis

Purpose: to find the *cost* of executing an algorithm (that solves a given problem)

(Archetypal problem: to sort a sequence of numbers)

Cost is typically *running time*, but can also be memory requirements, power consumption, etc.

To calculate the cost requires:

a *machine model.*

a *notation*, i.e., "programming language" for the machine.

Typically we are only interested in the *asymptotic complexity* of the algorithm.

"How fast does the execution time grow with the size of the input?".

An example: insertion sort

```
1 for j = 2 to length(A) do
2   key = A[j]
3   i=j-1
4   while i > 0 and A[i] > key do
5     A[i+1] = a[i]
6     i=i-1
7 A[i+1] = key
```

We want to find the execution time as a function of input size (`length (A)`)

Let us informally analyze insertion sort!

Assume execution time of a program is sum of the time of all executions of individual statements, *and* that the execution time of an individual statement is constant.

(How reasonable is this assumption, really?)

Thus, we can, for each statement take its execution time times the number of times it is executed, and then sum over all statements

Say statements 1 -7 have execution times $c_1,\ldots,c_7$ Each statement s is executed t_s times
Then total execution time is $\sum_{s=1}^{7} t_s \cdot c_s$

Let's calculate the different t_s on wyteboard and see what we get...

Results of analysis:

Best-case execution time (with n=length of A):

$c_1 n + (c_2 + c_3 + c_4 + c_7)(n-1)$ order $\Theta(n)$ (what do we mean by this?)

Worst-case execution time:

$$(\frac{c_4}{2} + \frac{c_5}{2} + \frac{c_6}{2})n^2 + (c_1 + c_2 + c_3 + \frac{c_4}{2} - \frac{c_5}{2} - \frac{c_6}{2} + c_7)n - (c_2 + c_3 + c_4 + c_7)$$

order $\Theta(n^2)$.

Average-case execution time:

order $\Theta(n^2)$.

What is $\Theta(n)$(and $\Theta(n^2)$)?

$\Theta(f)$ means *the set of functions that grow as fast as* f *when the argument becomes large enough*

Mathematically,

$$\Theta(f) = \{\, g \mid \exists c_1, c_2, n_0 > 0. \forall n \geq n_0. 0 \leq c_1 f(n) \leq g(n) \leq \; c_2 f(n) \,\}$$

"h is order $\Theta(f)$" means h $\Theta(f)$

What kind of mathematics did we use?

Proving h $\Theta(f)$is done by ordinary mathematical methods (reasoning about inequalities, deciding the existence of certain entities, ...).

We have used facts about sums, algebraic manipulations. Probability theory is used to get the average-case execution time. In short, we have used only traditional mathematics. Note, however, that certain details are swept under the carpet! In particular, implicit assumptions about semantics of loops etc. (for instance, how do we know the body of `for j = 2 to n` is executed exactly `n-1` times?).

Example 3: **Complexity Theory**

Deals with *problems*, or *classes* of problems, rather than single algorithms

Tries to find limits for how costly a certain problem (or class of problems) is on a certain machine model.

An example of a problem is sorting:

> $O(n \log n)$ algorithms are known (for sequential machine model)
>
> Not proved whether this is the ultimate lower limit!

5.6.2.1 A famous class of problems (a *complexity class*):

NP, the set of all problems that can be solved in polynomial time ($O(n^k)$ for some *k*) by a non-deterministic Turing machine

(*NP*   set of problems solvable by "brute force parallel search" in polynomial time)

A "hardest" problem in *NP* is known, 3-SAT: if 3-SAT always can be solved in polynomial time then *each* problem in *NP* can be solved in polynomial time

Proof by encoding: that each problem in *NP* can be translated into 3-SAT such that a solution of the translated problem solves the original problem (in polynomial time relative to the time to solve the translated problem in 3-SAT)

(Or the reverse: if there is *any* problem in *NP* that *cannot* be solved in polynomial time, then 3-SAT cannot either!)

3-SAT is *NP-complete*

Proof that *another* problem *Q* is *NP*-complete:

1 Show Q   *NP*

2 Show that if one can solve *Q* in (sequential) polynomial time then 3-SAT can be solved in polynomial time (via translation of 3-SAT into *Q*)

Complexity Theory uses *encodings* a lot.

Another famous complexity class:

P, the set of all problems that can be solved in polynomial time by a *deterministic* Turing machine (cf. *NP*)

The class of problems that can be solved *sequentially* in polynomial time (like, for instance, sorting).

Open question: is *P = NP*?

Generally assumed that *P=NP*, but has not been proved!

If indeed *P=NP*, then the concept of *NP*-completeness becomes quite meaningless

Example 4: Complete Partial Orders

A mathematical model that is good for describing the meaning of recursive definitions

It can be thought of as describing *information contents.*

The idea is that more and more information about the result of a computation becomes available as the computation proceeds.

We may want to describe infinite computations (e.g., a server computing an unbounded number of results). Therefore we need to "go to the limit" in the model.

A *complete partial order* (cpo) is a structure *D,* , where *D* is a set and is a binary relation on *D* such that:

it is a partial order; and

there is a *bottom element* D such that *d* for all *d* *D*; and

for each infinitely non-decreasing chain

d_0 d_1 … d_i …, there is a *least upper bound* $_{i=0}$ d_i.

0

Least upper bound means:

- $d_j \sqsubseteq \bigsqcup_{i=0}^{\infty} d_i$ for all j; and
- if $e \sqsubset \bigsqcup_{i=0}^{\infty} d_i$ for some $e \in D$ then there is a k such that, for all $j > k$, holds that $e \sqsubset d_j$.

Definition: a function $f\colon D \to D$ is *continuous* if $f(\bigsqcup_{i=0}^{\infty} d_i) = \bigsqcup_{i=0}^{\infty} f(d_i)$ for all least upper bounds $\bigsqcup_{i=0}^{\infty} d_i$

An interesting result in the theory of cpo's is *Kleene's Fixed-Point Theorem*:

Theorem:

Let f be continuous. Define $\mathit{fix}(f) = \bigsqcup_{i=0}^{\infty} f^i(\bot)$. Then $\mathit{fix}(f)$ is the least solution w.r.t. $\sqsubseteq$ of the equation $d = f(d)$.

We say it is the *least fixed point* of *f.* Many recursive definitions are of the form *d* = *f* (*d*). For instance, a recursively defined function in a simple functional language:
fac(n) = (if n == 0 then 0 else n * fac(n-1)).
This can be seen as an equation fac =*f*(fac).

With suitably chosen cpo, Kleene's fixed-point theorem gives a well-defined mathematical meaning to `fac.` Cpo's are central in *denotational semantics* of programming languages.

What is logic?Some would say the following:

$\wedge$	F	T
F	F	F
T	F	T

$$P \wedge Q \implies P$$

$$\forall x.p(x) \wedge q(x) \iff \forall x.p(x) \wedge \forall x.q(x)$$

But they are wrong!

Logic deals with *formal systems for derivations*, that is, "how to prove things", and *properties of derivations* (proofs).

Thus, logic is a *metatheory*, which deals with properties of other theories!

Example of a result in logic: "in all logical systems that can express arithmetic on whole numbers, it is possible to formulate statements that can neither be proved nor disproved"

Logical systems consist of: *Axioms*, which are assumed to hold without proof *Inference rules*, of the form "given these premises, this conclusion can be inferred"

Inference rules are often written on the form

$$\frac{\text{premise } 1 \quad \cdots \quad \text{premise } n}{\text{conclusion}}$$

Example (modus ponens in propositional logic):

$$\frac{P \quad P \implies Q}{\mathrm{Q}}$$

A proof of a statement in a logic is a *finite derivation of the statement as a conclusion, starting from axioms*

The set of provable statements is thus an inductively defined set, starting with the axioms as "base cases".

Even the set of *proofs* is an inductively defined set!

Induction principle for proofs (in some given logic). Show that the property P is true for all proofs:

1. Show that P holds for all axioms ("least possible proofs").
2. Show, for each derivation rule and its possible premises, that if P holds for each of the proofs of the premises, then P holds also for the proof of the conclusion.
3. Conclude that P holds for all proofs in the proof system.

In Computing Science, systems are sometimes modeled directly with logical inference systems!

Example:

- Operational semantics for programming languages
- Type systems

Properties of these can be proved with induction over derivations

We will consider a type system and an operational semantics for a small, typed language

Our language:

Three types: `int` for integers, `bool` for boolean values, `void` for programs

Constants `17`, `0`, `true`, `false`, ...

Identifiers (program variables) `X`, `Y`, `Z`, ...

Arithmetical expressions `17`, `X+99*Y`, ...

Boolean expressions `false`, `Y and Z`, `X > 17`, ...

Program statements:

> Assignments `X=e`, where `e` is an arithmetical or Boolean expression.
>
> Sequencing of statements `c1; c2`, where `c1`and `c2` are statements.
>
> Conditional statements `if b then c1 else c2`, where `b` is a Boolean expression, and `c1` and `c2` are statements.
>
> Looping statements `while b do c`, where `b` is a Boolean expression, and `c` is a statement

E.g.,: `X = 5; while X > 0 do {X = X -1}`

A type system for the language:

Statements of the form `e:type`, where `e` is a piece of a program

(a program statement, or arithmetical expression, or Boolean expression).

Axioms:

`n:int` for each integer constant `n`

`true: bool` and `false:bool`

(A function *type* from identifiers to {`int, bool`} that tells, for each identifier, what type it has)

`X:` *type*(`X`) for each identifier `X`

Derivation rules (typical examples):

$$\frac{\texttt{b1:bool} \quad \texttt{b2:bool}}{\texttt{b1 and b2:bool}} \quad \text{boolean operators}$$

$$\frac{\texttt{a1:int} \quad \texttt{a2:int}}{\texttt{a1 > a2:bool}} \quad \text{relational operators}$$

$$\frac{\texttt{a1:int} \quad \texttt{a2:int}}{\texttt{a1 + a2:int}} \quad \text{arithmetical operators}$$

$$\frac{\texttt{X:t} \quad \texttt{a:t}}{\texttt{X = a:void}} \quad \text{assignment}$$

$$\frac{\texttt{c1:void} \quad \texttt{c2:void}}{\texttt{c1;c2:void}} \quad \text{sequencing}$$

$$\frac{\texttt{b:bool} \quad \texttt{c1:void} \quad \texttt{c2:void}}{\texttt{if b then c1 else c2:void}} \quad \text{branching}$$

$$\frac{\texttt{b:bool} \quad \texttt{c:void}}{\texttt{while b do c:void}} \quad \text{repetition}$$

Exercise 1:

Derive a type for the expression (program)

`X = 5;`

`while X > 0 do {X = X -1}` when *type*(`X`) = `int`

(That is, construct a proof of a statement of the form `c:t`, where `c` is the program above)

Exercise 2:

Try to derive a type for the expression

`X = 5; while X > 0 do {X = false}` when *type*(X) = `int`

An example of a property one may wish to prove for the type system itself:

Theorem: if `e:t1` and `e:t2` then it holds that `t1 =t2` (each well-typed expression has a unique type)

Proof by induction over derivations!

A formal, operational semantics for the language:

We give the semantics as a logical inference system (just as the type system)

Derivable facts are statements about programs and program parts that tell what result(s) their executions can yield.

The facts use *states*

A state σ is a mapping from program variables to values (i.e., a description of the "current contents" in memory)

Executing a program, starting in a state σ, will transform the state into some new state σ'

Derivable facts are relations $\langle \mathrm{c}, \sigma \rangle \rightarrow \sigma'$

"Starting in state σ, executing program `c` can yield the state σ' "

Inference rules to derive facts of this form

For arithmetic and boolean expressions we have similar relations as derivable facts:

$\langle \mathrm{a}, \sigma \rangle \rightarrow n$ and $\langle \mathrm{b}, \sigma \rangle \rightarrow b$

Here, n is a natural number and b a boolean value

This is since evaluating pure expressions in our language does not change the state, only a value is returned

Read $\langle \mathrm{a}, \sigma \rangle \rightarrow n$ as "if a is evaluated when in state σ, then the number n can be returned"

Similarly for boolean expressions

Axioms and inference rules for arithmetic expressions:

Evaluation of numbers:

$\langle \mathrm{n}, \sigma \rangle \rightarrow n$

Evaluation of numeric program variable:

$\langle \mathrm{X}, \sigma \rangle \rightarrow \sigma(\mathrm{X})$

Evaluation of sums:

$$\frac{\langle \texttt{a1}, \sigma\rangle \rightarrow n_1 \quad \langle \texttt{a2}, \sigma\rangle \rightarrow n_2}{\langle \texttt{a1+a2}, \sigma\rangle \rightarrow n_1 + n_2}$$

Etc.

Axioms and inference rules for boolean expressions:

Evaluation of boolean constants:

$$\langle \texttt{true}, \sigma\rangle \rightarrow \mathit{true} \quad \langle \texttt{false}, \sigma\rangle \rightarrow \mathit{false}$$

Evaluation of boolean program variable:

$$\langle \texttt{X}, \sigma\rangle \rightarrow \sigma(\texttt{X})$$

Evaluation of inequality:

$$\frac{\langle \texttt{a1}, \sigma\rangle \rightarrow n_1 \quad \langle \texttt{a2}, \sigma\rangle \rightarrow n_2 \quad n_1 > n_2}{\langle \texttt{a1 > a2}, \sigma\rangle \rightarrow \mathit{true}}$$

$$\frac{\langle \texttt{a1}, \sigma\rangle \rightarrow n_1 \quad \langle \texttt{a2}, \sigma\rangle \rightarrow n_2 \quad n_1 \leq n_2}{\langle \texttt{a1 > a2}, \sigma\rangle \rightarrow \mathit{false}}$$

And similarly for the other relational operators …

Evaluation of negation:

$$\frac{\langle \texttt{b}, \sigma\rangle \rightarrow \mathit{false}}{\langle \texttt{not b}, \sigma\rangle \rightarrow \mathit{true}} \quad \frac{\langle \texttt{b}, \sigma\rangle \rightarrow \mathit{true}}{\langle \texttt{not b}, \sigma\rangle \rightarrow \mathit{false}}$$

Exercise: figure out suitable inference rules that give the semantics of boolean connectives `and`, `or`

Axioms and inference rules for execution of programs:

Execution of assignment of numeric variable:

$$\frac{\langle \texttt{a}, \sigma\rangle \rightarrow n}{\langle \texttt{X = a}, \sigma\rangle \rightarrow \sigma[n/\texttt{X}]}$$

$\sigma[n/X]$is state for which:

$\sigma[n/X](X)=n$

$\sigma[n/X](Y)=\sigma(Y)$, X != Y

Execution of sequenced programs:

$$\frac{\langle \texttt{c1}, \sigma\rangle \to \sigma'' \quad \langle \texttt{c2}, \sigma''\rangle \to \sigma'}{\langle \texttt{c1;c2}, \sigma\rangle \to \sigma'}$$

Execution of conditionals:

$$\frac{\langle \texttt{b}, \sigma\rangle \to \mathit{true} \quad \langle \texttt{c1}, \sigma\rangle \to \sigma'}{\langle \texttt{if b then c1 else c2}, \sigma\rangle \to \sigma'}$$

$$\frac{\langle \texttt{b}, \sigma\rangle \to \mathit{false} \quad \langle \texttt{c2}, \sigma\rangle \to \sigma'}{\langle \texttt{if b then c1 else c2}, \sigma\rangle \to \sigma'}$$

Execution of while-statement:

$$\frac{\langle \texttt{b}, \sigma\rangle \to \mathit{false}}{\langle \texttt{while b do c}, \sigma\rangle \to \sigma}$$

$$\frac{\langle \texttt{b}, \sigma\rangle \to \mathit{true} \quad \langle \texttt{c}, \sigma\rangle \to \sigma'' \quad \langle \texttt{while b do c}, \sigma''\rangle \to \sigma'}{\langle \texttt{while b do c}, \sigma\rangle \to \sigma'}$$

This one tends to require some thinking!

Exercise: Find state σ' such that

$$\langle \texttt{while X > 0 do X = X - 1}, \sigma\rangle \to \sigma'$$

if $\sigma(\texttt{X}) = 1$!

What can we do with this theory?

We can define a sensible *equivalence* between programs:

$$\texttt{c1} \sim \texttt{c2} \quad \text{iff} \quad \forall \sigma, \sigma'. \langle \texttt{c1}, \sigma\rangle \to \sigma' \iff \langle \texttt{c2}, \sigma\rangle \to \sigma'$$

"`C1`, when started in state σ, can yield state σ' precisely when `c2` can"

We can now use the equivalence to prove *correctness of program transformations* (as used by, e.g., an optimizing compiler)

Exercise: let `w = while b do c`. Show that `w~ if b then c; w else skip`

(Correctness of "loop unrolling")

(Extended language with `skip` statement, you figure out how to give its semantics)

5.7 Inductive Methods in Computing Science & engineering

5.7.1 Mathematical Induction

It is important to make a clear distinction between empirical induction and mathematical induction. In the empirical induction we try to establish the law. In the mathematical induction we have the law already formulated. We are trying to prove that it holds generally. The basis for mathematical induction is the property of the well-ordering principle for the natural numbers.

5.7.1.1 The Principle of Mathematical Induction

Suppose P(n) is a statement involving an integer n.
Than to prove that P(n) is true for every n $\geq n_0$ it is sufficient to show these two things:

1 $P(n_0)$ is true.
2 For any k $\geq n_0$, if P(k) is true, then P(k+1) is true.

A proof by induction is an application of this principle. The two parts of inductive proof are called the basis step and the induction step. In the induction step, we assume that statement is true in the case n = k, and we call this assumption the induction hypothesis.

5.7.1.2 The Strong Principle of Mathematical Induction

Suppose P(n) is a statement involving an integer n. In order to prove that P(n) is true for every $n \geq n_0$ it is sufficient to show these two things:

5. $P(n_0)$ is true.
6. For any k $\geq n_0$, if P(n) is true for every n satisfying $n_0 \leq n \leq k$, then P(k+1) is true.

A proof by induction using this strong principle follows the same steps as the one using the common induction principle. The only difference is in the form of induction hypothesis. Here the induction hypothesis is that k is some integer $k \geq n_0$ and that all the statements $P(n_0)$, $P(n_0 +1)$, …, P(k) are true.

Example 1. Proof by strong induction

P(n): n is either prime or product of two or more primes, for $n \geq 2$. Basic step. P(2) is true because 2 is prime. Induction hypothesis. k ≥ 2, and for every n satisfying 2 ≤ n ≤ k, n is either prime or a product of two or more primes.

Statement to be shown in induction step. If k+1 is prime, the statement P(k+1) is true. Otherwise, by definition of prime, k+1 = r*s, for some positive integers r and s, neither ofwhich is 1 or k+1. It follows that 2 ≤ r ≤ k and 2 ≤ s ≤ k. By the induction hypothesis, both r and s are either prime or product of two or more primes. Therefore, k+1 is the product of two or more primes, and P(k+1) is true.
The strong principle of induction is also referred to as the principle of complete induction, orcourse-of-values induction. It is as intuitively plausible as the ordinary induction principle; in fact, the two are equivalent.

As to whether they are true, the answer may seem a little surprising. Neither can be proved using standard properties of natural numbers. Neither can be disproved either!

This means essentially that to be able to use the induction principle, we must adopt it as an axiom. A well-known set of axioms for the natural numbers, the Peano axioms, includes one similarto the induction principle.

5.7.1.3 Peano's Axioms

1. N is a set and 1 is an element of N.
2. Each element x of N has a unique successor in N denoted x'.
3. 1 is not the successor of any element of N.
4. If x' = y' then x = y.
5. (Axiom of Induction) If M is a subset of N satisfying both: 1 is in M x in M implies x' in M then M = N.

The assertion that every nonempty subset on N (set of natural numbers) has a smallest element is usually referred to as the well-ordering principle for the natural numbers. It probably seems so obvious that no proof is necessary. It turns out that it is also equivalent to the two versions of the induction principle. In other words, it is also impossible to prove without using induction or something comparable.

Example 2 An inductive definition of natural numbers.

The natural numbers can be pulled, more or less, from the thin air, by the operation of collection (Kimber, Smith).

We start with no numbers, and collect what we have. This collection of nothing we write as {}. Call this representation zero. Now we have something, so we can collect our zero: {{}}. Call this one. Now we have two things to collect:{{}, {{}}}. This last collection, call it two, is the collection of two items:

The collection of nothing {}, and
The collection of collection of nothing {{}}.

Now we have three objects to collect, and the process continues, defining operations on all natural numbers in terms of operations on collection. It means that we have specified the base object (in this case the empty object) and an operation to produce new objects from previously defined ones (in this case, collection).

5.7.2. Recursive definitions and proofs by induction

Induction over natural numbers

Show that the property *P* is true for all natural numbers (whole numbers >= 0)

1. Show that *P* holds for 0
2. Show, for all natural numbers *n*, that if *P* holds for *n* then *P* holds also for *n*+1
3. Conclude that *P* holds for all *n*

Formulated in formal logic:

$$[P(0) \wedge \forall n.P(n) \implies P(n+1)] \implies \forall n.P(n)$$

Example: show that

$\sum_{i=0}^{n}(2i-1) = n^2 - 1$ for all natural numbers n.

Why does induction over the natural numbers work? The set of natural numbers N is an *inductively defined set.*

(A variation of) Peano's axiom:

- $0 \in \mathbf{N}$
- $\forall x.x \in \mathbf{N} \implies s(x) \in \mathbf{N}$
- $\forall x.0 \neq s(x)$
- $\forall x, y.x \neq y \implies s(x) \neq s(y)$

$s(x)$ "successor" to x, or $x+1$

$$\begin{array}{ccccccccc} 0 & \rightarrow & s(0) & \rightarrow & s(s(0)) & \rightarrow & s(s(s(0))) & \rightarrow & \cdots \\ 0 & & 1 & & 2 & & 3 & & \cdots \end{array}$$

Note how proofs by induction over the natural numbers follow the structure of their definition.

Also note that the definition of N is given a well-defined meaning by Kleene's fixed-point theorem:

$$\emptyset \subseteq \{0\} \subseteq \{0,1\} \subseteq \{0,1,2\} \subseteq \cdots$$

Inductively defined sets are typically sets of *infinitely* many fi*nite* objects

Entities in Computing Science are often finite (data structures, programs, ...)

Example: mathematical definition of the set of (finite) lists of integers, *List*

- $NIL \in List$
- $z \in \mathbf{Z} \wedge l \in List \implies z : l \in List$
- $\forall z, l.(z : l \neq NIL)$
- $\forall z, z', l, l'.(z : l = z' : l' \implies z = z' \wedge l = l')$

List is an *abstract data type* – the internal representation is hidden.

Need *not* be represented as linked structures in memory (but could be represented)
Typical elements in *List*: *NIL*, 3:(4:*NIL*) .

Note similarity with the set of natural numbers.

List is the set of finite (but arbitrarily long) lists of numbers.

We can define mathematical functions over lists. An example:

length(*NIL*) = 0

length (*z* : *l*) = 1 + *length*(*l*), for all $z \in \mathbf{Z}$ och $l \in List$

Defines *length* as a function *List* N

→

Recursive definition: *length* itself is used in the definition! (Seemingly circular definition, but note that *length* is not applied on the same argument in the right-hand side)

Exercise: show that *length* really is a well-defined partial function! (That is, that each function value is uniquely determined by the definition.)

Note the similarity with function definitions in some functional languages

Exercise: show $\forall l.length(l) \geq 0$

How to do this?

For each inductively defined set there is an *induction principle* that follows the inductive definition of the set. Induction is performed on the "pieces" of an entity that is built up from smaller entitites (e.g., a list).

Induction principle for *List*. Show that the property *P* is true for all lists of integers:

1 Show that *P* holds for *NIL*

2 Show, for all lists *l* and integers *z*, that if *P* holds for *l*, then *P* holds also for *z* : *l*

3 Conclude that *P* holds for all lists of integers

"Mathematical" lists, and functions like *length*, can be seen as *abstract specifications* of what lists are and how functions on them should work

Consider the following piece of C code:

```
#define NIL 0

struct list{
   int contents;
   list *succ;
}
int len(list *l){
 int length;
 length = 0;
 while(l != NIL) {length++; l = l -> l.succ;}
 return(length);
}
```

Interesting things to verify:

- That lists of `list`-structs represent "mathematical" lists in *List* correctly
- That `len(l)=` *length*(*l*) always, when `l` is the representation in C of *l*.

The verification requires that a *formal semantics* is defined for C programs, and that we define exactly what it means that a C entity represents a "mathematical" entity.

5.7.3. The Raw Image of Scientific Method

Usual Naïve Image of Scientific Method is nothing but the narrow inductivist conception of scientific enquiry. "If we try to imagine how a mind of superhuman power and reach[18], but normal so far as logical processes of its thought are concerned, … would use the scientific method, the process would be as follows: First, all facts would be observed and recorded, without selection or a priori guess as to their relative importance. Secondly, the observed and recorded facts would be analyzed, compared, and classified, without hypothesis or postulates other than those necessarily involved in the logic of thought. Third, from this analysis of the facts generalizations would be inductively drawn as to the relations, classificatory or causal, between them. Fourth, further research would be deductive as well as inductive, employing inferences from previously established generalizations."[19]

This narrow idea of scientific inquiry is groundless, for several reasons:

> A scientific investigation as described above could never get off the ground. Its first phase could never be carried out, for a collection of all facts would take infinite time, as there are infinite number of facts. Of course the only possible way to do data collection is to take only relevant facts. But in order to decide what is relevant and what is not, we have to have a theory or at least a hypothesis about what is it we are observing. Empirical facts or findings can be qualified as logically relevant or irrelevant only in reference to a given hypothesis, and not in reference to a given problem. A hypothesis is needed to give the direction to a scientific investigation.
>
> A set of empirical "facts" can be analyzed and classified in many different ways. Without hypothesis, analysis and classification are blind.
>
> Induction is sometimes imagined as a method that leads, by mechanical application of rules, from observed facts to corresponding general principles. Unfortunately, such rules do not exist!

We can mention one simple reason why it is not possible to derive hypothesis (theory) directly from the data. For example, theories about atoms contain terms like "atom", "electron", "proton", "psi-function", etc; yet what one actually measures are spectra (wave lengths, traces in bubble chambers, calorimetric data, etc. So the theory is formulated on a completely different (and more abstract) level than the observable data. The transition from data to theory requests creative imagination. Scientific hypothesis is formulated based on "educated guesses" at the connections between the phenomena under study, at regularities and patterns that might underlie their occurrence. Scientific guesses are completely different from any process of systematic inference. The discovery of important mathematical theorems, like the discovery of important theories in empirical science, requires inventive ingenuity. It calls for imaginative, insightful guessing. The interests of scientific objectivity are assured by the demand for an objective validation of such conjectures (hypotheses). And when e.g. a mathematical proposition has been proposed as a conjecture, its proof or disproof still requires inventiveness and ingenuity. For not even the rules of deductive inference provide a general mechanical procedure for constructing proofs. Their systematic role is rather the modest one of criteria of soundness of arguments offered as proofs. An argument will constitute a valid mathematical proof if it proceeds from axioms to the proposed theorem by a chain of steps each of which is valid according to one of the rules of deductive inference.

Scientific knowledge is not a result of applying some inductive inference procedure to the collected data, but rather by what is called hypotetico-deductive method, i. e. by inventing hypothesis as trial answer to a problem under study, and then subjecting hypothesis to empirical/logical test.

5.10 Induction vs Deduction, Hypothetico-Deductive Method

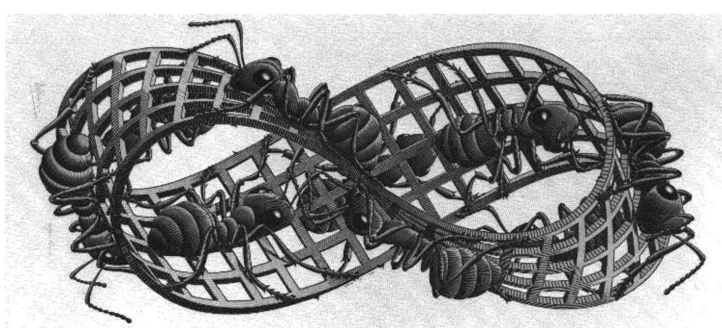
Talking about the deduction and induction as being a separate methods, we have not yet pointed out the fact that they actually occur as a part of the common hypothetico-deductive method, which can be simplified in the following scheme:

> Ask a question and formulate a hypothesis/educated guess (induction).
> Make predictions about the hypothesis (deduction).
> Test the hypothesis (induction).

Deduction, if applied correctly, leads to true conclusions. But deduction itself is based on the fact that we know something for sure. For example we know the general law which can be used to deduce some particular case, such as "All humans are mortal. Socrates is human. Therefore is Socrates mortal." How do we know that all humans are mortal? How have we arrived to the general rule governing our deduction? Again, there is no other method at hand but (empirical) induction.

In fact, the truth is that even induction implies steps following deductive rules. On our way from specific (particular) up to universal (general) we use deductive reasoning. We collect the observations or experimental results and extract the common patterns or rules or regularities by deduction. For example, in order to infer by induction the fact that all planets orbit the Sun, we have to analyse astronomical data using deductive reasoning.

In short: deduction and induction are - like two sides of a piece of paper - the inseparable parts of our thinking process.



Möbius strip, Maurits Cornelis Escher

5.11 Repetitions, Patterns, Identity

5.11.1 Repetitions and Patterns

"The choice is always the same. You can make your model more complex and more faithful to reality, or you can make it simpler and easier to handle. Only the most naïve scientist believes that the perfect model is the one that perfectly represents reality. Such a model would have the same drawbacks as a map depicting every park, every street, every building, every three, every pothole, every inhabitant, and every map[10]."

James Gleick, Chaos

Empirical method relies on observations and experiments, which lead to a collection of data describing phenomena.

In order to establish a pattern or regularity of behavior, we have to analyze (compare) the results (data) searching for similarities (repetitions) and differences.

All repetitions are approximate: the repetition B of an event A is not identical with A, or indistinguishable from A, but only similar to A.

As repetition is based upon similarity, it must be relative. Two things that are similar are always similar in certain respects.

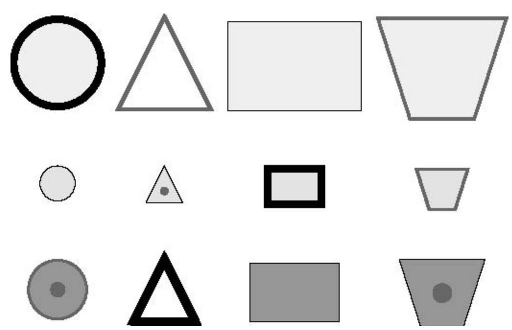


Figure 5 What do we call similar here?

We find that some of the figures above are similar with respect to shading, other are similar with respect to shape and some are similar with respect to edge or size.

Above diagrams illustrate the point that things may be similar in different respects, and that any two things which are from one point of view similar may be dissimilar from another point of view.

Generally, similarity, and consequently repetition, always presuppose the adoption of a point of view: some similarities or repetitions will strike us if we are interested in one problem, and others if we are interested in another problem. This shows how naïve it is to look upon repetition (pattern) as something ultimate, absolute or given by "direct observation”.

Searching for similarity and differences leads to classifications i.e. the division of objects or events in different groups/classes. The simplest tool by for classification is the binary opposition or dichotomy (dualism). When we use dichotomy, we only decide if an object is of kind A or of kind ~A. Examples of frequent dichotomies are yes/no, true/false, before/after, more/less, above/below, etc.

Dualism seems to be deeply rooted in the development of human categorization. Jakobson and Halle observe that 'the binary opposition is a child's first logical operation'. Whilst there are no opposites in 'nature', the binary oppositions we employ in our cultural practices have developed historically as they help to generate order out of the dynamic complexity of experience. At the most basic level of individual survival humans share with other animals the need to distinguish between own species and other, safe and dangerous, edible and inedible, dominance and submission, etc.

5.11.2 Identity

The basic feature of experimental method is its reproducibility: It must be possible to establish essentially the same experimental situation in order to obtain the same results. This means that the experimental arrangement can be made with essentially equivalent parts.

What we call “essentially equivalent” (or we can call it “essentially the same”) depends on situation. Even here the principle of information hiding helps us to get a practical “level of resolution” which means information hiding for all objects below that level.

For most purposes, one can, for example, assume that atoms (of certain kind, under well-defined conditions) are in practice identical, as well as their constituent parts. In fact, it is true as long as we are not interested in the structure of certan type of atom (or any other particle). If our focus is on the specific features of an atom, we will find that each atom of a certain type is a complex system of particles (protons, neutrons, electrons, etc) that can be in different energy states, in different spatial organizations etc.

So declaring two systems/particles/states as identical is entirely the matter of focus. For example if we focus on question of how many people in this country are vegetarians, we just treat all people as equal units. If we want to know how many women in this country are vegetarian, we discriminate between men and women in our analysis of people.

We can e.g. also assume that bacteria of particular sort are interchangeable (indistinguishable) in certain context. That enables us to make repeated experiments with different agents and treat all bacteria of the same type as equal. It does not mean that they are identical in the absolute sense. It only means that for our purpose the existing difference does not have any significance.

Let us take ancient atomic theory as another example. The problem of showing that one single physical body- say piece of iron is composed of atoms is at least as difficult as of showing that all swans are white. Our assertions go in both cases beyond all observational experience. The difficulty with these structural theories is not only to establish the universality of the law from repeated instances as to establish that the law holds even for one single instance.
A singular statement like “This swan here is white” may be said to be based on observation. Yet it goes beyond experience- not only because of the word “white”, but because of the word “swan”. For by calling something a “swan”, we attribute to it properties which go far beyond mere observation. So even the most ordinary singular statements are always the interpretations of the facts in the light of theories!

An illustration of identity: Snow myth (http://whyfiles.org/123snow/2.html) Is there any truth to the myth that every one of the countless trillions of snowflakes has a unique shape? Pao-Kuan Wang, an atmospheric physicist who studies flakes and ice in clouds at the University of Wisconsin-Madison, says it depends on how hard you look. "With a microscope, going down to the molecular level, of course they're all different, But without a microscope, at the superficial level, they may look alike."

5.11.3 Frege’s Puzzle about Identity Statements

Here are some examples of identity statements:

117 + 136=253.

The morning star is identical to the evening star.

Mark Twain is Samuel Clemens.

Bill is Debbie’s father.

Frege believed that these statements all have the form "a=b", where ‘a ’ and ‘b ’ are either names or descriptions that denote individuals. He naturally assumed that a sentence of the form "a=b" is true if and only if the object denoted by ‘a ’ is the same as the object denoted by ‘b ’. For example, "117 + 136=253" is true if and only if ‘117 + 136’ and ‘253’ denote the same number. And "Mark Twain is Samuel Clemens" is true if and only if ‘Mark Twain’ and ‘Samuel Clemens’ denote the same person. So the truth of "a=b" requires that the expressions flanking the identity sign denote the same object.

But Frege noticed that on this account of truth, the truth conditions for "a=b" are no different from the truth conditions for "a=a". For example, the truth conditions for "Mark Twain=Mark Twain" are the same as those for "Mark Twain=Samuel Clemens"; not only do the names flanking the identity sign denote the same object in each case, but the object is the same between the two cases. The problem is that the cognitive significance (or meaning) of the two sentences differ. We can learn that "Mark Twain=Mark Twain" is true simply by inspecting it; but we can’t learn the truth of "Mark Twain=Samuel Clemens" simply by inspecting it. Similarly, whereas you can learn that "117 + 136=117 + 136" and "the morning star is identical to the morning star" are true simply by inspection, you can’t learn the truth of "117 + 136=253" and "the morning star is identical to the evening star" simply by inspection. In the latter cases, you have to do some arithmetical work or astronomical investigation to learn the truth of these identity claims.

So the puzzle Frege discovered is: if we cannot appeal to a difference in denotation of the terms flanking the identity sign, how do we explain the difference in cognitive significance between "a=b" and "a=a"?

5.12 Causality and Determinism

5.12.1 Causality

Causality refers to the way of knowing that one thing causes another.

Early philosophers, as we mentioned before, concentrated on the ‘why’ of conceptual issues and questions. Later philosophers concentrated on the ‘how’ of more concrete issues and questions. The change in emphasis from conceptual to concrete coincides with the rise of empiricism.

Hume is probably the first philosopher to postulate a wholly empirical definition of causality. Of course, both the definition of "cause" and the "way of knowing" whether X and Y are causally linked have changed significantly over time. Some philosophers deny the existence of "cause" and some philosophers who accept its existence, argue that it can never be known by empirical methods. Modern scientists, on the other hand, define causality in limited contexts (e.g., in a controlled experiment).

Aristotle's Causality: Any discussion of causality begins with Aristotle's Metaphysics. [11] Aristoteles cause can be described as a set of necessary conditions needed to produce a certain effect.

There Aristotle defined four distinct types of cause (four answers to the question of why):

- the material cause is the stuff from which the thing is made;
- the formal cause is the pattern or structure it has;
- the efficient cause is the agent that imposed this form on that matter; and
- the final cause is the purpose for the thing.

To illustrate these definitions, think of a vase, made (originally) from clay by a potter, as the "effect" of some "cause," Aristotle would say that clay is the material cause of the vase. The vase's form is its formal cause. The energy invested by the potter is its efficient cause. And finally, the potter's intent is the final cause of the vase. Aristotle's final cause involves a teleological explanation and virtually all modern scientists reject teleology. Nevertheless, for Aristotle, all "effects" are purposeful; every thing comes into existence for some purpose (telos).

Modern scientists may also find Aristotle's material and formal causes curious. Can fuel "cause" a fire? Can a mold "cause" an ingot? On the other hand, Aristotle's efficient cause is quite close to what physicists mean by the phrase "X causes Y."

Indeed, this causal type well suited to modern science. An efficient cause ordinarily has an empirical correlate, for example; X is an event (a motion) producing another event, Y (for example another motion). Lacking any similar empirical correlates, material, formal, and (especially) final causes resist all attempts at empirical testing.

Galileo's Causality: Galileo was one of many Enlightenment scientists who wrote explicitly about causality. Galileo viewed cause as the set of necessary and sufficient conditions for an effect. If X and Y are causes of Z, in other words, then Z will occur whenever both X and Y occur; on the other hand, if only X or only Y occurs, then Z will not occur. We can state this more succinctly as "If and only if both X and Y occur, then Z occurs."

There is one problem with Galileo's definition. First, the list of causes for any Z would have to include every factor that made even the slightest difference in Z. This list could be so long that it would be impossible to find something that was not a cause of Z. This makes it

virtually impossible to test many causal hypotheses and, so, it makes Galileo's definition not practically useful to scientists.

Hume's Causality: David Hume's major philosophical work, A Treatise of Human Nature, lays the foundation for the modern view of causality. Hume rejected the existing rationalist concept of cause, arguing that causality was not a real relationship between two things but, rather, a perception. Accordingly, Hume's definition of causality emphasizes three elements that can be verified through observation. According to Hume, "X causes Y" if Precedence: X precedes Y in time. Contiguity: X and Y are contiguous in space and time. Constant Conjunction: X and Y always co-occur (or not occur).

At first glance, Hume's definition seems foolproof. But consider the causal proposition that "day causes night." This proposition satisfies all of Hume's three criteria but, yet, fails to satisfy our common expectation of causality. Day does not cause night and this highlights a potential flaw in Hume's definition. Indeed, each of Hume's three criteria poses special problems for the modern scientific method.

Contiguity (proximity): Spatial contiguity makes good common sense. If a cause occurs in one place, we should seek its effect in its vicinity. In historical context, however, Hume's criterion of spatial contiguity seemed to reject Newton's gravitational model. The orbits of planets, tides, and a range of other mechanical phenomena required action at a distance. In fact, contiguity is not amenable to empirical verification.

Precedence: Precedence also makes good common sense. If a cause occurs today, we should seek its effect tomorrow (or perhaps next week) but would not expect to see the effect yesterday (or perhaps last week). Causes should precede effects, not vice versa, and this implies further that there is a finite delay (maybe no longer than a nanosecond but a delay nevertheless) between cause and effect.

Kant offered the example of a lead ball resting on a cushion and causing an impression (dent) on the cushion. Did the lead ball (X) cause the impression (Y)? If so, X and Y occurred simultaneously.

(In fact there are nevertheless changes in quantum systems that appear instantaneously as a consequence of conservation laws, as e.g. in famous Einstein-Rosen-Podolsky paradox).

Constant Conjunction: The most controversial of Hume's three criteria is constant conjunction. The crux of this controversy can be illustrated by the hypothetical results of a simple experiment. We first culture 1000 bacterial colonies. We then treat 500 of the colonies (selected at random) with an anti-bacterial agent. The remaining 500 colonies are treated with a placebo agent. The controversy can be illustrated by the hypothetical result:

		Yes	No
X Occurs?	Yes	500	
	No		500

		Yes	No
X Occurs?	Yes	495	5
	No	5	495

is consistent with "X causes Y". is consistent with "X does not cause Y"

Though oversimplified, this hypothetical result demonstrates the problematic nature of the constant conjunction criterion. By Hume's criteria, there would be few causal relationships in the biological and social sciences.
Comparing the causality of Galileo and Hume gives an insight into the evolution of causal thought. Although Galileo was clearly a scientist, his definition of causality was not entirely empirical. Hume, on the other hand, is an empiricist.

Hume's causality is based on experiential or sensory relationships. To be sure, Hume argued that "X causes Y" could not be empirically verified but that a hypothetical causal relationship could be tested nevertheless. This sets the stage for an operationalized causality; i.e., a definition explicitly in terms of causality testing.

Mill's Causality: John Stuart Mill concentrated on the problems of operationalizing causality. Mill argued that causality could not be demonstrated without experimentation. His four general methods for establishing causation are

- the method of concurrent variation
  ["Whatever phenomenon varies in any manner, whenever another phenomenon varies in some particular manner, is either a cause or an effect of that phenomenon, or is connected with it through some fact of causation."];
- the method of difference
  ["If an instance in which the phenomenon under investigation occurs and an

instance in which it does not occur, have every circumstance in common except for one, that one occurring in the former; the circumstances in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause of the phenomena."];

- the method of residues
  ["Subduct from any phenomena such part as is known by previous inductions to be the effect of certain antecedents, and the residue of the phenomena is the effect of the remaining antecedents."]; and
- the method of agreement
  ["If two or more instances of a phenomena under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomenon."]. All modern experimental designs are based on one or more of these methods.

Probabilistic Causality: One approach to the practical problem posed by Hume's constant conjunction criterion is to make the criterion probabilistic. If we let P(Y | X) denote the probability that Y will occur given that X has occurred, then constant conjunction requires that
P(Y | X)=1 and P(Y | ~X)=0 where ~X indicates that X has not occurred. The problem, of course, is that biological and social phenomena virtually never satisfy this criterion. Probabilistic causalities address this problem by requiring only that the occurrence of X make the occurrence of Y more probable. In the same notation, if P(Y | X) > P(Y | ~X) then "X causes Y." While this makes the constant conjunction criterion more practical, however, it raises other problems. To illustrate, suppose that X has two effects, Y1 and Y2, and that Y1 precedes Y2. A widely used example is the atmospheric electrical event that causes lightening and thunder. Since we always see lightening (Y1) before we hear thunder (Y2), it appears that "lightening causes thunder. Indeed, Y1 and Y2 satisfy the probabilistic criterion

P(Y2 | Y1) > P (Y2) that we require of Y1=>Y2.

But in fact, lightening does not cause thunder. The foremost proponent of probabilistic causality, Patrick Suppes, solves this problem by requiring further that Y1 and Y2 have no common cause. As we discover at a later point, research designs constitute a method for ruling out common causes.

Design as Operational Causality: The history of causality can be broken into two eras. The first era begins with Aristotle and ends with Hume. The second era begins with John Stuart Mill and continues today. The difference between Hume and Mill may be unclear; after all, both were orthodox empiricists. But while Hume and Mill had much in common, Hume's causality was largely conceptual. Little attention was paid to the practical problem of implementing the concepts. Mill, on the other hand, described exactly how working scientists could implement (or operationalize) his causality. The most influential modern philosophers have followed Mill's example. Although the field of (experimental) design often deals with causality only implicitly, we can think of design as operationalized causality.

Rubin Causality: Many proposed causalities work well in one context but not in another.

To solve this problem, some modern philosophers have tried to limit their causalities to specific contexts, circumstances, or conditions. Accordingly, Rubin causality (named for Donald B. Rubin) is defined in the limited context of an experimental milieu. Under Rubin causality, only the relationship demonstrated in an experiment is a valid causal relationship.

To illustrate, suppose that we want to measure the effectiveness of an anti-bacterial soap. We apply the soap to a single bacterium. If the bacterium dies, the soap works. But if the bacterium dies, we still have this problem: sooner or later, all bacteria die; maybe this one died of natural causes. We eliminate this (and every other alternative hypothesis) by showing that a placebo treatment does not kill the bacterium. But since the bacterium is already dead, how is this possible?

The fundamental dilemma of causality, according to Rubin, is that, if we use an experimental unit (a bacterium, e.g.) to show that "X causes Y," we cannot use that same unit to show that some "non-X does not cause Y." We solve this dilemma by assuming that all units are more or less the same. This allows us to treat one bacterium with the anti-bacterial soap and another with a placebo. To make sure that the two bacteria are virtually indistinguishable, however, we randomly assign the bacteria to the soap and placebo. Since random assignment is unfeasible in some situations, Rubin causality holds that some variables (e.g., "race") cannot be causes.

Suppes Causality: Probabilistic causality, as proposed by Suppes, is another causality defined for a limited milieu. Where X and ~X denote the occurrence and nonoccurrence of X respectively, Suppes infers that X=>Y if two conditions are satisfied

$$P(Y \mid X) > P(Y \mid \sim X)$$
$$P(Y \mid X \text{ and } Z) \sim= P(Y \mid \sim X \text{ and } Z)$$

The first criterion ensures that the probability that Y will occur given that X has occurred is greater than the complementary probability that Y will occur given that X has not occurred. The second criterion ensures that X and Y are not asynchronous co-effects of Z.

5.12.2 Determinism

Determinism is the philosophical doctrine which regards everything that happens as determined by what preceded it. From the information given by a complete description of the world at time t, a determinist believes that the state of the world at time t + 1 can be deduced; or, alternatively, a determinist believes that every event is an instance of the operation of the laws of Nature.

The wide acceptance of this view, at least in the Western world, was a result of the work of mathematical physics in the 18th and 19th centuries. At this point, it looked as if Newton and his successors had reduced the universe to systems of equations, through which the position of any particle in the universe could be predicted forever, provided that sufficient information was known about the factors affecting its motion (in other words, providing that the position of all other particles in the universe was known).

Determinism is not a position of which Newton himself would have approved, since it denies any need for the existence of God, and seems to reduce human beings to the status of predictable machines. It was, however, regarded by others (for example the 18th-century mathematicians Laplace and Lagrange) as the triumph of science, showing that, through the laws of cause and effect, the future was as fixed as the past.

In science, determinism as a serious philosophy could not survive without the Newtonian (mechanicistic) view of the universe.

The quantum theory brought chance (probability) into physical theory at the beginning of the 20th century, as the uncertainty principle of Werner Heisenberg showed that some events were inherently unpredictable.

Einstein found indeterminacy of quantum theory deeply disturbing: 'God does not play dice', he insisted. Believing that apparent indeterminacy masks a deeper causality, Einstein found himself in a minority among physicists. Many philosophers however (for example Russell) began to feel that the concept of cause and effect was itself just an illusion. [12]

[12] In modern physics, instead of causal relations one is searching for "correlations" between phenomena, which is a symmetric relation, contrary to causation.Feedback processes are yet another example where it is not self-evident which is to be called the cause and which is the effect (as e.g. women's emancipation and education for women).

The new non-determinism which resulted from all this has led to the view, held by many scientists and others throughout the 20th century, that the universe is inherently chaotic, that actions do not lead to consequences. This view was a major influence on existentialism.

A third point of view is 'compatibilism', which holds that it is possible to give an account of human freedom without invoking non-determinist explanations of human action. As applied to political and historical thought determinism describes the views of those who think that individual 'choice' is

- (a) determined by conditions beyond the individual's control, and/or
- (b) that it is the result of free will.

Historical determinism emphasizes the limits imposed by long-term economic, political and social conditions. Economic determinism is the theory which attributes all major social and political interests and actions and their organization to the economic causes.

Economic determinism assumes that individuals are always seeking to maximize their utility, and are always engaged in 'rent-seeking'. In Marxist thought, economic determinism is supported by the doctrine of historical materialism which attempts to explain history as the product of changes in material conditions rather than as the product of changes in ideas, values and culture.

5.13 Limitations in Formal Logical Systems: Gödel’s Theorem

5.13.1 Axiomatic system of Euclid: Shaking up Geometry

Euclid showed that geometry can be built on a set of few axioms (ideas which are considered so elementary and manifestly obvious that they do not need to be proven as any proof would introduce more complex ideas).

When a system requires increasing number of axioms (as e.g. number theory does), doubts begin to arise. How many axioms are needed? How do we know that the axioms aren't mutually contradictory?

Until the 19th century no one was too worried about this. Geometry seemed rock solid. It had stood as conceived by Euclid for 2100 years. If Euclid's work had a weak point, it was his fifth axiom, the axiom about parallel lines. Euclid said that for a given straight line, one could draw only one other straight line parallel to it through a point somewhere outside it.

Around the mid-1800s mathematicians began to experiment with different definitions for parallel lines. Lobachevsky, Bolyai, Riemann and others created new geometries by assuming that there could be several parallel lines through the outside point or there could be no parallel lines. These geometries weren't mere intellectual constructs. In fact, it turns out that Riemann's geometry is better at describing the curvature of space than Euclid's. Consequently Einstein incorporated Riemann's ideas into relativity theory.

These new geometries became known as non-Euclidean, and they worried mathematicians a lot. The view of Kant that mathematics (and in particular Euclidean geometry) was a priori truth had been shattered in its ground. Scientists got far more inclined to accept that mathematical equations were not necessarily the arbiters of what goes on in the universe.

Euclids axioms had been unchanged for centuries. Since arithmetic is more complex than geometry, how could they be sure its axioms were trustworthy?

In a bit of brilliant work, German mathematician, David Hilbert, converted geometry to algebra, showing that if algebra was consistent, so was geometry. This served as a useful cross check but wasn't proof positive of either system.

5.13.2 Axiomatic system of Principia: Paradox in Set Theory

Uneasy mathematicians hoped that Hilbert's plan would fulfil its promise because axioms and definitions are based on commonsense intuition, but intuition was proving to be an unreliable guide. Not only had Riemann created a system of geometry which put commonsense notions on its head, but the philosopher-mathematician Bertrand Russell had found a serious paradox for set theory.

A set is one of the basic ideas of mathematics and logic. It is any collection of items chosen for some characteristic common for all its elements. Handling sets seemed fairly simple.

Russell's paradox was the following: There are two kinds of sets

- normal sets, which do not contain themselves, and
- non-normal sets, which are sets that do contain themselves.

The set of all apples is not an apple. Therefore it is a normal set. The set of all thinkable things is itself thinkable, so it is a non-normal set.

Let 'N' stand for the set of all normal sets. Is N a normal set? If it is a normal set, then by the definition of a normal set it cannot be a member of itself. That means that N is a non-normal set, one of those few sets which actually are members of themselves. But on the other hand…N is the set of all normal sets; if we describe it as a non-normal set, it cannot be a member of itself, because its members are, by definition, normal.

Russell did not feel that this paradox was insurmountable. By redefining the meaning of 'set' to exclude awkward sets, such as "the set of all normal sets," he felt that he could create a single self-consistent, self-contained mathematical system. Using improved symbolic logic, he and Alfred North Whitehead set out to do just that. The result was their masterful three volume Principia Mathematica. However, even before it was complete, Russell's expectations were dashed.

The man who showed once and for all that Russell's aim was impossible was Kurt Gödel. His revolutionary paper was titled "On Formally Undecidable Propositions of Principia Mathematica and Related Systems."

5.13.3 Gödel: Truth and Provability

Kurt Gödel actually proved two extraordinary theorems. They have revolutionized mathematics, showing that mathematical truth is more than logic and computation. Gödel has been called the most important logician since Aristotle. His two theorems changed logic and mathematics as well as the way we look at truth and proof.

Gödels first theorem proved that any formal system strong enough to support number theory has at least one undecidable statement. Even if we know that the statement is true, the system cannot prove it. This means the system is incomplete. For this reason, Gödel's first proof is called "the incompleteness theorem".

Gödel's second theorem is closely related to the first. It says that no one can prove, from inside any complex formal system, that it is self-consistent.

Hofstadter says, "Gödel showed that provability is a weaker notion than truth, no matter what axiomatic system is involved." In other words, we simply cannot prove some things in mathematics (from a given set of premises) which we nonetheless can know are true.

5.13.4 Implications of Gödel's Theorem

What do Gödel's theorems mean? First, Gödel shattered expectations that human thinking could be reduced to algorithms (a step by step, usually repetitive, procedures). Computers use algorithms. What it means is that our thought cannot be a strictly algorithmic process. [13]

In 1959 a disillusioned Russell lamented: "I wanted certainty in the kind of way in which people want religious faith. I thought that certainty is more likely to be found in mathematics than anywhere...But after some twenty years of arduous toil, I came to the conclusion that there was nothing more that I could do in the way of making mathematical knowledge indubitable. "

Gödel showed that "it is impossible to establish the internal logical consistency of a very large class of deductive systems - elementary arithmetic, for example - unless one adopts principles of reasoning so complex that their internal consistency is as open to doubt as that of the sytems themselves."

Observe however, that he did not prove a contradictory statement, that A = ~A, (the kind of thinking that occurs in many Eastern religions). Instead, he showed that no system can decide between a certain A and ~A, even where A is known to be true. Any finite system with sufficient power to support a full number theory cannot be self-contained.

[13] Roger Penrose makes much of this, arguing in Shadows of the Mind that computers will never be able to emulate the full depth of human thought. However, modern AI refute this sort of argument pointing out that computers can be made to switch between different formal systems as well as their logic can be made inconsistent (or not completely consistent).

Hence there are provable truths, unprovable truths, unprovable untruths and provable untruths. Alan Turing has shown that there are as many unprovable truths as there are provable truths.

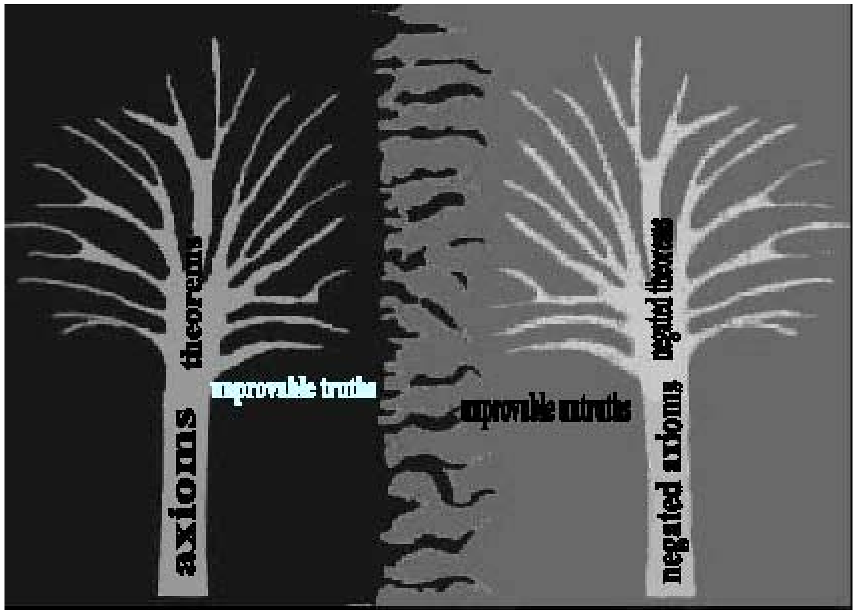


Figure 6 Truth vs. provability according to Gödel

The fact that a particular sentence is neither provable nor disprovable within a system means that it is logically independent of the axioms. Possible adding new axioms (or rules of inference) does not help in Gödel's case. Even if Gödel's sentence is added as an axiom, the new system would contain another unprovable sentence, saying of itself that it is not provable. Gödel's theorems uncover a fundamental limitation of formalization, and they say that this limitation could be overcome only at the price of consistency.

The theorems show that the supposed ideal of formalization - proving all and only all true sentences - is self-contradictory and actually undesirable:

- what good is a formalization that can prove a sentence which says that it is not provable (first theorem)?
- what good is a formalization that can prove its consistency when it would imply that it is not consistent (second theorem)?

Gödel's theorems show that the notions of truth and provability cannot coincide completely, which at first appears disturbing, since, as Quine says, we used to think that mathematical truth consisted in provability.
However, the notion of truth has a problem of its own, namely the classical liar paradox, of which Gödel's sentence is a paraphrase in proof-theoretic terms. Thus, Gödel's theorems do not actually establish any new discrepancy between provability and truth. Assuming consistency, Gödel's sentence is not simply true, because it is not always true i.e. not in all interpretations. If it were, it would be provable: provability is truth in all interpretations (according to the completeness theorem).

Gödel's theorem thus shows that there must always exist such unusual, unintended interpretations of the system; as Henkin says:

" We tend to reinterpret Gödel's incompleteness result as asserting not primarily a limitation on our ability to prove but rather on our ability to specify what we mean ... when we use a symbolic system in accordance with recursive rules [Gödel & the synthetic a priori]. "

Similarly, Polanyi says, though only in connection with the second theorem: " we never know altogether what our axioms mean. "

This characterization of formal language sounds more like something that might be said about ordinary, natural language. Thus, if we take as a characteristic of ordinary language its peculiar inexhaustibility and the frequent discrepancy between intended and expressed meaning, Gödel's theorems would show that, in this respect, formal languages are in fact not so far from natural ones.

Certain similarities between the self-reference in natural language and in Gödel's sentence and theorems have also been noticed at the lexical and pragmatic level [Hofstadter, p. 709]).

5.13.5. A Proof of Gödels Theorem

The proof of the incompleteness theorem is beyond the scope of this text, occupying, as it does at least 30 pages, but we can give a hint here.

We begin with logical theory of numbers. In this theory, there is a single constant, 0 (zero) and a single function S (the successor function). In the intended model, S(0) denotes 1, S(S(0)) denotes 2, and so on. The language has names for all the natural numbers.

The vocabulary also includes the function symbols +, × and Expt (exponentiation), and the usual set of logical connectives and quantifiers $(\wedge, \vee, \neg, \exists, \forall)$

The first step is to notice that the set of sentences that can be written in this language can be enumerated. (Imagine defining an alphabetical order on the symbols and then arranging in alphabetical order each of the sets of sentences of length 1, 2, and so on.)

We can then number each sentence α with #α (the Gödel number). This is important: number theory contains a name for each of its own sentences. Similarly, we can number each possible proof P with a Gödel number G(P), because a proof is simply a finite sequence of sentences.

Now suppose that we have a set A of sequences that are true statements about the natural numbers. Recalling that A can be named by a given set of integers, we can imagine writing in our language a sentence α(j,A) of the following sort:

$\forall i, i$ is not the Gödel number of a proof of the sentence whose Gödel number is j, where proof uses only premises in A.

Then let σ be the sentence α(#σ, A), that is, a sentence that states its own unprovability from A. (That this sentence always exists is true, but not entirely obvious.)

Now suppose that σ is provable from A. Then σ is false (because σ says it cannot be proved). But then we have a false sentence that is provable from A, so A cannot consist of only true sentences – a violation of our premise. Therefore σ is not provable from A. But this is exactly what σ itself claims. Hence σ is a true sentence.

So, we have shown that for any set of true sentences of number theory, and in particular for any set of basic axioms, there exist true sentences that cannot be proved from those axioms.

This establishes, among other things, that we can never prove all the theorems of mathematics within any given system of axioms. Clearly, it was an important discovery for mathematics. It can be summarized as follows:

For any non-trivial formal system F (a formal language, and a set of axioms and inference rules), it is possible to construct so-called "Gödel sentence" G(F) with the following properties:

G(F) is a sentence of F, but cannot be proved within F.
If F is consistent, then G(F) is true.

5.14 Fuzzy Logic

The precision, clarity and beauty of mathematics are the consequence of the fact that the logical basis of classical mathematics possesses the features of parsimony and transparency.

Classical logic owes its success in large part to the efforts of Aristotle and the philosophers who preceded him. In their endeavour to devise a concise theory of logic, and later mathematics, they formulated so-called "Laws of Thought".

One of these, the "Law of the Excluded Middle," states that every proposition must either be True or False.

Even when Parminedes proposed the first version of this law (around 400 B.C.) there were strong and immediate objections. For example, Heraclitus proposed that things could be simultaneously True and not True.

It was Plato who laid the foundation for what would become fuzzy logic, indicating that there was a third region (beyond True and False) where these opposites "tumbled about." Some among more modern philosophers follow the same path, particularly Hegel. But it was Lukasiewicz who first proposed a systematic alternative to the bi-valued logic of Aristotle.

In the early 1900's, Lukasiewicz described a three-valued logic, along with the corresponding mathematics. The third value he proposed can best be translated as the term "possible," and he assigned it a numeric value between True and False. Eventually, he proposed an entire notation and axiomatic system from which he hoped to derive modern mathematics.

Later, he declared that in principle there was nothing to prevent the derivation of an infinite-valued logic. Lukasiewicz felt that three- and infinite-valued logics were the most intriguing, but he ultimately settled on a four-valued logic because it seemed to be the most easily adaptable to Aristotelian logic.

Knuth proposed a three-valued logic similar to Lukasiewicz's, from which he speculated that mathematics would become even more elegant than in traditional bi-valued logic.

It was not until relatively recently that the notion of an infinite-valued logic took hold.

In 1965 Lotfi A. Zadeh published his seminal work "Fuzzy Sets" which described the mathematics of fuzzy set theory, and by extension fuzzy logic. This theory proposed making the membership function (or the values False and True) operate over the range of real numbers [0.0, 1.0].

New operations for the calculus of logic were proposed, and showed to be in principle a generalization of classic logic.

5.14.1. Fuzzy Logic Application

Fuzzy logic is a superset (generalization) of conventional (Boolean) logic that has been extended to handle the concept of partial truth - truth values between "completely true" and "completely false". It was introduced by Lotfi Zadeh of UC/Berkeley in the 1960's as a means to model the uncertainty of natural language.

Zadeh says that rather than regarding fuzzy theory as a single theory, we should regard the process of ``fuzzification" as a methodology to generalize ANY specific theory from a crisp (discrete) to a continuous (fuzzy) form.

Thus recently researchers have also introduced "fuzzy calculus", "fuzzy differential equations", and so on.

Fuzzy Subsets

Just as there is a strong relationship between Boolean logic and the concept of a subset, there is a similar strong relationship between fuzzy logic and fuzzy subset theory.

In classical set theory, a subset U of a set S can be defined as a mapping from the elements of S to the elements of the set {0, 1},

$$U: S \rightarrow \{0, 1\}$$

This mapping may be represented as a set of ordered pairs, with exactly one ordered pair present for each element of S. The first element of the ordered pair is an element of the set S, and the second element is an element of the set {0, 1}. The value zero is used to represent non-membership, and the value one is used to represent membership. The truth or falsity of the statement

x is in U

is determined by finding the ordered pair whose first element is x. The statement is true if the second element of the ordered pair is 1, and the statement is false if it is 0.

Similarly, a fuzzy subset F of a set S can be defined as a set of ordered pairs, each with the first element from S, and the second element from the interval [0,1], with exactly one ordered pair present for each element of S. This defines a mapping between elements of the set S and values in the interval [0,1]. The value zero is used to represent complete non-membership, the value one is used to represent complete membership, and values in between are used to represent intermediate DEGREES OF MEMBERSHIP. The set S is referred to as the universe of discourse for the fuzzy subset F. Frequently, the mapping is described as a function, the membership function of F. The degree to which the statement

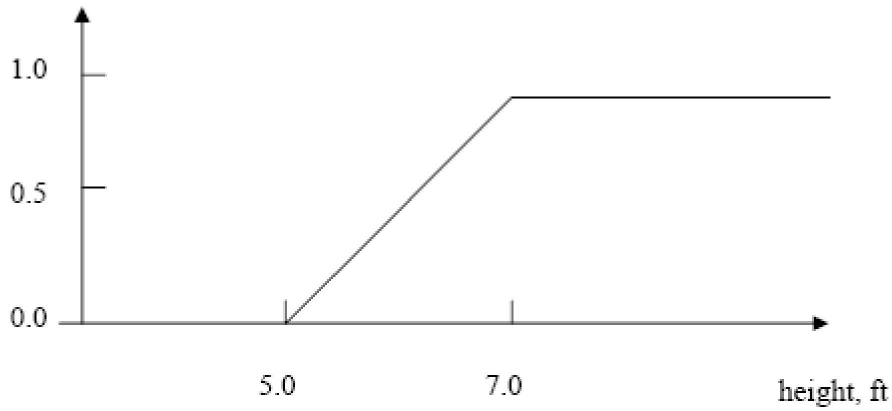
x is in F

is true is determined by finding the ordered pair whose first element is x. The DEGREE OF TRUTH of the statement is the second element of the ordered pair.

In practice, the terms "membership function" and fuzzy subset get used interchangeably. Here is an example. Let's talk about people and "tallness". In this case the set S (the universe of discourse) is the set of people. Let's define a fuzzy subset TALL, which will answer the question "to what degree is person x tall?" Zadeh describes TALL as a linguistic variable, which represents our category of "tallness". To each person in the universe of discourse, we have to assign a degree of membership in the fuzzy subset TALL. The easiest way to do this is with a membership function based on the person's height.

tall(x) = { 0, if height(x) < 5 ft., (height(x)-5ft.)/2ft., if 5 ft. <= height (x) <= 7 ft., 1, if height(x) > 7 ft. }

A graph of this looks like:



Given this definition, here are some example values:

Person	Height	degree of tallness	
Billy	3' 2"	0.00	[I think, 💻]
Fred	5' 5"	0.21	
Drew	5' 9"	0.38	
Erik	5' 10"	0.42	
Mark	6' 1"	0.54	
Kareem	7' 2"	1.00	[depends on who you ask]

Expressions like "A is X" can be interpreted as degrees of truth, e.g., "Drew is TALL" = 0.38.

Note: Membership functions used in most applications almost never have as simple a shape as tall(x). At minimum, they tend to be triangles pointing up, and they can be much more complex than that. Also, the discussion characterizes membership functions as if they always are based on a single criterion, but this isn't always the case, although it is quite common. One could, for example, want to have the membership function for TALL depend on both a person's height and their age (he's tall for his age).

This is perfectly legitimate, and occasionally used in practice. It's referred to as a two-dimensional membership function, or a "fuzzy relation". It's also possible to have even more criteria, or to have the membership function depend on elements from two completely different universes of discourse.

Logic Operations:

Now that we know what a statement like "X is LOW" means in fuzzy logic, how do we interpret a statement like

(X is LOW) and (Y is HIGH) or (not Z is MEDIUM)

The standard definitions in fuzzy logic are:

truth (not x) = 1.0 - truth (x)

truth (x and y) = minimum (truth(x), truth(y))

truth (x or y) = maximum (truth(x), truth(y))

Some researchers in fuzzy logic have explored the use of other interpretations of the AND and OR operations, but the definition for the NOT operation seems to be safe.

Note that if you plug just the values zero and one into these definitions, you get the same truth tables as you would expect from conventional Boolean logic. This is known as the EXTENSION PRINCIPLE, which states that the classical results of Boolean logic are recovered from fuzzy logic operations when all fuzzy membership grades are restricted to the traditional set {0, 1}. This effectively establishes fuzzy subsets and logic as a true generalization of classical set theory and logic. In fact, by this reasoning all crisp (traditional) subsets ARE fuzzy subsets of this very special type; and there is no conflict between fuzzy and crisp methods.

Some examples -- assume the same definition of TALL as above, and in addition, assume that we have a fuzzy subset OLD defined by the membership function:

old (x) = { 0,	if age(x) < 18 yr.
(age(x)-18 yr.)/42 yr.,	if 18 yr. <= age(x) <= 60 yr.
1,	if age(x) > 60 yr. }

And for compactness, let

a = X is TALL and X is OLD
b = X is TALL or X is OLD
c = not (X is TALL)

Then we can compute the following values.

Height	age	X is TALL	X is OLD	a	b	c
3' 2"	65	0.00	1.00	0.00	1.00	1.00
5' 5"	30	0.21	0.29	0.21	0.29	0.79
5' 9"	27	0.38	0.21	0.21	0.38	0.62
5' 10"	32	0.42	0.33	0.33	0.42	0.58
6' 1"	31	0.54	0.31	0.31	0.54	0.46
7' 2"	45	1.00	0.64	0.64	1.00	0.00
3' 4"	4	0.00	0.00	0.00	0.00	1.00

Here is a little conversion table to the metric system:

Feet+Inches = Meters	
3' 2"	0.9652
3' 4"	1.0160
5' 5"	1.6510
5' 9"	1.7526
5' 10"	1.7780
6' 1"	1.8542
7' 2"	2.1844

Fuzzy logic is supposed to be used for reasoning about inherently vague concepts. But it turns out that the useful applications are not in high-level artificial intelligence but rather in lower-level machine control, especially in consumer products.

Usually, fuzzy controllers are implemented as software running on standard microprocessors. A few special-purpose microprocessors have been built that do fuzzy operations directly in hardware, but even these use digital binary (0 or 1) signals at the lowest hardware level. There are some research prototypes of computer chips that use analog signals at the lowest level, but these chips simulate the operation of neurons rather than fuzzy logic.

Fuzzy systems, including fuzzy logic and fuzzy set theory, provide a rich and meaningful addition to standard logic. The mathematics generated by these theories is consistent, and fuzzy logic may be a generalization of classic logic. The applications which may be generated from or adapted to fuzzy logic are wide-ranging, and provide the opportunity for modeling of conditions which are inherently imprecisely defined, despite the concerns of classical logicians. Many systems may be modeled, simulated, and even replicated with the help of fuzzy systems, not the least of which is human reasoning itself.

5.15 Fallacies

What about not properly built arguments? Let us make the following distinction:

- A formal fallacy is a wrong formal construction of an argument.
- An informal fallacy is a wrong inference or reasoning.

5.15.1 Formal Fallacies

Let’s analyze following argument: "All fish swim. Kevin swims. Therefore Kevin is a fish", which appears to be a valid argument. It appears to be a modus ponens. But actually the argument does not follow the form of a modus ponens. The argument is a fallacious and invalid. If we were to use the symbols from the argument forms above, we could write:

If H is true, then so is I.
(As the evidence shows), I is true.
H is true

This form of reasoning, known as the fallacy of "affirming the consequent" is deductively invalid: its conclusion may be false even if premises are true. It is beyond the scope of this text to explore in more detail formal fallacies. We take just another typical example.

An auxiliary hypothesis is an assumption that is used in order to deduce the implications from the hypothesis.

First an example incorrect deduction when using auxiliary hypotheses:
If H and $A_1, A_2, \ldots, A_n$ is true, then so is I. But (As the evidence shows), I is not true. H and $A_1, A_2, \ldots, A_n$ are all false (Comment: One can be certain that H is false, only if one is certain that all of $A_1, A_2, \ldots, A_n$ are all true.)

And now again the fallacy of affirming the consequent:

If H is true, then so are $A_1, A_2, \ldots, A_n$. (As the evidence shows), $A_1, A_2, \ldots, A_n$ are all true.H is true

(Comment: $A_1, A_2, \ldots, A_n$ can be a consequence of some other premise, and not H.)

We shall leave our discussion of formal fallacies by concluding that if an argument is not built correctly, then the argument contains a formal fallacy.

5.15.2 Informal Fallacies

Usually we examine content rather than structure when we analyze an argument. An informal fallacy is a mistake in reasoning related to the content of an argument.

What follows is a brief listing of some of the more common informal fallacies.

Appeal to Authority
Arises when the appeal is made to an authority but the claim is outside of that authority's expertise. Professional athletes that appear in TV commercials have expertise in their

specific sports but not in food, fashion, or banking - unless, of course, they gain that expertise.

Ad Hominem
Arises when the character of your opponent is called into question instead of the validity of their position. A person's character has nothing to do with the validity of their position!

False Cause
Arises when one event is said to cause another simply because the one event follows the other in time.

Leading Question
Arises when the conclusion is assumed in the premises. Leading question is sort of a circular argument.

Appeal to Emotion
Arises when the persuasive point of the argument is not found within the truth of the argument but in how the argument "tugs at our heart strings." Appealing to pity is a very common fallacy. Appealing to fear is another "tried-and-true" fallacious method for winning hearts and minds.

Straw Man
Arises when the arguer attacks a position that is represented as the opponent's position which is really not the case. The common technique is to reinterpret the opponent's position usually in a way that is easy to dismiss, and defeat it, then credit yourself with victory. This type of argument is very common in political debate.

Equivocation
Arises when the meaning of a term becomes confused. When we equivocate we are employing a genuine meaning of a term but we are not employing the most appropriate meaning.

Composition
Arises when the attributes of the parts are applied to the whole. One example might be to assume that if all the parts of a watch are in perfect order then the watch itself must be in perfect order.

Division
Arises when the attributes of the whole are applied to the parts. For instance, one might argue that since a person was from a well-to-do family that the person in question was wealthy.

It is of course important to avoid both formal and informal fallacies within arguments. Every time a fallacy is invoked, the argument fails! Special attention should be paid to informal fallacies that can be difficult to detect. We should in the least examine the content of what we are saying before we embed our premises within our arguments. Isolating, identifying, and eliminating fallacies will help us understand certain position better and will help us defend that position more effectively.

5.15.3 Some not Entirely Uncommon "Proof Techniques"

Proof by vigorous handwaving
Works well in a classroom or seminar setting.

Proof by cumbersome notation
Best done with access to at least four alphabets and special symbols.

Proof by exhaustion
Proof around until nobody knows if the proof is over or not.

Proof by omission
'The reader may easily supply the details. "The other 253 cases are analogous.'

Proof by obfuscation
A long plotless sequence of true and/or meaningless syntactically related statements.

Proof by example
The author gives only the case n = 2 and suggests that it contains most of the ideas of the general proof.

Proof by funding
How could three different government agencies be wrong?

Proof by eminent authority
'I saw Karp in the elevator and he said it was probably NP-complete.'

Proof by personal communication
'Eight-dimensional colored cycle stripping is NP-complete [Karp, personal communication].'

Proof by reduction to the wrong problem
'To see that infinite-dimensional colored cycle stripping is decidable, we reduce it to the halting problem.'

Proof by reference to inaccessible literature
The author cites a simple corollary of a theorem to be found in a privately circulated memoir of the Slovenian Philological Society, 1883.

Proof by importance
A large body of useful consequences all follow from the proposition in question.
Proof by accumulated evidence
Long and diligent search has not revealed a counterexample.

Proof by cosmology
The negation of the proposition is unimaginable or meaningless. Popular for proofs of the existence of God.

Proof by mutual reference
In reference A, Theorem 5 is said to follow from Theorem 3 in reference B, which is shown from Corollary 6.2 in reference C, which is an easy consequence of Theorem 5 in reference A.

Proof by meta proof
A method is given to construct the desired proof. The correctness of the method is proved by any of these techniques.

Proof by picture
A more convincing form of proof by example. Combines well with proof by omission.

Proof by vehement assertion
It is useful to have some kind of authority in relation to the audience.

Proof by ghost reference
Nothing even remotely resembling the cited theorem appears in the reference given.

Proof by forward reference
Reference is usually to a forthcoming paper of the author, which is often not as forthcoming as at first.

Proof by semantic shift
Some standard but inconvenient definitions are changed for the statement of the result.

Proof by appeal to intuition
Cloud-shaped drawings frequently help here.

Proof by intimidation
'Trivial.'

Proof by exact naming:
'Let be p a point q and let's call it r.'

Proof by pacifism
'Before you battle about it, rather believe it.'

Proof by communication
Perhaps someone among you knows this?'

Proof by anti-question:
'Who really wants to know the proof?'

5.16 Evolution of Scientific Theory – A Brief Account

It appears that not much has happened since Socrates. Appearances are however deceptive! Let us mention only the recent developments of scientific theory.

5.16.1 Logical Positivism

During much of this century, "positivism" has dominated discussions of the scientific method. Positivism recognizes as valid only the knowledge based on experience.

During the 1920s positivism emerged as a philosophy of science in the form of logical positivism. Developed by the Vienna Circle, a group of scientists and philosophers, logical positivism accepted as its central doctrine Wittgenstein's verification theory of meaning. The verification theory holds that statements or propositions are meaningful only if they can be empirically verified. This criterion was adopted in an attempt to differentiate scientific (meaningful) statements from purely metaphysical (meaningless) statements.

According to logical positivists, universal scientific propositions are true according to whether they have been verified by empirical tests. (Yet no finite number of empirical tests can ever guarantee the truth of universal statements!). In short, empirical inductive inference can never be justified on purely logical grounds.

As a result of these difficulties, Carnap developed a more moderate version of positivism, which has come to be known as logical empiricism that became the "received view" in the philosophy of science for approximately next twenty years.

5.16.2 Logical Empiricism

Essentially, Carnap replaced the concept of verification with the idea of "gradually increasing confirmation". He argued that if verification is taken to mean the "complete and definitive establishment of truth," then universal statements could never be verified. However, they may be "confirmed" by the accumulation of successful empirical tests. Thus, science progresses through the accumulation of multiple confirming instances obtained under a wide variety of circumstances and conditions.

Logical empiricists believe that all knowledge begins with observation. This leads to empirical generalizations among observable entities. As our ideas progress, theories are formulated deductively to explain the generalizations, and new evidence is required to confirm or disconfirm the theories. Throughout the process, data are given precedence. Indeed, the entire process is viewed as essentially an inductive one.

Science in general and knowledge in particular are believed to emerge in an upward

fashion: from data to theory as “an ‘upward flow’ of meaning from the observational terms to the theoretical concepts,” and it is construed in a similar way by Hempel, Carnap and others logical empiricists.
Logical empiricism is characterized by the inductive statistical method. In this view, science begins with observation, and its theories are ultimately justified by the accumulation of further observations, which provide probabilistic support for its conclusion.

Of course, the logical empiricist’s use of a probabilistic interpretation does not avoid the problem of induction. It remains to be shown how a finite number of observations can lead to the logical conclusion that a universal statement is “probably true”.

Moreover, attempts to justify induction on the basis of experience are necessary circular.

The argument that induction has worked successfully in the past is itself an inductive argument and cannot be used to support the principle of induction (Chalmers).

In addition to the problem of (empirical) induction, logical empiricism encounters further difficulties because of its insistence that science rests on a secure observational base. There are at least two problems. The first is that observations are always subject to measurement error. The second, and even more significant, problem concerns the theory dependence of observation.

The fact that observation is theory impregnated does not, by itself, refute the logical empiricist position. It does, however, call into question the claim that science is securely anchored by the objective observation of “reality”.

In his development of falsificationism, Popper has offered an alternative method of theory justification that is designed to overcome some of the difficulties inherent in logical empiricism.

5.16.3 Popper and Falsificationism

Unlike positivists, Popper accepted the fact that “observation always presupposes the existence of some system of expectations”. For Popper, the scientific process begins when observations clash with existing theories or preconceptions. To solve this scientific problem, a theory is proposed and the logical consequences of the theory (hypotheses) are subjected to rigorous empirical tests.

The objective of testing is the refutation of the hypothesis. When a theory’s predictions are falsified, it has to be ruthlessly rejected. Those theories that survive falsification are said to be corroborated (= confirmed) and tentatively accepted.

In contrast to the gradually increasing confirmation of induction, falsificationism substitutes the logical necessity of deduction. Popper exploits the fact that a universal hypothesis can be falsified by a single negative instance. In Popper’s approach, if the deductively derived hypotheses are shown to be false, the theory itself is taken to be false.

Thus the problem of induction is seemingly avoided by denying that science rests on inductive inference. Note nevertheless that Popper's notion of corroboration itself depends on an inductive inference. According to falsificationism, then, science progresses by a process of "conjectures and refutations" (Popper).

Despite the apparent conformity of much scientific practice with the falsificationist account, serious problems remain with Popper's version of the scientific method. The most severe
problem is that it is impossible to conclusively refute a theory because realistic test situations depend on much more than just the theory under investigation.

Take for example all of the background assumptions that might be wrong - flaws in the equipment, the effects of unknown or wrongly disregarded physical processes, and the like. Any outcome can be rationally distrusted and explained away by ad hoc hypotheses that alter the background assumptions. Falsification can thus be regarded as particularly ambiguous.

The recognition that established theories often resist refutation by anomalies and some new theories frequently progress despite their empirical failures led a number of writers to challenge the positivistic views of Popper and the logical empiricists.

The scientific practice is often governed by a conceptual framework highly resistant to change.

In particular, Kuhn pointed out that the established framework is rarely, if ever, overturned by a single anomaly. Kuhn's model helped to initiate a new approach in the philosophy of science in which emphasis is placed on the conceptual frameworks that guide research activities.

5.16.4 Kuhn's Scientific Revolutions

Central to Kuhn's argument is the concept of a "paradigm", the world-view of a scientific community. The paradigm will include a number of specific theories that depend, in part, on the shared metaphysical beliefs of the community.

In Kuhn's view, the individual scientist's decision to pursue a new paradigm must be made on faith in its "future promise". Furthermore, in his view, science progresses through "paradigm shifts", but there is no guarantee that it progresses toward anything - least of all toward "the truth" (Kuhn).

Given its (seeming) advocacy of relativism, Kuhn's Structure of Scientific Revolutions became one of the most carefully analyzed and evaluated works in the philosophy of science.

In criticism of Kuhn, some writers have suggested alternative worldview models as for example “research tradition” concept, which attempts to restore rationality to theory selection by expanding the concept of rationality.

5.16.5. The Copernican Revolution

Greek philosopher Aristotle (the fourth century BC) had formulated the astronomy with the earth as the center of the universe, and sun, planets and stars orbiting it. The universe was divided into two regions: sub-lunar and super-lunar. All celestial objects in super-lunar region were made of incorruptible element called ether. Ether had a natural property to move around the center of the universe in perfect circles. Planets moved in circles, called epicycles, whose center moved, in circle around the Earth.

All substances in sub-lunar region were mixtures of four elements:

Air
Earth
Fire
Water

Each element had a natural place in the universe.
The natural place for the Earth was in the center of the universe; for Water on the surface of the Earth, for Air in the region above the surface of the Earth, and for the Fire above the atmosphere.

Consequently, each earthly object would have a natural place in the sub-lunar region depending on the proportion of four elements. All objects on Earth were thought to have natural property to move in strait lines upwards or downwards, towards their natural place. Thus stones have the natural motion straight downwards, towards the center of the earth, and flames have a natural motion straight upwards, striving towards the top of the atmosphere.

(Before Copernicus, it was the Pythagoreans who had first said that the "Sun is in the middle and the earth is one of the stars (planets) revolving round the Sun having day and night".)

All motion other than natural motion requires a force.
For instance, arrow needs to be propelled by a bow, and a wagon has to be drawn by horses.

In the second century AD Ptolemy developed within the Aristotelian physics a geocentric astronomical system that specified the orbits of the moon, the sun and all the planets. Ptolemy’s system was held as definite truth during the Antique and Middle Ages.

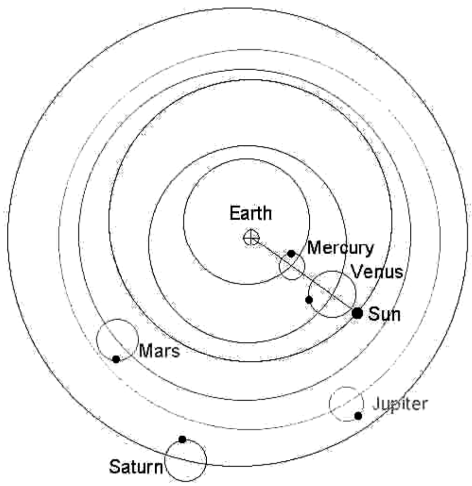


Figure 1 Ptolemy's system

In 1543 Copernicus published his new astronomy with sun in the center[21], and planets, including earth, orbiting the sun. The Copernican view was a radical negation of Aristotelian and Ptolemy's physics and astronomy.

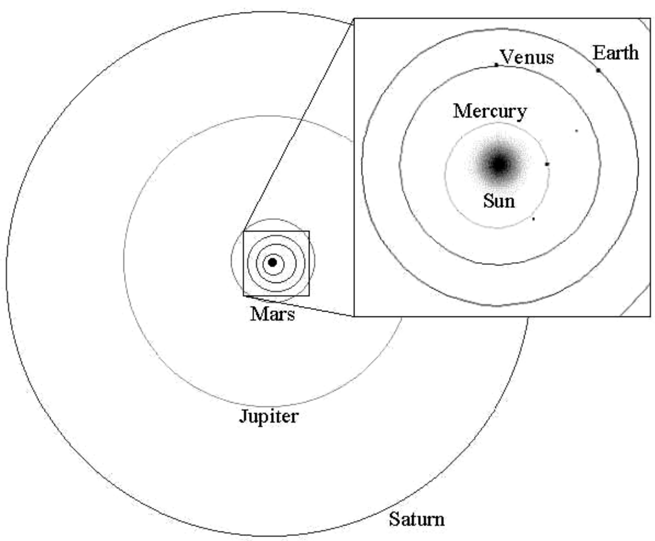


Figure 2 Copernican system

The idea of Copernicus was not really new! A sun-centered system had been proposed as early as about 200 B.C. by Aristarchus of Samos. However, it did not survive long under the weight of Aristotle's influence and "common sense". Aristarchus postulated that the planets orbited the sun – not the Earth – over a thousand years before Copernicus and Galileo made similar arguments. Aristarchus used deductive logic to estimate the size of the earth, the size and distance to the moon, the size and distance to the sun, then he even deduced that the points of light we see at night are not dots painted on some celestial sphere but stars like our sun at enormous distances. Aristarchus' discoveries remained truly unbelievable to the people of his time but stand today as pillars of deductive reasoning.

At first, the problem was that Copernicus had no alternative for Aristotelian physics, and hence had no strong enough arguments to defend his heliocentric system.

Copernicus model was against Aristotelian ideas of earth as natural center of the universe.

Moreover, there were number of other arguments against it, such as:

- Tower argument (the stone dropped from the top of a tower strikes the ground at the base of the tower, contrary to the hypothesis that the earth is spinning around its axes).
- Loose objects on the surface of the earth would be expected to flung from the earth surface in much the same ways stones would be flung from the rotating wheel.
- Absence of parallax in the observed positions of the stars
- Mars and Venus, as viewed by the naked eye, do not change size appreciably during the course of the year
- If the earth were moving through the universe one would expect wind blowing all the time…
- How to explain that the moon follows the earth on its journey through the universe?

Unfortunately, Copernicus had no adequate answers to those questions. He was himself deeply influenced by Aristotle's physics!

Even worse: the results obtained with circular, heliocentric orbits contradict observations, so Copernicus, like Ptolemy was forced to add epicycles (ad-hoc hypothesis)! Not even with this "improvement" could Copernicus show the results better than the old Ptolemy's.

From the point of view of theory of science, the interesting question is: Why should anyone leave the old well-established theory giving satisfactory predictions[22], theory in agreement with physical laws (of its own time) for a new one turning the whole worldview of physics and astronomy upside-down and promising no better agreement with observations?

Certain refinements have been made in Ptolemy's scheme and epicycles inside epicycles introduced in order to get better agreement with observations.

Nevertheless, a number of natural philosophers were attracted by Copernicus system. They have become more and more successful in defending it over the next hundred years or so.

The most prominent among those scientists was Galileo Galilei.

Galileo devised new mechanic to replace Aristotelian and so remove arguments against Copernicus.

He distinguished between the ideas of velocity and acceleration (change of velocity), and asserted that freely falling objects move with a constant acceleration that is independent of their weight.

He denied the Aristotelian claim that all motion requires a force and instead proposed circular law of inertia, according to which a moving body subject to no force will move indefinitely in a circle around the sun at uniform speed. (This law of inertia is later on replaced by Newton's linear law of inertia)

Galileo introduced the idea of relative motion and argued that the uniform motion could not be detected by mechanical means, without access to some reference point outside the system. He based his argument mainly on thought experiments, but he actually did perform a number of experiments as for example the famous ones with rolling spheres down inclined planes.

The new Galileo's mechanics helped to defend Copernican system.

An object held at the top of a tower and sharing its circular motion around the center of the earth will continue that motion (because of inertia) along with tower, after it is dropped and will strike the ground at the foot of the tower.

Galileo proposed the following experiment to show the correctness of the law of inertia. If we drop a stone from the mast of a uniformly moving ship on the sea, the stone will strike the deck at the foot of the mast!

Galileo also used telescope to observe celestial bodies. The discovery of the phases of Venus was another Galileo's contribution to a success of Copernican theory.

The next major support for Copernicus heliocentric scheme was from Kepler, Galileo's contemporary, who had at his disposal Tycho Brache's recordings of planetary positions, that were more accurate than those available to Copernicus.

Kepler discovered the following three (Kepler!) laws of planetary motion:

- LAW 1: The orbit of a planet/comet about the sun is an ellipse with the sun's center of mass at one focus. (That eliminated ad-hoc epicycles from Copernican model).

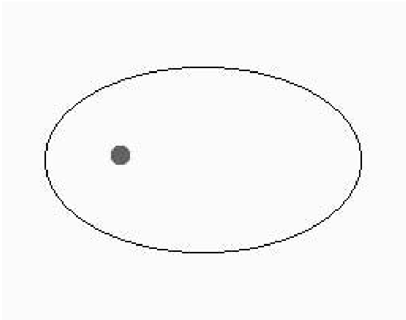


Figure 3 Keplers's first law

- LAW 2: A line joining a planet/comet and the Sun sweeps out equal areas in equal intervals of time

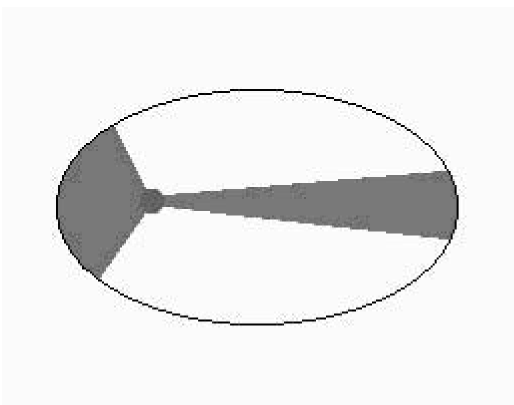


Figure 4 Keplers's second law

- LAW 3: The squares of the periods of the planets are proportional to the cubes of their semimajor axes:

$$Ta^2 / Tb^2 = Ra^3 / Rb^3$$

In 1687 Newton published his Principia. He defined force as the cause of acceleration rather than motion. He replaced Galileo's law of circular inertia with a law of linear inertia, according to which bodies continue to move in strait lines at uniform speed unless acted on by force.

In Principa Newton defines mass, velocity, and acceleration and his three laws of motion. From this and Kepler's Laws he attacked the problem of the planets devising the law of gravitation. He knew that the moon should fly away from the earth but somehow was held in orbit around the earth. But there was nothing connecting the moon to the earth. He showed that the force that causes Kepler's elliptical orbits is a central force, directed to the center of motion.

He also demonstrated that planets under the influence of this central force follow Kepler's second law. The gravitational law means that all masses attract all other masses. There is no repulsion in case of gravitation, contrary to magnetism.

Newton used the following thought experiment (gedanken experiment). Suppose we fire a cannon horizontally from a high mountain. The projectile will eventually fall to earth, because of the gravitational force directed toward the center of the earth. But as we increase the velocity, the projectile will travel further and further before returning to earth.

Finally, Newton reasoned that if the cannon projected the cannon ball with exactly the right velocity, the projectile would travel completely around the earth, always falling in the gravitational field but never reaching the earth, which is curving away at the same rate that the projectile falls. That is, the cannon ball would have been put into orbit around the earth.

Newton concluded that the orbit of the moon was of exactly the same nature: the moon continuously "fell" in its path around the earth because of the acceleration due to gravity, thus producing its orbit.

Utilizing Kepler's third law, Newton derived the law of gravitation.

Newton further realized that the mass of the sun must also be a factor, so he adjusted constants so to describe mass of the sun.

$$F_{12} = G\frac{M_1 M_2}{R^2}$$

Gravitational force is directly proportional to masses (M_S and M_P) and inversely proportional to the square of their distance R. Constant G is called gravitational constant.

Newton further generalized his theory to apply to any two masses, not only planets and sun.
Newton's three laws of motion constitute the basis for classical mechanics:

1. A body continues in a state of rest, or motion with a constant velocity, unless compelled to change by a force.

2. The acceleration of an object is directly proportional to the net force acting upon it and inversely proportional to its mass. F = ma, where F = net force, m = mass of the object, and a = the acceleration of the object = the time rate of change of the velocity. The direction of the acceleration is in the direction of the net force.

3. For every action force, there is an equal and opposite reaction force.

Newton's greatest contribution to support Copernicus was the law of gravitation. It enabled him to explain both Kepler's laws of planetary motion and Galileo's law of free fall.

In Newton's physics, the same universal laws govern both celestial bodies and earthly bodies.

That was the major scientific advancement compared to Aristotelian physics!

What can we learn from this historical example?

We can conclude that neither inductivists nor falsificationists can give a satisfactory explanation of Copernican revolution.

> The Copernican revolution did not take a place as a result of a new theory supported by experimental confirmation.
> New physical concepts of force, inertia and action on distance did not come in the first place as a result of observation and experiment.

Early formulations of the new theory, involving vaguely formulated novel conceptions, were preserved in spite of apparent falsifications! It was only due to the intellectual effort of number of scientists developing a new physics during several centuries, that the new theory could be satisfactory justified.

5.16.6 Hilbert's outstanding problems in Mathematics

Deciding which outstanding problems in mathematics are the most important is to decide the course of mathematics' future development. Perhaps the mathematician who had the greatest impact on the direction of 20th century mathematics – through naming problems that most wanted attention – was David Hilbert. At the second International Congress of Mathematicians, which met at Paris in 1900, Hilbert gave one of the leading addresses, in which he reviewed the basic trends of mathematical research at the end of the 19th century, and then formulated 23 problems, extending over all fields of mathematics, which he believed should occupy the attention of mathematicians in the following century.

Because of Hilbert's prestige, these problems were in focus of mathematicians, and many of them were solved. Some, however, have been solved only very recently, and still others continue to challenge matematicians. Each of these famous problems, with commentary where appropriate, is listed below.

1. The Continuum Hypothesis. Kurt Godel proved in 1938 that the generalized continuum hypothesis (GCH) is consistent relative to Zermelo Fraenkel set theory. In

1963, Paul Cohen showed that its negation is also consistent. Consequently, the axioms of mathematics as currently understood are unable to decide the GCH.

2. Whether the axioms of arithmetic are consistent. Kurt Gödel proved in 1933 that no theory (such as a arithmetic) is strong enough to prove its own consistency. However, Zermelo Fraenkel set theory can prove the consistency of (the Peano axioms) of arithmetic.
3. Whether two tetrahedra of equal base and altitude necessarily have the same volume. This was proved false by Max Dehn in 1900.
4. To construct all the metrics in which straight lines are geodesics.
5. How far Lie's conception of continuous groups of transformations is approachable without assuming that the transformations are differentiable. It was shown in the 1950's that every locally Euclidean group is a Lie group.
6. To axiomatize mathematical physics. Partially solved. In particular, John Von Neumann and others axiomatized quantum mechanics.
7. Whether a^b is transcendental, where a is algebraic and b is irrational. This stubborn problem remains unsolved.
8. The Riemann (zeta) Hypothesis. Unsolved.
9. To find the most general law of reciprocity in an algebraic number field. Solved by Artin in 1927 for abelian extensions of the rational numbers, but the non-abelian case remains open.
10. To find a method to determine whether a given Diophantine Equation is soluble. This "decidability" problem is kin to the larger problem pursued by the logicist program of decidability of theories in general. The particular case of Diophantine equations was finally dealt with in a direct way by Matijasevich in 1970, who showed that no such method exists.
11. The study of quadratic forms with algebraic coefficients.
12. The study of any algebraic number field extensions.
13. To show that the general equation of the seventh degree cannot be solved by means of functions of only two arguments.
14. Whether the ring K intersect $k[x_1, \ldots, x_n]$ is finitely generated over K, where K is a field, $k[x_1, \ldots, x_n]$ is a polynomial ring, and k is a subset of K, which is in turn a subset of $k(x_1, \ldots, x_n)$. Proved false by Nagata in 1959.
15. The rigorous foundation of Schubert's Enumerative Calculus.
16. The investigation of the topology of algebraic surfaces.
17. The expression of a definite rational function as a quotient of sums of squares. Artin showed in 1927 that a positive definite rational function is a sum of squares.
18. Whether there exist non-regular space-filling polyhedra.
19. Whether the solutions of Lagrangians are always analytic.
20. Whether every variational problem has a solution, provided suitable assumptions are made about boundary conditions.
21. To show that there always exists a linear differential equation of the Fuchsian class, with given singular points and monodromic groups. Solved by Deligne in 1970.
22. Development of the calculus of variations.

6 LANGUAGE AND COMMUNICATION

6.1Communication

Communication is imparting of information, interaction through signs/messages.
Information is the meaning that a human gives to signs by applying the known conventions used in their representation.
Sign is any physical event used in communication.
Language is a vocabulary and way of using it.

Semiotics, the science of signs, looks at how humans search for and construct meaning.

Three Levels of Semiotics (Theory of Signs)

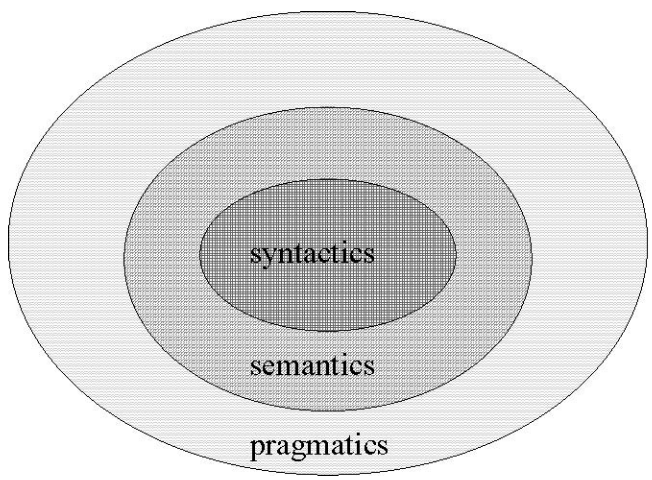


Figure : Semiotics Levels

Semiotics is important because it can help us not to take 'reality' for granted as something having a purely objective existence independent of human interpretation. It teaches us that reality is a system of signs.

Studying semiotics can assist us to become more aware of reality as a construction and of the roles played by ourselves and others in constructing it. It can help us to realize that information or meaning is not 'contained' in the world or in books, computers or audio-visual media. Meaning is not 'transmitted' to us -we actively create it according to a complex interplay of codes or conventions of which we are usually unaware.

The study of signs is the study of the construction and maintenance of reality.

'A sign... is something which stands to somebody for something in some respect or capacity' (Peirce). Sign takes a form of words, symbols, images, sounds, gestures, objects, etc.

Anything can be a sign as long as someone interprets it as 'signifying' something - referring to or standing for something.
A common 'dyadic' or two-part model of the sign defines (Saussure) a sign as composed of:

- a 'signifier' - the form which the sign takes and
- the 'signified' - the concept it represents.

The sign is the whole that results from the association of the signifier (a pointer) with the signified (that what pointer points to).

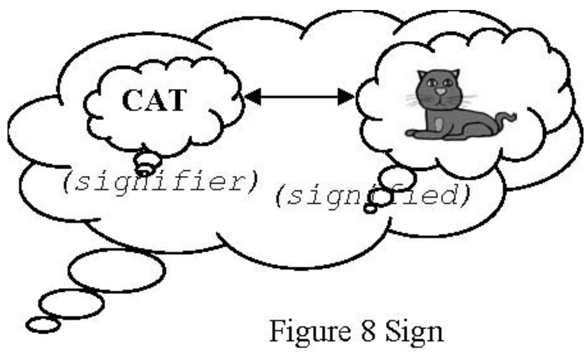


Figure 8 Sign

Although the signifier is treated by its users as 'standing for' the signified, semioticians emphasize that there is no necessary, intrinsic, direct or inevitable relationship between the signifier and the signified. 'That which we call a rose by any other name would smell as sweet', Shakespeare says.

Figure 9 “This is not a Magritte’s pipe”

Reality is divided up into arbitrary categories by every language and the conceptual world with which each of us is familiar could have been divided up very differently. Indeed, no two languages categorize reality in the same way. 'Languages differ by differentiating differently' (Passmore). Linguistic categories are not simply a consequence of some predefined structure in the world. There are no 'natural' concepts or categories which are simply 'reflected' in language. Language plays a crucial role in 'constructing reality'.

The idea that signs (e.g. words) are symbols for things, actions, qualities, relationships, etc, is naïve, gross simplification. The full meaning of a sign does not appear until it is placed in its context, and the context may serve an extremely subtle function- as with puns, or double entendre. And even then the ”meaning” will depend upon the listener, upon the speaker, upon their entire experience of the language, upon their knowledge of one another, and upon the whole situation.
'The whole depends on the parts, and the parts depend on the whole' (Saussure).

It should be noted however that whilst the relationships between signifiers and their signifieds are ontologically arbitrary (philosophically, it would not make any difference to the status of these entities in 'the order of things' if what we call 'black' had always been called 'white' and vice versa), this is not to suggest that signifying systems are socially or historically arbitrary. Natural languages are not, of course, arbitrarily established

As part of its social use within a code, every sign acquires a history and connotations of its own which are familiar to members of the sign-users' culture.

Although the signifier may seem to be freely chosen, from the point of view of the linguistic community it is imposed rather than freely chosen because a language is always an inheritance from the past which its users have no choice but to accept.

Structuralists seek to describe the overall organization of sign systems as 'languages' - as with Lévi-Strauss and myth, kinship rules and totemism.

6.2 Language and Thought

Different media and genres provide different frameworks for representing experience, facilitating some forms of expression and inhibiting others. The differences between media lead Emile Benveniste to argue that the 'first principle' of semiotic systems is that they are not 'synonymous': 'we are not able to say "the same thing" in systems based on different units in contrast to Hjelmslev, who asserted that 'in practice, language is a semiotic into which all other semiotics may be translated'.

We find corresponding two opposite discourses (Timothy Budd):

1 Translation is possible (linguistic realism). Applying Chomsky’s hypothesis that the basis of language is inborn capability of every human being, we can conclude that at least some parts of language must be common, and possible to map between different languages. We can also reason as follows. If there is a mapping from one language to the common world that we live in, as well as mapping from any other language to the

(essentially the same) world, then we can say that there is a mapping from the first language to the second language.[14] As a special case, "Fortran programs can be written in any language. (Timothy Budd) Church's conjecture: "Any computation for which there exists an effective procedure can be realized by an Turing machine."
2 Translation is essentially impossible. The Sapir-Whorf hypothesis (named after the linguists E Sapir and B L Whorf) can be described as relating two associated principles: linguistic determinism and linguistic relativism. Applying these two principles, the thesis is that people who speak languages

[14] Of course "common world" is not exactly the same in Africa and Antarctic. It is neither the same for the blind or deaf people. So parts of our language are not directly coupled to our actual experiences of the "common world". They are "translated" in terms of words for phenomena we have direct experience of.

with very different phonological, grammatical and semantic distinctions perceive and think about the world quite differently, their worldviews being shaped or determined by
their language.

6.3 The Sapir-Whorf Hypothesis

"It may be possible for any individual working in one language to imagine thoughts or to utter ideas that cannot in any way be translated, cannot even be understood, by individuals operating in a different linguistic framework. (…when the language have no equivalent words and lacks even concepts or categories for the ideas involved in thoughts…)."

Sapir argued in a classic passage that:

"The fact of the matter is that the 'real world' is to a large extent unconsciously built upon the language habits of the group. No two languages are ever sufficiently similar to be considered as representing the same social reality. The worlds in which different societies live are distinct worlds, not merely the same world with different labels attached... We see and hear and otherwise experience very largely as we do because the language habits of our community predispose certain choices of interpretation."

A very important question in semiotics and even in anthropology is whether the Sapir-Whorf hypothesis is true. While easy to formulate, it has proved hard to establish or disprove. The reason is that such research needs unusual circumstances - ideally two culturally identical groups using language that differs in one respect that affects a testable way of thinking.

A natural experiment however that fulfills the criteria for showing the existence of the Sapir-Whorf effect has recently been discovered, in the case of deaf infants raised by hearing/deaf parents in their understanding of mental state terms. This research shows that parents´ language has a strong impact upon infants´ cognition. Thus we can conclude that differences between languages can affect cognition as claimed by Sapir and Whorf.

Here it is important to point out that the lack of certain words or ideas is nothing permanent or definitive. Our understanding of the world is dynamic and ever adapting to new circumstances. Let alone mention new products on the market that force us all the time to accept new words and new concepts. Or take for example an African who has no experience of snow, coming to north and getting in contact with it; (s)he forms a new idea connected to that new experience. Even without any practical familiarity with some unknown phenomenon we can get an idea about it via analogy with well-known phenomena. For example we can describe chimerical visual effects of a fata morgana (mirage) in terms of some known optical phenomena.

Sapir-Whorf hypothesis has a tendency to underestimate non-verbal concepts i.e. the tacit knowledge. The language has certainly a strong inpact on our cognition, but the opposite is true as well: our cognition influences forming of our language, and thinking process can be very creative and lead to new concepts, that have not been present in language before.

> ”Language is called the
> garment[15] of thought;
> however, it should
> rather be, language is
> flesh-garment, the
> body of thought.”
> Carlyle, Past and
> Present

To conclude with, we can re-state (translate!) our original question about (im)possibility of translation to the following form: what is (an adequate) translation after all?

6.4 Reality and Communication

If a tree falls and the media aren't there, has it really fallen?

Hierarchical Structure of Language

Object-language Meta-language

In dictionaries on SCIENCE no definition of science!

The definition of SCIENCE can be found in PHILOSOPHY dictionaries.

6.5 Ambiguities of Language

To formulate and to communicate our thoughts, we use language. However, language is used not only to express thoughts, but also feelings, statements and like. So its function is complex and not strictly logical.

Building blocks of every language are words, structured in sentences according to some set of syntactic rules. The meaning of the word is affected by its context, and it is by no means unambiguous.

Some of the most important types of language ambiguity are:

1. Lexical ambiguity, where a word have more than one meaning. Let us compare different meanings of the word "meaning" in English and Swedish. We only take into account synonyms of the word "meaning" and synonyms of synonyms. There are two groups of synonyms, the first one centered on "sense", and the other one centered on "significance".

meaning (sense, connotation, denotation, import, gist; significance, importance, implication, value, consequence, worth)

sense (intelligence, brains, intellect, wisdom, sagacity, logic, good judgment; feeling) connotation (nuance, suggestion, implication, undertone, association, subtext, overtone) denotation (sense, connotation, import, gist) import (bring in, introduce, trade in, introduction, significance) gist (general idea, substance, idea, general picture, essence)

[15](clothes)
significance (meaning, implication, import, consequence, worth, connotation, importance) importance (significance, meaning, weight, consequence, magnitude, import, substance, value, worth) implication (insinuation, inference, suggestion, allegation, allegation, allusion, allusion,
proposition, connotation, inference) value (worth, price, cost, charge, rate, assessment, importance, worth, denomination) consequence (result, effect, outcome, end result, upshot, corollary, importance, significance) worth (value, merit, appeal, significance, attraction, importance, meaning, worth, worth)

Here is the Swedish translation.

meaning (mening, uppfattning betydelse, innebörd; tanke, avsikt)

mening (view, idea, intention, purpose, object, aim, sense, context, sentence)

uppfattning (opinion, convictions)

betydelse (meaning, signification, import, sense, significance, importance)

innebörd (meaning, signification, import, content, purport) tanke (thought, idea, opinion)

avsikt (intention, purpose, aim, object, end, design, motive, intent)

We see that there is a considerable difference between English and Swedish meaning of "meaning". The alternative centered on "significance" is missing. It would be of course even more pronounced if we continued to search for synonyms of synonyms.

Some more ambiguities…

1 Syntactic ambiguity like in "small dogs and cats" (are cats small?).
2 Semantic ambiguity comes often as a consequence of syntactic ambiguity. "Coast road" can be a road that follows the coast, or a road that leads to the coast.
3 Referential ambiguity is a sort of semantic ambiguity ("it" can refer to anything).
4 Pragmatic ambiguity (If the speaker says "I'll meet you next Friday", thinking that they are talking about 17^{th}, and the hearer think that they are talking about 24^{th}, then there is miscommunication.)
5 Local ambiguity occurs when sub-string can be parsed in several ways, but only one of those ways fits into a larger context. ("In English, the radio broadcasts" is a noun phrase in "the radio broadcasts inform", and a noun phrase followed by a verb in "the radio broadcasts information". It is possible for a phrase to be syntactically ambiguous but semantically unambiguous, as "S_1 and S_2 and S_3".
6 Vagueness is an important feature of natural languages. "It is warm outside" says something about temperature, but what does it mean? A warm winter day in Sweden is not the same thing as warm summer day in Kenya.

6.6 A Quest for an Ideal Language

The question is: is it possible, even in principle, to construct a non-ambiguous language? The language Leibniz envisioned, the one that could be used for perfectly correct formulation of logical reasoning?

Wittgenstein, Whitehead, Russell – have made an attempt to formulate a non-contradictory formal logical/philosophical system using well-defined elements. As the main problem in exact codifying and communicating of thoughts (logic) they identified ambiguities in natural languages. The ambition was to build up a perfect formal systems using purified and formalized unambiguous language. Bases of mathematics (as a generalization of ordinary logic) In Principia Mathematica by Russell and Whitehead was written almost entirely by mathematical symbols, without words.

However, Gödel theorem have shown that it was an utopia as within any sufficiently powerful logical system statements can be formulated which can neither be proved nor disproved within the system, unless the system itself is inconsistent.

6.7 The Evolution of Language

The question of origins and evolution of language is very interesting and disputed one. In 1866, The Société de Linguistique de Paris passed a by-law banning of all debate on the origin of language. However, outside the halls of that body, the discussion goes on. One theory is (Chomsky) that there is language module in the human brain, i.e. that language is something we have hard-coded as humans.

There is now considerable evidence that language use is made possible by a number of different skills that developed together with general cognitive capabilities, and that many traces of predecessors of human language can be found in other primates and even in the fossil record.

It is worthy of note that best trained and talented monkey have learned to use about thousand words, which indeed is a remarkable achievement, but no animal have ever been trained to communication capability comparable to four-years old human.

7. DISEMINATING KNOWLEDGE IN ENGINEERING

7.4 Writing a good thesis: Research Report Writing

This guide to thesis writing gives simple and practical advice on the problems of getting started, getting organised, dividing the huge task into less formidable pieces and working on those pieces. It also explains the practicalities of surviving the ordeal. It includes a set of suggested structures depending on the nature of your research and a guide to what should go in each section.

What is a thesis? For whom is it written? How should it be written?

Your thesis is a research report. The report concerns a problem or series of problems in your area of research and it should describe what was known about it previously, what you did towards solving it, what you think your results mean, and where or how further progress in the field can be made. Do not carry over your ideas from undergraduate assessment: a thesis is not an answer to an assignment question. One important difference is this: the reader of an assignment is usually the one who has set it. S/he already knows the answer (or one of the answers), not to mention the background, the literature, the assumptions and theories and the strengths and weaknesses of them. The readers of a thesis do not know what the "answer" is. If the thesis is for a PhD, the university requires that it make an original contribution to human knowledge: your research must discover something hitherto unknown.

Obviously your examiners will read the thesis. They will be experts in the general field of your thesis but, on the exact topic of your thesis, you are the world expert. Keep this in mind: you should write to make the topic clear to a reader who has not spent most of the last three years thinking about it.

Your thesis will also be used as a scientific report and consulted by future workers in your laboratory who will want to know, in detail, what you did. Theses are occasionally consulted by people from other institutions. More commonly theses are now stored in an entirely digital form. These may be stored as .pdf files on a server at your university. The advantage is that your thesis can be consulted much more easily by researchers around the world. Write with these possibilities in mind.

It is often helpful to have someone other than your adviser(s) read some sections of the thesis, particularly the introduction and conclusion chapters. It may also be appropriate to ask other members of staff to read some sections of the thesis which they may find relevant or of interest, as they may be able to make valuable contributions. In either case, only give them revised versions, so that they do not waste time correcting your grammar, spelling, poor construction or presentation.

Master's vs. PhD Thesis

There are different expectations for Master's theses and for Doctoral theses. This difference is not in format but in the significance and level of discovery as evidenced by the problem to be solved and the summary of contributions; a Doctoral thesis necessarily requires a more difficult problem to be solved, and consequently more substantial contributions.

The contribution to knowledge of a Master's thesis can be in the nature of an incremental improvement in an area of knowledge, or the application of known techniques in a new area. The Ph.D. must be a substantial and innovative contribution to knowledge.

How much detail?

The short answer is: rather more than for a scientific paper. Once your thesis has been assessed and your friends have read the first three pages, the only further readers are likely to be people who are seriously doing research in just that area. For example, a future research student might be pursuing the same research and be interested to find out exactly what you did. For important parts of apparatus, you should include workshop drawings, circuit diagrams and computer programs, usually as appendices. (By the way, the intelligible annotation of programs is about as frequent as porcine aviation, but it is far more desirable. You wrote that line of code for a reason: at the end of the line explain what the reason is.) You have probably read the theses of previous students in the lab where you are now working, so you probably know the advantages of a clearly explained, explicit thesis and/or the disadvantages of a vague one.

7.4.1 Getting Started

When you are about to begin, writing a thesis seems a long, difficult task. That is because it is a long, difficult task. Fortunately, it will seem less daunting once you have a couple of chapters done. Towards the end, you will even find yourself enjoying it---an enjoyment based on satisfaction in the achievement, pleasure in the improvement in your technical writing, and of course the approaching end. Like many tasks, thesis writing usually seems worst before you begin, so let us look at how you should make a start.

7.4.2 Thesis outline

First make up a thesis outline: several pages containing chapter headings, sub-headings, some figure titles (to indicate which results go where) and perhaps some other notes and comments. There is a section on chapter order and thesis structure at the end of this text. Once you have a list of chapters and, under each chapter heading, a reasonably complete list of things to be reported or explained, you have struck a great blow against writer's block. When you sit down to type, your aim is no longer a thesis---a daunting goal---but something simpler. Your new aim is just to write a paragraph or section about one of your subheadings. It helps to start with an easy one: this gets you into the habit of writing and gives you self-confidence. Often the Materials and Methods chapter is the easiest to write---just write down what you did; carefully, formally and in a logical order.

How do you make an outline of a chapter? For most of them, you might try the following method: Assemble all the figures that you will use in it and put them in the order that you would use if you were going to explain to someone what they all meant. You might as well rehearse explaining it to someone else---after all you will probably give several talks based on your thesis work. Once you have found the most logical order, note down the key words of your explanation. These key words provide a skeleton for much of your chapter outline.

Once you have an outline, discuss it with your adviser. This step is important: s/he will have useful suggestions, but it also serves notice that s/he can expect a steady flow of chapter drafts that will make high priority demands on his/her time. Once you and your adviser have agreed on a logical structure, s/he will need a copy of this outline for reference when reading the chapters which you will probably present out of order. If you have a co-adviser, discuss the outline with him/her as well, and present all chapters to both advisers for comments.

7.4.3 Organisation

It is encouraging and helpful to start a filing system. Open a word-processor file for each chapter *and one for the references.* You can put notes in these files, as well as text. While doing something for Chapter n, you will think "Oh I must refer back to/discuss this in Chapter m" and so you put a note to do so in the file for Chapter m. Or you may think of something interesting or relevant for that chapter. When you come to work on Chapter m, the more such notes you have accumulated, the easier it will be to write.

Make a back-up of these files and do so every day at least (depending on the reliability of your computer and the age of your disk drive). Do not keep back-up disks close to the computer in case the hypothetical thief who fancies your computer decides that s/he could use some disks as well.

A simple way of making a remote back-up is to send it as an email attachment to a consenting email correspondent, preferably one in a different location. You could also send it to yourself. In either case, be careful to dispose of superseded versions so that you don't waste disk space, especially if you have bitmap images or other large files.

You should also have a physical filing system: a collection of folders with chapter numbers on them. This will make you feel good about getting started and also help clean up your desk. Your files will contain not just the plots of results and pages of calculations, but all sorts of old notes, references, calibration curves, suppliers' addresses, specifications, speculations, letters from colleagues etc., which will suddenly strike you as relevant to one chapter or other. Stick them in that folder. Then put all the folders in a box or a filing cabinet. As you write bits and pieces of text, place the hard copy, the figures etc in these folders as well. Touch them and feel their thickness from time to time---ah, the thesis is taking shape.

If any of your data exist only on paper, copy them and keep the copy in a different location. Consider making a copy of your lab book. This has another purpose beyond security: usually the lab book stays in the lab, but you may want a copy for your own future use. Further, scientific ethics require you to keep lab books and original data for at least ten years, and a copy is more likely to be found if two copies exist.

While you are getting organised, you should deal with any university paperwork. Examiners have to be nominated and they have to agree to serve. Various forms are required by your department and by the university administration. Make sure that the rate limiting step is your production of the thesis, and not some minor bureaucratic problem.

Text Editors

One of the big FAQs for scientists: is there a word processor, ideally one compatible with MS Word, but which allows you to type mathematical symbols and equations conveniently? One solution is LaTeX, which is powerful, elegant, reliable, fast and *free* from http://www.latex-project.org/ or http://www.miktex.org/. As far as I know, the only equation editor for MS Word is slow and awkward. LyX, available free at http://www.lyx.org/, is a convenient front-end to LaTeX that has WYSIWYG.

A timetable

I strongly recommend sitting down with the adviser and making up a timetable for writing it: a list of dates for when you will give the first and second drafts of each chapter to your adviser(s). This structures your time and provides intermediate targets. If you merely aim "to have the whole thing done by [some distant date]", you can deceive yourself and procrastinate more easily. If you have told your adviser that you will deliver a first draft of chapter 3 on Wednesday, it focuses your attention.

You may want to make your timetable into a chart with items that you can check off as you have finished them. This is particularly useful towards the end of the thesis when you find there will be quite a few loose ends here and there.

Iterative solution

Whenever you sit down to write, it is very important to write *something.* So write something, even if it is just a set of notes or a few paragraphs of text that you would never show to anyone else. It would be nice if clear, precise prose leapt easily from the keyboard, but it usually does not. Most of us find it easier, however, to improve something that is already written than to produce text from nothing. So put down a draft (as rough as you like) for your own purposes, then clean it up for your adviser to read. Word-processors are wonderful in this regard: in the first draft you do not have to start at the beginning, you can leave gaps, you can put in little notes to yourself, and then you can clean it all up later.

Your adviser will expect to read each chapter in draft form. S/he will then return it to you with suggestions and comments. *Do not be upset if a chapter---especially the first one you write--- returns covered in red ink.* Your adviser will want your thesis to be as good as possible, because his/her reputation as well as yours is affected. Scientific writing is a difficult art, and it takes a while to learn. As a consequence, there will be many ways in which your first draft can be improved. So take a positive attitude to all the scribbles with which your adviser decorates your text: each comment tells you a way in which you can make your thesis better.

As you write your thesis, your scientific writing is almost certain to improve. Even for native speakers of English who write very well in other styles, one notices an enormous improvement in the first drafts from the first to the last chapter written. The process of writing the thesis is like a course in scientific writing, and in that sense each chapter is like an assignment in which you are taught, but not assessed. Remember, only the final draft is assessed: the more comments your adviser adds to first or second draft, the better.

Before you submit a draft to your adviser, run a spell check so that s/he does not waste time on those. If you have any characteristic grammatical failings, check for them.

Make it clear what is yours

If you use a result, observation or generalisation that is not your own, you must usually state where in the scientific literature that result is reported. The only exceptions are cases where every researcher in the field already knows it: dynamics equations need not be followed by a citation of Newton, circuit analysis does not need a reference to Kirchoff. The importance of this practice in science is that it allows the reader to verify your starting position. Physics in particular is said to be a vertical science: results are built upon results which in turn are built upon results etc. Good referencing allows us to check the foundations of your additions to the structure of knowledge in the discipline, or at least to trace them back to a level which we judge to be reliable. Good referencing also tells the reader which parts of the thesis are descriptions of previous knowledge and which parts are your additions to that knowledge. In a thesis, written for the general reader who has little familiarity with the literature of the field, this should be especially clear. It may seem tempting to leave out a reference in the hope that a reader will think that a nice idea or a nice bit of analysis is yours. I advise against this gamble. The reader will probably think: "What a nice idea! I wonder if it's original?". The reader can probably find out via the net or the library.

If you are writing in the passive voice, you must be more careful about attribution than if you are writing in the active voice. “The input image was created by applying noise generator…” does not make it clear whether you did this or your friend did it. “I created the input image…” is clear.

Style

The text must be clear. Good grammar and thoughtful writing will make the thesis easier to read. Scientific writing has to be a little formal---more formal than this text. Every one should remember that scientific English is an international language. Slang and informal writing will be harder for many to understand.

Short, simple phrases and words are often better than long ones. Some politicians use "at this point in time" instead of "now" precisely because it takes longer to convey the same meaning. They do not care about elegance or efficient communication. You should. On the other hand, there will be times when you need a complicated sentence because the idea is complicated. If your primary statement requires several qualifications, each of these may need a subordinate clause: "When [qualification], and where [proviso], and if [condition] then [statement]". Some lengthy technical words will also be necessary in many theses, particularly in fields like biochemistry. Do not sacrifice accuracy for the sake of brevity. "Black is white" is simple and catchy. An advertising copy writer would love it. "Objects of very different albedo may be illuminated differently so as to produce similar reflected spectra" is longer and uses less common words, but, compared to the former example, it has the advantage of being true. The longer example would be fine in a physics thesis because English speaking physicists will not have trouble with the words. (A physicist who did not know all of those words would probably be glad to remedy the lacuna either from the context or by consulting a dictionary.)

Sometimes it is easier to present information and arguments as a series of numbered points, rather than as one or more long and awkward paragraphs. A list of points is usually easier to write. You should be careful not to use this presentation too much: your thesis must be a connected, convincing argument, not just a list of facts and observations.

One important stylistic choice is between the active voice and passive voice. The active voice ("I measured the frequency...") is simpler, and it makes clear what you did and what was done by others. The passive voice ("The frequency was measured...") makes it easier to write ungrammatical or awkward sentences. If you use the passive voice, be especially wary of dangling participles. For example, the sentence "After considering all of these possible algorithms, algorithm-A was selected" implicitly attributes consciousness to algorithm-A. This choice is a question of taste: I prefer the active because it is clearer, more logical and makes attribution simple. The only arguments I have ever heard for avoiding the active voice in a thesis are (i) many theses are written in the passive voice, and (ii) some very polite people find the use of "I" immodest. Use the first person singular, not plural, when reporting work that you did yourself: the editorial 'we' may suggest that you had help beyond that listed in your acknowledgments, or it may suggest that you are trying to share any blame. On the other hand, retain plural verbs for "data": "data" is the plural of "datum", and lots of scientists like to preserve the distinction. Just say to yourself "one datum is ..", "these data are.." several times.

Presentation

There is no need for a thesis to be a masterpiece of desk-top publishing. Your time can be more productively spent improving the content than the appearance.

In many cases, a reasonably neat diagram can be drawn by hand faster than with a graphics package, and you can scan it if you want an electronic version. Either is usually satisfactory. A one bit (i.e. black and white), moderate resolution scan of a hand-drawn sketch will be bigger than a line drawing generated on a graphics package, but not huge. While talking about the size of files, we should mention that photographs look pretty but take up a lot of memory. There's another important difference, too. The photographer thought about the camera angle and the focus etc. The person who drew the schematic diagram thought about what components ought to be depicted and the way in which the components of the system interacted with each other. So the numerically small information content of the line drawing may be much more useful information than that in a photograph.

Another note about figures and photographs. In the digital version of your thesis, do not save ordinary photographs or other illustrations as bitmaps, because these take up a lot of memory and are therefore very slow to transfer. Nearly all graphics packages allow you to save in compressed format as .jpg or .gif files. Further, you can save space/speed things up by reducing the number of colours. In vector graphics (as used for drawings), shades or grey are often produced by black and white pixels, so one-bit colour is adequate.

In general, students spend too much time on diagrams; time that could have been spent on examining the arguments, making the explanations clearer, thinking more about the significance and checking for errors in the algebra. The reason, of course, is that drawing is easier than thinking.

I do not think that there is a strong correlation (either way) between length and quality. There is no need to leave big gaps to make the thesis thicker. Readers will not appreciate large amounts of vague or unnecessary text.

Approaching the end

A deadline is very useful in some ways. You must hand in the thesis, even if you think that you need one more draft of that chapter, or someone else's comments on this section, or some other refinement. If you do not have a deadline, or if you are thinking about postponing it, please take note of this: *A thesis is a very large work. It cannot be made perfect in a finite time*. There will inevitably be things in it that you could have done better. There will be inevitably be some typos. Indeed, by some law related to Murphy's, you will discover one when you first flip open the bound copy. No matter how much you reflect and how many times you proof read it, there will be some things that could be improved. There is no point hoping that the examiners will not notice: many examiners feel obliged to find some examples of improvements (if not outright errors) just to show how thoroughly they have read it. So set yourself a deadline and stick to it. Make it as good as you can in that time, and then hand it in! (In retrospect, there was an advantage in writing a thesis in the days before word processors, spelling checkers and typing programs. Students often paid a typist to produce the final draft and could only afford to do that once.)

How many copies?

Talk to your adviser about this. As well as those for the examiners, the university libraries and yourself, you should make some distribution copies. These copies should be sent to other researchers who are working in your field so that:

- they can discover what marvellous work you have been doing before it appears in journals;
- they can look up the fine details of methods and results that will or have been published more briefly elsewhere;
- they can realise what an excellent researcher you are. This realisation could be useful if a post- doctoral position were available in their labs. soon after your submission, or if they were reviewers of your research/post-doctoral proposal. Even having your name in their bookcases might be an advantage.

Whatever the University's policy on single or double-sided copies, the distribution copies could be double-sided paper, or digital, so that forests and postage accounts are not excessively depleted by the exercise. Your adviser could help you to make up a list of interested and/or potentially useful people for such a mailing list. A CD with your thesis will be cheaper than a paper copy. You don't have to burn them all yourself: companies make multiple copies for several dollars a copy.

Personal

In the ideal situation, you will be able to spend a large part---perhaps a majority---of your time writing your thesis. This may be bad for your physical and mental health.

Typing

Set up your chair and computer properly. The Health Service, professional keyboard users or perhaps even the school safety officer will be able to supply charts showing recommended relative heights, healthy postures and also exercises that you should do if you spend a lot of time at the keyboard. These last are worthwhile insurance: you do not want the extra hassle of back or neck pain. Try to intersperse long sessions of typing with other tasks, such as reading, drawing, calculating, thinking or doing research.

If you do not touch type, you should learn to do so for the sake of your neck as well as for productivity. There are several good software packages that teach touch typing interactively. If you use one for say 30 minutes a day for a couple of weeks, you will be able to touch type. By the time you finish the thesis, you will be able to touch type quickly and accurately and your six hour investment will have paid for itself. Be careful not to use the typing exercises as a displacement activity.

Exercise

Do not give up exercise for the interim. Lack of exercise makes you feel bad, and you do not need anything else making you feel bad while writing a thesis. 30-60 minutes of exercise per day is probably not time lost from your thesis: I find that if I do not get regular exercise, I sleep less soundly and longer. How about walking to work and home again? (Walk part of the way if your home is distant.) Many people opine that a walk helps them think, or clears the head. You may find that an occasional stroll improves your productivity.

Food

Do not forget to eat, and make an effort to eat healthy food. You should not lose fitness or risk illness at this critical time. Exercise is good for keeping you appetite at a healthy level. I know that you have little time for cooking, but keep a supply of fresh fruit, vegetables and bread. It takes less time to make a sandwich than to go to the local fast food outlet, and you will feel better afterwards.

Drugs

Thesis writers have a long tradition of using coffee as a stimulant and alcohol or marijuana as relaxants. (Use of alcohol and coffee is legal, use of marijuana is not.) Used in moderation, they do not seem to have ill effects on the quality of thesis produced. Excesses, however, are obviously counter-productive: several espressi and you will be buzzing too much to sit down and work; several drinks at night will slow you down next day.

Others

Other people will be sympathetic, but do not take them for granted. Spouses, lovers, family and friends should not be undervalued. Spend some time with them and, when you do, have a good time. Do not spend your time together complaining about your thesis: they already resent the thesis because it is keeping you away from them. If you can find another student writing a thesis, then you may find it therapeutic to complain to each other about advisers and difficulties. S/he need not be in the same discipline as you are.

Getting into the Real Business of Writing Thesis

So, you are preparing to write a Ph.D. dissertation in an experimental area of Computer Science. Unless you have written many formal documents before, you are in for a surprise: it's difficult!

Here are a few guidelines that may help you when you finally get serious about writing. The list goes on forever; you probably won't want to read it all at once. But, please read it before you write anything.

The General Idea:

1. A thesis is a hypothesis or conjecture.
2. A PhD dissertation is a lengthy, formal document that argues in defense of a particular thesis. (So many people use the term ``thesis'' to refer to the document that a current dictionary now includes it as the third meaning of ``thesis'').
3. Two important adjectives used to describe a dissertation are ``original'' and ``substantial.'' The research performed to support a thesis must be both, and the dissertation must show it to be so. In particular, a dissertation highlights original contributions.
4. The scientific method means starting with a hypothesis and then collecting evidence to support or deny it. Before one can write a dissertation defending a particular thesis, one must collect evidence that supports it. Thus, the most difficult aspect of writing a dissertation consists of organizing the evidence and associated discussions into a coherent form.
5. The essence of a dissertation is critical thinking, not experimental data. Analysis and concepts form the heart of the work.
6. A dissertation concentrates on principles: it states the lessons learned, and not merely the facts behind them.
7. In general, every statement in a dissertation must be supported either by a reference to published scientific literature or by original work. Moreover, a dissertation does not repeat the details of critical thinking and analysis found in

published sources; it uses the results as fact and refers the reader to the source for further details.

8. Each sentence in a dissertation must be complete and correct in a grammatical sense. Moreover, a dissertation must satisfy the stringent rules of formal grammar (e.g., no contractions, no colloquialisms, no slurs, no undefined technical jargon, no hidden jokes, and no slang, even when such terms or phrases are in common use in the spoken language). Indeed, the writing in a dissertaton must be crystal clear. Shades of meaning matter; the terminology and prose must make fine distinctions. The words must convey exactly the meaning intended, nothing more and nothing less.
9. Each statement in a dissertation must be correct and defensible in a logical and scientific sense. Moreover, the discussions in a dissertation must satisfy the most stringent rules of logic applied to mathematics and science.

What One Should Learn from the Exercise:

1. All scientists need to communicate discoveries; the PhD dissertation provides training for communication with other scientists.
2. Writing a dissertation requires a student to think deeply, to organize technical discussion, to muster arguments that will convince other scientists, and to follow rules for rigorous, formal presentation of the arguments and discussion.

A Rule of Thumb:

Good writing is essential in a dissertation. However, good writing cannot compensate for a paucity of ideas or concepts. Quite the contrary, a clear presentation always exposes weaknesses.

Definitions and Terminology:

1. Each technical term used in a dissertation must be defined either by a reference to a previously published definition (for standard terms with their usual meaning) or by a precise, unambiguous definition that appears before the term is used (for a new term or a standard term used in an unusual way).
2. Each term should be used in one and only one way throughout the dissertation.

3. The easiest way to avoid a long series of definitions is to include a statement: ``the terminology used throughout this document follows that given in [CITATION]." Then, only define exceptions.
4. The introductory chapter can give the intuition (i.e., informal definitions) of terms provided they are defined more precisely later.

Terms and Phrases to Avoid:

adverbs

Mostly, they are very often overly used. Use strong words instead. For example, one could say, ``Writers abuse adverbs."

jokes or puns

They have no place in a formal document.

``bad", ``good", ``nice", ``terrible", ``stupid"

A scientific dissertation does not make moral judgements. Use ``incorrect/correct" to refer to factual correctness or errors. Use precise words or phrases to assess quality (e.g., ``method A requires less computation than method B"). In general, one should avoid all qualitative judgements.

``true", ``pure",

In the sense of ``good" (it is judgemental).

``perfect"

Nothing is.

``an ideal solution"

You're judging again.

``today", ``modern times"

Today is tomorrow's yesterday.

``soon"

How soon? Later tonight? Next decade?

``we were surprised to learn..."

Even if you were, so what?

``seems'', ``seemingly'',

It doesn't matter how something appears;

``would seem to show''

all that matters are the facts.

``in terms of''

usually vague

``based on'', ``X-based'', ``as the basis of''

careful; can be vague

``different''

Does not mean ``various''; different than what?

``in light of''

colloquial

``lots of''

vague & colloquial

``kind of''

vague & colloquial

``type of''

vague & colloquial

``something like''

vague & colloquial

``just about''

vague & colloquial

``number of''

vague; do you mean ``some'', ``many'', or ``most''? A quantative statement is preferable.

``due to''

colloquial

``probably''

only if you know the statistical probability (if you do, state it quantatively

``obviously, clearly''

be careful: obvious/clear to everyone?

``simple''

Can have a negative connotation, as in ``simpleton''

``along with''

Just use ``with''

``actually, really''

define terms precisely to eliminate the need to clarify

``the fact that''

makes it a meta-sentence; rephrase

``this'', ``that''

As in ``This causes concern.'' Reason: ``this'' can refer to the subject of the previous sentence, the entire previous sentence, the entire previous paragraph, the entire previous section, etc. More important, it can be interpreted in the concrete sense or in the meta-sense. For example, in: *``X does Y. This means ...''* the reader can assume ``this'' refers to Y or to the fact that X does it. Even when restricted (e.g., ``this computation...''), the phrase is weak and often ambiguous.

``You will read about...''

The second person has no place in a formal dissertation.

``I will describe..."

> The first person has no place in a formal dissertation. If self-reference is essential, phrase it as ``Section 10 describes..."

``we" as in ``we see that"

> A trap to avoid. Reason: almost any sentence can be written to begin with ``we" because ``we" can refer to: the reader and author, the author and advisor, the author and research team, experimental computer scientists, the entire computer science community, the science community, or some other unspecified group.

``Hopefully, the program..."

> Computer programs don't hope, not unless they implement AI systems. By the way, if you are writing an AI thesis, talk to someone else: AI people have their own system of rules.

``...a famous researcher..."

> It doesn't matter who said it or who did it. In fact, such statements prejudice the reader.

Be Careful When Using ``few, most, all, any, every".

> A dissertation is precise. If a sentence says ``Most computer systems contain X", you must be able to defend it. Are you sure you really know the facts? How many computers were built and sold yesterday?

``must", ``always"

> Absolutely?

``should"

> Who says so?

``proof", ``prove"

> Would a mathematician agree that it's a proof?

``show"

> Used in the sense of ``prove". To ``show" something, you need to provide a formal proof.

``can/may''

Your mother probably told you the difference.

Voice:

Use active constructions. For example, say ``the operating system starts the device'' instead of ``the device is started by the operating system.''

Tense:

Write in the present tense. For example, say ``The system writes a page to the disk and then uses the frame...'' instead of ``The system will use the frame after it wrote the page to disk...''

Define Negation Early:

Example: say ``no data block waits on the output queue'' instead of ``a data block awaiting output is not on the queue.''

Grammar and Logic:

Be careful that the subject of each sentence really does what the verb says it does. Saying ``Programs must make procedure calls using the X instruction'' is not the same as saying ``Programs must use the X instruction when they call a procedure.'' In fact, the first is patently false! Another example: ``RPC requires programs to transmit large packets'' is not the same as ``RPC requires a mechanism that allows programs to transmit large packets.''

All computer scientists should know the rules of logic. Unfortunately the rules are more difficult to follow when the language of discourse is English instead of mathematical symbols. For example, the sentence ``There is a compiler that translates the N languages by...'' means a single compiler exists that handles all the languages, while the sentence ``For each of the N languages, there is a compiler that translates...'' means that there may be 1 compiler, 2 compilers, or N compilers. When written using mathematical symbols, the difference are obvious because ``for all'' and ``there exists'' are reversed.

Focus On Results And Not The People/Circumstances In Which They Were Obtained:

``After working eight hours in the lab that night, we realized..." has no place in the dissertation. It doesn't matter when you realized it or how long you worked to obtain the answer. Another example: ``Jim and I arrived at the numbers shown in Table 3 by measuring..." Put an acknowledgement to Jim in the dissertation, but do not include names (even your own) in the main body. You may be tempted to document a long series of experiments that produced nothing or a coincidence that resulted in success. Avoid it completely. In particular, do not document seemingly mystical influences (e.g., ``if that cat had not crawled through the hole in the floor, we might not have discovered the power supply error indicator on the network bridge"). Never attribute such events to mystical causes or imply that strange forces may have affected your results. Summary: stick to the plain facts. Describe the results without dwelling on your reactions or events that helped you achieve them.

Avoid Self-Assessment (both praise and criticism):

Both of the following examples are incorrect: ``The method outlined in Section 2 represents a major breakthrough in the design of distributed systems because..." ``Although the technique in the next section is not earthshaking,..."

References to Extant Work:

One always cites papers, not authors. Thus, one uses a singular verb to refer to a paper even though it has multiple authors. For example ``Johnson and Smith [J&S90] reports that..."

Avoid the phrase ``the authors claim that X". The use of ``claim" casts doubt on ``X" because it references the authors' thoughts instead of the facts. If you agree ``X" is correct, simply state ``X" followed by a reference. If one absolutely must reference a paper instead of a result, say ``the paper states that..." or ``Johnson and Smith [J&S 90] presents evidence that...".

Concept Vs. Instance:

A reader can become confused when a concept and an instance of it are blurred. Common examples include: an algorithm and a particular program that implements it, a programming language and a compiler, a general abstraction and its particular implementation in a computer system, a data structure and a particular instance of it in memory.

Terminology for Concepts and Abstractions

When defining the terminology for a concept, be careful to decide precisely how the idea translates to an implementation. Consider the following discussion:

VM systems include a concept known as an address space. The system dynamically creates an address space when a program needs one, and destroys an address space when the program that created the space has finished using it. A VM system uses a small, finite number to identify each address space. Conceptually, one understands that each new address space should have a new identifier. However, if a VM system executes so long that it exhausts all possible address space identifiers, it must reuse a number.

The important point is that the discussion only makes sense because it defines ``address space'' independently from ``address space identifier''. If one expects to discuss the differences between a concept and its implementation, the definitions must allow such a distinction.

Knowledge Vs. Data

The facts that result from an experiment are called ``data''. The term ``knowledge'' implies that the facts have been analyzed, condensed, or combined with facts from other experiments to produce useful information.

Cause and Effect:

A dissertation must carefully separate cause-effect relationships from simple statistical correlations. For example, even if all computer programs written in Professor X's lab require more memory than the computer programs written in Professor Y's lab, it may not have anything to do with the professors or the lab or the programmers (e.g., maybe the people working in professor X's lab are working on applications that require more memory than the applications in professor Y's lab).

Drawing Only Warranted Conclusions:

One must be careful to only draw conclusions that the evidence supports. For example, if programs run much slower on computer A than on computer B, one cannot conclude that the processor in A is slower than the processor in B unless one has ruled out all differences in the computers' operating systems, input or output devices, memory size, memory cache, or internal bus bandwidth. In fact, one must still refrain from judgement unless one has the results from a controlled experiment (e.g., running a set of several programs many times, each when the computer is otherwise idle). Even if the cause of some phenomenon seems obvious, one cannot draw a conclusion without solid, supporting evidence.

Commerce and Science:

In a scientific dissertation, one never draws conclusions about the economic viability or commercial success of an idea/method, nor does one speculate about the history of development or origins of an idea. A scientist must remain objective about the merits of an idea independent of its commercial popularity. In particular, a scientist never assumes that commercial success is a valid measure of merit (many popular products are neither well-designed nor well-engineered). Thus, statements such as ``over four hundred vendors make products using technique Y'' are irrelevant in a dissertation.

Politics and Science:

A scientist avoids all political influence when assessing ideas. Obviously, it should not matter whether government bodies, political parties, religious groups, or other organizations endorse an idea. More important and often overlooked, it does not matter whether an idea originated with a scientist who has already won a Nobel prize or a first-year graduate student. One must assess the idea independent of the source.

Canonical Organization:

In general, every dissertation must define the problem that motivated the research, tell why that problem is important, tell what others have done, describe the new contribution, document the experiments that validate the contribution, and draw conclusions. There is no canonical organization for a dissertation; each is unique. However, novices writing a dissertation in the experimental areas of CS may find the following example a good starting point:

- ### Chapter 1: Introduction

  An overview of the problem; why it is important; a summary of extant work and a statement of your hypothesis or specific question to be explored. Make it readable by anyone.

- ### Chapter 2: Definitions

  New terms only. Make the definitions precise, concise, and unambiguous.

- ### Chapter 3: Conceptual Model

  Describe the central concept underlying your work. Make it a ``theme'' that ties together all your arguments. It should provide an answer to the question posed in the introduction at a conceptual level. If necessary, add another chapter to give additional reasoning about the problem or its solution.

- ### Chapter 4: Experimental Measurements

  Describe the results of experiments that provide evidence in support of your thesis. Usually experiments either emphasize proof-of-concept (demonstrating the viability of a method/technique) or efficiency (demonstrating that a method/technique provides better performance than those that exist).

- ### Chapter 5: Corollaries And Consequences

  Describe variations, extensions, or other applications of the central idea.

- ### Chapter 6: Conclusions

  Summarize what was learned and how it can be applied. Mention the possibilities for future research.

- ### Abstract:

  A short (few paragraphs) summary of the the dissertation. Describe the problem and the research approach. Emphasize the original contributions.

Suggested Order For Writing:

The easiest way to build a dissertation is inside-out. Begin by writing the chapters that describe your research (3, 4, and 5 in the above outline). Collect terms as they arise and keep a definition for each. Define each technical term, even if you use it in a conventional manner.

Organize the definitions into a separate chapter. Make the definitions precise and formal. Review later chapters to verify that each use of a technical term adheres to its definition. After reading the middle chapters to verify terminology, write the conclusions. Write the introduction next. Finally, complete an abstract.

Key To Success:

By the way, there is a key to success: practice. No one ever learned to write by reading essays like this. Instead, you need to practice, practice, practice. Every day.

Parting thoughts:

We leave you with the following ideas to mull over. If they don't mean anything to you now, revisit them after you finish wirting a dissertation.

After great pain, a formal feeling comes.
-- Emily Dickinson

A man may write at any time, if he will set himself doggedly to it.
-- Samuel Johnson

Keep right on to the end of the road.
-- Harry Lauder

The average Ph.D. thesis is nothing but the transference of bones from one graveyard to another.
-- Frank J. Dobie

Conclusion

Keep going, you're nearly there! Most PhDs will admit that there were times when we thought about reasons for not finishing. But it would be crazy to give up at the writing stage, after years of work on the research, and it would be something to regret for a long time.

Writing a thesis is tough work. One anonymous post doctoral researcher told me: "You should tell everyone that it's going to be unpleasant, that it will mess up their lives, that they will have to give up their friends and their social lives for a while. It's a tough period for almost every student." She's right: it is certainly hard work, it will probably be stressful and you will have to adapt your rhythm to it. It is also an important rite of passage and the satisfaction you will feel afterwards is wonderful. On behalf of scholars everywhere, I wish you good luck!

Acknowledgements and Suggestions

This document will be continuously updated. If you have suggestions do send them. I thank all my research scholars and M.Tech. students who have inspired me to write this document and served as test domains for my experimentations. Opinions expressed in these documents are mine and do not necessarily reflect the policy of the Manonmaniam Sundaranar or of the Centre for Information Technology and Engineering.

A suggested thesis structure - Model 1:

The list of contents and chapter headings below is appropriate for some theses. In some cases, one or two of them may be irrelevant. Results and Discussion are usually combined in several chapters of a thesis. Think about the plan of chapters and decide what is best to report your work. Then make a list, in point form, of what will go in each chapter. Try to make this rather detailed, so that you end up with a list of points that corresponds to subsections or even to the paragraphs of your thesis. At this stage, think hard about the logic of the presentation: within chapters, it is often possible to present the ideas in different order, and not all arrangements will be equally easy to follow. If you make a plan of each chapter and section before you sit down to write, the result will probably be clearer and easier to read. It will also be easier to write.

Copyright waiver

Your institution may have a form for this (UNSW does). In any case, this standard page gives the university library the right to publish the work, possibly by microfilm or some other medium. (At UNSW, the Postgraduate Student Office will give you a thesis pack with various guide-lines and rules about thesis format. Make sure that you consult that for its formal requirements, as well as this rather informal guide.)

Declaration

Check the wording required by your institution, and whether there is a standard form. Many universities require something like: "I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text. (signature/name/date)"

Title page

This may vary among institutions, but as an example: Title/author/"A thesis submitted for the degree of Doctor of Philosophy in the Faculty of Science/The University of New South Wales"/date.

Abstract

Of all your thesis, this part will be the most widely published and most read because it will be published in Dissertation Abstracts International. It is best written towards the end, but not at the very last minute because you will probably need several drafts. It should be a distillation of the thesis: a concise description of the problem(s) addressed, your method of solving it/them, your results and conclusions. An abstract must be self-contained. Usually they do not contain references. When a reference is necessary, its details should be included in the text of the abstract. Check the word limit.

Acknowledgments

Most thesis authors put in a page of thanks to those who have helped them in matters scientific, and also indirectly by providing such essentials as food, education, genes, money, help, advice, friendship etc. *If any of your work is collaborative, you should make it quite clear who did which sections.*

Table of contents

The introduction starts on page 1, the earlier pages should have roman numerals. It helps to have the subheadings of each chapter, as well as the chapter titles. Remember that the thesis may be used as a reference in the lab, so it helps to be able to find things easily.

Introduction

What is the topic and why is it important? State the problem(s) as simply as you can. Remember that you have been working on this project for a few years, so you will be very close to it. Try to step back mentally and take a broader view of the problem. How does it fit into the broader world of your discipline?

Especially in the introduction, do not overestimate the reader's familiarity with your topic. You are writing for researchers in the general area, but not all of them need be specialists in your particular topic. It may help to imagine such a person---think of some researcher whom you might have met at a conference for your subject, but who was working in a different area. S/he is intelligent, has the same general background, but knows little of the literature or tricks that apply to your particular topic.

The introduction should be interesting. If you bore the reader here, then you are unlikely to revive his/her interest in the materials and methods section. For the first paragraph or two, tradition permits prose that is less dry than the scientific norm. If want to wax lyrical about your topic, here is the place to do it. Try to make the reader want to read the kilogram of A4 that has arrived uninvited on his/her desk. Go to the library and read several thesis introductions. Did any make you want to read on? Which ones were boring?

This section might go through several drafts to make it read well and logically, while keeping it short. For this section, I think that it is a good idea to ask someone who is not a specialist to read it and to comment. Is it an adequate introduction? Is it easy to follow? There is an argument for writing this section---or least making a major revision of it---towards the end of the thesis writing. Your introduction should tell where the thesis is going, and this may become clearer during the writing.

Literature review

Where did the problem come from? What is already known about this problem? What other methods have been tried to solve it?

Ideally, you will already have much of the hard work done, if you have been keeping up with the literature as you vowed to do three years ago, and if you have made notes about important papers over the years. If you have summarised those papers, then you have some good starting points for the review.

If you didn't keep your literature notes up to date, you can still do something useful: pass on the following advice to any beginning PhD students in your lab and tell them how useful this would have been to you. When you start reading about a topic, you should open a spread sheet file, or at least a word processor file, for your literature review. Of course you write down the title, authors, year, volume and pages. But you also write a summary (anything from a couple of sentences to a couple of pages, depending on the relevance). In other columns of the spread sheet, you can add key words (your own and theirs) and comments about its importance, relevance to you and its quality.

How many papers? How relevant do they have to be before you include them? Well, that is a matter of judgement. On the order of a hundred is reasonable, but it will depend on the field. You are the world expert on the (narrow) topic of your thesis: you must demonstrate this.

A political point: make sure that you do not omit relevant papers by researchers who are like to be your examiners, or by potential employers to whom you might be sending the thesis in the next year or two.

Middle chapters

In some theses, the middle chapters are the journal articles of which the student was major author. There are several disadvantages to this format.

One is that a thesis is both allowed and expected to have more detail than a journal article. For journal articles, one usually has to reduce the number of figures. In many cases, all of the interesting and relevant data can go in the thesis, and not just those which appeared in the journal. The degree of experimental detail is usually greater in a thesis. Relatively often a researcher requests a thesis in order to obtain more detail about how a study was performed.

Another disadvantage is that your journal articles may have some common material in the introduction and the "Materials and Methods" sections.

The exact structure in the middle chapters will vary among theses. In some theses, it is necessary to establish some theory, to describe the experimental techniques, then to report what was done on several different problems or different stages of the problem, and then finally to present a model or a new theory based on the new work. For such a thesis, the chapter headings might be: Theory, Materials and Methods, {first problem}, {second problem}, {third problem}, {proposed theory/model} and then the conclusion chapter. For other theses, it might be appropriate to discuss different techniques in different chapters, rather than to have a single Materials and Methods chapter.

Here follow some comments on the elements Materials and Methods, Theory, Results and discussion which may or may not correspond to thesis chapters.

Materials and Methods
This varies enormously from thesis to thesis, and may be absent in theoretical theses. It should be possible for a competent researcher to reproduce exactly what you have done by following your description. There is a good chance that this test will be applied: sometime after you have left, another researcher will want to do a similar experiment either with your gear, or on a new set-up in a foreign country. Please write for the benefit of that researcher.

In some theses, particularly multi-disciplinary or developmental ones, there may be more than one such chapter. In this case, the different disciplines should be indicated in the chapter titles.

Theory

When you are reporting theoretical work that is not original, you will usually need to include sufficient material to allow the reader to understand the arguments used and their physical bases. Sometimes you will be able to present the theory *ab initio,* but you should not reproduce two pages of algebra that the reader could find in a standard text. Do not include theory that you are not going to relate to the work you have done.

When writing this section, concentrate at least as much on the physical arguments as on the equations. What do the equations mean? What are the important cases?

When you are reporting your own theoretical work, you must include rather more detail, but you should consider moving lengthy derivations to appendices. Think too about the order and style of presentation: the order in which you did the work may not be the clearest presentation.

Suspense is not necessary in reporting science: you should tell the reader where you are going before you start.

Results and discussion

The results and discussion are very often combined in theses. This is sensible because of the length of a thesis: you may have several chapters of results and, if you wait till they are all presented before you begin discussion, the reader may have difficulty remembering what you are talking about. The division of Results and Discussion material into chapters is usually best done according to subject matter.

Make sure that you have described the conditions which obtained for each set of results. What was held constant? What were the other relevant parameters? Make sure too that you have used appropriate statistical analyses. Where applicable, show measurement errors and standard errors on the graphs. Use appropriate statistical tests.

Take care plotting graphs. The origin and intercepts are often important so, unless the ranges of your data make it impractical, the zeros of one or both scales should usually appear on the graph. You should show error bars on the data, unless the errors are very small. For single measurements, the bars should be your best estimate of the experimental errors in each coordinate. For multiple measurements these should include the standard error in the data. The errors in different data are often different, so, where this is the case, regressions and fits should be weighted (i.e. they should minimize the sum of squares of the differences weighted inversely as the size of the errors.) (A common failing in many simple software packages that draw graphs and do regressions is that they do not treat errors adequately. UNSW student Mike Johnston has written a plotting routine that plots data with error bars and performs weighted least square regressions. It is at http://www.phys.unsw.edu.au/3rdyearlab/graphing/graph.html). You can just 'paste' your data into the input and it generates a .ps file of the graph.

In most cases, your results need discussion. What do they mean? How do they fit into the existing body of knowledge? Are they consistent with current theories? Do they give new insights? Do they suggest new theories or mechanisms?

Try to distance yourself from your usual perspective and look at your work. Do not just ask yourself what it means in terms of the orthodoxy of your own research group, but also how other people in the field might see it. Does it have any implications that do not relate to the questions that you set out to answer?

Final chapter, references and appendices

Conclusions and suggestions for further work

Your abstract should include your conclusions in very brief form, because it must also include some other material. A summary of conclusions is usually longer than the final section of the abstract, and you have the space to be more explicit and more careful with qualifications. You might find it helpful to put your conclusions in point form.

It is often the case with scientific investigations that more questions than answers are produced. Does your work suggest any interesting further avenues? Are there ways in which your work could be improved by future workers? What are the practical implications of your work?

This chapter should usually be reasonably short---a few pages perhaps. As with the introduction, I think that it is a good idea to ask someone who is not a specialist to read this section and to comment.

References (See also under literature review)

It is tempting to omit the titles of the articles cited, and the university allows this, but think of all the times when you have seen a reference in a paper and gone to look it up only to find that it was not helpful after all.

Should you reference web sites and, if so, how? If you cite a journal article or book, the reader can go to a library and check that the cited document and check whether or not it says what you say it did. A web site may disappear, and it may have been updated or changed completely. So references to the web are usually less satisfactory. Nevertheless, there are some very useful and authoritative sources. So, *if the rules of your institution permit it*, it may be appropriate to cite web sites. (Be cautious, and don't overuse such citations. In particular, don't use a web citation where you could reasonably use a "hard" citation. Remember that your examiners are likely to be older and more conservative.) You should give the URL and also the date you downloaded it. If there is a date on the site itself (last updated on .....) you should included that, too.

Appendices

If there is material that should be in the thesis but which would break up the flow or bore the reader unbearably, include it as an appendix. Some things which are typically included in appendices are: important and original computer programs, data files that are too large to be represented simply in the results chapters, pictures or diagrams of results which are not important enough to keep in the main text.

A suggested thesis structure - Model 2 (Graduate):

Introduction

This note describes how to organize the written thesis which is the central element of your graduate degree. To know how to organize the thesis document, you first have to understand what graduate-level research is all about, so that is covered too. In other words, this note should be helpful when you are just getting started in your graduate program, as well as later when you start to write your thesis.

What Graduate Research is All About

The distinguishing mark of graduate research is *an original contribution to knowledge*. The thesis is a formal document whose sole purpose is to prove that you have made an original contribution to knowledge. Failure to prove that you have made such a contribution generally leads to failure.

To this end, your thesis must show two important things:

> you have identified a worthwhile problem or question which has not been previously answered,
> you have solved the problem or answered the question.

Your contribution to knowledge generally lies in your solution or answer.

What the Graduate Thesis is All About

Because the purpose of the graduate thesis is to prove that you have made an original and useful contribution to knowledge, the examiners read your thesis to find the answers to the following questions:

> what is this student's research question?
> is it a good question? (has it been answered before? is it a useful question to work on?)
> did the student convince me that the question was adequately answered?
> has the student made an adequate contribution to knowledge?

A very *clear* statement of the question is essential to proving that you have made an original and worthwhile contribution to knowledge. To prove the originality and value of your contribution, you must present a *thorough* review of the existing literature on the subject, and on closely related subjects. Then, by making *direct* reference to your literature review, you must *demonstrate* that your question (a) has not been previously answered, and (b) is worth answering. Describing how you answered the question is usually easier to write about, since you have been intimately involved in the details over the course of your graduate work.

If your thesis does not provide adequate answers to the few questions listed above, you will likely be faced with a requirement for major revisions or you may fail your thesis defence outright. For this reason, the generic thesis skeleton given below is designed to highlight the answers to those questions with appropriate thesis organization and section titles. The generic thesis skeleton can be used for any thesis. While some professors may prefer a different organization, the essential elements in any thesis will be the same. Some further notes follow the skeleton.

Always remember that a thesis is a *formal* document: every item must be in the appropriate place, and repetition of material in different places should be eliminated.

A Generic Thesis Skeleton

1. Introduction

This is a *general* introduction to what the thesis is all about -- it is *not* just a description of the contents of each section. Briefly *summarize* the question (you will be stating the question in detail later), some of the reasons why it is a worthwhile question, and perhaps give an overview of your main results. This is a birds-eye view of the answers to the main questions answered in the thesis (see above).

2. Background Information (optional)

A brief section giving background information may be necessary, especially if your work spans two or more traditional fields. That means that your readers may not have any experience with some of the material needed to follow your thesis, so you need to give it to them. A different title than that given above is usually better; e.g., "A Brief Review of Frammis Algebra."

3. Review of the State of the Art

Here you review the state of the art relevant to your thesis. Again, a different title is probably appropriate; e.g., "State of the Art in Zylon Algorithms." The idea is to *present* (critical analysis comes a little bit later) the major ideas in the state of the art right up to, but not including, your own personal brilliant ideas.

You organize this section *by idea*, and not by author or by publication. For example if there have been three important main approaches to Zylon Algorithms to date, you might organize subsections around these three approaches, if necessary:

3.1 Iterative Approximation of Zylons
3.2 Statistical Weighting of Zylons
3.3 Graph-Theoretic Approaches to Zylon Manipulation

4. Research Question or Problem Statement

Engineering theses tend to refer to a "problem" to be solved where other disciplines talk in terms of a "question" to be answered. In either case, this section has three main parts:

1. a concise statement of the question that your thesis tackles
2. justification, by *direct* reference to section 3, that your question is previously unanswered
3. discussion of why it is worthwhile to answer this question.

Item 2 above is where you *analyze* the information which you presented in Section 3. For example, maybe your problem is to "develop a Zylon algorithm capable of handling very large scale problems in reasonable time" (you would further describe what you mean by "large scale" and "reasonable time" in the problem statement). Now in your *analysis* of the state of the art you would show how each class of current approaches fails (i.e. can handle only small problems, or takes too much time). In the last part of this section you would explain why having a large-scale fast Zylon algorithm is useful; e.g., by describing applications where it can be used.

Since this is one of the sections that the readers are *definitely* looking for, highlight it by using the word "problem" or "question" in the title: e.g. "Research Question" or "Problem Statement", or maybe something more specific such as "The Large-Scale Zylon Algorithm Problem."

5. Describing How You Solved the Problem or Answered the Question

This part of the thesis is much more free-form. It may have one or several sections and subsections. But it all has only one purpose: to convince the examiners that you answered the question or solved the problem that you set for yourself in Section 4. So show what you did that is *relevant* to answering the question or solving the problem: if there were blind alleys and dead ends, do *not* include these, unless specifically relevant to the demonstration that you answered the thesis question.

6. Conclusions

You generally cover three things in the Conclusions section, and each of these usually merits a separate subsection:

1. Conclusions
2. Summary of Contributions
3. Future Research

Conclusions are *not* a rambling summary of the thesis: they are *short*, *concise* statements of the inferences that you have made because of your work. It helps to organize these as short numbered paragraphs, ordered from most to least important. All conclusions should be directly related to the research question stated in Section 4. Examples:

1. The problem stated in Section 4 has been solved: as shown in Sections ? to ??, an algorithm capable of handling large-scale Zylon problems in reasonable time has been developed.

2. The principal mechanism needed in the improved Zylon algorithm is the Grooty mechanism.

3. Etc.

The Summary of Contributions will be much sought and carefully read by the examiners. Here you list the contributions of *new* knowledge that your thesis makes. Of course, the thesis itself must substantiate any claims made here. There is often some overlap with the Conclusions, but that's okay. Concise numbered paragraphs are again best. Organize from *most* to *least* important. Examples:

1. Developed a much quicker algorithm for large-scale Zylon problems.

2. Demonstrated the first use of the Grooty mechanism for Zylon calculations.

3. Etc.

The Future Research subsection is included so that researchers picking up this work in future have the benefit of the ideas that you generated while you were working on the project. Again, concise numbered paragraphs are usually best.

7. References

The list of references is closely tied to the review of the state of the art given in section 3. Most examiners scan your list of references looking for the important works in the field, so make sure they are listed and referred to in section 3. Truth be known, most examiners also look for their own publications if they are in the topic area of the thesis, so list these too. Besides, reading your examiner's papers usually gives you a clue as to the type of questions they are likely to ask.

All references given *must* be referred to in the main body of the thesis. Note the difference from a Bibliography, which may include works that are not directly referenced in the thesis. Organize the list of references either alphabetically by author surname (preferred), or by order of citation in the thesis.

8. Appendices

What goes in the appendices? Any material which impedes the smooth development of your presentation, but which is important to justify the results of a thesis. Generally it is material that is of too nitty-gritty a level of detail for inclusion in the main body of the thesis, but which should be available for perusal by the examiners to convince them sufficiently. Examples include program listings, immense tables of data, lengthy mathematical proofs or derivations, etc.

Comments on the Skeleton

Again, the thesis is a formal document designed to address the examiner's two main questions. Sections 3 and 4 show that you have chosen a good problem, and section 5 shows that you solved it. Sections 1 and 2 lead the reader into the problem, and section 6 highlights the main knowledge generated by the whole exercise.

Note also that everything that *others* did is carefully separated from everything that *you* did. Knowing who did what is important to the examiners. Section 4, the problem statement, is the obvious dividing line. That's the main reason for putting it in the middle in this formal document.

Getting Started

The best way to get started on your thesis is to prepare an extended outline. You begin by making up the Table of Contents, listing each section and subsection that you propose to include. For each section and subsection, write a brief point-form description of the contents of that section. The entire outline might be 2 to 5 pages long. Now you and your thesis supervisor should carefully review this outline: is there unnecessary material (i.e. not directly related to the problem statement)? Then remove. Is there missing material? Then add. It is much less painful and more time-efficient to make such decisions early, during the outline phase, rather than after you've already done a lot of writing which has to be thrown away.

How Long Does it Take to Write a Thesis?

Longer than you think. Even after the research itself is all done -- models built, calculations complete -- it is wise to allow at least one complete term for writing the thesis. It's not the physical act of typing that takes so long, it's the fact that writing the thesis requires the complete organization of your arguments and results. It's during this formalization of your results into a well-organized thesis document capable of withstanding the scrutiny of expert examiners that you discover weaknesses. It's fixing those weaknesses that takes time.

This is also probably the first time that your supervisor has seen the formal expression of concepts that may have been approved previously in an informal manner. Now is when you discover any misunderstandings or shortcomings in the informal agreements. It takes time to fix these. Students for whom English is not the mother tongue may have difficulty in getting ideas across, so that numerous revisions are required. And, truth be known, supervisors are sometimes not quick at reviewing and returning drafts.

Bottom line: leave yourself enough time. A rush job has painful consequences at the defence.

Tips

Always keep the reader's backgrounds in mind. Who is your audience? How much can you reasonably expect them to know about the subject before picking up your thesis? Usually they are pretty knowledgeable about the general problem, but they haven't been intimately involved with the details over the last couple of years like you have: spell difficult new concepts out clearly. It sometimes helps to mentally picture a real person that you know who has the appropriate background, and to imagine that you are explaining your ideas directly to that person.

Don't make the readers work too hard! **This is fundamentally important. You know what few questions the examiners need answers for (see above). Choose section titles and wordings to clearly give them this information. The harder they have to work to ferret out your problem, your defence of the problem, your answer to the problem, your conclusions and contributions, the worse mood they will be in, and the more likely that your thesis will need major revisions.**

A corollary of the above: *it's impossible to be too clear*! Spell things out carefully, highlight important parts by appropriate titles etc. There's a huge amount of information in a thesis: make sure you direct the readers to the answers to the important questions.

Remember that *a thesis is not a story*: it usually doesn't follow the chronology of things that you tried. It's a formal document designed to answer only a few major questions.

Avoid using phrases like "Clearly, this is the case..." or "Obviously, it follows that ..."; these imply that, if the readers don't understand, then they must be stupid. They might not have understood because you explained it poorly.

Avoid *red flags*, claims (like "software is the most important part of a computer system") that are really only your personal opinion and not substantiated by the literature or the solution you have presented. Examiners like to pick on sentences like that and ask questions like, "Can you demonstrate that software is the most important part of a computer system?"

A Note on Computer Programs and Other Prototypes

The purpose of your thesis is to clearly document an original contribution to *knowledge*. You may develop computer programs, prototypes, or other tools as a means of proving your points, but remember, the thesis is *not* about the tool, it is about the contribution to knowledge. Tools such as computer programs are fine and useful products, but you can't get an advanced degree just for the tool. You must use the tool to demonstrate that you have made an original contribution to knowledge; e.g., through its use, or ideas it embodies.

Some sites with related material
How to survive a thesis defence
Research resources and links supplied by Deakin University
"Final year projects": a guide from Mike Hart at King Alfred's College, Winchester, UK
Postgraduate Student Resources supplied by University of Canberra
A useful aid to surviving meetings with management
The National Association of Graduate - Professional Students (USA)
Some relevant texts
Stevens, K. and Asmar, C (1999) 'Doing postgraduate research in Australia'. Melbourne University Press, Melbourne ISBN 0 522 84880 X.
Phillips, E.M and Pugh, D.S. (1994) 'How to get a PhD : a handbook for students and their supervisors'. Open University Press, Buckingham, England
Tufte, E.R. (1983) 'The visual display of quantitative information'. Graphics Press, Cheshire, Conn.
Tufte, E.R. (1990) 'Envisioning information' Graphics Press, Cheshire, Conn.

20 REFERENCES

1 What is This Thing Called Science?, Chalmers, A. Hackett Publishing Co., 1990
2 The Structure of Scientific Revolutions, Kuhn, T, Chicago: Univ. Chicago Press, 1962
3 Science and Its Fabrication, Chalmers, Minnesota Press, 1990
4 An Introduction to The Philosophy of Science, Carnap, R., NY: Basic Books, 1994
5 The Logic of Scientific Discovery, Popper, K.R., NY: Routledge, 1999
6 The Golem, What You Should Know about Science,Harry Collins and Trevor Pinch, Cambridge University Press, 2000
7 Objective Knowledge, Popper, K.R., Oxford, U.K.: Clarendon, 1972
8 Högskolans Lågvattenmärken, red. S O Hansson & P Sandin. Natur och Kultur, 2000
9 Against Method, Feyerabend, P., London, U.K.: Verso, 2000
10 Farewell to Reason, Feyerabend, P., London, UK: Verso, 1987
11 Filosofiska undersökningar, Ludwig Wittgenstein, sv. ö., Thales, 1992
12 Om visshet, Ludwig Wittgenstein, sv. ö. Thales, 1992

13 Introduktion till vetenskapsteorin, Lars-Göran Johansson, Thales, Stockholm 1999
14 Introduction to Philosophical Analysis, John Hospers, 4th ed London, Routledge, 1997
15 Philosophy of Natural Science, Carl Hempel, Englewood Cliffs, Prentice-Hall, 1966
16 Mr. Tompkins in Paperback, George Gamow, Cambridge University Press, 2000
17 Theory of Computing, Efim Kinber, Carl Smith. Prentice Hall, 2001
18 Introduction to Languages and the Theory of Computation, John C. Martin, McGraw-Hill, 1997
19. On Human Communication; A Review, a Survey, and a Criticism, Collin Cherry ,
M.I.T. Press, 1968
19 Semiotics: The Basics, Daniel Chandler, Routledge, 2001
20 Artificial Intelligence, S Russel & P Norvig, Prentice Hall, 1995

21 BACKGROUND LITERATURE

1 Philosophie, Gesellschaft und Physik, Robert Hofstetter, Verlag Hölder-Pichler-Tempsky, Wien 1992
2 Philosophy of Mathematics, Zvonimir Sikic, (in croatian), SK, Zgreb, 1995
3 Fuzzy Thinking, The New Science of Fuzzy Logic, Bart Kosko, HarperCollinsPublishers, London, 1994
4 The First ThreeMinutes (A modern view of the origin of the universe), Steven Weinberg, Collins Publishing Group, London, 1983
5 The Big Bang, The Creation and Evolution of the Universe, Joseph Silk, Freeman and Co., San Francisco, 1980.
6 Basic Concepts of Quantum Mechanics, L.V.Tarasov, Mir Publishers, Moscow, 1980
7 The Restless Universe, Max Born, Dover Publications, NY, 1951 (reprint)
8 Gödel, Escher, Bach, ett Evigt Gyllene Band, Douglas R. Hofstadter, Brombergs, 1992
9 The Emperor's New Mind Mind, Concerning Computers, Minds, and the Laws of Physics, Roger Penrose, Oxford University Press, 1989
10 Shadows of the Mind, A Search for the missing Science of Consciousness, Roger Penrose, Oxford University Press, 1994
11 The Mind's I, Douglas R. Hofstadter and Daniel Dennett, eds.,Bantam Books, 1982
12 Logic of Matemathical Discovery, Imre Lakatos, Cambridge University Press, 1979
13 Epistemology (in croatian), Ivan Supek, Mladost, Zagreb, 1971
14 Metaphysics, Aristoteles (in croatian), Hrvatska sveucilisna naklada, 1992
15 Parmenides et Philebus, Platon (in serbian),BIGZ, 1973
16 Einstein and Philosophical Problems of 20th Century Physics, Delokarov, (in russian), Nauka, 1979
17 Die Fragmente derVorsokratiker (Band I & II), Griechisch und Deutsch, Herman

Diels, Weidmann, 1974
18 Modern Experimental Physics and Epistemology (in croatian), Tomislav Petkovic, SK, Zgreb, 1990
19 Atom Leads the Game (Questions of Quantum Physics), (in croatian), Adamic et al. SK, Zgreb, 1973
20 Quamtum Dice, L Ponomarjov (in russian), Molodaja gvardija, 1971
21 Philosophers Confession, G. W. Leibniz (in serbian), Grafos, 1988
22 Nuclear Rites, (An Antropological Study of Lawrence Livermore National Lab Researcher Group), Hugh Gusterson, University of California Press, 1996
23 The Character of Physical Law, Richard Feyman, Cox and Wyman LTD, London, 1965
24 The Birth of Modern Mathematics, Z Sikic (in croatian), SK, Zgreb, 1989
25 Naturerklärung und Psyche, C. G. Jung, W Pauli, Walter-Vrlag AG, Olten, 1970
26 On the Past and the Future of the Universe, J Strnad (in croatian) SK, Zgreb, 1989
27 Filosofiska frågor, äventyr i teknikens värld, 1998, ISBN: 91-26-97860-1
28 Risker i tekniska system, Red. G. Grimvall, T. Thedèen, 1998, ISBN: 91-26-98200-5
29 Fermats gåta, Simon Singh, Norstedts, 1998
30 Kernenergie, Die weibliche Technik, J Tennenbaum, Dr. Böttiger Verlags-Gmbh, 1994
31 History of Physics, Ivan Supek, (in croatian), SK, Zgreb, 1990
32 "Subtle is the Lord…", The Science and the Life of Albert Einstein, Abraham Pais, Oxford University Press, 1982
33 The Paradigms of Programming, Robert W Floyd, Communications of the ACM, 22(8): 455-460, 1979
34 Relativism och rationalism (Relativity, Rationalism and the Sociology of Knowledge") B. Barnes & D. Bloor: Ed. M. Hollis & S. Lukes (1982).
35 "Den rationella övertygelsen (1968/1984) (Teknik och vetenskap som ideologi). Jürgen Habermas
36 The Universal Computer, The Road from Leibniz to Turing, Martin Davies, Norton, 2000

APPENDIX 12 Interesting Links

http://web.clas.ufl.edu/users/rhatch/pages/ (History of Science)
http://www.fordham.edu/halsall/science/sciencesbook.html (Internet History of Science Sourcebook) http://liinwww.ira.uka.de/bibliography/index.html (Computer Science Bibliographies) http://www.physlink.com (Physics and Astronomy on-line)
http://www.earlham.edu/~peters/philinks.htm (Guide to Philosophy on the Internet)
http://www.bib.mh.se/osd/nyfosd/omars00.htm (Mitthögskolans bibliotek, Östersund)
http://galileo.imss.firenze.it/vr/index.html (Galileo Museum Florenze)
http://philosophy.wisc.edu/forster/PhilSci/default.htm (Philosophy of Science Wisconsin)
http://www.utm.edu/research/iep/ (The Internet Encyclopedia of Philosophy)
http://www.nature.com/nature/ (Nature)
http://www.infidels.org/library/historical/charles_darwin/origin_of_species/ (Darwin)
http://www.kb.nl/dutchess/08/00/info-3551.html (Directory of On-Line Philosophy

Papers) http://www.kb.nl/dutchess/index.html (Search Engine)
http://www.intrepidsoftware.com/fallacy/toc.htm (Logical Fallacies)
http://www.intrepidsoftware.com/fallacy/syllog.htm (Syllogistic Fallacies)
http://www.websyte.com/alan/metamul.htm (Metaphysics)
http://www.bcc.ctc.edu/philosophy/jokes.htm (Philosophy Jokes)
http://www.netfunny.com/rhf/jokes/computer.html (Computer & Science Jokes)
http://www.u.arizona.edu/~chalmers/phil-humor.html (Philosophical humor)
http://www.newscientist.com/nsplus/insight/quantum/genious.html (New Scientist Quantum World) http://www.blackwellpublishers.co.uk/philos/#stars (Philosophy Resources)
http://digital.library.upenn.edu/webbin/book/subjectstart?QA (Online books: Math and Comp. Sci.) http://digital.library.upenn.edu/books/ (Books On-Line)
http://www.student.nada.kth.se/~d90-mst/geb/ (Mårten's GEB Page)
http://cns.pds.pvt.k12.ny.us/~jonathan/geb.html (Godel, Escher, Bach)
http://members.aol.com/McNelis/medsci_index.html (The Medieval Science Page)
http://encyclozine.com/Science/ (Worlds of Science)
http://encyclozine.com/Science/Method/ (What is Science?)
http://encyclozine.com/Science/Philosophy/ (The Philosophy of Science)
http://www.rhodes.edu/Philhtmls/philnet.html (Search of Philosophy)
http://www.earlham.edu/~peters/philinks.htm (Guide to Philosophy on the Internet)
http://www.geocities.com/SoHo/Museum/3828/escher.html (Esher) http://antioch-college.edu/~andrewc/pedagogy.html (Course Materials in Philosophy)
http://www.exploratorium.edu/origins/cern/place/index.html (CERN)
http://cern.web.cern.ch/CERN/ (CERN)
http://www.pbs.org/wnet/hawking/html/home.html (Stephen Hawking's Universe)
http://www.exploratorium.edu/origins/cern/ideas/bang2.html (CERN, Standard Model)
http://www.sciam.com/2000/0400issue/0400scicit1.html (Scientific American: Fireballs of Free Quarks)
http://www.ai.sri.com/~luong/research/proof.html (Dictionary of useful proof methods - humor) http://www.princeton.edu/~missouri/humor12.html (How to prove it - humor)
http://www.informatik.htw-dresden.de/~logic/conclusions/rule3.html (Logic)
http://www.sjsu.edu/depts/itl/graphics/induc/ind-ded.html (Induction and Deduction)
http://classroomextension.com/primers/criticalthinking/ (Critical Thinking)
http://mrrc.bio.uci.edu/se10/causality.html (Causality)
http://bayes.cs.ucla.edu/IJCAI99/mainframes.html (Reasoning with Cause and Effect)
http://dir.yahoo.com/Science/Research/Scientific_Method__The/
http://elib.cs.sfu.ca/Collections/CMPT/cs-journals/ (Computing Science Journals)
http://citeseer.nj.nec.com/ (Research Index)
http://www.cv.nrao.edu/fits/www/yp_comp_sci.html (Computer Science-related resources) http://www.cs.ucla.edu/csd/cs_catalog.html#scope (Computer Science Program and Courses) http://cdsweb.u-strasbg.fr/astroweb/comp_sci.html (AstroWeb: Computer Science Resources) http://www.cs.unb.ca/~alopez-o/comp-faq/faq.html (Comp.Theory FAQ)
http://www.brooklyn.liunet.edu/cwis/bklyn/library/resources/social/computer/compnotes.htm
(Computer Web) http://users.ox.ac.uk/~jrlucas/implic.html (The Implications of Gödel's Theorem) http://www.thymos.com/tat/machine.html (Thinking About Thought) http://nl.ijs.si/~damjan/g-m-c.html (Gödel)

http://www.geometry.net/goedel.html (K Gödel)
http://www.astr.ua.edu/4000WS/4000WS.html (Women in Science)
http://liinwww.ira.uka.de/bibliography/ (Computer Science Bibliographies)
http://www.cvc.org/science/kepler.htm (Kepler laws)
http://www.literature.org/authors/carroll-lewis/alices-adventures-in-wonderland/
(Alice's Adventures in Wonderland, Lewis Carroll)
http://www.xrefer.com/entry/551556 (Causality)
http://www.exploratorium.edu/origins/cern/ (CERN)
http://www.monitor.net/rachel/r586.html (Precautionary Principle)
http://www.perseus.tufts.edu/GreekScience/Students/Kristen/Aristarchus.html#three
(Astronomy. Aristarchus) http://members.aol.com/gca7sky/planets.htm Our Solar System http://www.scottlan.edu/lriddle/women/women.html (Women Mathematicians)
http://www-history.mcs.st-andrews.ac.uk/history/Mathematicians/ (Biographies of Mathematicians) http://plato.stanford.edu/entries/search.html (Stanford Encyclopedia of Philosophy) http://es.rice.edu/ES/humsoc/Galileo/Catalog/catalog.html (Catalog of the Scientific Community in the 16th and 17th Centuries)
http://www.hsa.brown.edu/~maicar/SearchGML.html (Greek Mythology Link)
http://hypatia.dcs.qmw.ac.uk/ (Hypatia Electronic Library in Computer Science and Pure Mathematics)
http://ucsu.colorado.edu/~brindell/soc-epistemology/Bibliographies/Feminist_Soc_Epis/philscience.htm Bibliography: Feminist Epistemology http://www.geocities.com/RainForest/3621/PHILINKS.HTM Links on the Philosophy of Technology http://math.rice.edu/~joel/NonEuclid/NonEuclid.html Non-Euclidean Geometry http://www.utm.edu/research/iep/text/descart/des-meth.htm Discourse on the Method Rene Descartes 1637 http://www.utm.edu/research/iep/ The Internet Encyclopedia of Philosophy

APPENDIX 13 Dictionary

Absolute-Modern science has abandoned the following myths of absolute:

- Absolute truth
- Absolute space (geometry)
- Absolute time
- Absolute motion/rest
- Absolute form of matter (mass/energy)
- Absolute determinism
- Absolute causality

Empiricism (Lat. empirismus, the standpoint based on experience). Primarily, the term signifies the theory that the phenomena of consciousness are simply the product of sensuous (physical) experience. Secondarily, in its logical (epistemological) usage, it designates the theory that all human knowledge is derived exclusively from experience, which means, either explicitly orimplicitly, external sense-percepts and internal representations and inferences. In this connection it is opposed to Intellectualism, Rationalism and Apriorism. Empiricism appears in the history of philosophy in three principal forms: Materialism,Sensism, and Positivism.

Evidence - the observations and measurements made to understand phenomena.

Epistemology – theory of knowledge.

Experiment [Lat. experimentum] A trial or special observation, made to confirm or disprove something; under conditions determined by the experimenter; an act or operation undertaken in order to discover someunknown principle or effect, or to test, establish, or illustrate some suggest or known truth; practical test; poof.

Feminist philosophy of science-Feminist philosophers have focused very much on scientific methodology. They have developed a number of new methodologies, including an adapted historical materialism (Nancy Hartsock), phenomenology (Dorothy E. Smith), and interactive observation processes (Evelyn Fox Keller). A principled rejection of a model of objectivity which relies on detachment is common theme.

Feminist philosophers bring a new perspective into study of scientific method. It is however worth to remind that the key quality of science is its generality and universality. Feminist contributions are to be evaluated in that perspective.
Hermeneutics Derived from a Greek word connected with the name of the god Hermes, the reputed messenger and interpreter of the gods.

Hermeneutics studies the principles of interpretation of the meaning of written texts. It includes the whole question of how a particular text is 'received,' especially considerations such as discerning the authors intentions as well as understanding the social context and the thought-forms of the period.

Grounded Theory is a strategy for handling research data within Social Sciences, providing modes of conceptualization for describing and explaining. It emphasizes the inductive process. Most research today is designed to verify existing theories, not to generate new ones.

Researchers seek out small gains of knowledge from existing "grand theories" rather than explore new areas. The existing research culture emphasizes good scientific, quantitative verification studies and downplays more qualitative studies whose objective is theory generation. Most of theory is thus generated through logical deduction from past studies and knowledge and not from the data itself. Grounded theory is an attempt to improve present situation towards more innovative theories.

One comment we can make here is that there is no observation or data collection without some (however simple but still) theoretical idea about what the goal is, what it might be we are observing and what we are looking for. The focus on data collection and systematization (which relies heavily on deduction of course) can be explained by the fact that the existing theories within Social Sciences have not been satisfactory. There is a trend (cultural phenomenon) to force these sciences to go towards more and more pronounced quantitative methods. So it is understandable that focus is set to data acquisition and systematization. There is nevertheless a question if this field of human research can ever be expressed quantitatively in the same way as natural sciences with their considerably simpler research objects.

The "exactness" of science is closely related to the language that is used to formulate scientific laws. Social sciences use sort of language that is almost as free and unrestricted as our everyday language. And it is not a mere coincidence. Social Sciences operate on our everyday life level. And that means they have to concentrate the whole of our everyday complexity in preferably the same form as geometry does with its abstract and simple idealized objects.

Hypothesis - a proposition explaining the occurrence of a phenomenon or phenomena, often asserted as a conjecture to guide further investigation.

Inference, deductive - a conclusion based on reasoning from accepted premises.

Inference, inductive (empirical induction)- a conclusion based on repeated observation.
Materialism in its basic shape was taught by the ancient atomists (Democritus, Leucippus, Epicurus, Lucretius), who, reducing all reality to atoms and motion, taught that experience is generated by images reflected from material objects through the sensory organs into the soul. The soul, a mere complex of the finest atoms, perceives not the objects but their effluent images. With modern materialists (Helvetius, Diderot, Feuerbach, etc.), knowledge is accounted for either by cerebral secretion or by motion. Avenarius, Mach, etc. subtilize this process so far as to reduce all experience to internal (empirio-criticism).

Metaphor – a form of expression where signs (words, pictures, etc) are not used in their ordinary meaning, but are referring to something else which has the qualities one wants to express. A crane[23] may be used as a metaphor of eternal life.

Multiple working hypotheses - a method of research where one considers not just a single hypothesis but instead several hypotheses that might explain the phenomenon under study. Many of these hypotheses will be contradictory, so that some, if not all, will prove to be false.

Myth In common jargon, a myth is a fiction - something which is untrue. Scholars of mythology define myth differently: a myth is a special kind of story which tries to interpret some aspect of the world around us. Robert W. Brockway, in his book Myth from the Ice Age to Mickey Mouse concisely summarizes a number of different scholarly ideas about the meaning of myth as follows:

"Myths are stories, usually, about gods and other supernatural beings (Frye). They are often stories of origins, how the world and everything in it came to be in illo tempore (Eliade). They are usually strongly structured and and their meaning is only discerned by linguistic analysis (Lévi-Strauss). Sometimes they are public dreams which, like private dreams, emerge from the unconscious mind (Freud). Indeed, they often reveal the archetypes of the collective unconscious (Jung). They are symbolic and metaphorical (Cassirer). They orient people to the metaphysical dimension, explain the origins and nature of the cosmos, validate social issues, and, on the psychological plane, address themselves to the innermost depths of the psyche (Campbell).

Some of myths are explanatory, pre-scientific attempts to interpret the natural world (Frazer). As such, they are usually functional and are the science of primitive peoples (Malinowski). Often, they are enacted in rituals (Hooke).

Natural law - a term rarely used today. Nineteenth-century science presumed that it could arrive at unchangeable, absolutely true, universal statements about nature, and these were to be "natural laws". Newton's ideas about gravitation, for example, were considered the "laws of gravity".

NP (nondeterministic polynomial time) <complexity> A set or property of computational decision problems solvable by a nondeterministic Turing Machine in a number of steps that is a polynomial function of the size of the input. The word "nondeterministic" suggests a method of generating potential solutions using some form of nondeterminism or "trial and error". This may take exponential time as long as a potential solution can be verified in polynomial time.

[23] a large bird with very long neck and long legs that eats fish, frequent in Japanese woodcuts.

NP-complete- (NPC, Nondeterministic Polynomial time complete) A set or property of computational decision problems which is a subset of NP (i.e. can be solved by a nondeterministic Turing Machine in polynomial time), with the additional property that it is also NP-hard. Thus a solution for one NP-complete problem would solve all problems in NP.

Positivism Positivists, following Comte, do not deny the supersensible[24]; they declare it unknowable. The only source of cognition, they claim, is sense-experience, experiment, and induction from phenomena. John Stuart Mill, following Hume, reduces all knowledge to series of conscious states linked by empirical associations and enlarged by inductive processes. The mind has no certainty of an external world, but only of "a permanent possibility of sensations". Spencer makes all knowledge relative. The actual existence of things is their persistence in our consciousness. Consciousness contains only subjective feelings. The relative supposes the absolute, but the latter is unknowable to us. Absolute is the object of faith and religion (Agnosticism[25]). All things, mind included, have resulted from a cosmic process of mechanical evolution in which they are still involved; hence all concepts and principles are in a continuous change.

Sensism By denying any essential difference between sensations and ideas (intellectual states), sensism logically involves materialism.

Sensism, which is found with Empedocles and Protagoras amongst the ancients, was given its first systematic form by Locke, Bacon and Hobbes. Locke derives all simple ideas from external experience (sensations), and all complex ideas from internal experience (reflection). Substance and cause are simply associations of subjective phenomena. Universal ideas are mere mental fabrication. Locke admits the existence, though he denies the demonstrability of the soul which is an immaterial and immortal principle.

Berkeley, accepting the teaching of Locke that ideas are only transformed sensations, subjectivizes not only the sensible or secondary qualities of matter (e.g. colour and sound), but also the primary qualities (extension, space, etc.), which Locke held to be objective. Berkeley denies the objective basis of universal ideas and indeed of the whole material universe. The reality of things he places in their being perceived, and this "perceivedness" is achieved in the mind by God, not by the object or subject. He still retains the substance-reality of the human soul and of spirits generally, God included.

Hume agrees with his two empiricist predecessors in teaching that the mind knows only its own subjective organic impressions, whereof ideas are but the images. The supersensible is therefore unknowable; the principle of causality is resolved into a mere feeling of successiveness of phenomena; its necessity is reduced to a subjective feeling resulting from uniform association experienced in consciousness, and the spiritual essence or substantial being of the soul is dissipated into a series of conscious states.

[24] what is beyond our senses, reality that is unobservable [25] skepticism

Theory - a coherent set of propositions that explain a class of phenomena that are supported by extensive factual evidence and that may be used for prediction of future observations. (John A. Cagle's definition) A theory is a scientific account of phenomena. At a minimum theory is a strategy for handling data in research, providing a conceptual system for describing and explaining. (Darnell's definition)A theory is a set of statements, including some law-like generalizations, systematically and logically related such that the set implies something about reality. It is an argument that purports to provide a necessary and sufficient explanation for a range of phenomena. It must be capable of corrigibility - that is, it must be possible to disconfirm or jeopardize it by making observations. A theory is valuable to the extent that it reduces the uncertainty about the outcome of a specific set of conditions.

Example of familiar theories:

Copernicus's theory of the heliocentric solar system,
Newton's theory of gravity,
Einstein's theory of relativity, and
Darwin's theory of natural selection

- Each of these theories draws on huge numbers of facts:
- observations of the passage of the sun and planets for the heliocentric theory;

the behavior of the planets, of projectiles, and rather famously of apples for the theory of gravity, and
the existence and location of fossils, as well as the modern distribution and reproduction of organisms, for the theory of natural selection.

Ockham's Razor (also spelled Occam's Razor even called law of economy, or law of parsimony) - a philosophical statement developed by William of Ockham, (1285–1347/49), a scholastic, that Pluralitas non est ponenda sine necessitate; "Plurality should not be assumed without necessity." The principle gives precedence to simplicity; of two competing theories, the simplest explanation of an entity is to be preferred. The principle is also expressed "Entities are not to be multiplied beyond necessity."

As a consequence of Ockham's Razor , theory should minimize unsupported assumptions.

Paradigm - a way of thinking, a conceptual world view.

Thomas Kuhn in The Structure of Scientific Revolution defined the concept of "paradigm shift". Kuhn argues that scientific progress is not evolutionary, but rather is a "series of peaceful interludes punctuated by intellectually violent revolutions", and in those revolutions "one conceptual world view is replaced by another".

In science, a major example of a change in paradigms was the change from Scholasticism to Scientific Empiricism, roughly around AD 1600.
Scholasticism, which assumed that answers to questions about nature could be deduced from ancient texts and philosophical principals, gave way to the modern view of science where induction from accumulated evidence is (or should be) the underpinning of theories. When Galileo was threatened by church authorities with torture for his claim that the earth orbits the sun, Galileo and his accusers were not only at odds about an astronomical theory. They were also arguing, if unwittingly, because they were using two very different paradigms: the churchmen were using scholasticism, and Galileo "scientific empiricism".

Parable – a short story told in order to make moral, religious or philosophical point.

Paradox –statement or condition that is strange because it involves two opposite facts or qualities, which appear impossible to be true at the same time (like rectangular circle or a barber who shaves all those who do not shave themselves).

Philosophy of Language In the end of the previous century, both Edmond Husserl and Gottlob Frege searched for such foundations of mathematics, which would allow understanding its subject matter without taking resort to psychology. Their ways of doing it parted; and each of them laid the foundation of an important philosophical school, Husserl of phenomenology and Frege of analytic philosophy.

Nowadays, however, the boundary between analytical philosophy and phenomenology seems to slowly fade away.

7. DISEMINATING KNOWLEDGE IN ENGINEERING

7.1 Searching for Research Papers

The following seven steps outline a simple and effective strategy for finding information for a research paper and documenting the sources you find. Depending on your topic and your familiarity with the library, you may need to rearrange or recycle these steps. Adapt this outline to your needs.

Step 1: Identify and Develop Your Topic

How to Find and Develop a Viable Research Topic

Step One: Identify a Topic.

State your topic idea as a question. For example, if you are interested in finding out about use of alcoholic beverages by college students, you might pose the question, "What effect does use of alcoholic beverages have on the health of college students?"

Identify the main concepts or keywords in your question. In this case they are alcoholic beverages, health, and college students.

Step Two: Test Your Topic.

Test the main concepts or keywords in your topic **by looking them up** in the appropriate background sources or **by using them as search terms** in the Library Catalog and in periodical indexes.

> If you are finding too much information and too many sources, narrow your topic by using the **and** operator: beer and health and college students, for example. Finding too little information may indicate that you need to broaden your topic. For example, look for information on students, rather than college students. Link synonymous search terms with or: alcoholic beverages or beer or wine or liquor. Using truncation with search terms also broadens the search and increases the number of items you find.

Once you have identified and tested your topic, you're ready to take **the next step, finding background information** on your research topic.

SUMMARY: State your topic as a question. For example, if you are interested in finding out about use of alcoholic beverages by college students, you might pose the question, "What effect does use of alcoholic beverages have on the health of college students?" Identify the main concepts or keywords in your question.

Step 2: Find Background Information

Once you have identified the main topic and keywords for your research, find one or more sources of background information to read. These sources will help you understand the broader context of your research and tell you in general terms what is known about your topic. The most common background sources are **encyclopedias and dictionaries** from the print and online reference collection. **Class textbooks** also provide background information.

Use Encyclopedias And Dictionaries

Reference Universe

> ***Reference Universe*** is a database you can search to find subject encyclopedias on any topic.

You can also find **encyclopedias and dictionaries** for specific topics by using the Library Catalog, by consulting a Subject Guide (an annotated bibliography of selected reference sources on a specific subject), by checking the General Interest and Reference section of the Library Gateway, or by **asking a reference librarian** to suggest appropriate titles. For more general background you may wish to consult Encyclopaedia Britannica Online.

Tip: Exploit Bibliographies

> Read the background information and note any useful sources (books, journals, magazines, etc.) listed in the bibliography at the end of the encyclopedia article or dictionary entry. The sources cited in the bibliography are good starting points for further research.
> Look up these sources in the Library Catalog and periodical indexes. Check the subject headings listed in the subject field of the online record for these books and articles. Then do subject searches using those subject headings to locate additional titles.

> Remember that many of the books and articles you find in the Library Catalog and periodical indexes will themselves have bibliographies. Check these bibliographies for additional relevant resources for your research.

By using this technique of routinely following up on sources cited in bibliographies, you can generate a surprisingly large number of books and articles on your topic in a relatively short time.

SUMMARY: Look up your keywords in the indexes to subject encyclopedias. Read articles in these encyclopedias to set the context for your research. Note any relevant items in the bibliographies at the end of the encyclopedia articles. Additional background information may be found in your lecture notes, textbooks, and reserve readings.

Step 3: Use Catalogs To Find Books And Media

Use The Library Catalog And Other Online Catalog

To find the location and call number of over seven million books (as well as millions of video, audio, microform, map, serial, and rare titles) owned by the University Library, use the **Library Catalog**.

Library of Congress Call Numbers

The Library uses Library of Congress call numbers. For a brief introduction, ask at reference for our **Library of Congress Classification** handout or see this web site: Library of Congress Classification Outline.

Other Ways of Finding Books (and more)

Catalogs of books and other matierials available at a large number of other libraries are under the **RLIN/Eureka** and **OCLC WorldCat** on the Library Gateway. Both **OCLC WorldCat** and **RLIN/Eureka** bring together the holdings of hundreds or thousands of libraries in a single searchable database. They use somewhat different searching methods from the Library Catalog, so follow their instructions carefully.

You can also search the online catalogs of many libraries around the world directly over the Internet using a Web site called **LibDex**.

SUMMARY: Use keyword searching for a narrow or complex search topic. Use subject searching for a broad subject. Print or write down the citation (author, title,etc.) and the location information (call number and library). Note the circulation status. When you pull the book from the shelf, scan the bibliography for additional sources. Watch for book-length bibliographies and annual reviews on your subject; they list citations to hundreds of books and articles in one subject area. Check the standard subject subheading "--BIBLIOGRAPHIES," or titles beginning with Annual Review of... in the Library Catalog.

Step 4: Use Indexes to Find Periodical Articles

What Are Periodicals?

Periodicals are continuous publications such as journals, newspapers, or magazines. They are issued regularly (daily, weekly, monthly, or quarterly).

The Library Catalog includes records for all the periodicals which are received.

The Library Catalog does not include information on individual articles in periodicals. **To find individual periodical articles by subject, article author, or article title, use periodical indexes.**

Finding Periodicals

To Find An Article, Use Periodical Indexes

When you don't have the citation to a specific article, but you do want to find articles on a subject, by a specific author or authors, or with a known article title, you need to use one or more **periodical indexes**. But how do you know **which** periodical index to use?

What kind of periodicals are you looking for?

- **scholarly journals?**
- **newspapers and substantive news sources?**
- **popular magazines?**
- **all three kinds?**

If you want articles from **scholarly, research, peer-reviewed journals**, ask a reference librarian to recommend an index/database for your topic. Some databases index journals exclusively. You can also use the subject menu in Find Databases on the Library Gateway to locate databases that index scholarly publications.
If you want **newspaper articles**, see this guide to newspaper indexes and full-text newspaper databases. First choice for finding newspaper articles is **Lexis Nexis Academic**
If you want **popular magazines**, use **Academic Search Premier** or **ProQuest Research Library.** A printed index, **Reader's Guide to Periodical Literature** covering popular magazines since 1890 is found in the Uris Library reference alcove. The online index **Reader's Guide Retrospective** indexes popular

magazines from 1890 to 1982 online. **Periodical Contents Index** covers some popular magazines for an even broader time period: 1770 to 1993.
If you want an index to **all three kinds** of articles, use **Academic Search Premier** or **ProQuest Research Library**. To find older articles, try **Periodical Contents Index**; it indexes periodicals from 1770 to 1993.
If you want to **search several databases simultaneously**, use **Find Articles** in the Library Gateway. **Find Articles** allows to choose one or more databases from a selected list of our subscription databases; most are periodical indexes. You can then search the selected databases simultaneously by keyword.

If you're not sure which kind of periodical you want or you're not sure which periodical index to use, or if you want help searching, ask a reference librarian**.**

Remember you can always review the titles of periodical indexes available online by searching or clicking on subject categories in the **Find Databases** section of **the Library Gateway.**

When You Have the Citation to a Specific Article, Use the Library Catalog

When you do have the citation or reference to a periodical article--if you know at least the title of the periodical and the issue date of the article you want--you can find its location by using the Library Catalog. Click on the Basic Search button, highlight "Journal Title" in the "Search By:" box, click in the search box, type in the title of the periodical in the search box, and press **<enter>**. Don't use the abbreviated titles that are often used in periodical indexes; remember to omit "a," "an" or "the" when you type in the periodical title.

Search Examples in the Library Catalog:

***** When searching for the title, *The Chronicle*

Type the following in the search box: **chronicle**
(Omit initial articles)

***** When searching for the title, *Journal of Modern History*

Type the following in the search box: **journal of modern history**
(Do not type j mod hist)

***** When searching for the title, *Annales Musicologiques: Moyen-Age et Renaissance*

You may type the following: **annales musicologiques moyen age**
(Omit punctuation)

Depending on the number of records your search retrieves from the Library Catalog, you will see either a list of entries or a single record for an individual periodical title. If there is a list of titles, scroll through it and click on the line that lists the journal title you want to see the call number and location information.

If the journal is available in electronic form, there will be a link following the field labelled "Electronic access:" in the catalog record. Click on this link. In most cases, this will take you to the opening screen for the journal, and you can choose the issue you want from there. If the link takes you to *ProQuest*, click on the the tab labeled "Periodicals" at the top of the *ProQuest* search page. Then search for the journal title you want.

If the journal is available in print form, record the call number and any additional location information in the catalog record. Now you're ready to find it on the shelf. Consult the local stack directory for the call number locations in individual libraries.

Locating Periodicals in Library

*** Current Periodicals:** Periodicals noted as "Current issues in Periodicals Room" in the Library Catalog are shelved by title in the Current Periodicals Room in Library. **Only a small selection of current periodicals is in this room**; all other current periodicals are shelved by call number in the stacks.

*** Back Periodicals** are shelved by call number in the stacks. Some back periodicals are shelved in specific subject rooms; watch for location notes in the Library Catalog record for the title you want.

Pay attention to the + and ++ indicators by the call number. Titles with the + and ++ (Oversize) designations and titles with no plus marks are each shelved in separate sections in Library.

Summary: Use periodical indexes and abstracts to find citations to articles. The indexes and abstracts may be in print or computer-based formats or both. Choose the indexes and format best suited to your particular topic; ask at the reference desk if you need help figuring out which index and format will be best. You can find periodical articles by the article author, title, or keyword by using the periodical indexes in the Library Gateway. If the full text is not linked in the index you are using, write down the citation from the index and search for the title of the periodical in the Library Catalog. The catalog lists the print, microform, and electronic versions of periodicals at .

Step 5: Find Internet Resources

Things to Know Before You Begin Searching....

What Are You Really Searching?

Finding the Web documents (a.k.a. Web "pages" or "sites") you want can be easy or seem impossibly difficult. This is in part due to the sheer size of the WWW, currently estimated to contain 50 billion or many more documents (no one really knows its size). Web searching difficulties also arise because the WWW is not indexed in any standard vocabulary. Unlike a library's catalogs, in which can use standardized Library of Congress subject headings to find books in most large, general libraries in the U.S., in Web searching you are always guessing what words will be in the pages you want to find or guessing what subject terms were chosen by someone to organize a web page or site covering some topic.

When you do what is called "searching the Web," you are NOT searching it directly. It is not possible to search the WWW directly. The Web is the totality of the many web pages which reside on computers (called "servers") all over the world. Your computer cannot find or go to them all directly. What you are able to do through your computer is access one or more of many intermediate search tools available now. You search a search tool's database or collection of sites -- a relatively small subset of the entire World Wide Web. The search tool provides you with hypertext links with URLs to other pages. You click on these links, and retrieve documents, images, sound, and more from individual servers around the world.

There is no way for anyone to search the entire Web, and any search tool that claims that it offers it all to you is distorting the truth. In addition to what you can find in search tools of various types, there is vast content that never appears in any general, free web search tool. This is called the Invisible Web or Deep Web.

Categories of search tools available now

At present, we find it useful to describe the kinds of intermediate search tools available to you in three categories. You use different strategies to find and exploit the potential of the tools in each class:

Types of Search Tools	Characteristics	Examples
Search Engines (& Meta-Search Engines)	Full-text of selected Web pages Search by keyword, trying to match exactly the words in the pages No browsing, no subject categories Databases compiled by "spiders" (computer-robot programs) with minimal human oversight Search-Engine size: from small and specialized to huge (about 20 billion	Search Engines recommended and described: Google, Yahoo Search, Ask.com Meta-Search Engines: Dogpile, Copernic, and others

	websites or pages) Meta-Search Engines quickly and superficially search several individual search engines at once and return results compiled into a sometimes convenient format. Caveat: They only catch about 1% of search results in any of the search engines they visit.	
Subject Directories	Human-selected sites picked by editors (sometimes experts in a subject) Often carefully evaluated and kept up to date, but not always -- frequently not if large and general Usually organized into hierarchical subject categories Often annotated with descriptions (not in Yahoo!) Can browse subject categories or search using broad, general terms **NO full-text of documents**. Searches need to be less specific than in search engines, because you are not matching on the words in the pages you eventually want. In Directories you are searching only the subject categories and descriptions you see in its pages.	Directories recommended and described: Librarians' Index, Infomine, Google Directory, About.com, AcademicInfo There are thousand more of Subject Directories on practically every topic you can think of.
Specialized Databases (The Invisible Web)	The Web provides access through a search box into the contents of a database in a computer somewhere	Locate specialized databases by looking for them in good Subject Directories like the Librarian's Index,

	Can be on any topic, can be trivial, commercial, task-specific, governmental, or a rich treasure devoted to your topic Also includes Also includes many pages generated as search results from libraries' online catalogs, and the many copyright-protected articles in the databases of journal and magazine publishers.	Yahoo!, or AcademicInfo; in special guides to searchable databases; and sometimes by keyword searching in general search engines

Where should I begin ?

In this chapter, we recommend beginning with different types of search tools, depending on what you know about your topic and what you want to know. Do you want broad information? Are you looking for something very specific or perhaps a unique term or phrase? Are you looking for a narrow aspect of a topic with a huge Web presence? When you search, are you overwhelmed by too many or off-target results? Are there a lot of synomyms of equivalent terms for what you seek? Each of these questions can give you a clue where to begin.

We encourage you to vary your search strategy and learn as you search, choosing the optimal approach for each stage of the process. Finding information on the Web is a process.

Recommended Search Strategy: Analyze Your Topic & Search With Peripheral Vision

The Five-Step Search Strategy We Recommend

Step #1. Analyze your topic to decide where to begin

Does your topic...

have distinctive words or phrases?

methernitha, unique meaning

"affirmative action", specific, accepted meaning in word cluster

have NO distinctive words or phrases you can think of? You have only common or general terms that get the "wrong" pages.

"order out of chaos", used in too many contexts to be useful

sundiata, retrieves a myth, a rock group, a person, etc.

seek an overview of a broad topic?

victorian literature, alternative energy sources

specify a narrow aspect of a broad or common topic?

> **automobile recyclability**, want current research, future designs, not how to recycle or oil recycling or other community efforts

have synonymous, equivalent terms, or variant spellings or endings that need to be included?

> **echinoderm OR echinoidea OR "sea urchin"**, any may be in useful pages
> **"cold fusion energy" OR "hydrogen energy"**, some use one term, some the other; you want both, although not precisely equivalent
> **millennium OR millennial OR millenium OR millenial OR "year 2000"**, etc. Pages you want may contain any or all.

Make you feel confused? Don't really know much about the topic yet? Need guidance?

Step #2. Pick the right starting place using this table:

YOUR TOPIC'S FEATURES:	**Search Engines**	**Subject Directories**	**Specialized Databases "Invisible Web"**	**Find an Expert**	**LUCK**
Distinctive or word or phrase?	Enclose phrases in **" "**. **Test run** your word or phrase in Google.	Search the broader concept, what your term is "about."	Want data? Facts? Statistics? All of something?	Look for a specialized subject directory on your topic.	Always on your side. Keep your mind open. Learn as you search.
NO distinctive words or phrases?	Use more than one term or phrase in " " to get fewer results.	Try to find distinctive terms in Subject Directories	One of many like things? Schedules? Maps?	E-mail the author of a good page you find.	
Seek an overview?	NOT RECOMMENDED	Look for a specialized Subject Directory focused on your topic	Look for a specialized database on the Invisible Web.	Ask a discussion group or blog. Never hurts to seek help.	
Narrow aspect of broad or common topic?	Boolean searching as in Yahoo! Search.	Look for a Directory focused on the broad subject.	Hard to predict what you might find.		
Synonyms, equivalent terms, variants	Choose search engines with Boolean OR, or Truncation, or Field limiting.	NOT RECOMMENDED			
Confused? Need more information?	NOT RECOMMENDED	Look for a Gateway Page (Subject Guide). Try an **encyclopedia.** iAsk at a **library** reference desk.			

Step #3. Learn as you go & VARY your approach with what you learn.
Don't assume you know what you want to find. Look at search results and see what you might use in addition to what you've thought of.

Step #4. Don't bog down in any strategy that doesn't work.
Switch from search engines to directories and back. Find specialized directories on your topic. Think about possible databases and look for them.

Step #5. Return to previous strategies better informed.

Search Strategies We Do NOT Recommend

Because of their inefficiency and often haphazard and frustrating results, we do not recommend either of the following two approaches to finding Web documents:

> **Browsing searchable directories.** If you can find a search box, search a directory. BROWSING is sometimes fun but rarely as efficient. The term "directories" refers here to any collection of web resources organized into subject categories or some other breakdown appropriate to the content (Subject Directories or directories of specialized databases). Browsing locates documents by your trying to match your topic in first the top, broadest layer of a subject hierarchy, then by choosing narrower sub-subject-categories in the hierarchy that you hope will lead to your target. Browsing encounters the difficulty of guessing under which subject category your topic is classified. The taxonomy in every directory differs, making browsing inconsistent from one search tool to another. The category "health" may contain documents on medicine, homeopathy, psychiatry, and fitness in one directory. In another "medicine" may include health, mental health, and alternative medicine, but not the term psychiatry and may classify fitness only under "lifestyle." *Searching (typing keywords in a search box) retrieves occurrences of your words no matter where they may be classified by subject. Use broad terms in searching any directory.*
>
> **Following links to sites recommended by heavy use or commercial interest.** Often in search engine results, you will see links to sites that are selected based on how often they are visited by others, or based on fees paid to the browser. Or you may see recommended "cool" sites. Use these with caution! Others may visit sites for reasons having no relation to your information interests, and the best sites for you may still be largely undiscovered by the vast public searching the Web. Taste varies and should vary. Make your own evaluations.

Table Matching Your Search Needs with the Features Search Engines Offer

The purpose of thinking about your topic before you start searching is to **determine what terms to search for** and **what features you need** to search successfully. The table below lists on the left features of many search inquiries. Use it to determine which features your searches need. On the left, the table describes search tool features designed to support each of the search needs listed on the left.

The links take you to the table of search engines -- so you can pick a search engine with the features you need.

Features of your search inquiry	**Matching Search Tools Features worth learning**
Are you looking for a **proper name** or a **distinct phrase** ? The name of an organization or society or movement A proper name or an individual A distinctive string of words generally associated with your topic Can you think of an organization, proper name, or phrase to search for? It might help zoom in on the pages you want.	**<u>PHRASE SEARCHING</u>** is a feature you want in every search tools you choose. Requires your terms all to appear in exactly the order you enter them. Enclose the phrase in double quotations " " Examples: **"affirmative action"** **"world health organization"** **"a person's name"** In , capitalizing initial letters will cause the terms to be searched as a phrase: **World Health Organization**
Are some of your terms **common words** with **many meanings and contexts** ? *Children* in conjunction with *television* and also *violence* *Censorship* as an aspect of *ethics* in *journalism*	**<u>BOOLEAN</u> AND** will help: **children AND television AND violence** **journalism AND ethics AND censorship** **Google** and **AllTheWeb**and most other search engines put AND in between words automatically (by default): **children television violence** **journalism ethics censorship**
Do you anticipate lots of search **results with terms you do not want** ? Your search for *biomedical engineering and cancer* brings you lots of academic programs, and you want research reports. So you try to exclude documents containing Department *of* or *School of*	**<u>BOOLEAN</u> AND NOT** will help: **"biomedical engineering" AND cancer AND NOT "Department of" AND NOT "School of"** or its **-EXCLUDES** near equivalent: **"biomedical engineering" cancer -"Department of" -"School of"**
Are there **synonyms**, **spelling variations**, or **foreign spellings** for some of your terms? *women, females* with *networking* *Sarajevo, Sarayevo* with *peace* *literature, litterature* with	**<u>BOOLEAN</u> OR** will help: **(women OR females) AND networking** **(Sarajevo OR Sarayevo) AND peace** **(literature OR litterature) AND (French or francaise)** In **Google**, capitalize OR (no need to type "and"):

French, francaise	**peace sarajevo OR sarayevo** **literature OR litterature** **french OR francaise** In AllTheWeb, use parentheses and omit the OR: **peace (sarajevo sarayevo)** **(literature litterature)** **(french francaise)**
Are you looking for **home pages** and/or other documents **primarily about** your term(s)? The home page of the *American Dietetic Association* Pages primarily about *Affirmative Action*	**LIMIT TO TITLE FIELD** IN DOCUMENTS **intitle:"American Dietetic Association"** **intitle:"affirmative action"** In **Google**, use **intitle:"affirmative action"**
Are you looking for terms with **many possible endings** ? *Feminism, feminist, feminine* *Children, child*	Some systems search word ending variants automatically (stemming). See the specific instructions for each of the recommended search tools. To be sure use OR searches: **children OR child**

Three Families or Types of Search Tools - Links to all search tools tables

1. Search engines defined (The BEST Search Engines to use - TABLE of features):

built by computer robot programs ("spiders") -- not by human selection
NOT organized by subject categories -- all pages are ranked by a computer algorithm
contain full-text (every word) of the web pages they link to -- you find pages by matching words in the pages you want
huge and often retrieve a lot of information -- for complex searches use ones that allow you to search within results (sub searching)
UNevaluated -- contain the good, the bad, and the ugly -- YOU must evaluate everything you find (more on evaluating.)

2. Subject directories defined (The BEST Subject Directories to use- TABLE of features):

built by human selection -- not by computers or robot programs
organized into subject categories, classification of pages by subjects -- subjects not standardized and vary according to the scope of each directory
NEVER contain full-text of the web pages they link to -- you can only search what you can see (titles, descriptions, subject categories, etc.) -- use broad or general terms

small and specialized to large, but smaller than most search engines -- huge range in size
often carefully evaluated and annotated (but not always!!)

3. Searchable database contents or the "Invisible Web" (How to find these databases in the Invisible Web):

Defined: Pages that cannot be found in search engines and rarely are in subject directories -- the "visible web" is what you can see using these tools -- Invisible Web is estimated to offer two to three times as many pages as the visible web.
WHY? There are many specialized searchable databases that the World Wide Web allows you to access through a search box in a web page (for example, the UCB Library Catalog Pathfinder, or any other library catalog; or some statistics databases searchable on the web). The terms you use in your search are sent into that specialized database, and are returned to you in another web page that is dynamically generated for your answer. It is not retained anywhere after your search.

- Search Engines cannot access such dynamically generated pages because the computer robots or spiders that build them *cannot type* the searches needed to generate the pages. Spiders find pages by visiting all the links in the pages they "know about." Unless there are links somewhere that the spiders can use to re-generate specialized database searches, the contents of the database is beyond them. Pages requiring passwords to access them are also closed to search engines, because spiders cannot type. (There are a few other types of pages that most search engines refuse to include; please see the discussion on the Invisible Web page.)
- Directories rarely have the contents of these pages, but, since directories are built by humans capable of typing, there is no reason directories cannot contain links which, if clicked, would cause a search in the database to be dynamically generated each time it is clicked.

What about MetaSearch Engines and Gateway Pages?

MetaSearch Engines (The BEST MetaSearch Engines or some of them TABLE of features) :

Defined: Utilities that search more than search engine and/or subject directory at once and then compile the results in a sometimes convenient display, sometimes consolidating all the results into a uniform format and listing. Some offer added value features like the ability to refine searches, customize which search engines or directories are queried, the time spent in each, etc. Some you must download and install on your computer, whereas most run as server-side applications.
Limitation: Like one-stop shopping, the idea has great appeal; but the implementation results in limiting your ability to control where you "shop." In my view, ***NONE of the MetaSearch engines is as good as YOU can be if you learn to search effectively!!*** They are a great idea but a disappointment in implementation.

Recommendation: This tutorial continues to monitor developing MetaSearch Engine technologies, but it ***no longer recommends them*** in its search strategy. They are quick and dirty, not thorough, most often omit Google (the best search engine around), always omit Northern Light (useful in academic research), and do not let you take advantage of advanced features in any search engines. Moreover, they are unpredictable both in how they will transmit a complex search, and you usually don't know what they will search beforehand.

Directories of Subject Guides and Gateway Pages defined (some good directories of this type):

a type of subject directory specializing in web pages compiled by some "expert" who spent a lot of time searching the web and assembling guides to a field, subject, discipline, etc.
of great value in doing academic research
locate them the way you locate other subject directories;

Search Engines - Comparison table of the best search engines UPDATED –

Recommended Search Engines: Tables of Features

Google appears still to have the largest database of Web pages, including many other types of Web documents (e.g., PDFs, Word, Excel, PowerPoint documents). Despite the presence of many advertisements and considerable clutter from blog sites and newsgroups, Google's popularity ranking often makes pages worth looking at rise near the top of search results.New "Googling to the Max" courses reflects our recognition that Google currently is the winning web search engine and so people need to learn to use it really well.

Google alone is often not sufficient, however. Less than half the searchable Web is fully searchable in Google. Overlap studies show that about half of the pages in any search engine database exist only in that database. Getting a second opinion is therefore often worth your time. For a second opinion, we recommend Teoma or Yahoo! Search. We no longer recommend using any meta-search engines for web searching, although some of the features in Dogpile and Vivisimo's Clusty may prove useful for some searches.

Features in common among the search engines we recommend. Search engines have become a little bit standardized, allowing us to use some of the same search techniques in all of them:

Things You CAN Do in Google, Yahoo!, and Ask.com	Things NOT Supported in Google, Yahoo!, or Ask.com
Phrase Searching by enclosing terms in double quotes **OR** searching with capitalized OR **- excludes, + requires** exact form of word **Limit** results by language in	**Truncation** - use OR searches for variants (airline OR airlines) **Case sensitivity** capitalization does not matter

Advanced Search

Some Ways the Recommended Search Engines Differ:

Search Engine	**Google** www.google.com	**Yahoo! Search** search.yahoo.com	**Ask.com** www.ask.com
Links to help	Google help pages	Yahoo! help pages	Ask help pages
Size, type Size varies frequently and widely. See tests and more charts.	HUGE. Size not disclosed in any way that allows comparison. Probably the biggest. Biggest in tests.	HUGE. Claims over 20 billion total "web objects."	LARGE. Claims to have 2 billion fully indexed, searchable pages. Strives to become #1 in size.
Noteworthy features and limitations	Popularity ranking using PageRank™. Indexes the first 101KB of a Web page, and 120KB of PDF's. ~ before a word finds synonyms sometimes (~help > FAQ, tutorial, etc.)	Shortcuts give quick access to dictionary, synonyms, patents, traffic, stocks, encyclopedia, and more.	Subject-Specific Popularity™ ranking. Suggests broader and narrower terms.
Phrase searching (term definition)	Yes. Use " ". Searches common "stop words" if in phrases in quotes.	Yes. Use " "	Yes. Use " ". Searches common "stop words" if in phrases in quotes.
Boolean logic (term definition)	Partial. AND assumed between words. Capitalize OR. - excludes. No ( ) or nesting. In Advanced Search, partial Boolean available in boxes.	Accepts AND, OR, NOT or AND NOT, and ( ). *Must be capitalized.* You **must** enclose terms joined by OR in parentheses (classic Boolean).	Partial. AND assumed between words. Capitalize OR. - excludes. No ( ) or nesting.
+Requires/ -Excludes (term definition)	- excludes + will allow you to retrieve "stop words" (e.g., +in)	- excludes + will allow you to search common words: "+in truth"	- excludes + will allow you to retrieve "stop words" (e.g., +in)
Sub-Searching (term definition)	Sort of . At bottom of results page, click "Search within results" and enter more terms. Adds terms.	Add terms.	Sort of . Add terms.
Results	Based on page	Automatic Fuzzy	Based on

Ranking (term definition)	popularity measured in links to it from other pages: high rank if a lot of other pages link to it. Fuzzy AND also invoked. Matching and ranking based on "cached" version of pages that may not be the most recent version.	AND.	Subject-Specific Popularity™, links to a page by related pages. More info.
Field limiting (term definition)	link: site: intitle: inurl: Advanced Search boxes for most of these. Offers Uncle Sam for US federal pages and other special searches.	link: site: intitle: inurl: url: hostname: (Explanation of these distinctions.)	intitle: inurl: site:
Truncation Stemming (term definition)	No truncation. Stems some words. Search variant endings and synonyms separately, separating with OR (capitalized): *airline OR airlines*	Neither. Search with OR as in Google.	Neither. Search with OR as in Google.
Case sensitivity (term definition)	No.	No.	No.
Language	Yes. Major Romanized and non-Romanized languages in Advanced Search.	Yes. Major Romanized and non-Romanized languages.	Yes. Major Romanized languages. Use Advanced Search to limit.
Limit by age of documents	In Advanced Search.	In Advanced Search.	In Advanced Search.
Translation	Yes, in Translate this page link following some pages. To and sometimes from English and major European languages and Chinese, Japanese, Korean.	Yes.	No.

What Makes a Search Engine Good?

All search engines consist of three parts: (1) a database of web documents, (2) a search engine operating on that database, and (3) a series of programs that determine how search results are displayed. Because the search engine business is competitive, most search engines also offer additional features that are convenient or fun. The table below shows what can vary within each of the three basic parts in search engines.

Parts of Search Engines	**Variables, and their implications for your searches**
1. Database of web documents	Size of database: How many documents does the search engine claim it has? How much of the total web are you able to search? Freshness ("up-to-dateness"): Search engine databases consist of copies of web pages and other documents that were made when their crawlers or spiders last visited each site. How often is the database refreshed to find new pages? How often do their crawlers update the copies of the web pages you are searching? Completeness of text: Is the database really "full" text, or only parts of the pages? Is every word indexed? Types of documents offered: All search engines offer web pages. Do they also have extensive PDF, Word, Excel, PowerPoint, and other formats like WordPerfect? Are they full-text searchable? Speed and consistency: How fast is it? How consistent is it? Do you get different results at different times?
2. The search engine's capabilitiesAll search engines let you enter some keywords and search on them. What happens inside? Can you limit in ways that will increase your chances of finding what you are looking for?	Basic Search options and limitations: Automatic default of AND assumed between words? Accepts " " to create phrases? Is there an easy way to allow for synonyms and equivalent terms (OR searching)? Can you OR phrases or just single words? Advanced Search options and limitations: Can you require your search terms in specific fields, such as the document title? Can you require some words in certain fields and others anywhere? Can you restrict to documents only from a certain domain (org, edu, gov, etc.)? Limit to more than one or only one? Can you limit by type of document (pdf or excel, etc.)? More than one? Can you limit by language? How reliably and easily can you limit to date last updated? General limitations and features: What do you have to do make it search on common or stop words? Maximum limit on search terms or on search complexity? Ability to search within previous results? Can you count on consistent results from search to search and from day to day? Can you customize the search or display? Is there a "family" filter? Does it work well? Is it easy to turn on or off?
3. Results displayAll search engines return a list of results it "thinks" are what you are looking for. How well does it "think like you expect it think"?	Ranking: Are they ranked by popularity or relevancy or both? Do pages with your words juxtaposed (like a phrase) rank highest? Do you get pages with only some of your words, perhaps in addition to pages with them all? Display: Are your keywords highlighted in context, showing excerpts from the web pages which caused the match? Some other excerpt from the page? Collapse pages from the same site: If it shows only one or a few pages from a site, does it show the one(s) with your terms? How easy is it to see all from the site? Can this be changed and saved as your preferred search method?
4. Other features	Search engine designers try to come up with all kinds of features and services that they hope will allure you to their services.

How do Search Engines Work?

Search Engines for the general web (like all those listed above) do not really search the World Wide Web directly. Each one searches a database of the full text of web pages selected from the billions of web pages out there residing on servers. When you search the web using a search engine, you are always searching a somewhat stale copy of the real web page. When you click on links provided in a search engine's search results, you retrieve from the server the current version of the page.

Search engine databases are selected and built by computer robot programs called spiders. Although it is said they "crawl" the web in their hunt for pages to include, in truth they stay in one place. They find the pages for potential inclusion by following the links in the pages they already have in their database (i.e., already "know about"). They cannot think or type a URL or use judgment to "decide" to go look something up and see what's on the web about it. (Computers are getting more sophisticated all the time, but they are still brainless.)

If a web page is never linked to in any other page, search engine spiders cannot find it. The only way a brand new page - one that no other page has ever linked to - can get into a search engine is for its URL to be sent by some human to the search engine companies as a request that the new page be included. All search engine companies offer ways to do this.

After spiders find pages, they pass them on to another computer program for "indexing." This program identifies the text, links, and other content in the page and stores it in the search engine database's files so that the database can be searched by keyword and whatever more advanced approaches are offered, and the page will be found if your search matches its content.

> Some types of pages and links are excluded from most search engines by policy. Others are excluded because search engine spiders cannot access them. Pages that are excluded are referred to as the "Invisible Web" -- what you don't see in search engine results. The Invisible Web is estimated to be two to three or more times bigger than the visible web.

Still Googling in 2006/2007

Google is still recognized as the best general web search engine.

Why Google?

- Google is the BIGGEST search engine database in the world
- PageRank™ often finds useful pages. It is one of the defaults that cannot be turned off in Google and is not for sale. Itworks on a unique combination of factors, some of which are:
  - Popularity - based on the number of links to a page and the importance of the pages that link
  - Importance - traffic, quality of links
  - Word proximity and occurrence in results
- Google has many useful ways to limit searches
- Google offers special "fuzzy" searches that are useful to search synonyms, find definitions, find similar/related pages, and more
- The shortcuts & special Google databases can enhance certain types of research
- Google Books and Google Scholar have great potential for university-level research using the web.

Some Google shortcomings:

Lots of "stop words" which you have to precede with a + to search or search in quotes. Try searching **"to be or not to be"** or **to be or not to be** without the quotes.
Full Boolean logic is not supported - OR, - for "not," and AND (implied as default). No parenthesese and no nesting. Yahoo! search DOES support full Boolean logic, with parentheses for nesting; be sure to capitalize OR, AND, and NOT.
Despite its default AND, Google sometimes returns pages that do not contain all of your terms. They occur in other pages that link to the page or in other pages on the same site. Google shows you these results because they are "important" (rank high) in Google. The only way to know whether your terms are in a page or why the page was provided is to look at Google's cached copy.

How to Find Subject-Focused Directories for a Specific Topic, Discipline, or Field

There are thousands of specialized directories on practically every subject. If you want an overview, or if you feel you've searched long enough, try to find one. Often they are done by experts -- self-proclaimed or heavily credentialed. Here are some ways to find them:

Use any of the Subject Directories above to find more specific directories. Here are some tips:

In the **Librarians' Index**, **Infomine,** or **Academic Info**, look for your subject as you would for any other purpose, and keep your eyes open for sites that look like directories. Read through the descriptions. Sometimes these resources are identified as "Directories, "Virtual Libraries," or "Gateway Pages." These three directories are small enough that you can read through the results of a search.
In **Yahoo!** and **Google** directories, try adding the terms ***web directories*** to your subject keyword term:

EXAMPLES:
civil war web directories
weddings web directories

In **About.com**, search by topic and look for pages that are described as "101" or "guides" or a "directory." About.com is written by "Guides" who, themselves, often are experts in the sections they manage. Sometimes they write excellent overviews of a topic.

Invisible Web - What it is, how to find it, and its inherent ambiguity (searchable databases on the Web)

What is the "INvisible Web", a.k.a. the "Deep Web"?

The **"visible web"** is what you see in the results pages from general web search engines. It's also what you see in almost all subject directories. The **"<u>in</u>visible web"** is what you cannot retrieve ("see") in the search results and other links contained in these types of tools.

The first version of this web page written in 2000, when this topic was new and baffling to many web searchers. Since then, search engines, crawlers and indexing programs have overcome many of the technical barriers that made it impossible for them to find and provide invisible web pages. These types of pages used to be invisible but can now be found in most search engine results:

Pages in non-HTML formats (pdf, Word, Excel, Corell suite, etc.) are "translated" into HTML now in most search engines and can "seen" in search results.
Script-based pages, whose links contain a ? or other script coding, no longer cause most search engines to exclude them.
Pages generated dynamically by other types of database software (e.g., Active Server Pages, Cold Fusion) can be indexed if there is a stable URL somewhere that search engine spiders can find. Once these were largely shunned by search engines. There are now many types of dynamically generated pages like these that are found in most general web search engines. There must a stable link to the page somewhere; see the next section.

Why?

There are still some hurdles search engine spiders cannot leap, and these still create a HUGE set of web pages not found in general search engines

Search engines still cannot type or think. If access to a web pages requires typing, web crawlers encounter a barrier they cannot go beyond. They cannot search our online catalogs and they cannot enter a password or login.

- o **The Contents of Searchable Databases.** Most of the invisible or deep web is made up of the contents of thousands of specialized searchable databases made available via the web. When you type a search in one of these databases, the search results are delivered to you in web pages that are generated just in answer to your search. Rarely are such pages stored anywhere: it is easier and cheaper to dynamically generate the answer page for each query than to store all the possible pages containing all the possible answers to all the possible queries people could make to the database.

- **Google Scholar** is a collection of citations with links to publishers or other sources where one can try to access the publication in full text. In many academic libraries (and some others), Google Scholar is providing convenient links to the online holdings of those libraries, purchased for exclusive use of their constituents. If you search Google Scholar, you find a lot of journal article references. **What you are seeing when you search Google Scholar is only tiny fraction of all the scholarly publications that exist online. Much more lurks in a new type of Invisible or Deep Web.**
- **WHY?** Google Scholar is only able to provide citations to journal contents for which its crawlers can find stable llinks. It cannot construct searches or enter passwords to go into passworded, copyright-protected articles in many publishers' databases. In some experiments conducted at UC Berkeley, we estimate that Google Scholar accesses about 10% of all we subscribe to for our students, faculty, staff, and users present on campus. Think about the millions of articles in Lexis/Nexis, the many thousands of articles indexed in

privately licensed databased libraries buy the rights for their users to read (e.g., Sociological Abstracts, ERIC, PscyhInfo, JSTOR, INSPEC). At Berkeley we subscribe ot about 200 of these.

Excluded Pages. There are some types of pages that search engine companies exclude by policy. There is no technical reason they could not include them if they wanted. It's a matter of selecting what and what not to include in databases that are already huge, expensive to operate, and whose search function is a low revenue producer.

- **Dynamically generated pages of little value beyond single use.** Think of the billions of possible web pages that can be generated by all the people who have looked for books in our online catalogs. Each of them is creating a results page in reponse to their specific need. Search engines do not want all of these pages in their web databases. They would be clutter of little interest to anyone.
- **Many databases in this category.** There are many thousands of public-record, official, and special-purpose databases containing government, financial, logistical, and other types of information that is needed to answer very specific inquiries of interest to very few people. Even if stable links existed to such pages, search engines would not want them. More clutter.

How to Find the Invisible Web

Simply think "databases" and keep your eyes open. You can find searchable databases containing invisible web pages in the course of routine searching in most general web directories. Of particular value in academic research are

Librarians Index
AcademicInfo
Infomine

Use Google and other search engines to locate searchable databases by searching a subject term and the word "database". If the database uses the word database in its own pages, you are likely to find it in Google. The word "database" is also useful in searching a topic in the Google Directory or the Yahoo! directory, because they sometimes use the term to describe searchable databases in their listings.

EXAMPLES for Google & Yahoo:
plane crash database
languages database
toxic chemicals database

Remember that the Invisible Web exists. Remember that, in addition to what you find in search engine results (including Google Scholar) and most web directories, there are these gold mines you have to search directly. This includes all of the licensed article, magazine, reference, news archives, and other research resources that libraries and some industries buy for those authorized to use them. The contents of these are not freely available: libraries and corporations buy the rights for their authorized users to view the contents. If they appear free, it's because you are somehow authorized to search and read the contents (library card holder, member of the company, etc.).

As part of your wise web search strategy, spend a little time looking for databases in your field or topic of study or research. Remember, however, that all proprietary information -- most of the journals, magazines, news, and books -- are not freely available. Publishers and authors control them under copyright and other distribution rules. You will be prompted to pay or enter a password to see full text. A library you have the rights to use may have access to what you want, however.

The Ambiguity Inherent in the Invisible Web:

It is very difficult to predict what sites or kinds of sites or portions of sites will or won't be part of the Invisible Web. There are several factors involved:

- Which sites replicate some of their content in static pages (hybrid of visible and invisible in some combination)?
- Which replicate it all (visible in search engines if you construct a search matching terms in the page)?
- Which databases replicate none of their dynamically generated pages in links and must be searched directly (totally invisible)?
- Search engines can change their policies on what the exclude and include.

Evaluating Web Pages: Why and How

Evaluating web pages skillfully requires you to do two things at once:

1. Train your eye and your fingers to employ a series of **techniques** that help you quickly find what you need to know about web pages;
2. Train your mind to think critically, even suspiciously, by **asking a series of questions** that will help you decide how much a web page is to be trusted.

This page is organized to combine the two techniques into a process that begins with looking at your search results from a search engine or other source, follows through by investigating the content of page, and extends beyond the page to what others may say about the page or its author(s).

1. What can the URL tell you?

Techniques for Web Evaluation :

1. Before you leave the list of search results -- before you click and get interested in anything written on the page -- glean all you can from the URLs of each page.
2. Then choose pages most likely to be reliable and authentic.

Questions to ask:	What are the implications?
Is it somebody's personal page? Read the URL***** carefully: o Look for a personal name (e.g., *jbarker* or *barker*) following a tilde (**~**), a percent sign (%), or or the words	Personal pages are not necessarily "bad," but you need to investigate the author very carefully. For personal pages, there is no publisher or domain owner vouching

"users," "members," or "people." o Is the server a commercial ISP* or other provider mostly of web page hosting (like aol.com or geocities.com	for the information in the page.
What type of domain does it come from ? (educational, nonprofit, commercial, government, etc.) Is the domain appropriate for the content? o Government sites: look for .gov, .mil, .us, or other country code o Educational sites: look for .edu o Nonprofit organizations: look for .org If from a foreign country, look at the country code and read the page to be sure who published it.	Look for a appropriateness, fit. What kind of information source do you think is most reliable for your topic?
Is it published by an entity that makes sense? Who "published" the page? In general, the publisher is the agency or person operating the "server" computer from which the document is issued. o The server is usually named in first portion of the URL (between *http://* and the first /) Have you heard of this entity before? Does it correspond the name of the site? Should it?	You can rely more on information that is published by the source: Look for New York Times news from **www.nytimes.com** Look for health information from any of the agencies of the National Institute of Health on sites with **nih** somewhere in the domain name.

2. Scan the perimeter of the page, looking for answers to these questions:

Techniques for Web Evaluation :

1. Look for links that say "**About us**," "**Philosophy**," "**Background**," "**Biography**," "**Who am I**," etc.
2. If you cannot find any links like these, you can often find this kind of information if you **Truncate back the URL**.
   INSTRUCTIONS for Truncating back a URL: In the top Location Box, delete the end characters of the URL stopping just before each / (leave the slash). Press enter to see if you can see more about the author or the origins/nature of the site providing the page.
   Continue this process, one slash (/) at a time, until you reach the first single / which is preceded by the domain name portion. This is the page's server or "publisher."
3. Look for the date "last updated" - usually at the bottom of a web page.
   Check the date on all the pages on the site.
   Do not rely on a date given in IE's File|Properties or Netscape/Mozilla's View|Page Info displays. These dates can be automatically kept current and are useless in critical evaluation.

Questions to ask:	**What are the implications?**
Who wrote the page? Look for the name of the author, or the name of the organization, institution, agency, or whatever who is responsible for the page o An e-mail contact is not enough If there is no personal author, look for an agency or organization that claims responsibility for the page. o If you cannot find this, locate the publisher by truncating back the URL (see technique above). Does this publisher claim responsibility for the content? Does it explain why the page exists in any way?	Web pages are all created with a purpose in mind by some person or agency or entity. They do not simply "grow" on the web like mildew grows in moist corners. You are looking for someone who claims accountability and responsibility for the content. An e-mail address with no additional information about the author is not sufficient for assessing the author's credentials. If this is all you have, try e-mailing the author and asking politely for more information about him/her.
Is the page dated? Is it current enough? Is it "stale" or "dusty" information on a time-sensitive or evolving topic? **CAUTION:** Undated factual or statistical information is no better than anonymous information. Don't use it.	How recent the date needs to be depends on your needs. For some topics you want current information. For others, you want information put on the web near the time it became known. In some cases, the importance of the date is to tell you whether the page author is still maintaining an interest in the page, or has abandoned it.
What are the author's credentials on this subject? Does the purported background or education look like someone who is qualified to write on this topic? Might the page be by a hobbyist, self-proclaimed expert, or enthusiast? o Is the page merely an opinion? Is there any reason you should believe its content more than any other page? o Is the page a rant, an extreme view, possibly distorted or exaggerated? If you cannot find strong, relevant credentials, look very closely at documentation of sources (next section).	Anyone can put anything on the web for pennies in just a few minutes. Your task is to distinguish between the reliable and questionable. Many web pages are opinion pieces offered in a vast public forum. You should hold the author to the same degree of credentials, authority, and documentation that you would expect from something published in a reputable print resource (book, journal article, good newspaper).

3. Look for indicators of quality information:

Techniques for Web Evaluation :

1. Look for a link called "links," "additional sites," "related links," etc.
2. In the text, if you see little footnote numbers or links that might refer to documentation, take the time to explore them.

What kinds of publications or sites are they? Reputable? Scholarly?

Are they real? On the web (where no publisher is editing most pages), it is possible to create totally fake references.

3. Look at the publisher of the page (first part of the URL).

Expect a journal article, newspaper article, and some other publications that are recent to come from the original publisher IF the publication is available on the web.

Look at the bottom of such articles for copyright information or permissions to reproduce.

Questions to ask:	**What are the implications?**
Are sources documented with footnotes or links? Where did the author get the information? o As in published scholarly/academic journals and books, you should expect documentation. If there are links to other pages as sources, are they to reliable sources? Do the links work?	In scholarly/research work, the credibility of most writings is proven through footnote documentation or other means of revealing the sources of information. Saying what you believe without documentation is not much better than just expressing an opinion or a point of view. What credibility does your research need? An exception can be journalism from highly reputable newspapers. But these are not scholarly. Check with your instructor before using this type of material. Links that don't work or are to other weak or fringe pages do not help strengthen the credibility of your research.
If reproduced information (from another source), is it complete, not altered, not fake or forged? Is it retyped? If so, it could easily be altered. Is it reproduced from another publication? o Are permissions to reproduce and copyright information provided? o Is there a reason there are not links to the original source if it is online (instead of reproducing it)?	You may have to find the original to be sure a copy of something is not altered and is complete. Look at the URL: is it from the original source? If you find a legitimate article from a reputable journal or other publication, it should be accompanied by the copyright statement and/or permission to reprint. If it is not, be suspicious. Try to find the source. If the URL of the document is not to the original source, it is likely that it is illegally reproduced, and the text could be altered, even with the copyright information present.
Are there links to other resources on the topic?	Many well developed pages offer links to other pages on the same topic that they

Are the links well chosen, well organized, and/or evaluated/annotated? Do the links work? Do the links represent other viewpoints? Do the links (or absence of other viewpoints) indicate a bias?	consider worthwhile. They are inviting you compare their information with other pages. Links that offer opposing viewpoints as well as their own are more likely to be balanced and unbiased than pages that offer only one view. Anything not said that could be said? And perhaps would be said if all points of view were represented? Always look for bias. Especially when you agree with something, check for bias.

4. What do others say?

Techniques for Web Evaluation :

1. Find out what other web pages link to this page.

a. Use alexa.com URL information:
Simply paste the url into alexa.com's search box.
You will see, depending on the volume of traffic to the page:

Traffic rank
Subjective reviews
"Site statistics" including some page history, sites that link to the page
Contact/ownership info for the domain name
A link to the Internet archive of website history "Wayback Machine"

b. Do a **link:** search in Google, Yahoo!, or another search engine where this can be done.
INSTRUCTIONS for doing a **link:** search in Google or Yahoo!:
1. Copy the URL of the page you are investigating (Ctrl+C in Windows).
2. Go to the search engine site, and type **link:** in the search box.
3. Paste the URL of the investigated site into the search box immediately following **link:** (no space after the colon).
The pages listed all contain one or more links to the page you are looking for.
If you find no links, try a shorter portion of the URL, stopping after each /.

2. Look the page up in a reputable directory that evaluates its contents (Librarians' Index, Infomine, About.com, AcademicInfo, or a specialized directory you trust).
INSTRUCTIONS: Go to the directory and search for the title of the site you are investigating. Look for the publisher if you can't find a specific page from a larger site.

3. Look up the author's name in Google or Yahoo!
INSTRUCTIONS in Google: Search the name three ways:
a. without quotes - **Jabberwocky Webauthor**
b. enclosed in quotes as a phrase - **"Jabberwocky Webauthor"**
c. enclosed in quotes with ***** between the first and last name - **"Jabberwocky * Webauthor"** (The ***** can stand for any middle initial or name in Google only).

Questions to ask:	What are the implications?

Who links to the page? Are there many links? What kinds of sites link to it? What do they say? Are any of them directories? Try looking at what directories say.	Sometimes a page is linked to only by other parts of its own site (not much of a recommendation). Sometimes a page is linked to by its fan club, and by detractors. Read both points of view. If a page or its site is in a bona fide directory, think about whether there is much critical evaluation of the links in the directory.
Is the page listed in one or more reputable directories or pages?	Good directories include a tiny fraction of the web, and inclusion in a directory is therefore noteworthy. But read what the directory says! It may not be 100% positive.
What do others say about the author or responsible authoring body?	"Googling someone" (new term for this) can be revealing. Be sure to consider the source. If the viewpoint is radical or controversial, expect to find detractors. Think critically about all points of view.

5. Does it all add up?

Techniques for Web Evaluation :

1. Step back and think about all you have learned about the page. Listen to your gut reaction. Think about why the page was created, the intentions of its author(s).

   If you have doubts, ask your instructor or come to one of the library reference desks and ask for advice.

2. Be sensitive to the possibility that you are the victim of irony, spoof, fraud, or other falsehood.
3. Ask yourself if the web is truly the best place to find resources for the research you are doing.

Questions to ask:	So what? What are the implications?
Why was the page put on the web? Inform, give facts, give data? Explain, persuade? Sell, entice? Share? Disclose?	These are some of the reasons to think of. The web is a public place, open to all. You need to be aware of the entire range of human possibilities of intentions behind web pages.
Might it be ironic? Satire or parody? Think about the "tone" of the page. Humorous? Parody? Exaggerated? Overblown arguments? Outrageous photographs or juxtaposition of unlikely images? Arguing a viewpoint with examples that suggest that what is argued is ultimately not possible.	It is easy to be fooled, and this can make you look foolish in turn.
Is this as good as resources I could find if I used the library, or some of the web-based indexes available through the library, or other print	What is your requirement (or your instructor's requirement) for the quality of reliability of your information?

resources? Are you being completely fair? Too harsh? Totally objective? Requiring the same degree of "proof" you would from a print publication? Is the site good for some things and not for others? Are your hopes biasing your interpretation?	In general, published information is considered more reliable than what is on the web. But many, many reputable agencies and publishers make great stuff available by "publishing" it on the web. This applies to most governments, most institutions and societies, many publishing houses and news sources. But take the time to check it out.

WHY? Rationale for Evaluating What You Find on the Web

The World Wide Web can be a great place to accomplish research on many topics. But putting documents or pages on the web is easy, cheap or free, unregulated, and unmonitored (at least in the USA). There is a famous Steiner cartoon published in the *New Yorker* (July 5, 1993) with two dogs sitting before a terminal looking at a computer screen; one says to the other "On the Internet, nobody knows you're a dog." The great wealth that the Internet has brought to so much of society is the ability for people to express themselves, find one another, exchange ideas, discover possible peers worldwide they never would have otherwise met, and, through hypertext links in web pages, suggest so many other people's ideas and personalities to anyone who comes and clicks. There are some real "dogs" out there, but there's also great treasure.

Therein lies the rationale for evaluating carefully whatever you find on the Web. The burden is on you - the reader - to establish the validity, authorship, timeliness, and integrity of what you find. Documents can easily be copied and falsified or copied with omissions and errors -- intentional or accidental. In the general World Wide Web there are no editors (unlike most print publications) to proofread and "send it back" or "reject it" until it meets the standards of a publishing house's reputation. Most pages found in general search engines for the web are self-published or published by businesses small and large with motives to get you to buy something or believe a point of view. Even within university and library web sites, there can be many pages that the institution does not try to oversee. *The web needs to be free like that!! And you, if you want to use it for serious research, need to cultivate the habit of healthy skepticism, of questioning everything you find with critical thinking.*

More About Evaluating Web Sources

Evaluating Information Found on the Internet
<http://www.library.jhu.edu/researchhelp/general/evaluating/>
An excellent series of pages on this subject (from the Milton Library at Johns Hopkins University).
Critical Evaluation of Resources
<http://www.lib.berkeley.edu/TeachingLib/Guides/Evaluation.html>
A broader, more theoretical look at the criteria and importance of evaluating all types of resources (also from the UC Berkeley Teaching Library).

For annotated descriptions of many other good guides to evaluating web pages, search the subject "Evaluation of Internet Resources" in the **Librarians' Index to the Internet** <*http://lii.org*>.

Citation Styles, Style Guides, and Avoiding Plagiarism

Why cite sources?
Whenever you quote or base your ideas on another person's work, you must document the source you used. Even when you do not quote directly from another work, if reading that source contributed to the ideas presented in your paper, you must give the authors proper credit.

Citations allow readers to locate and further explore the sources you consulted, show the depth and scope of your research, and give credit to authors for their ideas. Citations provide evidence for your arguments and add credibility to your work by demonstrating that you have sought out and considered a variety of resources. In written academic work, citing sources is standard practice and shows that you are responding to this person, agreeing with that person, and adding something of your own. Think of documenting your sources as providing a trail for your reader to follow to see the research you performed and discover what led you to your original contribution.

By following these guidelines, you avoid plagiarism.

How do you cite sources?
The means to identify sources is to provide citations within your text linking appropriate passages to relevant resources consulted or quoted. This can be done through in-text parenthetic notes, footnotes, or endnotes. In addition, a bibliography or list of works cited, is almost always placed at the end of your paper. The citation system and format you use will be determined by the **citation style** you choose.

How do you choose a style?
Ask your instructor which style sheet he or she wishes you to use and if there are other special formatting instructions you should follow.

Where do I find the most authoritative information about these styles?
If you have questions or citations not covered by the Library's guides, please consult one of the following official style manuals. If you consult other, less official manuals or online style guides that purport to explain these style, please be aware that these sometimes contain errors which conflict with the official guides:

APA Style

Publication Manual of the American Psychological Association. 5th ed. Washington, DC: American Psychological Association, 2001.
The official APA style guide.
Electronic Reference Formats Recommended by the American Psychological Association
An official excerpt from the APA Publication Manual above.

From the American Psychological Association website, http://www.apastyle.org/elecref.html

MLA Style

Gibaldi, Joseph. *MLA Handbook for Writers of Research Papers.* 6th ed. New York: Modern Language Association of America, 2003.
A somewhat simplified guide, adequate for undergraduate and most other research papers.
Gibaldi, Joseph *The MLA Style Manual and Guide to Scholarly Publishing.* 2nd ed. New York: Modern Language Association of America, 1998.
For graduate students, scholars, and professional writers (more depth on copyright, legal issues, and writing theses, dissertations, and scholarly publishing).

Turabian Style

Turabian, Kate L. *A Manual for Writers of Term Papers, Theses, and Dissertations*, 6th edition. Chicago: University of Chicago Press, 1996.

Chicago Style

The Chicago Manual of Style. 15th ed. Chicago: University of Chicago Press, 2003.

What is plagiarism?

Plagiarism is defined in the University of California, Berkeley, Campus Code of Student Conduct as follows:

"Plagiarism is defined as the use of intellectual material produced by another person without acknowledging its source. This includes, but is not limited to:

> *(a.) Copying from the writings or works of others into one's academic assignment without attribution, or submitting such work as if it were one's own;*
> *(b.) Using the views, opinions, or insights of another without acknowledgment; or*
> *(c.) Paraphrasing the characteristic or original phraseology, metaphor, or other literary device of another without proper attribution."*

Plagiarism is a serious violation of academic and student conduct rules and is punishable at most universities and colleges with a failing grade and possibly more severe action.

Glossary of Internet & Web Jargon

BACK / FORWARD

Buttons in most browsers' Tool Button Bar, upper left. BACK returns you to the document previously viewed. FORWARD goes to the next document, after you go BACK.
If it seems like the BACK button does not work, check if you are in a new browser window; some Web pages are programmed to open a new window when you click on some links. Each window has its own short-term search HISTORY. If this does not work, right click on the BACK button to select the page you want (some Web pages are programmed to disable BACK).

BLOG or **WEB LOG**

A blog (short for "web log") is a type of web page that serves as a publicly accessible personal journal (or log) for an individual. Typically updated daily, blogs often reflect the personality of the author. Blog software usually has an archive of old blog postings. Many blogs can be searched for terms in the archive. Blogs have become a vibrant, fast-growing medium for communication in professional, poltical, news, trendy, and other specialized web communities. Many blogs provide RSS feeds, to which one can subscribe and receive alerts to new postings in selected blogs.

BOOKMARK/FAVORITES

Way in browsers to store in your computer direct links to sites you wish to return to. Netscape, Mozilla, and Firefox use the term Bookmarks. The equivalent in Internet Explorer (IE) is called a "Favorite." To create a bookmark, click on BOOKMARKS or FAVORITES, then ADD. Or left-click on and drag the little bookmark icon to the place you want a new bookmark filed. To visit a bookmarked site, click on BOOKMARKS and select the site from the list.

You can download a bookmark file to diskette and install it on another computer. In most browsers now, you can do this with an Import... and Export... set of commands which can be found under FILE or in the Manage Bookmarks window's FILE.

BOOLEAN LOGIC

Way to combine terms using "operators" such as "AND," "OR," "AND NOT" and sometimes "NEAR." AND requires all terms appear in a record. OR retrieves records with either term. AND NOT excludes terms. Parentheses may be used to sequence operations and group words. Always enclose terms joined by OR with parentheses. Which search engines have this?

See -REJECT TERM and FUZZY AND. Want a more extensive explanation of Boolean logic, with illustrations?

BROWSE

To follow links in a page, to shop around in a page, exploring what's there, a bit like window shopping. The opposite of browsing a page is **searching** it. When you search a page, you find a search box, enter terms, and find all occurrences of the terms throughout the site. When you browse, you have to guess which words on the page pertain to your interests. Searching is usually more efficient, but sometimes you find things by browsing that you might not find because you might not think of the "right" term to search by.

BROWSERS

Browsers are software programs that enable you to view WWW documents. They "translate" HTML-encoded files into the text, images, sounds, and other features you see. Microsoft Internet Explorer (called simply IE), Mozilla, Firefox, Safari, and Opera are examples of "graphical" browsers that enable you to view text and images and many other WWW features. They are software that must be installed on your computer. For more information about browsers, consult the introductory pages of the Teaching Library tutorial.

CACHE

In browsers, "cache" is used to identify a space where web pages you have visited are stored in your computer. A copy of documents you retrieve is stored in cache. When you use GO, BACK, or any other means to revisit a document, the browser first checks to see if it is in cache and will retrieve it from there because it is much faster than retrieving it from the server.

CACHED LINK

In search results from Google, Yahoo! Search, and some other search engines, there is usually a Cached link which allows you to view the version of a page that the search engine has stored in its database. The live page on the web might differ from this cached copy, because the cached copy dates from whenever the search engine's spider last visited the page and detected modified content. Use the cached link to see when a page was last crawled and, in Google, where your terms are and why you got a page when all of your search terms are not in it.

CASE SENSITIVE

Capital letters (upper case) retrieve only upper case. Most search tools are not case sensitive or only respond to initial capitals, as in proper names. It is always safe to key all lower case (no capitals), because lower case will always retrieve upper case. Which search engines have this?

CGI

"Common Gateway Interface," the most common way Web programs interact dynamically with users. Many search boxes and other applications that result in a page with content tailored to the user's search terms rely on CGI to process the data once it's submitted, to pass it to a background program in JAVA, JAVASCRIPT, or another programming language, and then to integrate the response into a display using HTML.

COOKIE

A message from a WEB SERVER computer, sent to and stored by your browser on your computer. When your computer consults the originating server computer, the cookie is sent back to the server, allowing it to respond to you according to the cookie's contents. The main use for cookies is to provide customized Web pages according to a profile of your interests. When you log onto a "customize" type of invitation on a Web page and fill in your name and other information, this may result in a cookie on your computer which that Web page will access to appear to "know" you and provide what you want. If you fill out these forms, you may also receive e-mail and other solicitation independent of cookies.

CRAWLER or WEBCRAWLER

Same as Spider.

DOMAIN, TOP LEVEL DOMAIN (TLD)

Hierarchical scheme for indicating logical and sometimes geographical venue of a web-page from the network. In the US, common domains are .edu (education), .gov (government agency), .net (network related), .com (commercial), .org (nonprofit and research organizations). Outside the US, domains indicate country: ca (Canada), uk (United Kingdom), au (Australia), jp (Japan), fr (France), etc. Neither of these lists is exhaustive.

DOMAIN NAME, DOMAIN NAME SERVER (DNS)ENTRY

Any of these terms refers to the initial part of a URL, down to the first /, where the domain and name of the host or SERVER computer are listed (most often in reversed order, name first, then domain). The domain name gives you who "published" a page, made it public by putting it on the Web.

A domain name is translated in huge tables standardized across the Internet into a numeric IP address unique the host computer sought. These tables are maintained on computers called "Domain Name Servers." Whenever you ask the browser to find a URL, the browser must consult the table on the domain name server that particular computer is networked to consult.

"Domain Name Server entry" frequently appears a browser error message when you try to enter a URL. If this lookup fails for any reason, the "lacks DNS entry" error occurs. The most common remedy is simply to try the URL again, when the domain name server is less busy, and it will find the entry (the corresponding numeric IP address). For more information, see "All About Domain Names."

DOWNLOAD
To copy something from a primary source to a more peripheral one, as in saving something found on the Web (currently located on its server) to diskette or to a file on your local hard drive. More information.

EXTENSION or FILE EXTENSION
In Windows, DOS and some other operating systems, one or several letters at the end of a filename. Filename extensions usually follow a period (dot) and indicate the type of file. For example, *this.txt* denotes a plain text file, *that.htm* or *that.html* denotes an HTML file. Some common image extensions are *picture.jpg* or *picture.jpeg* or *picture.bmp* or *picture.gif*

FAVORITES
In the Internet Explorer browser, a means to get back to a URL you like, similar to Bookmarks.

FEED READER
A software package that enables you to easily read the XML code in which RSS feeds are written. Bloglines is currently the most popular feed reader but there are many competitors.

FIELD SEARCHING
Ability to limit a search by requiring word or phrase to appear in a specific field of documents (e.g., title, url, link). See LIMITING TO FIELD.

FIND
Tool in most browsers to search for word(s) keyed in document in screen only. Useful to locate a term in a long document. Can be invoked by the keyboard command, Ctrl+F.

FRESHNESS
How up-to-date a search engine database is, based primarily on how often its spiders recirculate around the Web and update their copies of the web pages they hold, and discover new ones. Also determined by how quickly they integrate new sites that web authors send to them. Two weeks is about as good as most search engines do, but some update certain selected web sites more frequently, even daily.

FRAMES
A format for web documents that divides the screen into segments, each with a scroll bar as if it were as "window" within the window. Usually, selecting a category of documents in one frame shows the contents of the category in another frame. To go BACK in a frame, position the cursor in the frame an press the right mouse button, and select "Back in frame" (or Forward).
You can adjust frame dimensions by positioning the cursor over the border between frames and dragging the border up/down or right/left holding the mouse button down over the border.

FTP
File Transfer Protocol. Ability to transfer rapidly entire files from one computer to another, intact for viewing or other purposes.

FUZZY AND
In ranking of results, documents with all terms (Boolean AND) are ranked first, followed by documents containing any terms (Boolean OR) are retrieved. The farther down, the fewer the terms, although at least one should always be present.

HEAD or HEADER (of HTML document)

The top portion of the HTML source code behind Web pages, beginning with <HEAD> and ending with </HEAD>. It contains the Title, Description, Keywords fields and others that web page authors may use to describe the page. The title appears in the title bar of most browsers, but the other fields cannot be seen as part of the body of the page. To view the <HEAD> portion of web pages in your browser, click VIEW, Page Source. In Internet Explorer, click VIEW, Source. Some search engines will retrieve based on text in these fields.

HISTORY, Search History

Available by using the combined keystrokes CTRL + H, a more permanent record of sites you have visited/retrieved than GO. You can set how many days your browser retains history in Edit | Preferences, or in Tools | Options.

HOST

Computer that provides web-documents to clients or users. See also server.

HTML

Hypertext Markup Language. A standardized language of computer code, imbedded in "source" documents behind all Web documents, containing the textual content, images, links to other documents (and possibly other applications such as sound or motion), and formatting instructions for display on the screen. When you view a Web page, you are looking at the product of this code working behind the scenes in conjunction with your browser. Browsers are programmed to interpret HTML for display.

HTML often imbeds within it other programming languages and applications such as SGML, XML, Javascript, CGI-script and more. It is possible to deliver or access and execute virtually any program via the WWW.

You can see HTML by selecting the View pop-down menu tab, then "Document Source."

HYPERTEXT

On the World Wide Web, the feature, built into HTML, that allows a text area, image, or other object to become a "link" (as if in a chain) that retrieves another computer file (another Web page, image, sound file, or other document) on the Internet. The range of possibilities is limited by the ability of the computer retrieving the outside file to view, play, or otherwise open the incoming file. It needs to have software that can interact with the imported file. Many software capabilities of this type are built into browsers or can be added as "plug-ins."

INTERNET (Upper case I)

The vast collection of interconnected networks that all use the TCP/IP protocols and that evolved from the ARPANET of the late 60's and early 70's. An "internet" (lower case i) is any computers connected to each other (a network), and are not part of the Internet unless the use TCP/IP protocols. An "intranet" is a private network inside a company or organization that uses the same kinds of software that you would find on the public Internet, but that is only for internal use. An intranet may be on the Internet or may simply be a network.

IP Address or IP Number

(Internet Protocol number or address). A unique number consisting of 4 parts separated by dots, e.g. 165.113.245.2

Every machine that is on the Internet has a unique IP address. If a machine does not have an IP number, it is not really on the Internet. Most machines also have one or more Domain Names that are easier for people to remember.

ISP or Internet Service Provider

A company that sells Internet connections via modem (examples: aol, Mindspring - thousands of ISPs to choose from; not easy to evaluate). Faster, more expensive Internet connectivity is available via cable, DSL, ISDN, or web-TV. Often these companies also provide Web page hosting service (free or relatively inexpensive web pages -- the origin of many personal pages).

JAVA

A network-oriented programming language invented by Sun Microsystems that is specifically designed for writing programs that can be safely downloaded to your computer through the Internet and immediately run without fear of viruses or other harm to our computer or files. Using small Java programs (called "Applets"), Web pages can include functions such as animations, calculators, and other fancy tricks. We can expect to see a huge variety of features added to the Web using Java, since you can write a Java program to do almost anything a regular computer program can do, and then include that Java program in a Web page. For more information search any of these jargon terms in the PC Webopedia.

JAVASCRIPT

A simple programming language developed by Netscape to enable greater interactivity in Web pages. It shares some characteristics with JAVA but is independent. It interacts with HTML, enabling dynamic content and motion.

KEYWORD(S)

A word searched for in a search command. Keywords are searched in any order. Use spaces to separate keywords in simple keyword searching. To search keywords exactly as keyed (in the same order).

LIMITING TO A FIELD

Requiring that a keyword or phrase appear in a specific field of documents retrieved. Most often used to limit to the "Title" field in order to find documents primarily about one or more keywords. (Can be used for other fields. See the table summarizing search tools features.)

LINK

The URL imbedded in another document, so that if you click on the highlighted text or button referring to the link, you retrieve the outside URL. If you search the field "link:", you retrieve on text in these imbedded URLs which you do not see in the documents.

LINK "ROT"

Term used to describe the frustrating and frequent problem caused by the constant changing in URLs. A Web page or search tool offers a link and when you click on it, you get an error message (e.g., "not available") or a page saying the site has moved to a new URL. Search engine spiders cannot keep up with the changes. URLs change frequently because the documents are moved to new computers, the file structure on the computer is reorganized, or sites are discontinued. If there is no referring link to the new URL, there is little you can do but try to search for the same or an equivalent site from scratch.

LISTSERVERS

A discussion group mechanism that permits you to subscribe and receive and participate in discussions via e-mail. Blogs and RSS feeds provide some of the communication functionality of listservers.

META-SEARCH ENGINE

Search engines that automatically submit your keyword search to several other search tools, and retrieve results from all their databases. Convenient time-savers for relatively simple keyword searches (one or two keywords or phrases in " "). See Meta-Search Engines page for complete descriptions and examples.

NESTING

A term used in Boolean searching to indicate the sequence in which operations are to be performed. Enclosing words in parentheses identifies a group or "nest." Groups can be within other groups. The operations will be performed from the innermost nest to the outmost, and then from left to right.

NEWSGROUP

A discussion group operated through the Internet. Not to be confused with LISTSERVERS which operate through e-mail.

PERSONAL PAGE

A web page created by an individual (as opposed to someone creating a page for an institution, business, organization, or other entity). Often personal pages contain valid and useful opinions, links to important resources, and significant facts. One of the greatest benefits of the Web is the freedom it as given almost anyone to put his or her ideas "out there." But frequently personal pages offer highly biased personal perspectives or ironical/satirical spoofs, which must be evaluated carefully. The presence in the page's URL of a personal name (such as "jbarker") and a ~ or % or the word "users" or "people" or "members" very frequently indicate a site offering personal pages.

PACKET, PACKET JAM

When you retrieve a document via the WWW, the document is sent in "packets" which fit in between other messages on the telecommunications lines, and then are reassembled when they arrive at your end. This occurs using TCP/IP protocol. The packets may be sent via different paths on the networks which carry the Internet. If any of these packets gets delayed, your document cannot be reassembled and displayed. This is called a "packet jam." You can often resolve packet jams by pressing STOP then RELOAD. RELOAD requests a fresh copy of the document, and it is likely to be sent without jamming.

PDF or .pdf or pdf file

Abbreviation for Portable Document Format, a file format developed by Adobe Systems, that is used to capture almost any kind of document with the formatting in the original. Viewing a PDF file requires Acrobat Reader, which is built into most browsers and can be downloaded free from Adobe.

PHRASE

More than one KEYWORD, searched exactly as keyed (all terms required to be in documents, in the order keyed). Enclosing keywords in quotations " " forms a phrase in AltaVista, , and some other search tools. Some times a phrase is called a "character string."

PLUG-IN

An application built into a browser or added to a browser to enable it to interact with a special file type (such as a movie, sound file, Word document, etc.)

POPULARITY RANKING of search results

Some search engines rank the order in which search results appear primarily by how many other sites link to each page (a kind of popularity vote based on the assumption that other pages would create a link to the "best" pages). Google is the best example of this. See also Subject-Based Ranking.

+REQUIRE or -REJECT A TERM OR PHRASE

Insert + immediately before a term (no space) to limit search to documents containing a term. Insert - immediately before a term (no space) to exclude documents containing a term. Can be used immediately (no space) before the " " delimiting a phrase.

Functions partially like basic BOOLEAN LOGIC. If + precedes more than one term, they are required as with Boolean AND. If - is used, terms are excluded as with Boolean AND NOT. If neither + no - is used, the default if Boolean OR. However, full Boolean logic allows parentheses to group and sequence logical operations, and +/- do not. Which search engines have this?

RELEVANCY RANKING of search results

The most common method for determining the order in which search results are displayed. Each search tool uses its own unique algorithm. Most use "fuzzy and" combined with factors such as how often your terms occur in documents, whether they occur together as a phrase, and whether they are in title or how near the top of the text. Popularity is another ranking system.

RSS or RSS feeds

Short for "Really Simple Synication" (a.k.a. Rich Site Summary or RDF Site Summary), refers ti a group of XML based web-content distribution and republication (Web syndication) formats primarily used by news sites and weblogs (blogs). Any website can issue an RSS feed. By subscribing to an RSS feed, you are alerted to new additions to the feed since you last read it. In order to read RSS feeds, you must use a "feed reader," which formats the XML code into an easily readable format (feed readers are to XML and RSS feeds as web browsers are to HTML and web pages.

SCRIPT

A script is a type of programming language that can be used to fetch and display Web pages. There are may kinds and uses of scripts on the Web. They can be used to create all or part of a page, and communicate with searchable databases. Forms (boxes) and many interactive links, which respond differently depending on what you enter, all require some kind of script language. When you find a question marke (?) in the URL of a page, some kind of script command was used in generating and/or delivering that page. Most search engine spiders are instructed not to crawl pages from scripts, although it is usually technically possible for them to do so (see Invisible Web for more information).

SERVER, WEB SERVER

A computer running that software, assigned an IP address, and connected to the Internet so that it can provide documents via the World Wide Web. Also called HOST computer. Web servers are the closest equivalent to what in the print world is called the "publisher" of a print document. An important difference is that most print publishers carefully edit the content and quality of their publications in an effort to market them and future publications. This convention is not required in the Web world, where anyone can be a publisher; careful evaluation of Web pages is therefore mandatory. Also called a "Host."

SERVER-SIDE

Something that operates on the "server" computer (providing the Web page), as opposed to the "client" computer (which is you or someone else viewing the Web page). Usually it is a program or command or procedure or other application causes dynamic pages or animation or other interaction.

SHTML, usually seen as **.shtml**

An file name extension that identifies web pages containing SSI commands.

SITE or WEB-SITE

This term is often used to mean "web page," but there is supposed to be a difference. A web page is a single entity, one URL, one file that you might find on the Web. A "site," properly speaking, is an location or gathering or center for a bunch of related pages linked to from that site. For example, the site for the present tutorial is the top-level page "Internet Resources." All of the pages associated with it branch out from there -- the web searching tutorial and all its pages, and more. Together they make up a "site." When we estimate there are 5 billion web pages on the Web, we do not mean "sites." There would be far fewer sites.

SPIDERS

Computer robot programs, referred to sometimes as "crawlers" or "knowledge-bots" or "knowbots" that are used by search engines to roam the World Wide Web via the Internet, visit sites and databases, and keep the search engine database of web pages up to date. They obtain new pages, update known pages, and delete obsolete ones. Their findings are then integrated into the "home" database.

Most large search engines operate several robots all the time. Even so, the Web is so enormous that it can take six months for spiders to cover it, resulting in a certain degree of "out-of-datedness" (link rot) in all the search engines. For more information, read about search engines.

SPONSOR (of a Web page or site)

Many Web pages have organizations, businesses, institutions like universities or nonprofit foundations, or other interests which "sponsor" the page. Frequently you can find a link titled "Sponsors" or an "About us" link explaining who or what (if anyone) is sponsoring the page. Sometimes the advertisers on the page (banner ads, links, buttons to sites that sell or promote something) are "sponsors." *WHY is this important?* Sponsors and the funding they provide may, or may not, influence what can be said on the page or site -- can bias what you find, by excluding some opposing viewpoint or causing some other imbalanced information. The site is not bad because of sponsors, but you they should alert you to the need to evaluate a page or site very carefully.

SSI commands

SSI stands for "server-side include," a type of HTML instruction telling a computer that serves Web pages to dynamically generate data, usually by inserting certain variable contents into a fixed template or boilerplate Web page. Used especially in database searches.

STEMMING

In keyword searching, word endings are automatically removed (*lines* becomes *line*); searches are performed on the stem + common endings (*line* or *lines* retrieves *line, lines, line's, lines', lining, lined).* Not very common as a practice, and not always disclosed. Can usually be avoided by placing a term in " ".

STOP WORDS

In database searching, "stop words" are small and frequently occurring words like *and, or, in, of* that are often ignored when keyed as search terms. Sometimes putting them in quotes " " will allow you to search them. Sometimes + immediately before them makes them searchable. See Table of Search Engine features.

SUBJECT-BASED POPULARITY RANKING of search results

A variation on popularity ranking in which the links in pages on the same subject are used to in ranking search results.

SUBJECT DIRECTORY

An approach to Web documents by a lexicon of subject terms hierarchically grouped. May be browsed or searched by keywords. Subject directories are smaller than other searchable databases, because of the human involvement required to classify documents by subject.

SUB-SEARCHING

Ability to search only within the results of a previous search. Enables you to refine search results, in effect making the computer "read" the search results for you selecting documents with terms you sub-search on. Can function much like RESULTS RANKING.

TCP/IP

(Transmission Control Protocol/Internet Protocol) -- This is the suite of protocols that defines the Internet. Originally designed for the UNIX operating system, TCP/IP software is now available for every major kind of computer operating system. To be truly on the Internet, your computer must have TCP/IP software. See also IP Address.

TELNET

Internet service allowing one computer to log onto another, connecting as if not remote.

THESAURUS

In some search tools, the terms you choose to search on can lead you to other terms you may not have thought of. Different search tools have different ways of presenting this information, sometimes with suggested words you may choose among and sometimes automatically. The terms are based on the terms in the results of your search, not on some dictionary-like thesaurus.

TITLE (of a document)

The official title of a document from the "meta" field called title. The text of this meta title field may or may not also occur in the visible body of the document. It is what appears in the top bar of the window when you display the document and it is the title that appears in search engine results. The "meta" field called title is not mandatory in HTML coding. Sometimes you retrieve a document with "No Title" as its supposed title; this is caused when the meta-title field is left blank.

In Alta Vista and some other search tools, *title:* search also matches on the "meta" field, which contains document descriptors not displayed on the Web. See also LIMITING TO A FIELD.

TRUNCATION

In a search, the ability to enter the first part of a keyword, insert a symbol (usually *), and accept any variant spellings or word endings, from the occurrence of the symbol forward. (E.g., *femini** retrieves *feminine, feminism, feminism,* etc.)

URL

Uniform Resource Locator. The unique address of any Web document. May be keyed in a browser's OPEN or LOCATION / GO TO box to retrieve a document. There is a logic the layout of a URL:

Anatomy of a URL:

Type of file (could say ftp:// or telnet://)	**Domain name (computer file is on and its location on the Internet)**	**Path or directory on the computer to this file**	**Name of file, and its file extension (usually ending in .html or .htm)**
http://	www.lib.berkeley.edu/	TeachingLib/Guides/Internet/	FindInfo.html

USENET

Bulletinboard-like network featuring thousands of "newsgroups." Google incorporates the historic file of Usenet Newsgroups (bzck to 1981) into its Google Groups. Yahoo Groups offers a similar service, but does not include the old "Usenet Newsgroups." Blogs are replacing some of the need for this type of community sharing and information exchange.

WORD VARIANTS

Different word endings (such as *-ing, -s, es, -ism, -ist,*etc.) will be retrieved only if you allow for them in your search terms. One way to do this TRUNCATION, but few systems accept truncation. Another way is to enter the variants either separated by BOOLEAN OR (and grouped in parentheses). In +REQUIRE/-REJECT non-Boolean systems, enter the variant terms preceded with neither + nor -, because this will allow documents containing any of them to retrieved.

XHTML

A variant of HTML. Stands for Extensible Hypertext Markup Language is a hybrid between HTML and XML that is more universally acceptable in Web pages and search engines than XML.

XML

Extensible Markup Language, a dilution for Web page use of SGML (Standard General Markup Language), which is not readily viewable in ordinary browsers and is difficult to apply to Web pages. XML is very useful (among other things) for pages emerging from databases and other applications where parts of the page are standardized and must reappear many times.

Can't find the term you want?

Search almost any computer jargon in the PC Webopedia <http://www.pcwebopedia.com/>, "the #1 online encyclopedia and search engine dedicated to computer technology" from PC Magazine's experts. Another excellent encyclopedia for computer jargon and many other topics is Wikipedia, "The Free Encyclopedia." Excellent context and background for many topics.

Summary: Use search engines and subject directories to locate materials on the Web. Check to see if your class has a bibliography or research guide created by librarians.

Step 6: Evaluate What You Find

How to Evaluate the Information Sources You Find

Evaluating the authority, usefulness, and reliability of the information you find is a crucial step in the process of library research. The questions you ask about books, periodical articles, multimedia titles, or Web pages are similar whether you're looking at a citation to the item, a physical item in hand, or an electronic version on a computer.

All Sources:

Critically Analyzing Information Sources lists some of the questions you should ask when you consider the appropriateness of a particular book, article, media resource, or Web site for your research.

Books:

Use **book reviews** to gather critical information about books. Three quick ways to access them online [users only]:

Periodical Abstracts. [1986- ; some full-text reviews]
Book Review Digest. [1983- ; excerpts from some reviews]
Bowker's Books in Print. [in-print books from any year; full-text of short reviews]

More sources for book reviews at Book Reviews: A Finding Guide.

Periodicals:

Distinguishing Scholarly from Non-Scholarly Periodicals: A Checklist of Criteria shows how to **evaluate periodicals** by looking at their format, intended audience, and appearance.

Web Sites:

Evaluating Web Pages: Questions to Ask & Strategies for Getting the Answers is an excellent guide from UC Berkeley.

Five Criteria for Evaluating Web Sites is a brief table of tips and questions to ask.

For another approach, see Evaluating Web Sites: Criteria and Tools.

Step 7: Cite What You Find Using a Standard Format

Give credit where credit is due; cite your sources.

Citing or documenting the sources used in your research serves two purposes, it gives proper credit to the authors of the materials used, and it allows those who are reading your work to duplicate your research and locate the sources that you have listed as references.

Knowingly representing the work of others as your own is plagarism. Use one of the styles listed below or another style approved by your instructor.

Style guides in print (book) format:

- Gibaldi, Joseph. **MLA Handbook for Writers of Research Papers.** 6th ed. New York: MLA, 2003.
  (Olin Ref Z 253 .M68 2003; also Uris Ref, others)

This handbook is based on the *MLA Style Manual* (Olin and Uris Ref PN 147 .G444x 1998) and is intended as an aid for college students writing research papers. Included here is information on selecting a topic, researching the topic, note taking, the writing of footnotes and bibliographies, as well as sample pages of a research paper. Useful for the beginning researcher.

- **Publication Manual of the American Psychological Association**. 5th ed. Washington: APA, 2001. **(Olin Ref BF 76.7 .P83x 2001; also Uris Ref, Mann Ref, others)**

  The authoritative style manual for anyone writing in the field of psychology. Useful for the social sciences generally. Chapters discuss the content and organization of a manuscript, writing style, the American Psychological Association citation style, and typing, mailing and proofreading.

Research Tips:

Work From The General To The Specific.
Find background information first, then use more specific and recent sources.

Record What You Find And Where You Found It.
Record the complete citation for each source you find; you may need it again later.

Translate Your Topic Into The Subject Language Of The Indexes And Catalogs You Use.
Check your topic words against a thesaurus or subject heading list.

7.2 Writing and Presentation of a Research paper for a Conference or Journal.

7.2.1 Components of a Research Article: Purpose-Content-Style

Title

Purpose: To attract readers interested in this field of study. The importance of the title cannot be overstated as it is a major determinant of whether the paper will be read. It is the ***only*** aspect of the article that appears in tables of contents and in many of the databases used for literature searches.

Content: The title should indicate the focus of the paper, and should contain enough relevant “keywords” (i.e., search terms) to enable readers to find this paper when searching a relevant database.

Style: There are two types of titles:

1. *Descriptive* titles, which states the focus of the study:

The effect of heat on ice.

2. *Conclusion* titles, which provide the authors’ main conclusion from their study.

Heat melts ice.

For research articles, we prefer a descriptive title. It “lets the data speak for themselves” (an important concept in research) and allows the author to provide the necessary restrictions on the conclusions, restrictions that usually cannot be accommodated in a title. (For example, that heat melts ice only as long as sufficient time is provided to allow the ice molecules to reach the melting point of water, which is 0 C for pure water at sea level.)

There is another reason for preferring descriptive titles: Should evidence later come to light that indicates the authors’ conclusion is incorrect, the authors’ curriculum vitae will not contain a permanent reminder of this mistake! In some cases, a journal’s *Instructions to* A*uthors* will specify which style of title to use.

Often authors are asked to provide a second, briefer title. Referred to as the *running title*, this title will appear as a header on every other page of the printed article. Like the full title, the running title should be chosen with care.

Length: A title should be short yet provide enough keywords so that individuals will be able to find it when searching relevant databases. A general recommendation is 5-10 words.

Authors

Purpose: To specify the individuals responsible for the research presented in the paper.

Content: The definition of who should be an author (and in what order the list should be provided, see below) varies with the field, the culture, and even the research group. Because of this potential for ambiguity, the rules to be used for determining authorship, including the order of authors, should be clearly agreed upon at the outset of a research project.

Our belief is that authorship denotes an “intellectual contribution” to the work, and that an author should be able to explain and defend the work. This definition of authorship is probably the most common one among researchers and journal editors. Note that within this framework, “honorary authorship” listing someone as an author who has **not** made an intellectual contribution (e.g., the head of the department or that individual who provided the funds) would be considered unethical.

In many areas of science, students typically publish papers in which their advisor is an author. However, in some other fields, e.g., education and many areas of the humanities, students frequently are the sole author on their papers.

Independent of the method used to determine authorship, it is *essential* that all authors have given their consent to be designated as such and have approved the final version of the manuscript. One author is chosen as the "corresponding author." The editor of the journal will direct all correspondence to this individual who then has responsibility for keeping the other authors up-to-date with regard to the status of the manuscript.

Style: Like authorship, the *order* in which the authors are listed varies with field and culture. Often the order of the authors denotes their contribution to the work. For example, in some fields the first author has made the largest contribution, the remaining authors are listed in descending order of effort. On the other hand, in many fields the senior author (the one responsible for overseeing the project) is listed last, and the person that did most of the day-to-day work on the project is listed first. A third possibility, no longer very common, is to list the authors alphabetically.

When possible, it is advisable for researchers to pick a format for their name that they will continue to use throughout their professional life. Thus, John A. Jones, John Jones, and J. A. Jones are each sufficiently different that it might be difficult for someone to know that each name refers to the same author. Changing ones name because of marriage provides a particular challenge in this regard. Some individuals choose not to change their professional name under these circumstances.

Addresses

Purpose: To indicate the institution(s) at which the research was performed, and to provide readers with a way of contacting the authors.

Content: There are three types of addresses listed on a paper:

1. The *main addresses* listed under the names of the authors indicate where the work was performed.

2. The *corresponding address* is usually listed as a footnote to the list of authors. This indicates to the publisher and, later, to the readers, where the "corresponding author" can be contacted regarding any questions related to the article. Recently, many authors have been including multiple forms of "addresses," including regular mailing addresses, fax numbers, and email addresses.

3. *Current addresses* are often listed for any authors who have moved to a different institution between the time that the research was performed and the manuscript is published.

Note that if an author has moved since completing the research, it is inappropriate to list their current address as the main address. One of the criteria for evaluating an institution is the nature of the work that is done under their auspices, and thus an accurate indication of this credit should be provided.

Style: If the research was conducted at more than one department or institution, then superscripts should be used to indicate which author worked in which location.

Abstract

Purpose: To provide a brief summary of the paper. Along with the title, this is one of the most important components of a research article. After reading the title, researchers commonly scan the abstract to determine what the authors found, and based on this information they often decide if they will read the rest of the paper.

Content: The abstract is written as a mini-paper, i.e., it contains the following information in this order:

1. *introduction*: a few sentences to provide background information on the problem investigated
2. *Methods*: techniques used
3. *Results*: the major results presented in the paper; provide quantitative information when possible.
4. *Discussion*: the authors' interpretation of the results presented
5. *Final summary*: the major conclusions and "big picture" implications. Note that this is the most important part of the abstract, as researchers will often read this part of the abstract first, to see if the implications of the study are important enough to warrant the reading of the full paper.

In most cases, *abbreviations* are not permitted in an abstract.

Searchable databases and on-line journals now make it relatively easy to obtain titles and abstracts of scientific publications. One of the implications of this is that it can no longer be assumed that only specialists in ones discipline will read the abstract. Indeed, the readership can now be assumed to include policy makers, and both advocates and opponents of the research performed. Thus, authors are advised to take this into account when writing their abstract by (1) making it as intelligible as possible to a general readership, and (2) taking extra care to avoid stating things that might be misconstrued by readers who are uninformed and/or unsupportive of the type of research described.

Tense: The abstract should be written in the past tense for the authors' work, present tense for general knowledge and other researchers' work.

Length: The number of words permitted in the abstract is usually specified in a journal's *Instructions to Authors*. Commonly it is 150-300 words.

Introduction

Purpose: To provide the reader with background on the research described in the paper.

Content: The introduction consists primarily of the following types of information, generally provided in this order:

1. *Why the study was undertaken*: What gap in the knowledge of the field were the authors trying to fill by undertaking this study? What problem were the authors trying to address?

2. *The nature of the work performed*: The variables that were investigated and the methods that were used.

3. *The state of the problem at the end of the study*: A brief statement of the major findings presented in the paper, and implications of the study for example, how the work contributes to "the big picture," questions left unanswered, new questions that have emerged.

Note: whereas the information from #1 and #2 are essential components of an introduction, some individuals believe that the information from #3 is optional. We strongly recommend including this information as it helps the reader to evaluate more accurately what they are reading in the sections that follow. This point helps to highlight one of the major distinctions between scientific writing and many other types of prose writing *in scientific writing everything is done to avoid mystery or suspense*.

It is essential that the Introduction should provide a brief but scholarly review of the relevant literature with appropriate references (see "References," below). Authors should neither unduly inflate their contributions nor obscure controversy by ignoring papers that have already been published.

Abbreviations should generally not be used in the title or abstract, and only sparingly in other parts of the manuscript. A rule of thumb is not to use an abbreviation unless the word is used at least 10 times or is best recognized in its abbreviated form. When an abbreviation is to be employed, it should be defined the first time it is used and then always used thereafter. For example,

> *The questionnaire was given to 100 employees of the Environmental Protection Agency (EPA), as well as 50 former EPA employees.*

Most journals will provide a list of abbreviations that do not need to be defined because they are commonly used. Consult the *Instructions for Authors*.

Tense: What the authors did/found is in the past tense; everything else is in the present tense.

Length: Generally not more than 500 words.

Methods

Purpose: The methods section serves two functions: (1) to enable readers to evaluate the work performed (e.g., did the authors use the most appropriate and accurate techniques for their work?), and (2) to permit readers to replicate the study if they desire to do so.

Content: All aspects of the methodology used in the study must be described thoroughly enough so that scientists working in that field would be able to replicate the work. This includes both what was used as well as what was done. Thus, sometimes this section is referred to as "*Methods and Materials*."

If the method has already been published in the scientific literature (whether or not it was written by the same authors), readers should be referred to the original description for the details of the method. However, it is important to include enough information so that readers are able to evaluate the work being presented without having to refer to another publication. This means specifying the critical variables for that type of work, for example, how long the samples were incubated, how many minutes subjects were allowed to work on a task, or what strain of laboratory rats were used. In addition, it is essential to indicate any deviations from the method cited.

Often the company (including city and state) that manufactures a particular reagent is specified to reduce any ambiguity about what was used; likewise, the model number for a piece of equipment is often indicated.

Tense: Past tense.

Style: If several different procedures are described, it is preferable to subdivide the methods section via the use of headings. This enables readers to refer quickly back to a specific method.

Results

Purpose: To provide the data collected.

Content: The contents of the Results section depend on type of article:

In ***full-length research articles***, the more common type of article, only data - what was observed - is included in this section. Interpretations should be reserved for the discussion section. The idea behind this format is to "let the data speak for themselves." However, some authors like to include some introductory or transition material to help the flow of this section.

In ***short research articles*** (sometimes called "short" or "brief" communications), results and discussion (interpretation) are sometimes mixed. Refer to the journals' *Instructions to Authors* for guidance.

Tense: Authors' results should be in past tense, and general statements in present tense.

Style: If the results of several different experiments are described, it is preferable to subdivide this section via the use of headings.

Length: This usually is the shortest section of a manuscript.

Discussion

Purpose: To provide the reader with a plausible interpretation of the data reported and to relate these findings to what other investigators have found.

Content: The section provides the following information, generally in this order:

1. *Summary of conclusions*: what the authors conclude from their data, for example, relationships between variables, trends, etc.

2. *Relation to other results*: the relation of these findings to previous work, e.g. "supports the findings of Alvarez et al., (1994)" or "is contrast to …"

3. *Aberrant results:* any abnormalities or exceptions inherent in the data or in relation to with respect to the scientific literature, and if possible, explanations for these aberrations. (Note: item #3 and #4 may be intermixed.)

4. *Implications:* theoretical or practical implications of the work, i.e., "the big picture"

5. *Grand summary:* a summary of the results and conclusions reported in the paper

Tense: Current knowledge is stated in present tense, the author's work is stated in past tense.

Style: If headings were used in the results section, it is very convenient for the reader if the relevant portion of the discussion is presented under the same headings.

Length: generally, up to 1500 words.

Acknowledgments

Purpose: To recognize and thank those individuals and organizations whose contributions to the work presented should be acknowledged but are not extensive enough to merit authorship.

Content: When applicable, the following information is presented in this order:

1. Individuals other than authors who made a significant contribution to the research by donating important reagents or materials, collecting data, providing extensive advice on drafts of the manuscript, etc. Typically, the nature of the contribution is noted; for example

> *The authors thank Dr. Marcia Jones for providing the genetically modified mice used in these studies, and Mr.*

David Wendall for his assistance in analyzing the tissue samples.

2. If the work has been presented at a conference, then this is often noted. For example,

Portions of this work were presented at the 25th Annual Society for Neuroscience Meeting, November 11-16, 1996, San Diego, CA.

3. Organizations that funded the research. The general format for this information is

This work was supported by U.S. Public Health Service Grant MH43947.

Note that it is ***essential*** to get permission from any individual whose help is acknowledged. Also, many scientific societies and journals are indicating that it is essential to disclose any financial support that has been provided for the work.

Length: Limit to significant contributors.

References

Purpose: To provide the full citation for article referenced in the text.

Content: A complete reference includes all of the authors' names, the title of the article, the journal name, the volume number, page numbers, and the year of publication.

Style: A wide range of styles is used for citing references in the text and bibliography. Check the journal's *Instructions to Authors* for information about the content and formatting of references.

Within the text, articles are cited by providing the author and year of the article (e.g., Fischer and Zigmond, 1996). When there are more than two authors, the first author is provided together with *e. al.* (e.g., Fischer et al..., 1996). If more than one reference is cited for a given point, they are usually listed in chronological order (e.g., Zigmond and Fischer, 1995; Fischer et al., 1996). If there is any ambiguity, a letter can be added to the year of publication (e.g., Fischer et al., 1995a; 1995b).

At the end of the paper a list of references, or bibliography, is provided. This list must be limited to the references cited within the text and most often is provided in alphabetical order.

In some cases citations appear in the text as numbers, usually a superscript, which then refer to a particular item in the reference list.

It is the obligation of the authors to provide a scholarly listing of the primary references of relevance to the paper. Authors are obliged to do a thorough review of the key areas of the scientific literature as part this process. In general, original *research* articles rather than *review* articles should be cited, and the research articles should be the earliest ones that made the particular finding.

It is essential that authors check each reference that they cite. Simply copying a reference from the bibliography of a published paper is inadequate since errors in referencing are very common. In checking a reference, authors must not only make sure that the citation is accurate but also that the text actually supports the point for which it being used as a reference.

References that are not readily available or are in a language not understood by the author present a particular challenge. In the former case, most libraries provide a service that enables authors to obtain papers from a wide range of other libraries. An alternative that is sometimes available is to contact directly the author of the article in question and request a copy.

Articles in foreign languages sometimes provide enough information in their tables and figures to permit an accurate comprehension of their results, even if the language itself is not understood. In this regard it is helpful that a number of scientific terms are the same in English as in many other languages. Alternatively, it usually is possible to have an article translated by a local service.

If a reference cannot be checked by the author, the only alternative to not citing it is to cite it as a secondary reference (e.g., Hooke, 1665, as cited in Fischer, 1995).

If citations are needed for more than one point in a sentence, it is helpful for the reader if the citations appear throughout the sentence, rather than as a collection at the end. For example,

Previous studies have shown that this compound can exist in a solid (Wang and Beauford, 1993), liquid (Jones et al., 1992), or gaseous (Diaz, 1995) state.

Length: Ideally, a paper will list all the references necessary to document each point that is made by the authors. In practical terms, however, most journals will impose a limit in order to conserve space. A rule of thumb is no more than 6 references for a particular point and no more than 100 references per paper.

Tables and Figures

Purpose: To report data that are too numerous or complicated to be described adequately in the text; to reveal trends or patterns in the data.

Content: Tables; possible figures include graphs (bar, line, scatter), diagrams, cartoons (i.e., chemical structures or mechanisms), and photographs.

Style: Figures are usually in black and white. Color is extremely expensive to publish, and should only be used when it provides unique information. (Note: For further details see "Construction of Tables and Figures," which will be posted soon.)

Number: Limit the number of tables and figures to those that provide essential information that could not adequately be presented in text.

Table and Figure Legends

Purpose: To provide a knowledgeable reader with the information required for understanding the table or figure.

Content: The composition of a legend depends on the item it refers to. It should provide information regarding the conditions of the experiment, but not give a summary or interpretation of the results. In addition, statistical information is often provided. This may include

1. The number of times an experiment was performed or a condition was tested.
2. What the values in the table or figure represent, for example *mean S.E.M.* (standard error of the mean)
3. The statistical test used in analyzing the data
4. Whether the test was "one-tailed" or "two-tailed" (if relevant)
5. The *p* value that was used in determining significance
6. If an asterisk or other mark is used in the table or graph to denote statistically significant results, then this mark should be defined.

For example, the statistics portion of a figure legend might look like this,

> *n=5 for each condition. Values represent mean S.E.M. Data were analyzed using a one-tailed Student's t-test. * denotes significance, p 0.05.*

Tense: Past tense.

Style: Each table and figure should be understandable on its own, without reference to the text.

Within a manuscript, the placement of the legend varies depending on whether it refers to a table or figure:

Table: The title, table, and legend should appear on the same page, in the order listed.

Figure: Each figure should appear on a separate page. The numbered legends are listed one after another (i.e., several to a page). The title for a figure comprises the first sentence in the figure legend.

Bibliography of Resources

There are many useful books about writing research articles. We have listed below a selection of the best, those that we believe deserve a place in every technical writer's library.

1. Briscoe, M.H. *Preparing Scientific Illustrations: A Guide to Better Posters, Presentations, and Publications, 2nd Edition.* New York: Springer, 1996.

   We cannot overstate the usefulness of this book. If you have questions about constructing tables and figures, the answer is probably in here.

2. Council of Biology Editors. *Scientific Style and Format: The CBE Manual for Authors, Editors, and Publishers, 6th Edition.* NY: Cambridge University Press, 1994.

   Excellent reference text.

3. Woodford FP (1999) *How to Teach Scientific Communication.* Reston, VA: Council of Biology Editors

An outstanding guide to how to teach writing. Includes a number of "before and after" examples.

4. Day, R. A. *How to Write and Publish a Scientific Paper, 5th Edition.* Phoenix: Oryx Press, 1998.

There are a great many books on the subject; this is the best — wise and witty, takes you from creating the title to checking the galley proofs. Robert Day has been our teacher and the inspiration for these workshops.

7.2.2 Journal-Style Scientific Writing

7.2.2.1 Overview

A critical aspect of the scientific process is the reporting of new results in scientific journals in order to disseminate that information to the larger community of scientists. Communication of your results contributes to the pool of knowledge within your discipline (and others!) and very often provides information that helps others interpret their own experimental results. Most journals accept papers for publication only after peer review by a small group of scientists who work in the same field and who recommend the paper be published (usually with some revision).

The format and structure presented here is a general one; the various scientific journals, and oftentimes specific disciplines, utilize slightly different formats and/or writing styles. Mastery of the format presented here will enable you to adapt easily to most journal- or discipline-specific formats. While these guidelines are a *necessary* tool of learning the scientific writing style and format, they are not *sufficient,* by themselves, to make you an accomplished writer. This guide will not teach you how to write in the English language, i.e., it is not a grammar book. You, the writer, must **practice *writing* and *thinking*** within this structure, ***and***, learn by example from the writings of others; learning the nuances of this style and format will be enhanced as you **read the scientific literature - pay attention to how professional scientists write about their work**. You *will* see improvement in your own scientific writing skills by repeatedly practicing reading, writing, and critiquing of other's writing.

7.2.2.2 The four major aspects of writing journal-style scientific papers

(1) Fundamental style considerations;

(2) A suggested strategy for efficiently writing up research results;

(3) The nuts and bolts of format and content of each section of a paper (part of learning to write a scientific paper is learning how to follow instructions precisely), and,

(4) Basic information regarding peer critiques of scientific writing.

ALL journals have a set of instructions for authors which explicitly state how their paper should be formatted for submission. Consider this guide to be your instructions when writing lab reports for Bio 201, 270, and s42. We encourage you to follow the directions carefully and to make full use of this guide as you prepare your papers. Please ask for help if you have questions about format, style, or content. Above all, remember to write with **precision**, **clarity**, and **economy**.

7.2.2.3 Getting Started

The first task to accomplish as you begin the process of writing is to order and organize the information you wish to present. Some people work well from an outline, others do not. Some people write first to discover the points, then rearrange them using an after-the-fact outline. Whatever process you may use, be aware that scientific writing requires special attention to order and organization. Because the paper will be divided into sections, you need to know what information will go into each. If you don't normally work from an outline, this may be an occasion when you'll at least want to develop a list of the major points to be included in each section, before you begin to write. If the paper has multiple authors, then this is a good time to work (and negotiate!) with your collaborators to insure that all the points the group wants to make get listed.

Audience: Who will be reading your paper? Usually you will be writing to your peers. Simple advice: address your paper to another interested student, or lab group, in this course or major, and assume they have *at least* the same knowledge and expertise base as you. Knowing your audience helps you to decide what information to include--you would write a very different article for a narrow, highly technical, disciplinary journal vs. one that went out to a broad range of disciplines. Similarly, you would write a paper for an audience of other subject majors very differently than one you would write for a cross section of the college. **Do not** write your paper specifically for your instructor.

7.2.2.4 Prose

Your writing should be in complete sentences and easily understood. It should conform to the conventions of standard written English (sentence form, grammar, spelling, etc.). Your ideas will have little impact, no matter how good the research, if they are not communicated well. Remember always that scientific terminology very often has precise meaning. Be certain you choose your words correctly and wisely.

It is important to write clearly and concisely. **Make sure that every paragraph has a clear topic sentence and that the paragraph content supports the topic.** The goal is to report your findings and conclusions clearly, and with as few words as necessary. Your audience (other scientists usually) are not interested in flowery prose, they want to know your findings. **Remember:** Writing and thinking are closely linked enterprises - many people have noted that, "*fuzzy writing reflects fuzzy thinking.*" When people have difficulty translating their ideas into words, they generally do not know the materials as well as they think.

7.2.2.5 Style Considerations

Be clear and concise: Write briefly and to the point. Say what you mean **clearly** and avoid embellishment with unnecessary words or phrases. **Brevity** is very important. Use of the active voice alone shortens sentence length considerably.

Precise word use is critical: Scientific terminology carries specific meaning - learn to use it *appropriately* and use it *consistently*. A critical function of technical terminology is to say a lot with a few words, i.e., ***economy***. This applies as well to appropriate *acronyms* (e.g., PCR) and *abbreviations*. Direct your paper toward the average reader in your intended audience. If writing for a highly technical journal, you will necessarily use the technical jargon. If writing for a general science audience you would limit the jargon.

Some things to avoid:

> You do not have to try to impress people by using words most people have never heard of. Many published articles are like this, and they are poor papers on account of it.
> *Do not* use colloquial speech, slang, or "childish" words or phrases.
> *Do not* use contractions: **for example**, "*don't*" must be "*do not*" and *"isn't*" must be "*is not*" etc.

Abbreviations: Do not use abbreviations in the text *except* for units of measure. Always abbreviate these when using them with data (2 mm; 10 min.). Except for temperature units (F,C, K) never abbreviate units of measure when using them in a non-data context (e.g., "we measured length in millimeters"; "time was recorded in minutes"; "temperature was measured in F (or C)"; "100 years have passed since Mendel did..."). A list of common abbreviations and conversions is provided.

Use Past Tense: Research papers reflect work that has been completed, therefore use the past tense throughout your paper (including the Introduction) when referring to the *actual work* that you did, including statements about your expectations or hypotheses. Use the past tense, as well, when referring to the work of others that you may cite.

First vs. Third Person: If there is one stylistic area where scientific disciplines and journals vary widely, it is the use of first vs. third person constructions. Some disciplines and their journals have moved away from a very strict adherence to the third person construction, and permit limited use of the first person in published papers. Other disciplines, especially the biomedical fields, still prefer the third person constrcution. Limit your use of first person construction (i.e., " *I (or we) undertook this study* ....): usually it is most acceptable in the Introduction and Discussion sections, and then only to a limited extent. Use first person in the methods *sparingly* if at all, and avoid its use in the results.

Use Active Verbs: Use active verbs whenever possible; writing that overly uses passive verbs (is, was, has, have, had) is deadly to read and almost always results in more words than necessary to say the same thing.

ACTIVE: "*the mouse consumed oxygen at a higher rate*..."
PASSIVE: "*oxygen was consumed by the mouse at a higher rate*.."

The clarity and effectiveness of your writing will improve dramatically as you increase the use of the active voice.

Other specific comments on style are also included for each section of the paper.
Remember: precise word use, past tense, active voice, brevity.

References References to the research findings of others are an integral component of any research paper. The usual practice is to summarize the finding or other information in your own words and then cite the source. Any ideas or other information that are not your own must be substantiated by a reference that is cited in the text. **As a rule, in research papers, direct quotation and footnoting are not practiced - simply restate the author's ideas or findings in your own words and provide a citation.**

Plagiarism (use of others words, ideas, images, etc. without citation) is not to be tolerated and can be easily avoided by adequately referencing any and all information you use from other sources. In the strictest sense, plagiarism is representation of the work of others as being *your* work. Paraphrasing other's words too closely may be construed as plagiarism in some circumstances. In journal style papers there is virtually no circumstance in which the findings of someone else cannot be expressed in your own words with a proper citation of the source. If you are unclear about what constitutes plagiarism, please confer with your instructor.

7.2.3 The Structure, Format, Content, and Style of a Journal-Style Scientific Paper

Why a Scientific Format?

The scientific format may seem confusing for the beginning science writer due to its rigid structure which is so different from writing in the humanities. One reason for using this format is that it is a means of efficiently communicating scientific findings to the broad community of scientists in a uniform manner. Another reason, perhaps more important than the first, is that this format allows the paper to be read at several different levels. For example, many people skim Titles to find out what information is available on a subject. Others may read only titles and Abstracts. Those wanting to go deeper may look at the Tables and Figures in the Results, and so on. The take home point here is that the scientific format helps to insure that at whatever level a person reads your paper (beyond title skimming), they will likely get the key results and conclusions.

The Sections of the Paper

Most journal-style scientific papers are subdivided into the following sections: Title, Authors and Affiliation, Abstract, Introduction, Methods, Results, Discussion, Acknowledgments, and Literature Cited, which parallel the experimental process. This is the system we will use. This website describes the style, content, and format associated with each section.

The sections appear in a journal style paper in the following prescribed order:

Experimental process	Section of Paper
What did I do in a nutshell?	Abstract
What is the problem?	Introduction
How did I solve the problem?	Materials and Methods
What did I find out?	Results
What does it mean?	Discussion
Who helped me out?	Acknowledgments (optional)
Whose work did I refer to?	Literature Cited
Extra Information	Appendices (optional)

Section Headings:

Main Section Headings: Each main section of the paper begins with a heading which should be ***capitalized***, ***centered*** at the beginning of the section, and ***double spaced*** from the lines above and below. **Do not underline the section heading OR put a colon at the end.**

Example of a main section heading:

INTRODUCTION

Subheadings: When your paper reports on more than one experiment, use subheadings to help organize the presentation. Subheadings should be ***capitalized*** (first letter in each word), ***left justified,*** and either ***bold italics*** OR ***underlined***.

Example of a subheading:

Effects of Light Intensity on the Rate of Electron Transport

Title, Authors' Names, and Institutional Affiliations

1. **Function**: Your paper should begin with a **Title** that succinctly describes the *contents* of the paper. Use descriptive words that you would associate strongly with the content of your paper: the molecule studied, the organism used or studied, the treatment, the

location of a field site, the response measured, etc. A majority of readers will find your paper via electronic database searches and those search engines key on words found in the title.

2. **Format**:

> The **title** should be centered at the top of page 1 (DO NOT use a title page - it is a waste of paper for our purposes); the title is NOT underlined or italicized.
> the **authors' names** (PI or primary author first) and **institutional affiliation** are *double-spaced from and centered below* the title. When more then two authors, the names are separated by commas except for the last which is separated from the previous name by the word "and".

For example:

Ducks Over-Winter in Colorado Barley Fields in Response to Increased Daily Mean Temperature

Ima Mallard, Ura Drake, and Woodruff Ducque
Department of Informatics, University of Colorado - Boulder

The title is not a section, but it is necessary and important. The title should be short and unambiguous, yet be an adequate description of the work. A general rule-of-thumb is that the title should contain the **key words describing the work** presented. Remember that the title becomes the basis for most on-line computer searches - if your title is insufficient, few people will find or read your paper. For example, in a paper reporting on an experiment involving dosing mice with the sex hormone estrogen and watching for a certain kind of courtship behavior, *a poor title would be:*

Mouse Behavior

Why? It is very general, and could be referring to any of a number of mouse behaviors. *A better title would be*:

The Effects of Estrogen on the Nose-Twitch Courtship Behavior in Mice

Why? Because the key words identify a specific behavior, a modifying agent, and the experimental organism. If possible, give the key result of the study in the title, as seen in the first example. Similarly, the above title could be restated as:

Estrogen Stimulates Intensity of Nose-Twitch Courtship Behavior in Mice

3. **Strategy for Writing Title.**

ABSTRACT

1. **Function**: An abstract summarizes, in one paragraph (usually), the major aspects of the entire paper in the following prescribed sequence:

the ***question(s) you investigated*** (or purpose), (**from Introduction**)
- state the purpose very clearly in the first or second sentence.

the ***experimental design*** *and* ***methods*** used, (**from Methods**)
- clearly express the basic design of the study.
- Name or briefly describe the basic methodology used without going into excessive detail-be sure to indicate the key techniques used.

the ***major findings*** including ***key quantitative results***, or ***trends*** (**from Results**)
- report those results which answer the questions you were asking
- identify trends, relative change or differences, etc.

a brief summary of your ***interpetations*** and ***conclusions***. (from **Discussion**)
- clearly state the implications of the answers your results gave you.

Whereas the Title can only make the simplest statement about the content of your article, the Abstract allows you to elaborate more on each major aspect of the paper. The length of your Abstract should be kept to about 200-300 words maximum (a typical standard length for journals.) Limit your statements concerning each segment of the paper (i.e. purpose, methods, results, etc.) to two or three sentences, if possible. The Abstract helps readers decide whether they want to read the rest of the paper, or it may be the only part they can obtain via electronic literature searches or in published abstracts. Therefore, enough key information (e.g., summary results, observations, trends, etc.) must be included to make the Abstract useful to someone who may to reference your work.

How do you know when you have enough information in your Abstract? A simple rule-of-thumb is to imagine that you are another researcher doing an study similar to the one you are reporting. If your Abstract was the only part of the paper you could access, would you be happy with the information presented there?

2. **Style**: The Abstract is ONLY text. Use the active voice when possible, but much of it may require passive constructions. Write your Abstract using concise, but complete, sentences, and get to the point quickly. **Use past tense**. Maximum length should be 200-300 words, usually in a single paragraph.

The Abstract **SHOULD NOT** contain:

lengthy background information,
references to other literature,

elliptical (i.e., ending with ...) or incomplete sentences,
abbreviations or terms that may be confusing to readers,
any sort of illustration, figure, or table, or references to them.

3. **Strategy**: Although it is the first section of your paper, the Abstract, by definition, must be written last since it will summarize the paper. To begin composing your Abstract, take whole sentences or key phrases from each section and put them in a sequence which summarizes the paper. Then set about revising or adding words to make it all cohesive and clear. As you become more proficient you will most likely compose the Abstract from scratch.

4. **Check your work**: *Once you have the completed abstract, check to make sure that the information in the abstract completely agrees with what is written in the paper. Confirm that* ***all*** *the information appearing the abstract actually appears in the body of the paper.*

INTRODUCTION

1. **Function**: The function of the Introduction is to:

Establish the context of the work being reported. This is accomplished by discussing the relevant primary research literature (with citations) and summarizing our current understanding of the problem you are investigating;
State the purpose of the work in the form of the hypothesis, question, or problem you investigated; and,
Briefly explain your rationale and approach and, whenever possible, the possible outcomes your study can reveal.

Quite literally, the Introduction must answer the questions, "*What was I studying*? *Why was it an important question? What did we know about it before I did this study? How will this study advance our knowledge?*"

2. **Style**: Use the active voice as much as possible. Some use of first person is okay, but do not overdo it.

3. **Structure**: The structure of the Introduction can be thought of as an inverted triangle - the broadest part at the top representing the most general information and focusing down to the specific problem you studied. Organize the information to present the more general aspects of the topic early in the Introduction, then narrow toward the more specific topical information that provides context, finally arriving at your statement of purpose and rationale. A good way to get on track is to sketch out the Introduction *backwards*; start with the specific purpose and then decide what is the scientific context in which you are asking the question(s) your study addresses. Once the scientific context is decided, then you'll have a good sense of what level and type of general information with which the Introduction should begin.

Here is the information should flow in your Introduction:

Begin your Introduction by clearly identifying the subject area of interest. Do this by using *key words* from your Title in the first few sentences of the Introduction to get it focused directly on topic at the appropriate level. This insures that you get to the primary subject matter quickly without losing focus, or discussing information that is too general. For example, in the mouse behavior paper, the words *hormones* and *behavior* would likely appear within the first one or two sentences of the Introduction.

Establish the *context* by providing a brief and balanced review of the pertinent published literature that is available on the subject. The key is to summarize (for the reader) what we knew about the specific problem *before* you did your experiments or studies. This is accomplished with a general review of the *primary research literature* (with citations) but should not include very specific, lengthy explanations that you will probably discuss in greater detail later in the Discussion. The judgment of what is general or specific is difficult at first, but with practice and reading of the scientific literature you will develop e firmer sense of your audience. In the mouse behavior paper, for example, you would begin the Introduction at the level of mating behavior in general, then quickly focus to mouse mating behaviors and then hormonal regulation of behavior. Lead the reader to your statement of purpose/hypothesis by focusing your literature review from the more general context (the big picture e.g., hormonal modulation of behaviors) to the more specific topic of interest to you (e.g., role/effects of reproductive hormones, especially estrogen, in modulating specific sexual behaviors of mice.)

What literature should you look for in your review of what we know about the problem? Focus your efforts on the *primary research journals* - the journals that publish original research articles. Although you may read some general background references (encyclopedias, textbooks, lab manuals, style manuals, etc.) to get yourself acquianted with the subject area, do not cite these, becasue they contain information that is considered fundamental or "common" knowledge wqithin the discipline. Cite, instead, articles that reported specific results relevant to your study. Learn, as soon as possible, how to find the *primary literature* (research journals) and *review articles* rather than depending on reference books. The articles listed in the Literature Cited of relevant papers you find are a good starting point to move *backwards* in a line of inquiry. Most academic libraries support the **Citation Index** - an index which is useful for tracking a line of inquiry *forward* in time. Some of the newer search engines will actually send you alerts of new papers that cite particular articles of interest to you. *Review articles* are particularly useful because they summarize all the research done on a narrow subject area over a brief period of time (a year to a few years in most cases).

Be sure to clearly state the purpose and /or hypothesis that you investigated. When you are first learning to write in this format it is okay, and actually preferable, to use a pat statement like, "The purpose of this study was to...." or "We investigated three possible mechanisms to explain the ... (1) blah, blah..(2) etc. It is most usual to place the statement of purpose near the end of the Introduction, often as the topic sentence of the final paragraph. It is not necessary (or even desirable) to use the words "hypothesis" or "null hypothesis", since these are usually implicit if you clearly state your purpose and expectations.

Provide a clear statement of the rationale for your approach to the problem studied. For example: State briefly how you approached the problem (e.g., you studied oxidative respiration pathways in isolated mitochondria of cauliflower). This will usually follow your statement of purpose in the last paragraph of the Introduction. Why did you choose this kind of experiment or experimental design? What are the scientific merits of this particular *model* system? What advantages does it confer in answering the particular question(s) you are posing? Do not discuss here the actual *techniques* or *protocols* used in your study (this will be done in the Materials and Methods); your readers will be quite familiar with the usual techniques and approaches used in your field. If you are using a *novel* (new, revolutionary, never used before) technique or methodology, the merits of the new technique/method versus the previously used methods *should be* presented in the Introduction.

MATERIALS AND METHODS

This section is variously called **Methods** or **Methods and Materials**.

1. **Function**: In this section you explain *clearly* how you carried out your study in the following *general* structure and organization (details follow below):

the **subjects used** (plant, animal, human, etc.) and their pre-experiment handling and care, and when and where the study was carried out (if location and time are important factors);
if a field study, a **description of the study site**, including the physical and biological features, and precise location;
the **experimental OR sampling design** (i.e., how the experiment or study was structured. For example, controls, treatments, the variable(s) measured, how many samples were collected, replication, etc.);
the **protocol for collecting data**, i.e., how the experimental procedures were carried out, and,
how the data were analyzed (statistical procedures used).

Organize your presentation so your reader will understand the logical flow of the experiment(s); **subheadings** work well for this purpose. Each experiment or procedure should be presented as a unit, even if it was broken up over time. In general, provide enough quantitative detail (how much, how long, when, etc.) about your experimental protocol such that other scientists could reproduce your experiments. You should also indicate the statistical procedures used to analyze your results, including the probability level at which you determined significance (usually at 0.05 probability).

2. **Style**: The style in this section should read as if you were verbally describing the conduct of the experiment. You may use the active voice to a certain extent, although this section requires more use of third person, passive constructions than others. Avoid use of the first person in this section. Remember to use the **past tense** throughout. The Methods section ***is not*** a step-by-step, directive, protocol as you might see in your lab manual.

Describe the organism(s) used in the study. This includes giving the *source* (supplier or *where* and *how* collected), *size*, *how they were handled* before the experiment, what they were fed, etc. In genetics studies include the strains or genetic stocks used.

Describe the site where your field study was conducted. The description must include both *physical* and *biological* characteristics of the site pertinant to the study aims. Include the date(s) of the study (e.g., 10-15 April 1994) and the exact location of the study area. Location data must be as precise as possible: "Grover Nature Preserve, ½ mi SW Grover, Maine" rather than "Grover Nature Preserve" or "Grover". When possible, give the actual latitude and longitude position of the site (the WWW has sites which provide this service). It is most often a good idea to include a **map** (labeled as a Figure) showing the location in relation to some larger more recognizable geographic area. Someone else should be able to go to the exact location of your study if they want to repeat or check your work, or just visit your study area.

> **NOTE: For laboratory studies you should *not* report the date and location of the study *UNLESS* it is relevant. Most often it is *not*.**

Describe your experimental design clearly. Be sure to include the *hypotheses* you tested, *controls*, *treatments*, *variables* measured, how many *replicates* you had, what you actually *measured*, what form the *data* take, etc. Always identify treatments by the variable or treatment name, NOT by an ambiguous, generic name or number (e.g., use "2.5% saline" rather than "test 1".) When your paper includes more than one experiment, use subheadings to help organize your presentation by experiment. A general experimental design worksheet is available to help plan your experiments in the core courses.

Describe the protocol for your study in sufficient detail that other scientists could repeat your work to verify your findings. Foremost in your description should be the "quantitative" aspects of your study - the masses, volumes, incubation times, concentrations, etc., that another scientist needs in order to duplicate your experiment. When using standard lab or field methods and instrumentation, it is not always necessary to explain the procedures (e.g., serial dilution) or equipment used (e.g., autopipetter) since other scientists will likely be familiar with them already. You may want to identify certain types of equipment by brand or category (e.g., ultracentrifuge vs. prep centrifuge). It is appropriate to give the source for reagents used parenthetically, e.g., "....poly-l-Lysine (Sigma #1309)." When using a method described in another published source, you can save time and words by referring to it and providing the relevant citation to the source. Always make sure to describe any modifications you have made of a standard or published method.

Describe how the data were summarized and analyzed. Here you will indicate what types of data summaries and analyses were employed to answer each of the questions or hypotheses tested.

The information should include:

how the data were **summarized** (Means, percent, etc) and how you are reporting **measures of variability** (SD,SEM, etc)

- this lets you avoid having to repeatedly indicate you are using mean ± SD.

data transformation (e.g., to normalize or equalize variances)

statistical tests used with reference to the particular questions they address, e.g.,

"A Paired t-test was used to compare mean flight duration before and after applying stablizers to the glider's wings."

"One way ANOVA was used to compare mean weight gain in weight-matched calves fed the three different rations."

any other **numerical** or **graphical techniques** used to analyze the data

Here is some additional advice on particular problems common to new scientific writers.

Problem*: The Methods section is prone to being wordy or overly detailed.*

Avoid repeatedly using a single sentence to relate a single action; this results in very lengthy, wordy passages. A related sequence of actions can be combined into one sentence to improve clarity and readability:

Problematic Example: This is a very long and wordy description of a common, simple procedure. It is characterized by single actions per sentence and lots of unnecessary details.

"The petri dish was placed on the turntable. The lid was then raised slightly. An inoculating loop was used to transfer culture to the agar surface. The turntable was rotated 90 degrees by hand. The loop was moved lightly back and forth over the agar to spread the culture. The bacteria were then incubated at 37 C for 24 hr."

Improved Example: Same actions, but all the important information is given in a single, concise sentence. Note that superfluous detail and otherwise obvious information has been deleted while important missing information was added.

"Each plate was placed on a turntable and streaked at opposing angles with fresh overnight E. coli culture using an inoculating loop. The bacteria were then incubated at 37 C for 24 hr."

Best: Here the author assumes the reader has basic knowledge of microbiological techniques and has deleted other superfluous information. The two sentences have been combined because they are related actions.

"Each plate was streaked with fresh overnight E. coli culture and incubated at 37 C for 24 hr."

> **Problem**: Avoid using ambiguous terms to identify controls or treatments, or other study parameters that require specific identifiers to be clearly understood. Designators such as Tube 1, Tube 2, or Site 1 and Site 2 are completely meaningless out of context and difficult to follow in context.

Problematic example: In this example the reader will have no clue as to what the various tubes represent without having to constantly refer back to some previous point in the Methods.

"A Spec 20 was used to measure A_{600} of Tubes 1,2, and 3 immediately after chloroplasts were added (Time 0) and every 2 min. thereafter until the DCIP was completely reduced. Tube 4's A_{600} was measured only at Time 0 and at the end of the experiment."

Improved example: Notice how the substitution (in red) of treatment and control identifiers clarifies the passage both in the context of the paper, and if taken out of context.

"A Spec 20 was used to measure A_{600} of the reaction mixtures exposed to light intensities of 1500, 750, and 350 uE/m2/sec immediately after chloroplasts were added (Time 0) and every 2 min. thereafter until the DCIP was completely reduced. The A_{600} of the no light control was measured only at Time 0 and at the end of the experiment."

RESULTS

1. **Function**: The function of the Results section is to objectively present your key results, *without* interpretation, in an orderly and logical sequence using both illustrative materials (Tables and Figures) and text. Summaries of the statistical analyses may appear either in the text (usually parenthetically) or in the relevant Tables or Figures (in the legend or as footnotes to the Table or Figure). The Results section should be organized around a series of Tables and/or Figures sequenced to present your key findings in a logical order. The text of the Results section follows this sequence and highlights the answers to the questions/hypotheses you investigated. Important negative results should be reported, too. Authors usually write the text of the results section based upon this sequence of Tables and Figures.

2. **Style**: Write the text of the Results section concisely and objectively. The passive voice will likely dominate here, but use the active voice as much as possible. Use the **past tense**. Avoid repetitive paragraph structures. Do not interpret the data here. The transition into interpretive language can be a slippery slope. Consider the following two examples:

> This example highlights the trend/difference that the author wants the reader to focus:

The duration of exposure to running water had a pronounced effect on cumulative seed germination percentages (Fig. 2). Seeds exposed to the 2-day treatment had the highest cumulative germination (84%), 1.25 times that of the 12-h or 5-day groups and 4 times that of controls.

> In contrast, this example strays subtly into interpretation by referring to optimality (a conceptual model) and tieing the observed result to that idea:

The results of the germination experiment (Fig. 2) suggest that the optimal time for running-water treatment is 2 days. This group showed the highest cumulative germination (84%), with longer (5 d) or shorter (12 h) exposures producing smaller gains in germination when compared to the control group.

Things to consider as you write your Results section:

What are the "results"?: When you pose a testable hypothesis that can be answered experimentally, or ask a question that can be answered by collecting samples, you accumulate observations about those organisms or phenomena. Those observations are then analyzed to yield an answer to the question. In general, the answer is the " key result".

The above statements apply regardless of the complexity of the analysis you employ. So, in an introductory course your analysis may consist of visual inspection of figures and simple calculations of means and standard deviations; in a later course you may be expected to apply and interpret a variety of statistical tests. You instructor will tell you the level of analysis that is expected.

For example, **suppose you asked the question**, "*Is the average height of male students the same as female students in a pool of Biology majors*?" You would first collect height data from large random samples of male and female students. You would then calculate the descriptive statistics for those samples (mean, SD, n, range, etc) and plot these numbers. In a course where statistical tests are not employed, you would visually inspect these plots. Suppose you found that male Biology majors are, on average, 12.5 cm taller than female majors; this is the answer to the question.

> Notice that the outcome of a statistical analysis is not a key result, but rather an analytical *tool* that helps us understand *what is* our key result.

Organize the results section based on the sequence of Table and Figures you'll include. Prepare the Tables and Figures as soon as all the data are analyzed and arrange them in the sequence that best presents your findings in a logical way. A good strategy is to note, on a draft of each Table or Figure, the one or two key results you want to addess in the text portion of the Results. Simple rules to follow related to Tables and Figures:

Tables and Figures are assigned numbers separately and in the sequence that you will refer to them from the text.

- The first Table you refer to is Table 1, the next Table 2 and so forth.
- Similarly, the first Figure is Figure 1, the next Figure 2, etc.

Each Table or Figure must include a brief description of the results being presented and other necessary information in a legend.

- **Table legends go above the Table**; tables are read from top to bottom.
- **Figure legends go below the figure**; figures are usually viewed from bottom to top.

When referring to a Figure *from the text*, "Figure" is abbreviated as Fig.,e.g., **Fig. 1**. Table is never abbreviated, e.g., **Table 1**.

The body of the Results section is a text-based presentation of the key findings which includes references to each of the Tables and Figures. The text should guide the reader through your results stressing the key results which provide the answers to the question(s) investigated. A major function of the text is to provide clarifying information. You must refer to each Table and/or Figure individually and in sequence, and clearly indicate for the reader the key results that each conveys. Key results depend on your questions, they might include obvious trends, important differences, similarities, correlations, maximums, minimums, etc.

Some things to avoid:

Do not reiterate each value from a Figure or Table - only the key result or trends that each conveys.
Do not present the same data in both a Table and Figure - this is considered redundant and a waste of space and energy. Decide which format best shows the result and go with it.
Do not report raw data values when they can be summarized as means, percents, etc.

Statistical test summaries (test name, *p*-value) are usually reported parenthetically in conjunction with the biological results they support. Always report your results with parenthetical reference to the statistical conclusion that supports your finding (if statistical tests are being used in your course). This parenthetical reference should include the statistical test used and the level of significance (test statistic and DF are optional). For example, if you found that the mean height of male Biology majors was significantly larger than that of female Biology majors, you might report this result (in blue) and your statistical conclusion (shown in red) as follows:

"Males (180.5 ± 5.1 cm; n=34) averaged 12.5 cm taller than females (168 ± 7.6 cm; n=34) in the AY 1995 pool of Biology majors (two-sample t-test, t = 5.78, 33 d.f., $p < 0.001$)."

Note that the report of the key result (shown in blue) would be identical in a paper written for a course in which statistical testing is not employed - the section shown in red would simply not appear.

- Avoid devoting whole sentences to report a statistical outcome alone.
- Two notes about the use of the word ***significant(ly)***.
    - In scientific studies, the use of this word implies that a statistical test was employed to make a decision about the data; in this case the test indicated a larger difference in mean heights than you would expect to get by chance alone. Limit the use of the word "significant" to this purpose only.
    - If your parenthetical statistical information includes a p-value that is significant, it is unncecssary (and redundant) to use the word "significant" in the body of the sentence (see example above).

Present the results of your experiment(s) in a sequence that will logically support (or provide evidence against) the hypothesis, or answer the question, stated in the Introduction. For example, in reporting a study of the effect of an experimental diet on the skeletal mass of the rat, consider first giving the data on skeletal mass for the rats fed the *control* diet and then give the data for the rats fed the *experimental* diet.

Report *negative* results - they are important! If you did not get the anticipated results, it may mean your hypothesis was incorrect and needs to be reformulated, or perhaps you have stumbled onto something unexpected that warrants further study. In either case, your results may be of importance to others even though they did not support your hypothesis. Do not fall into the trap of thinking that results contrary to what you expected are necessarily "bad data". If you carried out the work well, they are simply your results and need interpretation. Many important discoveries can be traced to "bad data".

Always enter the appropriate units when reporting data or summary statistics.

- for an ***individual value*** you would write, "the mean length was 10 m", or, "the maximum time was 140 min."
- When including a measure of variability, place the unit *after* the error value, e.g., "...was 10 ± 2.3 m".

Likewise place the unit after the last in a ***series of numbers*** all having the same unit. For example: "lengths of 5, 10, 15, and 20 m", or "no differences were observed after 2, 4, 6, or 8 min. of incubation".

Discussion

1. Function: The function of the Discussion is to interpret your results in light of what was already known about the subject of the investigation, and to explain our new understanding of the problem after taking your results into consideration. The Discussion will always connect to the Introduction by way of the question(s) or hypotheses you posed and the literature you cited, but it does not simply repeat or rearrange the Introduction. Instead, it tells how your study has moved us forward from the place you left us at the end of the Introduction.

Fundamental questions to answer here include:

Do your results provide answers to your testable hypotheses? If so, how do you interpret your findings?
Do your findings agree with what others have shown? If not, do they suggest an alternative explanation or perhaps a unforseen design flaw in your experiment (or theirs?)
Given your conclusions, what is our new understanding of the problem you investigated and outlined in the Introduction?
If warranted, what would be the next step in your study, e.g., what experiments would you do next?

2. Style: Use the active voice whenever possible in this section. Watch out for wordy phrases; be concise and make your points clearly. Use of the first person is okay, but too much use of the first person may actually distract the reader from the main points.

3. Approach: Organize the Discussion to address each of the experiments or studies for which you presented results; discuss each in the same sequence as presented in the Results, providing your interpretation of what they mean in the larger context of the problem. Do not waste entire sentences restating your results; if you need to remind the reader of the result to be discussed, use "bridge sentences" that relate the result to the interpretation:

"The slow response of the lead-exposed neurons relative to controls suggests that...[interpretation]".

You will necessarily make reference to the findings of others in order to support your interpretations.Use subheadings, if need be, to help organize your presentation. Be wary of mistaking the reiteration of a result for an interpretation, and make sure that no new results are presented here that rightly belong in the results.

You must relate your work to the findings of other studies - including previous studies you may have done and those of other investigators. As stated previously, you may find crucial information in someone else's study that helps you interpret your own data, or perhaps you will be able to reinterpret others' findings in light of yours. In either case you should discuss reasons for similarities and differences between yours and others' findings. Consider how the results of other studies may be combined with yours to derive a new or perhaps better substantiated understanding of the problem. Be sure to state the conclusions that can be drawn from your results in light of these considerations. You may also choose to briefly mention further studies you would do to clarify your working hypotheses. Make sure to reference any outside sources as shown in the Introduction section.

Do not introduce new results in the Discussion. Although you might occasionally include in this section tables and figures which help explain something you are discussing, they must not contain new data (from your study) that should have been presented earlier. They might be flow diagrams, accumulation of data from the literature, or something that shows how one type of data leads to or correlates with another, etc. For example, if you were studying a membrane-bound transport channel and you discovered a new bit of information about its mechanism, you might present a diagram showing how your findings helps to explain the channel's mechanism.

Acknowledgments (included as needed)

If, in your experiment, you received any significant help in thinking up, designing, or carrying out the work, or received materials from someone who did you a favor by supplying them, you must acknowledge their assistance and the service or material provided. Authors *always* acknowledge **outside reviewers** of their drafts (in PI courses, this would be done *only* if an instructor or other individual critiqued the draft prior to evaluation) and any **sources of funding** that supported the research. Although usual style requirements (e.g., 1st person, objectivity) are relaxed somewhat here, Acknowledgments are always brief and never flowery.

Place the **Acknowledgments** between the Discussion and the Literature Cited.

Literature Cited

1. Function: The Literature Cited section gives an alphabetical listing (by first author's last name) of the references that you actually cited in the body of your paper. Instructions for writing full citations for various sources are given in on separate page. A complete format list for virtually all types of publication may be found in Huth and others(1994)**.**

NOTE: ***Do not*** label this section "Bibliography". A bibliography contains references that you may have read but have not specifically cited in the text. Bibliography sections are found in books and other literary writing, but not scientific journal-style papers.

Appendices

Function: An Appendix contains information that is non-essential to understanding of the paper, but may present information that further clarifies a point without burdening the body of the presentation. An appendix is an *optional* part of the paper, and is only rarely found in published papers.

Headings: Each Appendix should be identified by a Roman numeral in sequence, e.g., Appendix I, Appendix II, etc. Each appendix should contain different material.

Some examples of material that might be put in an appendix (not an exhaustive list):

raw data
maps (foldout type especially)
extra photographs
explanation of formulas, either already known ones, or especially if you have "invented" some statistical or other mathematical procedures for data analysis.
specialized computer programs for a particular procedure
full generic names of chemicals or compounds that you have referred to in somewhat abbreviated fashion or by some common name in the text of your paper.
diagrams of specialized apparati.

Figures and Tables in Appendices

Figures and Tables are often found in an appendix. These should be formatted as discussed previously, but are numbered in a separate sequence from those found in the body of the paper. So, the first Figure in the appendix would be Figure 1, the first Table would be Table 1, and so forth. In situations when multiple appendices are used, the Table and Figure numbering must indicate the appendix number as well.

7.2.4 Critical Reading Towards Critical Writing

Critical writing depends on critical reading. Most of the papers you write will involve reflection on written texts - the thinking and research that has already been done on your subject. In order to write your own analysis of this subject, you will need to do careful critical reading of sources and to use them critically to make your own argument. The judgments and interpretations you make of the texts you read are the first steps towards formulating your own approach.

Critical Reading: What is It?

To read critically is to make judgments about **how** a text is argued. This is a highly reflective skill requiring you to "stand back" and gain some distance from the text you are reading. (You might have to read a text through once to get a basic grasp of content before you launch into an intensive critical reading.) THE KEY IS THIS:

don't read looking only or primarily for **information**
do read looking for **ways of thinking** about the subject matter

When you are reading, highlighting, or taking notes, avoid extracting and compiling lists of evidence, lists of facts and examples. Avoid approaching a text by asking "What information can I get out of it?" Rather ask "How does this text work? How is it argued? How is the evidence (the facts, examples, etc.) used and interpreted? How does the text reach its conclusions?

How Do I Read Looking for Ways of Thinking?

1. First determine the **central claims** or **purpose** of the text (its thesis). A critical reading attempts to assess how these central claims are developed or argued.
2. Begin to make some judgements about **context** . What audience is the text written for? Who is it in dialogue with? (This will probably be other scholars or authors with differing viewpoints.) In what historical context is it written? All these matters of context can contribute to your assessment of what is going on in a text.
3. Distinguish the **kinds of reasoning** the text employs. What concepts are defined and used? Does the text appeal to a theory or theories? Is any specific methodology laid out? If there is an appeal to a particular concept, theory, or method, how is that concept, theory, or method then used to organize and interpret the data? You might also examine how the text is organized: how has the author analyzed (broken down) the material? Be aware that different disciplines (i.e. history, sociology, philosophy, biology) will have different ways of arguing.
4. Examine the **evidence** (the supporting facts, examples, etc) the text employs. Supporting evidence is indispensable to an argument. Having worked through Steps 1-3, you are now in a position to grasp how the evidence is used to develop the argument and its controlling claims and concepts. Steps 1-3 allow you to see evidence in its context. Consider the kinds of evidence that are used. What counts as evidence in this argument? Is the evidence statistical? literary? historical? etc. From what sources is the evidence taken? Are these sources primary or secondary?
5. Critical reading may involve **evaluation**. Your reading of a text is already critical if it accounts for and makes a series of judgments about how a text is argued. However, some essays may also require you to assess the strengths and weaknesses of an argument. If the argument is strong, why? Could it be better or differently supported? Are there gaps, leaps, or inconsistencies in the argument? Is the method of analysis problematic? Could the evidence be interpreted differently? Are the conclusions warranted by the evidence presented? What are the unargued assumptions? Are they problematic? What might an opposing argument be?

Some Practical Tips

1. Critical reading occurs after some preliminary processes of reading. Begin by skimming research materials, especially introductions and conclusions, in order to strategically choose where to focus your critical efforts.
2. When highlighting a text or taking notes from it, teach yourself to highlight argument: those places in a text where an author explains her analytical moves, the concepts she uses, how she uses them, how she arrives at conclusions. Don't let yourself foreground and isolate facts and examples, no matter how interesting they may be. First, look for the large patterns that give purpose, order, and meaning to those examples. The opening sentences of paragraphs can be important to this task.
3. When you begin to think about how you might use a portion of a text in the argument you are forging in your own paper, try to remain aware of how this portion fits into the whole argument from which it is taken. Paying attention to context is a fundamental critical move.
4. When you quote directly from a source, use the quotation critically. This means that you should not substitute the quotation for your own articulation of a point. Rather, introduce the quotation by laying out the judgments you are making about it, and the reasons why you are using it. Often a quotation is followed by some further analysis.
5. Critical reading skills are also critical listening skills. In your lectures, listen not only for information but also for ways of thinking (see the Health Sciences Writing Centre's link on Making Notes from Lectures). Your instructor will often explicate and model ways of thinking appropriate to a discipline.

A Strategy for Writing Up Research Results

Get Organized: **Lists, Outlines, Notecards, etc.** Before starting to write the paper, take the time to think about and develop a list of points to be made in the paper. As you progress, use whichever strategy works for you to begin to order and to organize those points and ideas into sections.

A. Balanced Review of the Primary Research Literature: Do an in-depth, balanced review of the primary research literature relevant to your study prior to designing and carrying out the experiments. This review will help you learn what is known about the topic you are investigating and may let you avoid unnecessarily repeating work done by others. This literature will form the basis of your Introduction and Discussion. Training in *on-line searches* is available through the librarians and is included in Biology 201. Do your search early enough to take advantage of the *Interlibrary Loan System* if need be.

B. Write the Introduction: Once your hypothesis has been refined for testing, you will draft the Introduction to your paper. In PI courses you will bring the Introduction to lab the day of the experiment for critique by an instructor or TWA (Technical Writing Assistant).

C. Design and Conduct the Experiment: Keep careful notes on procedures used during the experiment . You should write the Materials and Methods section upon completion of the experiment.

D Analyze and Interpret the Results: Once the data are collected, you must analyze and interpret the results. Analysis will include data summaries (e.g., calculating means and variances) and statistical tests to verify conclusions. Most scientists lay out their Tables and Figures upon completion of the data analysis before writing the Results section. Write the Table and Figure legends. It is good practice to note the one or two key results that each Table or Figure conveys and use this information as a basis for writing the Results section. Sequence and number the Tables and Figures in the order which best enables the reader to reach your conclusions.

E. Write the Results Section: Remember that the Results section has both text and illustrative materials (Tables and Figures). Use the text component to guide the reader through your key results, i.e., those results which answer the question(s) you investigated. Each Table and Figure must be referenced in the text portion of the results, and you must tell the reader what the key result(s) is that each Table or Figure conveys.

F. Write the Discussion: Interpretation of your results includes discussing how your results modify and fit in with what we previously understood about the problem. Review the literature again at this time. After completing the experiments you will have much greater insight into the subject, and by going through some of the literature again, information that seemed trivial before, or was overlooked, may tie something together and therefore prove very important to your own interpretation. Be sure to cite the works that you refer to.

G. Write the Abstract and Title: The Abstract is always the last section written because it is a concise summary of the entire paper and should include a clear statement of your aims, a brief description of the methods, the key findings, and your interpretation of the key results. The Title will probably be written earlier, but is often modified once the the final form of the paper clearly known.

H. Self-Revise Your Paper: Most authors revise their papers *at least* 2-3x before giving it out for peer review. Go back over your paper now and read it carefully; **read it aloud**. Does it say what you wanted it to say? Do any ideas, experiments, or interpretations need to be moved around within the text to enhance the logical flow of your arguments? Can you shorten long sentences to clarify them?

Can you change passive verbs to active forms? Do the Tables and Figures have sufficient information to stand alone outside the context of the paper? Use your dictionary to correct spelling and your spell checker to catch typos.

I. Peer Review: Have knowledgable colleagues critique your paper. Use their comments to revise your paper yet again.

J. Prepare the Final Draft: Carefully proof-read your final draft to make sure its as well done as possible. Double check that you've properly cited all your sources in the text and in the Literature Cited. Check the formatting one last time. The instructors LOVE to give full credit for format issues whenever possible, but will not hesitate to take points off for sloppy work.

7.2.5 Revising Your Research Paper

Self-Revision by the Author(s)

Revision of your writing is an on-going process from the time you begin until the final copy is submitted. A strategy that works for many people is to write out an initial draft in total without substantial revision and then let it sit for a day. Come back to it then and begin revising your paper working from a ***global perspective*** (overall organization) to ***paragraph content and organization*** and finally down to ***sentence level line editing***. Implicit in these instructions is the assumption that you are checking the content for scientific correctness and accuracy.

GLOBAL

- check the sequence of ideas/background/content in each section for logical progression (*your topic sentences should do this*).
- check for a strong relationship of ideas between the Introduction (*what we knew before our study*) and the Discussion (*how our study changes or supports our previous understanding*).

PARAGRAPH

- check that each paragraph has a **coherent topic sentence**, most often as the lead sentence.
- in each paragraph do the **other sentences support the topic sentence**?
- check the **transitions between paragraphs** to ensure they are *logical* and *smooth.*

LINE EDITING

- check for *consistent* and *correct* use of **terminology.**
- can you change a passive verb construction to an **active verb**?
- eliminate **superfluous lead phrases** (*Once that was done,* ..).
- remove all **colloquial language.**
- check for **redundancy** (i.e., places where you repeat what you have said elsewhere).
- read each senetence closely for **clarity** and **brevity. Can you say the same thing with fewer words?**
- **READ THE PAPER ALOUD** to find those quirky sentences that you wrote while still half asleep - if doesn't sound correct when spoken aloud, it will read even more oddly.

MISCELLANEOUS

- check that all of your **sources are cited correctly** in the text.
- check the **numbering sequence of your tables and figures.**
- check the **Literature Cited** for completeness and correct format.
- check the **line spacing between headings and text**, and Tables and Figures and text.
- check the **page breaks** to make sure you do not split tables or figures.
- are the **authors' names** spelled correctly?
- run spell check on the document to find **typographical errors** and read carefully for **spelling** and **grammatical errors**.
- check your main headings and subheadings for proper case and placement.

Revision After Peer Critique

After reading carefully the comments and suggestions to improve your paper, discuss them with the reviewer (when possible) to get clarification or to argue your point, if you should disagree. In general, you will make the changes *as suggested* by the reviewer unless you have good, and justifiable, reasons not to.

Once you are clear on the changes to be made, approach the revision using the same *global*, *paragraph*, *line editing* strategy.

Make the global changes first and recheck the items listed previously.
Make the paragraph level changes and recheck list.
Make the line edit changes and recheck the list.
Recheck the miscelleaneous items.

Final Revision

If possible, have your reviewer examine the paper again (cookies help!) one last time. For PI courses, this is the opportunity for co-authors to check the final draft to make sure it satisfies their expectations. If all the changes have been made to everyone's satisfaction, make one last check of overall appearance of the document to catch recalcitrant page breaks, etc.

7.2.6 Presenting Your Research Paper Effectively

Introduction to Effective Presentations

An essential aspect of any research project is dissemination of the findings arising from the study. The most common ways to make others aware of your work is by publishing the results in a journal article, or by giving an oral or poster presentation (often at a regional or national meeting). While efforts are made to teach the elements of writing a journal article in many graduate school curricula, much less attention is paid to teaching those skills necessary to develop a good oral or poster presentation - even though these arguably are the most common and most rapid ways to disseminate new findings.

7.2.6.1 Developing an Effective Oral Presentation

Initial Planning

This is where you begin to tailor the talk to the situation, and for that reason this stage is very important for a successful presentation. Talk to your host and clarify these points before you spend much preparation time. If the environment and audience are unfamiliar to you, this is a critical stage. You may even want to do a literature search on potential audience members to identify areas of common interest or potential questions which may arise. Begin this stage early - the more lead time you allow yourself, the more time you will have to think up novel approaches to the topic and the more interesting and substantial your presentation will be.

Before you begin preparing the presentation, you'll need to determine:

1. The type of talk you will be expected to give
    - will this be an informal chat, a seminar discussion, or a more formal presentation?
    - different talks have different purposes; the intent of a conference presentation is not the same as a job talk. When in doubt, ask for guidance from your host.
2. The composition of the audience
    - will you be speaking to a general audience or specialists?
    - how many people are expected to attend?
    - is this likely to be a friendly audience? An interactive audience?
3. The time allotted for the talk
    - the longer the talk, the more freedom you will have to explore the topic
    - a short talk needs to be very clear and to address the topic directly
    - is question time included?
4. Expectations for information content
    - is there a specific purpose for having you give a talk? Clarify the expectations beforehand and plan to address them during the presentation.
    - will you be presenting novel concepts to this audience, or building upon their prior knowledge? Either way, make sure you cover the basics clearly, and early in the talk, to avoid loosing the audience.

Preparation

Once you have a general idea of what you want to say, you'll have to decide how to say it. Unlike a conversation or a written document, a talk is a one-shot attempt to make a point. By contrast, a conversation consists of repetitions and clarification's based on questions and immediate feedback, while a written paper allows a reader to puzzle through its contents as often as necessary. It is essential that your talk be well-constructed and tidy, and that your points be presented to the audience both a logical sequence and unambiguously. This all takes a fair amount of preparation. **Start early!**

Here are a few pointers to get you started:

1. Start preparing far in advance by thinking through what needs to be said. Collect material which may relate to the topic from unusual sources, and sleep on these ideas. The final product will be more fully-developed and interesting.
2. Using big letters and a bold pen, write a clear statement of the problem and its importance, and then pin that statement on the wall above your desk.
3. Develop this theme into one jargon-free sentence that will catch the attention of the audience. Next, identify the issues you plan to address (brainstorm, then trim back; see the portion of this tutorial on outlining).
4. Arrange these issues in a logical sequence (which may change as you develop the talk). This process is easier if you use index cards to organize your talk, with one idea per card.
5. Computer-based presentation programs (PowerPoint, Persuasion, etc.) can be wonderful time-savers. The time invested in learning to use these programs is rewarded by the speed with which a presentation can be created, even by a moderately-skilled user. These programs are good tools for organizing your presentation (an electronic version of the index cards idea), they can be used to create visuals for the presentation (e.g., slides and transparencies), and even project those visuals during the presentation.
6. Avoid using lists (First ..., Second ...); you may confuse listing systems (First ..., Point B..., and another thing ...), or you may discover later in the talk that you've missed a point entirely, and then you'll be forced to backtrack. Both of these problems tend to distract your audience away from the points you are trying to make, and both give the appearance of poor organizational skills.
7. Retention of information by the audience is reduced as a talk proceeds, so if you do want to make a series of points, organize them from the most to the least important. That way, the audience is more likely to remember the important points later. You may even find that the less important points become irrelevant to the focus of the talk as you practice.
8. Determine transition elements which will help your audience to follow the link from one issue to the next. These should be logical, and may presented by posing a question, or explaining your own discovery of the link's existence.
9. Use short sentences with simple constructions. The concept will be made more clear, and the sentence structure is more similar to conversational styles.
10. Run through the talk once, early. Go back and re-think the sequencing. Discard non-essential elements.
11. Don't assume the audience will be familiar with basic concepts that form the foundation of your talk. Outline these concepts briefly but clearly early in the talk to avoid confusion.
12. Attempt to identify problems or questions the audience may have and address them in the talk, before the audience has a chance to think of these things themselves.
13. Determine which elements would benefit by being presented with visual aids. Spend time working out the best way to present the material. Head on over to the accompanying tutorials for information on presenting material in an effective way using visual aids.

14. Prepare thumbnails sketches of these visual aids, then run through the talk again. Re-work the most appropriate and essential visual aids and discard the rest. Don't forget to proof-read your visuals! Do so while there is plenty of time to re-print that critical slide with the glaring typo.
15. The earlier you start on the visuals, the better they will be. On the other hand, avoid fine tuning each visual endlessly; if you find yourself diddling the details, go on to do something more productive instead.
16. When in doubt about which presentation medium to use (transparencies, slides, videos, multimedia, etc.), choose the format which is the least complex which remains consistent with both clarity and content of the presentation. Keep in mind that the more technology you use, the more things there will be which can go wrong. These technological difficulties may develop into a gruesome presentation experience, particularly if you are giving the talk in an unfamiliar setting!
17. If you do need to use multimedia technology in your presentation, call ahead to make sure the technology you require is supported in the room where you'll be talking!
18. **The most important preparation factor is to REHEARSE!** Do so in private at first. Then for a real acid test, videotape yourself and watch the results with a critical eye. It's often a painful and humbling experience, but the results will be worth it.
19. You can then try the presentation out in front of a few colleagues. Ask for feedback, then *act on that information*. Select those who know a little about your topic, and not those who know a lot. This will focus your attention on attempting to explain why you did what you did in simple terms, rather than encouraging attention to details only specialists care about.
20. If you start preparing early, you'll have plenty of time to refine the presentation based on your colleagues' feedback. This is always a useful process.
21. Don't waste your colleagues' time; if you are sincere about wanting that feedback, don't wait until the night before the presentation to ask for other people's input.
22. Remember, the shorter the talk, the more difficult it will be to cover the material clearly and completely. Be strict about including only what is essential information for the presentation, and removing *all* the non-essential tidbits.

Outlining

The primary purpose of a presentation is to provide information which the audience will then remember at a later date. Detailed referencing of material or extensive review of data won't be remembered - and may put the audience to sleep!

One way to maintain interest is to organize and present the material in a novel manner. Using a non-standard ordering of material will help to keep the audience interested. Similarly, organizing your material in a new way (rather than re-working an old talk) will help to keep your own interest in the topic, and will result in a talk which is more fresh and exciting.

The importance of outlining is often stressed in preparing written and oral presentations, but an outline following a linear format (headings, subheadings, etc.) may be restrictive. A list of terms and ideas can be daunting, and tends to focus attention on the final items. Consider creating an 'idea network' as an alternative approach for organizing your presentation (or your written paper).

A question that often is asked at this stage is "Is this enough material for the talk?". Actually, you'll probably find that you have far too much material. It is important to develop a realistic view of how much material is appropriate, and the ability to be ruthless in eliminating non-essential material. These abilities vary widely among presenters, and are important factors in determining the quality of the presentation. Here are a few guidelines for helping determine how much material is enough . . .

Important Elements

Keeping these elements in mind as you prepare and practice the presentation will reduce the amount of re-working you'll have to do as it evolves, and will result in a more streamlined and effective end product.

1. **Rate:** The optimal rate for a scientific talk is about 100 words per minute. Any faster and the audience can't absorb the additional information. Use pauses, and repeat critical information.
2. **Opening:** The opening should catch the interest and attention of the audience immediately, while avoiding trite filler phrases (Thank you for having me . . .) and technical jargon.
3. **Transitions:** The link between successive elements of the talk should be planned carefully, smooth, and logical. You should make the relation between successive elements clear to the audience.
4. **Conclusion:** Summarize the main concepts you've discussed, and how your work relates to issues you've raised. Aim to help your audience achieve high retention of this final information. Signal that the summary is beginning ("In summary, ..."), but don't begin the summary too soon or else the audience will start to leave before you finish!
5. **Length:** *Don't run over!* **Ever!** Shorten your talk by removing details, concepts, and information, not by eliminating words. If it becomes absolutely essential to supply details, supplement your presentation with a handout. Make about 10% more handouts than you think you'll need. Always leave time for a few questions at the end of the talk.

Remember that there is no point in giving a presentation if the audience isn't listening. You should make a big effort to help them be interested in what you have to say. It therefore is appropriate to use techniques to retain audience interest, provided these techniques don't detract from the content or professionalism of the talk.

Practice makes perfect

You've probably heard this before, but that doesn't diminish its importance. **Practice is the single most important factor contributing to a good presentation.** No matter how rushed you might be, make time for at least a few practice runs. The effects of practice will be apparent, and a poorly presented talk reflects upon both you and your attitude towards the material and audience. Don't be fooled by people who claim to be able to throw together a talk at a moment's notice. Generally, their talks fall into two categories - talks which are disjointed and awkward, and talks which have had the rough edges removed by numerous prior presentations (i.e., dull and unexciting).

One problem is that you can waste a tremendous amount of time by practicing all the wrong parts of your talk. It is necessary to run through the talk a few times to get an idea of how the talk will flow. After that, seek some outside feedback to make sure you are on the right track. Finally, practice all parts of the talk equally. If you always start at the beginning and work until you run into problems, the beginning of the talk will be great, but the final portion of the talk will be relatively more weak. Begin one out of every few practice runs in the middle or at the end of the talk.

Yes, that means running through the talk once or twice isn't enough, particularly if the material is new to you. If the presentation is important, treat it that way. **Practice.** My own rule of thumb is a minimum of 10 practice runs for any one presentation. This can be a big commitment of time, but consider what's riding on a successful job talk . . .

OK, so here are a few hints on how to manage the practice talks, and how to deal with making and using notes.

Presentation

Having spent all that time preparing the talk itself, there are still a few things you can do at the last minute which will help ensure a successful presentation. Or, if you are the nervous type, help fill time . . .

1. Before the day begins, or last thing the night before, run through your talk once more. Use a mirror or visualize standing in front of an audience as you practice. If you've brought a slide carousel with you (a good idea), check their arrangement. You probably won't have time to do this later. Remember to seat the locking ring properly!
2. If possible, take a tour the room you'll use for the presentation early in the day. Look for potential problems with line of sight due to furniture, dark spots due to dead overhead lights, intruding sound from ventilation - these all can be fixed with a bit of prior warning and a polite request.
3. If you need specialized equipment, make sure it is available ahead of time - don't spring that information on your host at the last minute.
4. Check again to see that your slides are oriented properly in the carousel. Lock that ring!

5. Make sure the focus switch works, and determine who will be controlling the slide advance. Do the slide advance, reverse, and focus features all work?
6. It's your show, so ask for help with the equipment if you need it; it's better to ask for help then fumble around during the presentation. Determine who will be controlling equipment for you.
7. Computer presentations introduce a whole host of potential issues - here are a few to consider:
    - Is the host software compatible with your presentation? Are the fonts, bullets, colors, etc. the same?
    - Is there a sound card in the host computer? Is the sound system operational - but not too painfully loud?
    - Back-up your presentation before you leave using an alternate medium, then bring it with you separately from the one you plan to use (e.g., packed in a different suitcase), or e-mail it to yourself as an attachment - you may be able to access it from your destination if needed.
    - Alternatively, e-mail it to your host and ask that her or she download the file and test it on the computer you'll be using - BEFORE you depart for the trip!
    - Did you include all the required files and resources for your presentation?
    - You might consider making a set of 35mm presentation slides from your electronic presentation, then bringing the stack of slides along as your fail-safe backup - this strategy may depend upon your degree of compulsiveness and/or paranoia, or how important the presentation is to you.
    - Keep in mind that failures of technology can be devestating, but that 1) the embrassment is greater on the part of the host if their equipment is at fault, and 2) the host is usually impressed if you provide an alternate solution to the problem - suggesting you are a proactive and prepared person ....
8. Irrespective of what your presentation medium might be, letting your presentation slides, disk, CD, etc. out of your sight before the presentation begins can lead to disaster.
9. If the room is large, or your voice small, use a microphone. Try it out before the audience arrives (blowing into the mike or counting '1-2-3' they have arrived is tacky, so don't do it).
10. Check to see that accessories are present; chalk, eraser, markers, and especially a pointer. If it is a laser pointer, does it have fresh batteries loaded? Keep in mind that green wavelength lasers DEVOUR batteries!
11. Avoid standing behind a lectern or desk during the presentation. Stand to one side of the projection screen or blackboard, and closer to the audience if possible.
12. Moderate movement and hand gestures are OK, but avoid pacing and flapping.
13. Don't be afraid to insist on a few minutes to yourself prior to the talk; 15 to 30 minutes is standard. If you have an itinerary, check to see that you've had time allotted for preparation. If you are running behind, see if someone is willing to meet with you after the talk, then use that time to prepare instead. Use this time to double-check your materials, and your introductory and summary statements. Don't allow yourself to be distracted by audience members coming up to chat.

14. Don't wait until the very last minute to make that run to the bathroom, and remember to check carefully your appearance - including zippers, buttons and other closures - before you reappear.

The Moment of Truth

Gulp. So you are sitting there, about to be introduced. Now what?

1. Take several deep breaths as you are being introduced (but don't sigh!). Visualize your rehearsed opening statement; don't improvise at the last moment.
2. State your objectives at start of your talk, then restate them again at the end of the talk. In between, discuss how your material relates to these objectives.
3. Unless you intentionally have had experience as a stand-up comic, avoid making jokes. The results can be disappointing, and may suggest an unprofessional attitude.
4. Choose a natural, moderate rate of speech and use automatic, moderate gestures.
5. Monitor your behavior, and avoid habitual behaviors (pacing, fumbling change in pocket, twirling hair).
6. Laser pointers are wonderful pointing devices, but remember not to point them at the audience. They are best used by flashing the pointer on and off, so that the place you are indicating is illuminated briefly. Don't swirl the laser around and around one place on the projection screen, or sweep it from place to place across the screen. This is very distracting for the audience, and they will end up watching the pointer and not listening to what you are saying.
7. Likewise, and for the same reasons, avoid using the cursor as in pointer in your computer presentations.
8. Also try to avoid pointing things out on the computer's monitor at the podium - although the presenter's natural tendency is to concentrate on the monitor's screen, the audience will be looking over your shoulder at the projection screen and won't be able to see where your finger is pointing. If you find yourself doing this, power-off or disable the monitor to force yourself to concentrate on the projection screen!
9. Enthusiasm for your topic is contagious, but don't overdo it - you'll alienate the audience.
10. Converse with your audience. Involve them in the process of the presentation by posing questions and making eye contact. Be patient if you ask a question - answers sometimes take time to formulate.
11. Keep an eye on your time, and don't run over your limit. **Ever.**
12. Be prepared for interruptions (late arrivals, cell phones or pagers, burned out projector bulbs, fire drills, *etc.*).
13. If you must turn down the room lights, don't turn them off entirely. Don't leave the lights down any longer than necessary - remember to turn them back up! Of course, the snores from the sleeping audience may remind you to turn the lights back on if you've forgotten.
14. Don't apologize for any aspect of your presentation. This should be your very best effort; if you have to apologize, you haven't done your job properly.

15. Don't criticize aspects of the trip, city, facilities, *etc.* during your talk. This is another way to alienate your audience quickly. For instance, they may or may not have chosen to live in this horrible climate, but it isn't your place to remind them how horrible it is. Remember that you are a guest and it is impolite to exhibit your prejudices publically.
16. Strive to have a prepared and memorable summary. If nothing else, the take home message is what the audience will remember after you leave.
17. When you reach the summary and are about to finish, resist the temptation to add a few last impromptu words. They will be unpracticed, and will be the last thing many of your audience will hear you say. End your talk with the insightful, firm summary statement you have prepared.
18. Don't be afraid to give yourself credit for your own work, but do remember to give others credit where due. I prefer to do this early, others may prefer doing it later in the talk. If planned for later in the talk, don't forget to acknowledge these people's efforts, even if you have to skip a statement or two to remain within your time frame. The best friend of one of these contibutors may be in your audience! If you include slides borrowed from other people, or slides which include other people's data or figures, always give credit to these people right on that slide. This shows a professional attitude, and (better yet) can save you many words of explanation.

Handling Questions

Your presentation doesn't end once you've finished what you have to say. The question period often is the part of the talk which influences the audience the most. After all, you've had time to practice the rest of the talk. This is the part of the presentation where your ability to interact with the audience will be evaluated. Since you can't always predict the what you'll be asked, how can you prepare for the questioning? Here are a few guidelines:

1. Always repeat each question so the entire audience knows what you've been asked.
2. Before you answer, take a moment to reflect on the question. By not rushing to give an answer, you show a degree of respect for the questioner, and you give yourself time to be sure you are answering the question that actually was asked. If you are unsure, restate the question or ask for a clarification.
3. Above all, wait for the questioner to finish asking the question before you begin your answer! The only exception is when it becomes necessary to break in on a vague, rambling question; this is your show, and you have only a limited time to make your presentation. It is essential, however, that you break in tactfully. Say something like "So, are you asking ....?" This will focus the question and give you a place to begin an answer. Remember that your ability to interact with an audience also is being evaluated.
4. If a question is asked during the talk, and it will clarify an ambiguity, answer it immediately.

5. Postpone questions aimed at resolving specific problems (or arcane knowledge) until the end of the talk, or private discussion. This is particularly important if the answer will distract either you or the audience away from the flow of your presentation.
6. Avoid prolonged discussions with one person, extended answers, and especially arguments.
7. If you can't answer a question, just say so. Don't apologize. You then may:
   - Offer to research an answer, then get back to the questioner later.
   - Suggest resources which would help the questioner to address the question themselves.
   - Ask for suggestions from the audience.
8. Finish your answer by asking the person who asked that question whether or not you answered the question sufficiently for them. This acknowledges and thanks the questioner, it lets the rest of the audience feel comfortable asking questions (because it shows you are genuinely interested in addressing audience issues, not just in lecturing to them), and it gives you a chance to more fully answer the question if your first effort was not quite on target. If the questioner says you didn't answer it and you believe you did, either ask them to clarity the question or suggest that the two of you go into more detail at a break or after the presentation. (My thanks to Dr. Bruce Bloom for this very astute suggestion!).

The following information may be presented elsewhere in these tutorials, but are worth considering again here!

The Question & Answer slide.
This is the slide that follows your conclusions and remains in the background as you answer questions from the audience. A very good idea is to carefully select the most important images from your prior slides, shrink them so they all fit on this slide, then arrange them so that they are up for the audience to see as they Q & A session goes on. This will allow them to consider your data and interpretations without having to recall details, it gives them a chance to reconsider the information after you've delivered the 'big question' answer, and it helps you to guide the content of the Q&A session to meet your own ends.

Avoid:

turning off the projector (you'll just have to turn it on again - and wait while it warms up).
projecting a blank white (dazzling) or blank black (too dark) slide
leaving your conclusion slide in place as you answer questions - just not as interesting or as provocative as the strategy I suggest above.

The Acknowledgements slide.
This is an important slide! No one works in a vacumn, nor are many scientists sufficiently

wealthy to fund their own research without outside sources! On the other hand, avoid reading a list of names and agencies as this:

takes time
typically is information of little interest to the audience
may give the appearance of 'name dropping' (*e.g., look at all the important people at important places whith whom I've worked*)
no one will remember names of people they don't know anyway

But, as I said, this is important - particularly if a collaborator's best friend is sitting in the audience! Consider using a collage of photos of your collaborators in their natural settings for this purpose. You may still point out the primary collaborators, but a visual approach is inherently more interesting for the audience. Moreover, at a crowded conference a collaborator is more likely to be recognized in the hallway or at a restaurant if your audience has seen his or her photo during your talk. You should consider carefully the content of the photo, however, if you are presenting it in a professional setting!

Finally, consider presenting the acknowledgement slide at the front end of the talk. You'll need to restrain yourself so you don't spend too much time on this slide, but that seldom is a problem at the beginning of a talk. This is better than not being able to present the information at all if the slide is at the end of the talk and you've run out of time....

7.3 Review and opposition of engineering/scientific research papers

7.3.1 Effective Peer Reviewing

After many months or years of effort, a manuscript is finally prepared and submitted to a scientific journal. The author anticipates the reviewers' decision: accept or reject. If the article wins positive reviews and the paper is accepted for publication, all is well. However, rejection and/or negative reviews are disappointing, sometimes infuriating.
All scientists have experienced these feelings. Therefore, when we are asked to review other scientists' work, it is important to provide useful and objective feedback that is constructive and not personal. We are responsible for sending the message to authors that It's all about advancing science. Based on experience as reviewer and "reviewee", several steps are offered to the potential reviewer to help ensure that a review is completed effectively:

1. Assess the Ability to Review the Manuscript

Is there time to do a good review?

The editorial office generally allows two weeks to a month to complete a review. A good block of time during that period is required to review, because it is difficult to review a little each day. If that time is not available, the review may be rushed and the reviewer may neglect to give the submission the attention it deserves. Is the reviewer familiar enough with the specific field to review the submission? Usually, requests are to review submissions on topics with which the potential reviewer is familiar. If not, the reviewer must invest time to accumulate the information necessary to make a fair judgment about the quality of the manuscript and the science it describes. Is there enough time to research the subject to do it justice?

Is there a conflict of interest for reviewing this submission?
Although it is important to be familiar with the field, it is not fair to accept a request to review if the reviewer cannot be objective about the material presented. Perhaps the author and reviewer are currently working on the same project in a competitive field, or perhaps the author is someone with whom the reviewer and/or the reviewer's lab has had a disagreement.

2. Review the Material Using Appropriate Criteria
Journals usually provide reviewers with specific criteria for evaluating submissions. Although many of the criteria are similar from one journal to another, it is necessary to judge whether the quality of the work reaches the level of the journal's expectation. Some journals prefer a mechanistic analysis to a simple, descriptive one. If the work is a continuation of a study reported by the same author, does the current manuscript contain sufficient new data to justify publication? Is the text clear? Are the data fairly interpreted and not overreaching? Is the quality of figures sufficient to support conclusions? Are there no twists in logic?

3. Provide Constructive Feedback
The reviewer is accountable for providing a response that gives the author enough information to improve both the article and the research. There is almost always useful feedback that can be offered, even for excellent manuscripts that need no major revision. Indeed, this is where the reviewer is best serving the purpose of advancing good science: by providing constructive feedback whether or not the article is being recommended for acceptance. In a typical 1-2 page review, it may be useful to include three sections:

1. Summary (2-3 sentences).
2. Overall evaluation of the research and recommendation (1-2 paragraphs).
3. Details supporting the evaluation (varies).

The review begins with a brief summary of the study stating the new and important finding(s) reported in the manuscript. To assure the author that the reviewer understands the manuscript correctly, the author may wish to phrase this section in his/her own words rather than cutting and pasting sentences from the manuscript. Then, evaluate the overall scientific merit by pointing out the major significance and/or critical problems of the study. Enumerate not only weaknesses but also strengths of the manuscript, which will encourage the author. The recommendation of whether the manuscript is published, is revised before being accepted, or is rejected, is made based on these criteria. If the submission is not being recommended for acceptance, it is important to clearly state in this section why.

If the manuscript is not accepted for publication, provide details to support the evaluation. This is the most important section of the review. Pay particular attention to the following two points. First, clarify what the author could do to improve the work, providing scientific reasons, as concrete as possible, for each part of the manuscript about which there is concern. Second, avoid pushing your own scientific views onto the submission, instead focusing only on evaluating the science of the submitted manuscript. Avoid nitpicking over minor points and language, as long as they don't affect the overall scientific value of the manuscript.

Reviewing articles allows reviewers to learn about new and exciting science. The review should be done impartially without consideration of who produced the manuscript. Comments should be clear, concise, focused on the science, and respectful of the author. Everyone who reviews is also reviewed. The Golden Rule applies: when a review arrives with helpful comments to improve the work, pay the same respect to the next author whose manuscript you review.

7.3.2 Reporting a negative feed back

One of the hardest things about getting started with peer reviewing is dealing with your reluctance to give negative feedback. After all, we're all socialized not to say mean things to people, and purely negative commentary usually doesn't end up helping the writer anyway. Many reviewers consider the job of peer review as unpleasant one. You should find ways to get around this problem by

1) remembering to **give positive commentary** where a writer has done well, and by,

2) **turning negative feedback into productive feedback**.

1. **When reviewing, it is always important to note a paper's strengths, so that the author will not lose these in the process of revision.** Never assume an author will automatically know which parts of a paper work well... remember, they have been immersed in it too long to be objective. The peer review form may ask you to list the three major strengths of the paper - but remember to do this throughout the paper too, writing marginal comments like "good paragraph" when you read a part that flows well.

2. **But how to deal with the parts that really do have problems? The key is to make sure the comments you write are substantive comments.** As we read, we all have reactions to problematic parts of a paper: "Huh? This is unclear"... "Gosh, this is disorganized!"... " What is this person trying to say here???". But these reactions are only the first step in the process of constructing helpful commentary, and writing down these initial reactions as comments is not usually useful to the writer.

How can you turn these unhelpful comments into helpful ones? You need to go a step beyond your initial reaction, and ask yourself *why* you are reacting negatively to that sentence or paragraph. Why, for instance, does a paragraph seem disorganized? Are several topics mixed together in one paragraph? Or is a single topic treated, but presented out of logical sequence, so that the reader is constantly grasping for information not yet given? Or does the writer seem to start with one idea or position, but then reverse him/herself later in the paragraph?

You can see that this process will take some work on your part, because you need to reflect on your reactions and read in a very involved way. Below are some examples of unhelpful “reaction-type” comments that have been turned into helpful comments by this process of reflection.

Example 1:

Unhelpful comment: "This section needs work."

Helpful Comment: Combine the related actions into a single sentence in Methods, eg, "Flies were assigned randomly to 5 treatment groups of 25, and were weighed, sexed, and marked with non-toxic paint before behaviorial trials began"

Example 2:

Unhelpful comment: "Disorganized!"

Helpful Comment: "This section discusses both animal-rearing conditions and experimental methods, but the two are mixed together. Could you separate each into its own paragraph?"

Example 3:

Unhelpful comment: "How are these references relevant?"

Helpful Comment: "The background and references given in poaragraph 2 don't seem directly relevant to our hypothesis. I think we need references on how light has been shown to affect flowering (in sunflower or any species), and less on other factors that promote or inhibit flowering."

Example 4:

Unhelpful comment: "Unclear."

Helpful Comment: "I'm not sure what your interpretation is after these two paragraphs: does the experiment show that mung beans cure cancer, or not? Which are we concluding? If the sample size is too small, we need to discuss that when we suggest future research, but that does not change our results here."

7.3.2

7.4 Writing a good thesis: Research Report Writing

This guide to thesis writing gives simple and practical advice on the problems of getting started, getting organised, dividing the huge task into less formidable pieces and working on those pieces. It also explains the practicalities of surviving the ordeal. It includes a set of suggested structures depending on the nature of your research and a guide to what should go in each section.

Most students find that doing the research for the thesis is the most challenging part of Ph.D. research. They often budget their time to allow a very short period for the actual writing of the thesis.

This plan invariably leads to an unpleasant surprise: writing results in a form that other people can understand is a **very** slow process! Here are some of the often unanticipated reasons:

In order to get a well-written paper, the first 2 or 3 drafts must often be completely discarded!

In the course of your several years of research, you have probably changed notation several times, developed new points of view on your work, and developed many results that looked significant at the time but now seem to contribute nothing toward your final product. Sorting through all of your work and reorganizing it is a lengthy process.

Even if you have several technical reports, conference papers, or journal articles discussing partial results, the audience for your thesis is different, and thus the style of exposition must be significantly changed. A research paper is addressed to a group of experts in the field, who presumably know the literature and the background issues quite well. A thesis is written more for the generalist. A thorough literature review must be included, as well as an evaluation of where your work fits into the scheme of things.

All the small details that were put off and forgotten must now be filled in. Citations must be checked, the historical progression of various results must be carefully documented, the ``trivial cases'' must be worked through, the documentation of your methods must be complete.

Your thesis advisor will probably have strong opinions on how the work should be presented. Adapting your style to these requirements will take some flexibility and thought.

Your committee members, your first detached readers, will often find undefined jargon or symbols, holes in your arguments (or at least in your presentation of them), and other deficiencies.

Even after you are on track, you will probably find that a ``good'' day of writing produces about 5 pages, leading to an overall average of perhaps a quarter page per day.

7.4.1 The art of minimizing the pain of thesis writing

Some habits begun early in your research will help:

> Keep careful notes about your work. You might choose to keep bound logbooks (square ruled paper is helpful) or on-line notes. Write your notes regularly: write up every new result, but make an entry at least weekly even if you believe that nothing of significance has been accomplished. Even noting what you are thinking about can be helpful.
>
> If possible, write up each piece of the work for publication as it is completed. This makes the final writing easier because each piece is documented at its completion time rather than months or years later, and the early write-ups give a basis for organizing the thesis. In addition, it establishes your reputation early and makes the job search much easier.
>
> As you read other theses and published works, be a student of technical writing styles. Find out what works and what doesn't. Study a good writing manual.

A student who has developed skill at writing non-technical term papers as an undergraduate will have an easier time of learning to be a good technical writer, but there is one additional skill that must be added: you must also be a good teacher!

7.4.2 What is a thesis? For whom is it written? How should it be written?

Your thesis is a research report. The report concerns a problem or series of problems in your area of research and it should describe what was known about it previously, what you did towards solving it, what you think your results mean, and where or how further progress in the field can be made.

When you write a term project, you are explaining the work of others. You have a good idea of what is immediately obvious and what is more difficult to grasp, since you recently went through the exercise of grasping the material yourself.

Do not carry over your ideas from undergraduate assessment: a thesis is not an answer to an assignment question. One important difference is this: the reader of an assignment is usually the one who has set it. S/he already knows the answer (or one of the answers), not to mention the background, the literature, the assumptions and theories and the strengths and weaknesses of them. The readers of a thesis do not know what the "answer" is. If the thesis is for a PhD, the university requires that it make an original contribution to human knowledge: your research must discover something hitherto unknown.

Obviously your examiners will read the thesis. They will be experts in the general field of your thesis but, on the exact topic of your thesis, you are the world expert. Keep this in mind: you should write to make the topic clear to a reader who has not spent most of the last three years thinking about it. It is easy to be fooled into thinking that since something is now obvious to you after several years of study it is also obvious to your reader. The most difficult part of thesis writing is organizing and presenting your material in an understandable way.

An important early step is to develop a tentative outline. The outline will probably change several times, but it is important always to have a current one foremost in your mind so that you can make the pieces fit together smoothly.

A typical outline will be of the form:

Chapter 1: Introduction
What is the problem?
Why is it important?
What have other people done?
What is central idea(s) of my approach?
How is the rest of the thesis organized?

Chapter 2: The problem

Define the problem.
Introduce the jargon.
Discuss the basic properties.

Chapter 3: Big idea 1

...

Chapter *k*+2: Big idea *k*

Chapter *k*+3: Conclusion

Recapitulate what has been accomplished.

Discuss ideas for future work.

Don't think that the thesis must be written starting at page 1 and continuing until the end. Most often, the presentation of the ``big ideas'' shapes the presentation of ``the problem.'' The introduction is often written (or at least rewritten) last. The important thing is to jump in and begin writing something, and make notes along the way of how other sections need to be adapted so that they all work together.

One way to organize each chapter is to present the material to a group of fellow students. (If you cannot find an audience, then present to an imaginary one.) If you can organize your ideas into a coherent hour lecture, using chalkboard or overhead projector, on a level understandable by your fellow students, you are probably ready to write a chapter.

Your thesis will also be used as a scientific report and consulted by future workers in your laboratory who will want to know, in detail, what you did. Remember that the style of thesis writing is expository: you are trying to communicate your ideas, their significance, and their limitations. It is not the compressed style of a page-limited conference paper or journal article. Don't make your reader work too hard! At the same time, don't talk down to the reader, wasting time with repetition or adding unnecessary filler. Committee members and later readers will resent such tactics.

Theses are occasionally consulted by people from other institutions. More commonly theses are now stored in an entirely digital form. These may be stored as .pdf files on a server at your university. The advantage is that your thesis can be consulted much more easily by researchers around the world. Write with these possibilities in mind.

It is often helpful to have someone other than your adviser(s) read some sections of the thesis, particularly the introduction and conclusion chapters. It may also be appropriate to ask other members of staff to read some sections of the thesis which they may find relevant or of interest, as they may be able to make valuable contributions. In either case, only give them revised versions, so that they do not waste time correcting your grammar, spelling, poor construction or presentation.

7.4.3 Master's vs. PhD Thesis

There are different expectations for Master's theses and for Doctoral theses. This difference is not in format but in the significance and level of discovery as evidenced by the problem to be solved and the summary of contributions; a Doctoral thesis necessarily requires a more difficult problem to be solved, and consequently more substantial contributions.

The contribution to knowledge of a Master's thesis can be in the nature of an incremental improvement in an area of knowledge, or the application of known techniques in a new area. The Ph.D. must be a substantial and innovative contribution to knowledge.

How much detail?

The short answer is: rather more than for a scientific paper. Once your thesis has been assessed and your friends have read the first three pages, the only further readers are likely to be people who are seriously doing research in just that area. For example, a future research student might be pursuing the same research and be interested to find out exactly what you did. For important parts of apparatus, you should include workshop drawings, circuit diagrams and computer programs, usually as appendices. (By the way, the intelligible annotation of programs is about as frequent as porcine aviation, but it is far more desirable. You wrote that line of code for a reason: at the end of the line explain what the reason is.) You have probably read the theses of previous students in the lab where you are now working, so you probably know the advantages of a clearly explained, explicit thesis and/or the disadvantages of a vague one.

Getting Started

When you are about to begin, writing a thesis seems a long, difficult task. That is because it is a long, difficult task. Fortunately, it will seem less daunting once you have a couple of chapters done. Towards the end, you will even find yourself enjoying it---an enjoyment based on satisfaction in the achievement, pleasure in the improvement in your technical writing, and of course the approaching end. Like many tasks, thesis writing usually seems worst before you begin, so let us look at how you should make a start.

7.4.4 Thesis outline

First make up a thesis outline: several pages containing chapter headings, sub-headings, some figure titles (to indicate which results go where) and perhaps some other notes and comments. There is a section on chapter order and thesis structure at the end of this text. Once you have a list of chapters and, under each chapter heading, a reasonably complete list of things to be reported or explained, you have struck a great blow against writer's block. When you sit down to type, your aim is no longer a thesis---a daunting goal---but something simpler. Your new aim is just to write a paragraph or section about one of your subheadings. It helps to start with an easy one: this gets you into the habit of writing and gives you self-confidence. Often the Materials and Methods chapter is the easiest to write---just write down what you did; carefully, formally and in a logical order.

How do you make an outline of a chapter? For most of them, you might try the following method: Assemble all the figures that you will use in it and put them in the order that you would use if you were going to explain to someone what they all meant. You might as well rehearse explaining it to someone else---after all you will probably give several talks based on your thesis work. Once you have found the most logical order, note down the key words of your explanation. These key words provide a skeleton for much of your chapter outline.

Once you have an outline, discuss it with your adviser. This step is important: s/he will have useful suggestions, but it also serves notice that s/he can expect a steady flow of chapter drafts that will make high priority demands on his/her time. Once you and your adviser have agreed on a logical structure, s/he will need a copy of this outline for reference when reading the chapters which you will probably present out of order. If you have a co-adviser, discuss the outline with him/her as well, and present all chapters to both advisers for comments.

Organisation

It is encouraging and helpful to start a filing system. Open a word-processor file for each chapter *and one for the references.* You can put notes in these files, as well as text. While doing something for Chapter n, you will think "Oh I must refer back to/discuss this in Chapter m" and so you put a note to do so in the file for Chapter m. Or you may think of something interesting or relevant for that chapter. When you come to work on Chapter m, the more such notes you have accumulated, the easier it will be to write.

Make a back-up of these files and do so every day at least (depending on the reliability of your computer and the age of your disk drive). Do not keep back-up disks close to the computer in case the hypothetical thief who fancies your computer decides that s/he could use some disks as well.

A simple way of making a remote back-up is to send it as an email attachment to a consenting email correspondent, preferably one in a different location. You could also send it to yourself. In either case, be careful to dispose of superseded versions so that you don't waste disk space, especially if you have bitmap images or other large files.

You should also have a physical filing system: a collection of folders with chapter numbers on them. This will make you feel good about getting started and also help clean up your desk. Your files will contain not just the plots of results and pages of calculations, but all sorts of old notes, references, calibration curves, suppliers' addresses, specifications, speculations, letters from colleagues etc., which will suddenly strike you as relevant to one chapter or other. Stick them in that folder. Then put all the folders in a box or a filing cabinet. As you write bits and pieces of text, place the hard copy, the figures etc in these folders as well. Touch them and feel their thickness from time to time---ah, the thesis is taking shape.

If any of your data exist only on paper, copy them and keep the copy in a different location. Consider making a copy of your lab book. This has another purpose beyond security: usually the lab book stays in the lab, but you may want a copy for your own future use. Further, scientific ethics require you to keep lab books and original data for at least ten years, and a copy is more likely to be found if two copies exist.

While you are getting organised, you should deal with any university paperwork. Examiners have to be nominated and they have to agree to serve. Various forms are required by your department and by the university administration. Make sure that the rate limiting step is your production of the thesis, and not some minor bureaucratic problem.

Text Editors

One of the big FAQs for scientists: is there a word processor, ideally one compatible with MS Word, but which allows you to type mathematical symbols and equations conveniently? One solution is LaTeX, which is powerful, elegant, reliable, fast and *free* from http://www.latex-project.org/ or http://www.miktex.org/. As far as I know, the only equation editor for MS Word is slow and awkward. LyX, available free at http://www.lyx.org/, is a convenient front-end to LaTeX that has WYSIWYG.

Each university has a set of style requirements for the thesis. These requirements often give rules for the use of different fonts, the format for bibliographies, the width of the margins, etc. Check around and see if your department or university has a style file compatible with your typesetting system, so that you can satisfy these rules easily. If not, be prepared to iterate a few times to make the style-checkers happy.

A timetable

I strongly recommend sitting down with the adviser and making up a timetable for writing it: a list of dates for when you will give the first and second drafts of each chapter to your adviser(s). This structures your time and provides intermediate targets. If you merely aim "to have the whole thing done by [some distant date]", you can deceive yourself and procrastinate more easily. If you have told your adviser that you will deliver a first draft of chapter 3 on Wednesday, it focuses your attention.

You may want to make your timetable into a chart with items that you can check off as you have finished them. This is particularly useful towards the end of the thesis when you find there will be quite a few loose ends here and there.

Iterative solution

Whenever you sit down to write, it is very important to write *something.* So write something, even if it is just a set of notes or a few paragraphs of text that you would never show to anyone else. It would be nice if clear, precise prose leapt easily from the keyboard, but it usually does not. Most of us find it easier, however, to improve something that is already written than to produce text from nothing. So put down a draft (as rough as you like) for your own purposes, then clean it up for your adviser to read. Word-processors are wonderful in this regard: in the first draft you do not have to start at the beginning, you can leave gaps, you can put in little notes to yourself, and then you can clean it all up later.

Your adviser will expect to read each chapter in draft form. S/he will then return it to you with suggestions and comments. *Do not be upset if a chapter---especially the first one you write--- returns covered in red ink.* Your adviser will want your thesis to be as good as possible, because his/her reputation as well as yours is affected. Scientific writing is a difficult art, and it takes a while to learn. As a consequence, there will be many ways in which your first draft can be improved. So take a positive attitude to all the scribbles with which your adviser decorates your text: each comment tells you a way in which you can make your thesis better.

As you write your thesis, your scientific writing is almost certain to improve. Even for native speakers of English who write very well in other styles, one notices an enormous improvement in the first drafts from the first to the last chapter written. The process of writing the thesis is like a course in scientific writing, and in that sense each chapter is like an assignment in which you are taught, but not assessed. Remember, only the final draft is assessed: the more comments your adviser adds to first or second draft, the better.

Before you submit a draft to your adviser, run a spell check so that s/he does not waste time on those. If you have any characteristic grammatical failings, check for them.

Make it clear what is yours

If you use a result, observation or generalisation that is not your own, you must usually state where in the scientific literature that result is reported. The only exceptions are cases where every researcher in the field already knows it: dynamics equations need not be followed by a citation of Newton, circuit analysis does not need a reference to Kirchoff. The importance of this practice in science is that it allows the reader to verify your starting position. Physics in particular is said to be a vertical science: results are built upon results which in turn are built upon results etc. Good referencing allows us to check the foundations of your additions to the structure of knowledge in the discipline, or at least to trace them back to a level which we judge to be reliable. Good referencing also tells the reader which parts of the thesis are descriptions of previous knowledge and which parts are your additions to that knowledge. In a thesis, written for the general reader who has little familiarity with the literature of the field, this should be especially clear. It may seem tempting to leave out a reference in the hope that a reader will think that a nice idea or a nice bit of analysis is yours. I advise against this gamble. The reader will probably think: "What a nice idea! I wonder if it's original?". The reader can probably find out via the net or the library.

If you are writing in the passive voice, you must be more careful about attribution than if you are writing in the active voice. “The input image was created by applying noise generator…” does not make it clear whether you did this or your friend did it. “I created the input image…” is clear.

Style

The text must be clear. Good grammar and thoughtful writing will make the thesis easier to read. Scientific writing has to be a little formal---more formal than this text. Every one should remember that scientific English is an international language. Slang and informal writing will be harder for many to understand.

Short, simple phrases and words are often better than long ones. Some politicians use "at this point in time" instead of "now" precisely because it takes longer to convey the same meaning. They do not care about elegance or efficient communication. You should. On the other hand, there will be times when you need a complicated sentence because the idea is complicated. If your primary statement requires several qualifications, each of these may need a subordinate clause: "When [qualification], and where [proviso], and if [condition] then [statement]". Some lengthy technical words will also be necessary in many theses, particularly in fields like biochemistry. Do not sacrifice accuracy for the sake of brevity. "Black is white" is simple and catchy. An advertising copy writer would love it. "Objects of very different albedo may be illuminated differently so as to produce similar reflected spectra" is longer and uses less common words, but, compared to the former example, it has the advantage of being true. The longer example would be fine in a physics thesis because English speaking physicists will not have trouble with the words. (A physicist who did not know all of those words would probably be glad to remedy the lacuna either from the context or by consulting a dictionary.)

Sometimes it is easier to present information and arguments as a series of numbered points, rather than as one or more long and awkward paragraphs. A list of points is usually easier to write. You should be careful not to use this presentation too much: your thesis must be a connected, convincing argument, not just a list of facts and observations.

One important stylistic choice is between the active voice and passive voice. The active voice ("I measured the frequency...") is simpler, and it makes clear what you did and what was done by others. The passive voice ("The frequency was measured...") makes it easier to write ungrammatical or awkward sentences. If you use the passive voice, be especially wary of dangling participles. For example, the sentence "After considering all of these possible algorithms, algorithm-A was selected" implicitly attributes consciousness to algorithm-A. This choice is a question of taste: I prefer the active because it is clearer, more logical and makes attribution simple. The only arguments I have ever heard for avoiding the active voice in a thesis are (i) many theses are written in the passive voice, and (ii) some very polite people find the use of "I" immodest. Use the first person singular, not plural, when reporting work that you did yourself: the editorial 'we' may suggest that you had help beyond that listed in your acknowledgments, or it may suggest that you are trying to share any blame. On the other hand, retain plural verbs for "data": "data" is the plural of "datum", and lots of scientists like to preserve the distinction. Just say to yourself "one datum is ..", "these data are.." several times.

Presentation

There is no need for a thesis to be a masterpiece of desk-top publishing. Your time can be more productively spent improving the content than the appearance.

In many cases, a reasonably neat diagram can be drawn by hand faster than with a graphics package, and you can scan it if you want an electronic version. Either is usually satisfactory. A one bit (i.e. black and white), moderate resolution scan of a hand-drawn sketch will be bigger than a line drawing generated on a graphics package, but not huge. While talking about the size of files, we should mention that photographs look pretty but take up a lot of memory. There's another important difference, too. The photographer thought about the camera angle and the focus etc. The person who drew the schematic diagram thought about what components ought to be depicted and the way in which the components of the system interacted with each other. So the numerically small information content of the line drawing may be much more useful information than that in a photograph.

Another note about figures and photographs. In the digital version of your thesis, do not save ordinary photographs or other illustrations as bitmaps, because these take up a lot of memory and are therefore very slow to transfer. Nearly all graphics packages allow you to save in compressed format as .jpg or .gif files. Further, you can save space/speed things up by reducing the number of colours. In vector graphics (as used for drawings), shades or grey are often produced by black and white pixels, so one-bit colour is adequate.

In general, students spend too much time on diagrams; time that could have been spent on examining the arguments, making the explanations clearer, thinking more about the significance and checking for errors in the algebra. The reason, of course, is that drawing is easier than thinking.

I do not think that there is a strong correlation (either way) between length and quality. There is no need to leave big gaps to make the thesis thicker. Readers will not appreciate large amounts of vague or unnecessary text.

Approaching the end

A deadline is very useful in some ways. You must hand in the thesis, even if you think that you need one more draft of that chapter, or someone else's comments on this section, or some other refinement. If you do not have a deadline, or if you are thinking about postponing it, please take note of this: *A thesis is a very large work. It cannot be made perfect in a finite time*. There will inevitably be things in it that you could have done better. There will be inevitably be some typos. Indeed, by some law related to Murphy's, you will discover one when you first flip open the bound copy. No matter how much you reflect and how many times you proof read it, there will be some things that could be improved. There is no point hoping that the examiners will not notice: many examiners feel obliged to find some examples of improvements (if not outright errors) just to show how thoroughly they have read it. So set yourself a deadline and stick to it. Make it as good as you can in that time, and then hand it in! (In retrospect, there was an advantage in writing a thesis in the days before word processors, spelling checkers and typing programs. Students often paid a typist to produce the final draft and could only afford to do that once.)

How many copies?

Talk to your adviser about this. As well as those for the examiners, the university libraries and yourself, you should make some distribution copies. These copies should be sent to other researchers who are working in your field so that:

- they can discover what marvellous work you have been doing before it appears in journals;
- they can look up the fine details of methods and results that will or have been published more briefly elsewhere;
- they can realise what an excellent researcher you are. This realisation could be useful if a post- doctoral position were available in their labs. soon after your submission, or if they were reviewers of your research/post-doctoral proposal. Even having your name in their bookcases might be an advantage.

Whatever the University's policy on single or double-sided copies, the distribution copies could be double-sided paper, or digital, so that forests and postage accounts are not excessively depleted by the exercise. Your adviser could help you to make up a list of interested and/or potentially useful people for such a mailing list. A CD with your thesis will be cheaper than a paper copy. You don't have to burn them all yourself: companies make multiple copies for several dollars a copy.

7.4.5 Dealing with Your Doctoral Thesis Committee

Ideally, you have chosen your committee members because of their interest in your research area and in you. Ideally, the members have followed your research over the course of a year or more, and understand your problem and your approach. Ideally, they all get along well, and egos are not a factor. And ideally, they are willing to take the time to read your thesis in detail and give you valuable feedback.

But the world is not always ideal. You might be very lucky to find one professor other than your advisor who is willing to listen and read and comment meaningfully. Other committee members may prefer a less active role, at one extreme, simply showing up for your oral exam and questioning you. Rules or reality may have dictated that some committee members have little interest in your research area, or little time to devote to mentoring.

Whatever the situation, draw your committee members as much into the process as they wish to be. If the committee is established early, then stop by or send a brief update to them two or three times a year so that they can follow your progress. If it is established after the thesis is written, give them plenty of time to read the thesis, and then contact each one, asking whether it would be helpful if you stopped by to answer questions or discuss your work. You don't want to be surprised at the oral exam by a very unhappy committee member.

After the oral exam, it is courteous to give a bound copy of the final version of the thesis to each committee member, and to express gratitude for the time they spent on your committee. Their participation should be noted in your thesis acknowledgements.

Personal

In the ideal situation, you will be able to spend a large part---perhaps a majority---of your time writing your thesis. This may be bad for your physical and mental health.

Typing

Set up your chair and computer properly. The Health Service, professional keyboard users or perhaps even the school safety officer will be able to supply charts showing recommended relative heights, healthy postures and also exercises that you should do if you spend a lot of time at the keyboard. These last are worthwhile insurance: you do not want the extra hassle of back or neck pain. Try to intersperse long sessions of typing with other tasks, such as reading, drawing, calculating, thinking or doing research.

If you do not touch type, you should learn to do so for the sake of your neck as well as for productivity. There are several good software packages that teach touch typing interactively. If you use one for say 30 minutes a day for a couple of weeks, you will be able to touch type. By the time you finish the thesis, you will be able to touch type quickly and accurately and your six hour investment will have paid for itself. Be careful not to use the typing exercises as a displacement activity.

Exercise
Do not give up exercise for the interim. Lack of exercise makes you feel bad, and you do not need anything else making you feel bad while writing a thesis. 30-60 minutes of exercise per day is probably not time lost from your thesis: I find that if I do not get regular exercise, I sleep less soundly and longer. How about walking to work and home again? (Walk part of the way if your home is distant.) Many people opine that a walk helps them think, or clears the head. You may find that an occasional stroll improves your productivity.

Food
Do not forget to eat, and make an effort to eat healthy food. You should not lose fitness or risk illness at this critical time. Exercise is good for keeping you appetite at a healthy level. I know that you have little time for cooking, but keep a supply of fresh fruit, vegetables and bread. It takes less time to make a sandwich than to go to the local fast food outlet, and you will feel better afterwards.

Drugs
Thesis writers have a long tradition of using coffee as a stimulant and alcohol or marijuana as relaxants. (Use of alcohol and coffee is legal, use of marijuana is not.) Used in moderation, they do not seem to have ill effects on the quality of thesis produced. Excesses, however, are obviously counter-productive: several espressi and you will be buzzing too much to sit down and work; several drinks at night will slow you down next day.

Others
Other people will be sympathetic, but do not take them for granted. Spouses, lovers, family and friends should not be undervalued. Spend some time with them and, when you do, have a good time. Do not spend your time together complaining about your thesis: they already resent the thesis because it is keeping you away from them. If you can find another student writing a thesis, then you may find it therapeutic to complain to each other about advisers and difficulties. S/he need not be in the same discipline as you are.

7.4.6 Getting into the Real Business of Writing Thesis

So, you are preparing to write a Ph.D. dissertation in an experimental area of Computer Science. Unless you have written many formal documents before, you are in for a surprise: it's difficult!

Here are a few guidelines that may help you when you finally get serious about writing. The list goes on forever; you probably won't want to read it all at once. But, please read it before you write anything.

The General Idea:

10. A thesis is a hypothesis or conjecture.
11. A PhD dissertation is a lengthy, formal document that argues in defense of a particular thesis. (So many people use the term ``thesis'' to refer to the document that a current dictionary now includes it as the third meaning of ``thesis'').
12. Two important adjectives used to describe a dissertation are ``original'' and ``substantial.'' The research performed to support a thesis must be both, and the dissertation must show it to be so. In particular, a dissertation highlights original contributions.
13. The scientific method means starting with a hypothesis and then collecting evidence to support or deny it. Before one can write a dissertation defending a particular thesis, one must collect evidence that supports it. Thus, the most difficult aspect of writing a dissertation consists of organizing the evidence and associated discussions into a coherent form.
14. The essence of a dissertation is critical thinking, not experimental data. Analysis and concepts form the heart of the work.
15. A dissertation concentrates on principles: it states the lessons learned, and not merely the facts behind them.
16. In general, every statement in a dissertation must be supported either by a reference to published scientific literature or by original work. Moreover, a dissertation does not repeat the details of critical thinking and analysis found in published sources; it uses the results as fact and refers the reader to the source for further details.
17. Each sentence in a dissertation must be complete and correct in a grammatical sense. Moreover, a dissertation must satisfy the stringent rules of formal grammar (e.g., no contractions, no colloquialisms, no slurs, no undefined technical jargon, no hidden jokes, and no slang, even when such terms or phrases are in common use in the spoken language). Indeed, the writing in a dissertaton must be crystal clear. Shades of meaning matter; the terminology and prose must make fine distinctions. The words must convey exactly the meaning intended, nothing more and nothing less.
18. Each statement in a dissertation must be correct and defensible in a logical and scientific sense. Moreover, the discussions in a dissertation must satisfy the most stringent rules of logic applied to mathematics and science.

What One Should Learn from the Exercise:

3. All scientists need to communicate discoveries; the PhD dissertation provides training for communication with other scientists.
4. Writing a dissertation requires a student to think deeply, to organize technical discussion, to muster arguments that will convince other scientists, and to follow rules for rigorous, formal presentation of the arguments and discussion.

A Rule of Thumb:

Good writing is essential in a dissertation. However, good writing cannot compensate for a paucity of ideas or concepts. Quite the contrary, a clear presentation always exposes weaknesses.

Definitions and Terminology:

5. Each technical term used in a dissertation must be defined either by a reference to a previously published definition (for standard terms with their usual meaning) or by a precise, unambiguous definition that appears before the term is used (for a new term or a standard term used in an unusual way).
6. Each term should be used in one and only one way throughout the dissertation.
7. The easiest way to avoid a long series of definitions is to include a statement: ``the terminology used throughout this document follows that given in [CITATION]." Then, only define exceptions.
8. The introductory chapter can give the intuition (i.e., informal definitions) of terms provided they are defined more precisely later.

Terms and Phrases to Avoid:

adverbs

> Mostly, they are very often overly used. Use strong words instead. For example, one could say, ``Writers abuse adverbs."

jokes or puns

> They have no place in a formal document.

``bad", ``good", ``nice", ``terrible", ``stupid"

> A scientific dissertation does not make moral judgements. Use ``incorrect/correct" to refer to factual correctness or errors. Use precise words or phrases to assess quality (e.g., ``method A requires less computation than method B"). In general, one should avoid all qualitative judgements.

``true'', ``pure'',

In the sense of ``good'' (it is judgemental).

``perfect''

Nothing is.

``an ideal solution''

You're judging again.

``today'', ``modern times''

Today is tomorrow's yesterday.

``soon''

How soon? Later tonight? Next decade?

``we were surprised to learn...''

Even if you were, so what?

``seems'', ``seemingly'',

It doesn't matter how something appears;

``would seem to show''

all that matters are the facts.

``in terms of''

usually vague

``based on'', ``X-based'', ``as the basis of''

careful; can be vague

``different''

Does not mean ``various''; different than what?

``in light of''

colloquial

``lots of''

vague & colloquial

``kind of''

vague & colloquial

``type of''

vague & colloquial

``something like''

vague & colloquial

``just about''

vague & colloquial

``number of''

vague; do you mean ``some'', ``many'', or ``most''? A quantative statement is preferable.

``due to''

colloquial

``probably''

only if you know the statistical probability (if you do, state it quantatively

``obviously, clearly''

be careful: obvious/clear to everyone?

``simple''

Can have a negative connotation, as in ``simpleton''

``along with''

Just use ``with''

``actually, really''

> define terms precisely to eliminate the need to clarify

``the fact that''

> makes it a meta-sentence; rephrase

``this'', ``that''

> As in ``This causes concern.'' Reason: ``this'' can refer to the subject of the previous sentence, the entire previous sentence, the entire previous paragraph, the entire previous section, etc. More important, it can be interpreted in the concrete sense or in the meta-sense. For example, in: *``X does Y. This means ...''* the reader can assume ``this'' refers to Y or to the fact that X does it. Even when restricted (e.g., ``this computation...''), the phrase is weak and often ambiguous.

``You will read about...''

> The second person has no place in a formal dissertation.

``I will describe...''

> The first person has no place in a formal dissertation. If self-reference is essential, phrase it as ``Section 10 describes...''

``we'' as in ``we see that''

> A trap to avoid. Reason: almost any sentence can be written to begin with ``we'' because ``we'' can refer to: the reader and author, the author and advisor, the author and research team, experimental computer scientists, the entire computer science community, the science community, or some other unspecified group.

``Hopefully, the program...''

> Computer programs don't hope, not unless they implement AI systems. By the way, if you are writing an AI thesis, talk to someone else: AI people have their own system of rules.

``...a famous researcher...''

> It doesn't matter who said it or who did it. In fact, such statements prejudice the reader.

Be Careful When Using ``few, most, all, any, every''.

A dissertation is precise. If a sentence says ``Most computer systems contain X'', you must be able to defend it. Are you sure you really know the facts? How many computers were built and sold yesterday?

``must'', ``always''

Absolutely?

``should''

Who says so?

``proof'', ``prove''

Would a mathematician agree that it's a proof?

``show''

Used in the sense of ``prove''. To ``show'' something, you need to provide a formal proof.

``can/may''

Your mother probably told you the difference.

Voice:

Use active constructions. For example, say ``the operating system starts the device'' instead of ``the device is started by the operating system.''

Tense:

Write in the present tense. For example, say ``The system writes a page to the disk and then uses the frame...'' instead of ``The system will use the frame after it wrote the page to disk...''

Define Negation Early:

Example: say ``no data block waits on the output queue" instead of ``a data block awaiting output is not on the queue."

Grammar and Logic:

Be careful that the subject of each sentence really does what the verb says it does. Saying ``Programs must make procedure calls using the X instruction" is not the same as saying ``Programs must use the X instruction when they call a procedure." In fact, the first is patently false! Another example: ``RPC requires programs to transmit large packets" is not the same as ``RPC requires a mechanism that allows programs to transmit large packets."

All computer scientists should know the rules of logic. Unfortunately the rules are more difficult to follow when the language of discourse is English instead of mathematical symbols. For example, the sentence ``There is a compiler that translates the N languages by..." means a single compiler exists that handles all the languages, while the sentence ``For each of the N languages, there is a compiler that translates..." means that there may be 1 compiler, 2 compilers, or N compilers. When written using mathematical symbols, the difference are obvious because ``for all" and ``there exists" are reversed.

Focus On Results And Not The People/Circumstances In Which They Were Obtained:

``After working eight hours in the lab that night, we realized..." has no place in the dissertation. It doesn't matter when you realized it or how long you worked to obtain the answer. Another example: ``Jim and I arrived at the numbers shown in Table 3 by measuring..." Put an acknowledgement to Jim in the dissertation, but do not include names (even your own) in the main body. You may be tempted to document a long series of experiments that produced nothing or a coincidence that resulted in success. Avoid it completely. In particular, do not document seemingly mystical influences (e.g., ``if that cat had not crawled through the hole in the floor, we might not have discovered the power supply error indicator on the network bridge"). Never attribute such events to mystical causes or imply that strange forces may have affected your results. Summary: stick to the plain facts. Describe the results without dwelling on your reactions or events that helped you achieve them.

Avoid Self-Assessment (both praise and criticism):

Both of the following examples are incorrect: ``The method outlined in Section 2 represents a major breakthrough in the design of distributed systems because..." ``Although the technique in the next section is not earthshaking,..."

References to Extant Work:

One always cites papers, not authors. Thus, one uses a singular verb to refer to a paper even though it has multiple authors. For example ``Johnson and Smith [J&S90] reports that..."

Avoid the phrase ``the authors claim that X". The use of ``claim" casts doubt on ``X" because it references the authors' thoughts instead of the facts. If you agree ``X" is correct, simply state ``X" followed by a reference. If one absolutely must reference a paper instead of a result, say ``the paper states that..." or ``Johnson and Smith [J&S 90] presents evidence that...".

Concept Vs. Instance:

A reader can become confused when a concept and an instance of it are blurred. Common examples include: an algorithm and a particular program that implements it, a programming language and a compiler, a general abstraction and its particular implementation in a computer system, a data structure and a particular instance of it in memory.

Terminology for Concepts and Abstractions

When defining the terminology for a concept, be careful to decide precisely how the idea translates to an implementation. Consider the following discussion:

VM systems include a concept known as an address space. The system dynamically creates an address space when a program needs one, and destroys an address space when the program that created the space has finished using it. A VM system uses a small, finite number to identify each address space. Conceptually, one understands that each new address space should have a new identifier. However, if a VM system executes so long that it exhausts all possible address space identifiers, it must reuse a number.

The important point is that the discussion only makes sense because it defines ``address space" independently from ``address space identifier". If one expects to discuss the differences between a concept and its implementation, the definitions must allow such a distinction.

Knowledge Vs. Data

The facts that result from an experiment are called ``data". The term ``knowledge" implies that the facts have been analyzed, condensed, or combined with facts from other experiments to produce useful information.

Cause and Effect:

A dissertation must carefully separate cause-effect relationships from simple statistical correlations. For example, even if all computer programs written in Professor X's lab require more memory than the computer programs written in Professor Y's lab, it may not have anything to do with the professors or the lab or the programmers (e.g., maybe the people working in professor X's lab are working on applications that require more memory than the applications in professor Y's lab).

Drawing Only Warranted Conclusions:

One must be careful to only draw conclusions that the evidence supports. For example, if programs run much slower on computer A than on computer B, one cannot conclude that the processor in A is slower than the processor in B unless one has ruled out all differences in the computers' operating systems, input or output devices, memory size, memory cache, or internal bus bandwidth. In fact, one must still refrain from judgement unless one has the results from a controlled experiment (e.g., running a set of several programs many times, each when the computer is otherwise idle). Even if the cause of some phenomenon seems obvious, one cannot draw a conclusion without solid, supporting evidence.

Commerce and Science:

In a scientific dissertation, one never draws conclusions about the economic viability or commercial success of an idea/method, nor does one speculate about the history of development or origins of an idea. A scientist must remain objective about the merits of an idea independent of its commercial popularity. In particular, a scientist never assumes that commercial success is a valid measure of merit (many popular products are neither well-designed nor well-engineered). Thus, statements such as ``over four hundred vendors make products using technique Y'' are irrelevant in a dissertation.

Politics and Science:

A scientist avoids all political influence when assessing ideas. Obviously, it should not matter whether government bodies, political parties, religious groups, or other organizations endorse an idea. More important and often overlooked, it does not matter whether an idea originated with a scientist who has already won a Nobel prize or a first-year graduate student. One must assess the idea independent of the source.

Canonical Organization:

In general, every dissertation must define the problem that motivated the research, tell why that problem is important, tell what others have done, describe the new contribution, document the experiments that validate the contribution, and draw conclusions. There is no canonical organization for a dissertation; each is unique. However, novices writing a dissertation in the experimental areas of CS may find the following example a good starting point:

- ### Chapter 1: Introduction

  An overview of the problem; why it is important; a summary of extant work and a statement of your hypothesis or specific question to be explored. Make it readable by anyone.

- ### Chapter 2: Definitions

  New terms only. Make the definitions precise, concise, and unambiguous.

- ### Chapter 3: Conceptual Model

  Describe the central concept underlying your work. Make it a ``theme'' that ties together all your arguments. It should provide an answer to the question posed in the introduction at a conceptual level. If necessary, add another chapter to give additional reasoning about the problem or its solution.

- ### Chapter 4: Experimental Measurements

  Describe the results of experiments that provide evidence in support of your thesis. Usually experiments either emphasize proof-of-concept (demonstrating the viability of a method/technique) or efficiency (demonstrating that a method/technique provides better performance than those that exist).

- ### Chapter 5: Corollaries And Consequences

  Describe variations, extensions, or other applications of the central idea.

- ### Chapter 6: Conclusions

  Summarize what was learned and how it can be applied. Mention the possibilities for future research.

- ### Abstract:

  A short (few paragraphs) summary of the the dissertation. Describe the problem and the research approach. Emphasize the original contributions.

Suggested Order For Writing:

The easiest way to build a dissertation is inside-out. Begin by writing the chapters that describe your research (3, 4, and 5 in the above outline). Collect terms as they arise and keep a definition for each. Define each technical term, even if you use it in a conventional manner.

Organize the definitions into a separate chapter. Make the definitions precise and formal. Review later chapters to verify that each use of a technical term adheres to its definition. After reading the middle chapters to verify terminology, write the conclusions. Write the introduction next. Finally, complete an abstract.

Key to Success:

By the way, there is a key to success: practice. No one ever learned to write by reading essays like this. Instead, you need to practice, practice, practice. Every day.

Parting thoughts:

We leave you with the following ideas to mull over. If they don't mean anything to you now, revisit them after you finish wirting a dissertation.

> After great pain, a formal feeling comes.
> -- Emily Dickinson

> A man may write at any time, if he will set himself doggedly to it.
> -- Samuel Johnson

> Keep right on to the end of the road.
> -- Harry Lauder

> The average Ph.D. thesis is nothing but the transference of bones from one graveyard to another.
> -- Frank J. Dobie

Conclusion

Keep going, you're nearly there! Most PhDs will admit that there were times when we thought about reasons for not finishing. But it would be crazy to give up at the writing stage, after years of work on the research, and it would be something to regret for a long time.

Writing a thesis is tough work. One anonymous post doctoral researcher told me: "You should tell everyone that it's going to be unpleasant, that it will mess up their lives, that they will have to give up their friends and their social lives for a while. It's a tough period for almost every student." She's right: it is certainly hard work, it will probably be stressful and you will have to adapt your rhythm to it. It is also an important rite of passage and the satisfaction you will feel afterwards is wonderful. On behalf of scholars everywhere, I wish you good luck!

Acknowledgements and Suggestions

This document will be continuously updated. If you have suggestions do send them. I thank all my research scholars and M.Tech. students who have inspired me to write this document and served as test domains for my experimentations. Opinions expressed in these documents are mine and do not necessarily reflect the policy of the Manonmaniam Sundaranar or of the Centre for Information Technology and Engineering.

7.4.7 A suggested thesis structure - Model 1(Ph.D)

The list of contents and chapter headings below is appropriate for some theses. In some cases, one or two of them may be irrelevant. Results and Discussion are usually combined in several chapters of a thesis. Think about the plan of chapters and decide what is best to report your work. Then make a list, in point form, of what will go in each chapter. Try to make this rather detailed, so that you end up with a list of points that corresponds to subsections or even to the paragraphs of your thesis. At this stage, think hard about the logic of the presentation: within chapters, it is often possible to present the ideas in different order, and not all arrangements will be equally easy to follow. If you make a plan of each chapter and section before you sit down to write, the result will probably be clearer and easier to read. It will also be easier to write.

Copyright waiver

Your institution may have a form for Copyright. In any case, this standard page gives the university library the right to publish the work, possibly by microfilm or some other medium.

Declaration

Check the wording required by your institution, and whether there is a standard form. Many universities require something like: "I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text. (signature/name/date)"

Title page

This may vary among institutions, but as an example: Title/author/"A thesis submitted for the degree of Doctor of Philosophy in the Faculty of Science/Manonmaniam Sundaranar University of Tirunelveli"/date.

Abstract

Of all your thesis, this part will be the most widely published and most read because it will be published in Dissertation Abstracts International. It is best written towards the end, but not at the very last minute because you will probably need several drafts. It should be a distillation of the thesis: a concise description of the problem(s) addressed, your method of solving it/them, your results and conclusions. An abstract must be self-contained. Usually they do not contain references. When a reference is necessary, its details should be included in the text of the abstract. Check the word limit.

Acknowledgments

Most thesis authors put in a page of thanks to those who have helped them in matters scientific, and also indirectly by providing such essentials as food, education, genes, money, help, advice, friendship etc. *If any of your work is collaborative, you should make it quite clear who did which sections.*

Table of contents

The introduction starts on page 1, the earlier pages should have roman numerals. It helps to have the subheadings of each chapter, as well as the chapter titles. Remember that the thesis may be used as a reference in the lab, so it helps to be able to find things easily.

Introduction

What is the topic and why is it important? State the problem(s) as simply as you can. Remember that you have been working on this project for a few years, so you will be very close to it. Try to step back mentally and take a broader view of the problem. How does it fit into the broader world of your discipline?

Especially in the introduction, do not overestimate the reader's familiarity with your topic. You are writing for researchers in the general area, but not all of them need be specialists in your particular topic. It may help to imagine such a person---think of some researcher whom you might have met at a conference for your subject, but who was working in a different area. S/he is intelligent, has the same general background, but knows little of the literature or tricks that apply to your particular topic.

The introduction should be interesting. If you bore the reader here, then you are unlikely to revive his/her interest in the materials and methods section. For the first paragraph or two, tradition permits prose that is less dry than the scientific norm. If want to wax lyrical about your topic, here is the place to do it. Try to make the reader want to read the kilogram of A4 that has arrived uninvited on his/her desk. Go to the library and read several thesis introductions. Did any make you want to read on? Which ones were boring?

This section might go through several drafts to make it read well and logically, while keeping it short. For this section, I think that it is a good idea to ask someone who is not a specialist to read it and to comment. Is it an adequate introduction? Is it easy to follow? There is an argument for writing this section---or least making a major revision of it---

towards the end of the thesis writing. Your introduction should tell where the thesis is going, and this may become clearer during the writing.

Literature review

Where did the problem come from? What is already known about this problem? What other methods have been tried to solve it?

Ideally, you will already have much of the hard work done, if you have been keeping up with the literature as you vowed to do three years ago, and if you have made notes about important papers over the years. If you have summarised those papers, then you have some good starting points for the review.

If you didn't keep your literature notes up to date, you can still do something useful: pass on the following advice to any beginning PhD students in your lab and tell them how useful this would have been to you. When you start reading about a topic, you should open a spread sheet file, or at least a word processor file, for your literature review. Of course you write down the title, authors, year, volume and pages. But you also write a summary (anything from a couple of sentences to a couple of pages, depending on the relevance). In other columns of the spread sheet, you can add key words (your own and theirs) and comments about its importance, relevance to you and its quality.

How many papers? How relevant do they have to be before you include them? Well, that is a matter of judgement. On the order of a hundred is reasonable, but it will depend on the field. You are the world expert on the (narrow) topic of your thesis: you must demonstrate this.

A political point: make sure that you do not omit relevant papers by researchers who are like to be your examiners, or by potential employers to whom you might be sending the thesis in the next year or two.

Middle chapters

In some theses, the middle chapters are the journal articles of which the student was major author. There are several disadvantages to this format.

One is that a thesis is both allowed and expected to have more detail than a journal article. For journal articles, one usually has to reduce the number of figures. In many cases, all of the interesting and relevant data can go in the thesis, and not just those which appeared in the journal. The degree of experimental detail is usually greater in a thesis. Relatively often a researcher requests a thesis in order to obtain more detail about how a study was performed.

Another disadvantage is that your journal articles may have some common material in the introduction and the "Materials and Methods" sections.

The exact structure in the middle chapters will vary among theses. In some theses, it is necessary to establish some theory, to describe the experimental techniques, then to report what was done on several different problems or different stages of the problem, and then finally to present a model or a new theory based on the new work. For such a thesis, the chapter headings might be: Theory, Materials and Methods, {first problem}, {second problem}, {third problem}, {proposed theory/model} and then the conclusion chapter. For other theses, it might be appropriate to discuss different techniques in different chapters, rather than to have a single Materials and Methods chapter.

Here follow some comments on the elements Materials and Methods, Theory, Results and discussion which may or may not correspond to thesis chapters.

Materials and Methods

This varies enormously from thesis to thesis, and may be absent in theoretical theses. It should be possible for a competent researcher to reproduce exactly what you have done by following your description. There is a good chance that this test will be applied: sometime after you have left, another researcher will want to do a similar experiment either with your gear, or on a new set-up in a foreign country. Please write for the benefit of that researcher.

In some theses, particularly multi-disciplinary or developmental ones, there may be more than one such chapter. In this case, the different disciplines should be indicated in the chapter titles.

Theory

When you are reporting theoretical work that is not original, you will usually need to include sufficient material to allow the reader to understand the arguments used and their physical bases. Sometimes you will be able to present the theory *ab initio,* but you should not reproduce two pages of algebra that the reader could find in a standard text. Do not include theory that you are not going to relate to the work you have done.

When writing this section, concentrate at least as much on the physical arguments as on the equations. What do the equations mean? What are the important cases?

When you are reporting your own theoretical work, you must include rather more detail, but you should consider moving lengthy derivations to appendices. Think too about the order and style of presentation: the order in which you did the work may not be the clearest presentation.

Suspense is not necessary in reporting science: you should tell the reader where you are going before you start.

Results and discussion

The results and discussion are very often combined in theses. This is sensible because of the length of a thesis: you may have several chapters of results and, if you wait till they are all presented before you begin discussion, the reader may have difficulty remembering what you are talking about. The division of Results and Discussion material into chapters is usually best done according to subject matter.

Make sure that you have described the conditions which obtained for each set of results. What was held constant? What were the other relevant parameters? Make sure too that you have used appropriate statistical analyses. Where applicable, show measurement errors and standard errors on the graphs. Use appropriate statistical tests.

Take care plotting graphs. The origin and intercepts are often important so, unless the ranges of your data make it impractical, the zeros of one or both scales should usually appear on the graph. You should show error bars on the data, unless the errors are very small. For single measurements, the bars should be your best estimate of the experimental errors in each coordinate. For multiple measurements these should include the standard error in the data. The errors in different data are often different, so, where this is the case, regressions and fits should be weighted (i.e. they should minimize the sum of squares of the differences weighted inversely as the size of the errors.) (A common failing in many simple software packages that draw graphs and do regressions is that they do not treat errors adequately. UNSW student Mike Johnston has written a plotting routine that plots data with error bars and performs weighted least square regressions. It is at http://www.phys.unsw.edu.au/3rdyearlab/graphing/graph.html). You can just 'paste' your data into the input and it generates a .ps file of the graph.

In most cases, your results need discussion. What do they mean? How do they fit into the existing body of knowledge? Are they consistent with current theories? Do they give new insights? Do they suggest new theories or mechanisms?

Try to distance yourself from your usual perspective and look at your work. Do not just ask yourself what it means in terms of the orthodoxy of your own research group, but also how other people in the field might see it. Does it have any implications that do not relate to the questions that you set out to answer?

Final chapter, references and appendices

Conclusions and suggestions for further work

Your abstract should include your conclusions in very brief form, because it must also include some other material. A summary of conclusions is usually longer than the final section of the abstract, and you have the space to be more explicit and more careful with qualifications. You might find it helpful to put your conclusions in point form.

It is often the case with scientific investigations that more questions than answers are produced. Does your work suggest any interesting further avenues? Are there ways in which your work could be improved by future workers? What are the practical implications of your work?

This chapter should usually be reasonably short---a few pages perhaps. As with the introduction, I think that it is a good idea to ask someone who is not a specialist to read this section and to comment.

References (See also under literature review)

It is tempting to omit the titles of the articles cited, and the university allows this, but think of all the times when you have seen a reference in a paper and gone to look it up only to find that it was not helpful after all.

Should you reference web sites and, if so, how? If you cite a journal article or book, the reader can go to a library and check that the cited document and check whether or not it says what you say it did. A web site may disappear, and it may have been updated or changed completely. So references to the web are usually less satisfactory. Nevertheless, there are some very useful and authoritative sources. So, *if the rules of your institution permit it*, it may be appropriate to cite web sites. (Be cautious, and don't overuse such citations. In particular, don't use a web citation where you could reasonably use a "hard" citation. Remember that your examiners are likely to be older and more conservative.) You should give the URL and also the date you downloaded it. If there is a date on the site itself (last updated on .....) you should included that, too.

Appendices
If there is material that should be in the thesis but which would break up the flow or bore the reader unbearably, include it as an appendix. Some things which are typically included in appendices are: important and original computer programs, data files that are too large to be represented simply in the results chapters, pictures or diagrams of results which are not important enough to keep in the main text.

A suggested thesis structure - Model 2 (Graduate):

Introduction

This note describes how to organize the written thesis which is the central element of your graduate degree. To know how to organize the thesis document, you first have to understand what graduate-level research is all about, so that is covered too. In other words, this note should be helpful when you are just getting started in your graduate program, as well as later when you start to write your thesis.

What Graduate Research is All About

The distinguishing mark of graduate research is *an original contribution to knowledge*. The thesis is a formal document whose sole purpose is to prove that you have made an original contribution to knowledge. Failure to prove that you have made such a contribution generally leads to failure.

To this end, your thesis must show two important things:

> you have identified a worthwhile problem or question which has not been previously answered,
> you have solved the problem or answered the question.

Your contribution to knowledge generally lies in your solution or answer.

What the Graduate Thesis is All About

Because the purpose of the graduate thesis is to prove that you have made an original and useful contribution to knowledge, the examiners read your thesis to find the answers to the following questions:

> what is this student's research question?
> is it a good question? (has it been answered before? is it a useful question to work on?)
> did the student convince me that the question was adequately answered?
> has the student made an adequate contribution to knowledge?

A very *clear* statement of the question is essential to proving that you have made an original and worthwhile contribution to knowledge. To prove the originality and value of your contribution, you must present a *thorough* review of the existing literature on the subject, and on closely related subjects. Then, by making *direct* reference to your literature review, you must *demonstrate* that your question (a) has not been previously answered, and (b) is worth answering. Describing how you answered the question is usually easier to write about, since you have been intimately involved in the details over the course of your graduate work.

If your thesis does not provide adequate answers to the few questions listed above, you will likely be faced with a requirement for major revisions or you may fail your thesis defence outright. For this reason, the generic thesis skeleton given below is designed to highlight the answers to those questions with appropriate thesis organization and section titles. The generic thesis skeleton can be used for any thesis. While some professors may prefer a different organization, the essential elements in any thesis will be the same. Some further notes follow the skeleton.

Always remember that a thesis is a *formal* document: every item must be in the appropriate place, and repetition of material in different places should be eliminated.

A Generic Thesis Skeleton

1. Introduction

This is a *general* introduction to what the thesis is all about -- it is *not* just a description of the contents of each section. Briefly *summarize* the question (you will be stating the question in detail later), some of the reasons why it is a worthwhile question, and perhaps give an overview of your main results. This is a birds-eye view of the answers to the main questions answered in the thesis (see above).

2. Background Information (optional)

A brief section giving background information may be necessary, especially if your work spans two or more traditional fields. That means that your readers may not have any experience with some of the material needed to follow your thesis, so you need to give it to them. A different title than that given above is usually better; e.g., "A Brief Review of Frammis Algebra."

3. Review of the State of the Art

Here you review the state of the art relevant to your thesis. Again, a different title is probably appropriate; e.g., "State of the Art in Zylon Algorithms." The idea is to *present* (critical analysis comes a little bit later) the major ideas in the state of the art right up to, but not including, your own personal brilliant ideas.

You organize this section *by idea*, and not by author or by publication. For example if there have been three important main approaches to Zylon Algorithms to date, you might organize subsections around these three approaches, if necessary:

3.1 Iterative Approximation of Zylons
3.2 Statistical Weighting of Zylons
3.3 Graph-Theoretic Approaches to Zylon Manipulation

4. Research Question or Problem Statement

Engineering theses tend to refer to a "problem" to be solved where other disciplines talk in terms of a "question" to be answered. In either case, this section has three main parts:

1. a concise statement of the question that your thesis tackles
2. justification, by *direct* reference to section 3, that your question is previously unanswered
3. discussion of why it is worthwhile to answer this question.

Item 2 above is where you *analyze* the information which you presented in Section 3. For example, maybe your problem is to "develop a Zylon algorithm capable of handling very large scale problems in reasonable time" (you would further describe what you mean by "large scale" and "reasonable time" in the problem statement). Now in your *analysis* of the state of the art you would show how each class of current approaches fails (i.e. can handle only small problems, or takes too much time). In the last part of this section you would explain why having a large-scale fast Zylon algorithm is useful; e.g., by describing applications where it can be used.

Since this is one of the sections that the readers are *definitely* looking for, highlight it by using the word "problem" or "question" in the title: e.g. "Research Question" or "Problem Statement", or maybe something more specific such as "The Large-Scale Zylon Algorithm Problem."

5. Describing How You Solved the Problem or Answered the Question

This part of the thesis is much more free-form. It may have one or several sections and subsections. But it all has only one purpose: to convince the examiners that you answered the question or solved the problem that you set for yourself in Section 4. So show what you did that is *relevant* to answering the question or solving the problem: if there were blind alleys and dead ends, do *not* include these, unless specifically relevant to the demonstration that you answered the thesis question.

6. Conclusions

You generally cover three things in the Conclusions section, and each of these usually merits a separate subsection:

1. Conclusions
2. Summary of Contributions
3. Future Research

Conclusions are *not* a rambling summary of the thesis: they are *short*, *concise* statements of the inferences that you have made because of your work. It helps to organize these as short numbered paragraphs, ordered from most to least important. All conclusions should be directly related to the research question stated in Section 4. Examples:

1. The problem stated in Section 4 has been solved: as shown in Sections ? to ??, an algorithm capable of handling large-scale Zylon problems in reasonable time has been developed.

2. The principal mechanism needed in the improved Zylon algorithm is the Grooty mechanism.

3. Etc.

The Summary of Contributions will be much sought and carefully read by the examiners. Here you list the contributions of *new* knowledge that your thesis makes. Of course, the thesis itself must substantiate any claims made here. There is often some overlap with the Conclusions, but that's okay. Concise numbered paragraphs are again best. Organize from *most* to *least* important. Examples:

1. Developed a much quicker algorithm for large-scale Zylon problems.

2. Demonstrated the first use of the Grooty mechanism for Zylon calculations.

3. Etc.

The Future Research subsection is included so that researchers picking up this work in future have the benefit of the ideas that you generated while you were working on the project. Again, concise numbered paragraphs are usually best.

7. References

The list of references is closely tied to the review of the state of the art given in section 3. Most examiners scan your list of references looking for the important works in the field, so make sure they are listed and referred to in section 3. Truth be known, most examiners also look for their own publications if they are in the topic area of the thesis, so list these too. Besides, reading your examiner's papers usually gives you a clue as to the type of questions they are likely to ask.

All references given *must* be referred to in the main body of the thesis. Note the difference from a Bibliography, which may include works that are not directly referenced in the thesis. Organize the list of references either alphabetically by author surname (preferred), or by order of citation in the thesis.

8. Appendices

What goes in the appendices? Any material which impedes the smooth development of your presentation, but which is important to justify the results of a thesis. Generally it is material that is of too nitty-gritty a level of detail for inclusion in the main body of the thesis, but which should be available for perusal by the examiners to convince them sufficiently. Examples include program listings, immense tables of data, lengthy mathematical proofs or derivations, etc.

Comments on the Skeleton

Again, the thesis is a formal document designed to address the examiner's two main questions. Sections 3 and 4 show that you have chosen a good problem, and section 5 shows that you solved it. Sections 1 and 2 lead the reader into the problem, and section 6 highlights the main knowledge generated by the whole exercise.

Note also that everything that *others* did is carefully separated from everything that *you* did. Knowing who did what is important to the examiners. Section 4, the problem statement, is the obvious dividing line. That's the main reason for putting it in the middle in this formal document.

Getting Started

The best way to get started on your thesis is to prepare an extended outline. You begin by making up the Table of Contents, listing each section and subsection that you propose to include. For each section and subsection, write a brief point-form description of the contents of that section. The entire outline might be 2 to 5 pages long. Now you and your thesis supervisor should carefully review this outline: is there unnecessary material (i.e. not directly related to the problem statement)? Then remove. Is there missing material? Then add. It is much less painful and more time-efficient to make such decisions early, during the outline phase, rather than after you've already done a lot of writing which has to be thrown away.

How Long Does it Take to Write a Thesis?

Longer than you think. Even after the research itself is all done -- models built, calculations complete -- it is wise to allow at least one complete term for writing the thesis. It's not the physical act of typing that takes so long, it's the fact that writing the thesis requires the complete organization of your arguments and results. It's during this formalization of your results into a well-organized thesis document capable of withstanding the scrutiny of expert examiners that you discover weaknesses. It's fixing those weaknesses that takes time.

This is also probably the first time that your supervisor has seen the formal expression of concepts that may have been approved previously in an informal manner. Now is when you discover any misunderstandings or shortcomings in the informal agreements. It takes time to fix these. Students for whom English is not the mother tongue may have difficulty in getting ideas across, so that numerous revisions are required. And, truth be known, supervisors are sometimes not quick at reviewing and returning drafts.

Bottom line: leave yourself enough time. A rush job has painful consequences at the defence.

Tips

Always keep the reader's backgrounds in mind. Who is your audience? How much can you reasonably expect them to know about the subject before picking up your thesis? Usually they are pretty knowledgeable about the general problem, but they haven't been intimately involved with the details over the last couple of years like you have: spell difficult new concepts out clearly. It sometimes helps to mentally picture a real person that you know who has the appropriate background, and to imagine that you are explaining your ideas directly to that person.

Don't make the readers work too hard! **This is fundamentally important. You know what few questions the examiners need answers for (see above). Choose section titles and wordings to clearly give them this information. The harder they have to work to ferret out your problem, your defence of the problem, your answer to the problem, your conclusions and contributions, the worse mood they will be in, and the more likely that your thesis will need major revisions.**

A corollary of the above: *it's impossible to be too clear*! Spell things out carefully, highlight important parts by appropriate titles etc. There's a huge amount of information in a thesis: make sure you direct the readers to the answers to the important questions.

Remember that *a thesis is not a story*: it usually doesn't follow the chronology of things that you tried. It's a formal document designed to answer only a few major questions.

Avoid using phrases like "Clearly, this is the case..." or "Obviously, it follows that ..."; these imply that, if the readers don't understand, then they must be stupid. They might not have understood because you explained it poorly.

Avoid *red flags*, claims (like "software is the most important part of a computer system") that are really only your personal opinion and not substantiated by the literature or the solution you have presented. Examiners like to pick on sentences like that and ask questions like, "Can you demonstrate that software is the most important part of a computer system?"

A Note on Computer Programs and Other Prototypes

The purpose of your thesis is to clearly document an original contribution to *knowledge.* You may develop computer programs, prototypes, or other tools as a means of proving your points, but remember, the thesis is *not* about the tool, it is about the contribution to knowledge. Tools such as computer programs are fine and useful products, but you can't get an advanced degree just for the tool. You must use the tool to demonstrate that you have made an original contribution to knowledge; e.g., through its use, or ideas it embodies.

7.4.7 TIPS for surviving a thesis defence

The thesis defence or viva is like an examination in some ways. It is different in many ways, however. The chief difference is that *the candidate usually knows more about the syllabus than do the examiners*.

Some questions will be sincere questions: the asker asks because s/he doesn't know and expects that the candidate will be able to rectify this. Students often expect questions to be difficult and attacking, and answer them accordingly. Often the questions will be much simpler than you expect.

In a curious relativistic effect, time expands in the mind of the student. A few seconds pause to reflect before answering seems eminently reasonable to the panel, but to the defender it seems like minutes of mute failure. *Take your time*.

For the same reason, let them take their time. Let them finish the question.

The phrase "That's a good question" is exceedingly useful. It flatters the asker and may get him/her onside, or less offside; it gives you time to think; it implies that you have understood the question and assessed it already and that you have probably thought about it before. If necessary, it can be followed by a bit more stalling "Now the answer to that is not obvious/straightforward..." which has the same advantages.

If the nightmare ever did come true, and some questioner found a question that put something in the work in doubt... mind you this is thankfully very rare.... then what? Well the first thing would be to concede that the question imposes a serious limitation on the applicability of the work "Well you have identified a serious limitation in this technique, and the results have to be interpreted in the light of that observation". The questioner is then more likely to back off and even help answer it, whereas a straight denial may encourage him/her to pursue more ardently. Then go through the argument in detail - showing listeners how serious it is while giving yourself time to find flaws in it or to limit the damage that will ensue. In the worst caese, one would then think of what can be saved. But all this is hypothetical because this won't happen.

What usually happens is that the examiners have read the work perhaps twice, and looked closely at some parts that interested them most. These are usually the good bits. They are not out to fail you. *It is a lot more complicated to fail you than to pass you.* In general, they feel good about the idea of a new, fresh researcher coming into their area. You are no immediate threat to them. They have to show that they have read it and they have to give you the opportunity to show that you understand it (you do, of course). And they usually have a genuine interest in the work. Some of them may feel it is necessary to maintain their image as senior scholars and founts of wisdom. Judicious use of the "Good question", "Yes, you're right of course", "Good idea.." and "Thanks for that" will allow that with a minimum of fuss and a maximum of time for champagne drinking.

If one of the examiners is a real ..., your thesis defence is probably not the best place and time in which to do anything about it, except perhaps for allowing him/her to demonstrate it clearly and thus to establish the support of the rest of the panel. If you want a major dispute, save it up for when you are on even ground, unless you are very very sure of yourself and think that you have nothing to lose.

Be ready for a 'free kick'. It is relatively common that a panel will ask one (or more) questions that, whatever the actual wording may be, are essentially an invitation to you to tell them (briefly) what is important, new and good in your thesis. You ought not stumble at this stage, so you should rehearse this. You should be able to produce on demand (say) a one minute speech and a five minute speech, and be prepared to extend them if invited by further questions. Do not try to recite your abstract: written and spoken styles should be rather different. Rather, rehearse answers to the questions: "What is your thesis about, and what have you done that merits a PhD?".

Read points the first two bullet points again. Keep calm - and good luck!

WORKSHOP – GRAMMER AND PUNCTUATIONS

Common Errors: Editing Checklist

1. Sentence Fragment

Make sure each word group you have punctuated as a sentence contains a grammatically complete and independent thought that can stand alone as an acceptable sentence.

Sentence Fragment	How to fix it
Tests of the Shroud of Turin have produced some curious findings. For example, the pollen of forty-eight plants native to Europe and the Middle East. [2nd sentence = fragment]	Tests of the Shroud of Turin have produced some curious findings. For example, the cloth contains the pollen of forty-eight plants native to Europe and the Middle East.
Scientists report no human deaths due to excessive caffeine consumption. Although caffeine does cause convulsions and death in certain animals. [2nd sentence = fragment]	Scientists report no human deaths due to excessive caffeine consumption, although caffeine does cause convulsions and death in certain animals.

2. Sentence Sprawl

Too many equally weighted phrases and clauses produce tiresome sentences.

Sentence sprawl	How to fix it
The hearing was planned for Monday, December 2, but not all of the witnesses could be available, so it was rescheduled for the following Friday, and then all the witnesses could attend. [There are no grammatical errors here, but the sprawling sentence does not communicate clearly and concisely.]	The hearing, which had been planned for Monday, December 2, was rescheduled for the following Friday so that all witnesses would be able to attend.

3. Misplaced and dangling modifiers

Place modifiers near the words they describe; be sure the modified words actually appear in the sentence.

Misplaced or Dangling Modifier	How to fix it
When writing a proposal, an original task is set for research.	When writing a proposal, a scholar sets an original task for research.
Many tourists visit Arlington National Cemetery, where veterans and military personnel are buried every day from 9:00 a.m. until 5:00 p.m.	Every day from 9:00 a.m. until 5:00 p.m., many tourists visit Arlington National Cemetery, where veterans and military personnel are buried.

4. Faulty parallelism

Be sure you use grammatically equal sentence elements to express two or more matching ideas or items in a series.

Faulty parallelism	How to fix it

The candidate's goals include winning the election, a national health program, and the educational system.	The candidate's goals include winning the election, enacting a national health program, and improving the educational system.
Some critics are not so much opposed to capital punishment as postponing it for so long.	Some critics are not so much opposed to capital punishment as they are to postponing it for so long.

5. Unclear pronoun reference

All pronouns must clearly refer to definite referents [nouns].

Use **it**, **they**, **this**, **that**, **these**, **those**, and **which** carefully to prevent confusion.

Unclear pronoun reference	How to fix it
Einstein was a brilliant mathematician. This is how he was able to explain the universe.	Einstein, who was a brilliant mathematician, used his ability with numbers to explain the universe.
Because Senator Martin is less interested in the environment than in economic development, he sometimes neglects it.	Because of his interest in economic development, Senator Martin sometimes neglects the environment.

6. Incorrect pronoun case

Determine whether the pronoun is being used as a subject, or an object, or a possessive in the sentence, and select the pronoun form to match.

Incorrect pronoun case	How to fix it
Castro's communist principles inevitably led to an ideological conflict between he and President Kennedy.	Castro's communist principles inevitably led to an ideological conflict between him and President Kennedy.
Because strict constructionists recommend fidelity to the Constitution as written, no one objects more than them to judicial reinterpretation.	Because strict constructionists recommend fidelity to the Constitution as written, no one objects more than they [do] to judicial reinterpretation.

7. Omitted commas

Use commas to signal nonrestrictive or nonessential material, to prevent confusion, and to indicate relationships among ideas and sentence parts.

Omitted commas	How to fix them
When it comes to eating people differ in their tastes.	When it comes to eating, people differ in their tastes.
The Huns who were Mongolian invaded Gaul in 451.	The Huns, who were Mongolian, invaded Gaul in 451. ["Who were Mongolian" adds information but does not change the core meaning of the sentence because Huns were a Mongolian people; this material is therefore nonrestrictive or nonessential.]

8. Superfluous commas

Unnecessary commas make sentences difficult to read.

Superfluous commas	How to fix them
Field trips are required, in several courses, such as, botany and geology.	Field trips are required in several courses, such as botany and geology.
The term, "scientific illiteracy," has become almost a cliche, in educational circles.	The term "scientific illiteracy" has become almost a cliche in educational circles

9. Comma splices

Do not link two independent clauses with a comma (unless you also use a coordinating conjunction: **and**, **or**, **but**, **for**, **nor**, **so**, **yet**).

Instead use a period or semicolon, or rewrite the sentence.

Comma splice	How to fix it
In 1952 Japan's gross national product was one third that of France, by the late 1970s it was larger than the GNPs of France and Britain combined.	In 1952 Japan's gross national product was one third that of France. By the late 1970s it was larger than the GNPs of France and Britain combined.
Diseased coronary arteries are often surgically bypassed, however half of all bypass grafts fail within ten years.	Diseased coronary arteries are often surgically bypassed; however, half of all bypass grafts fail within ten years.

10. Apostrophe errors

Apostrophes indicate possession for nouns ("**Jim's hat**," "**several years' work**") but not for personal pronouns (**its**, **your**, **their**, and **whose**).

Apostrophes also indicate omissions in contractions ("**it's**" = "**it is**").

In general, they are not used to indicate plurals.

Apostrophe error	How to fix it
In the current conflict its uncertain who's borders their contesting.	In the current conflict it is [it's] uncertain whose borders they are [they're] contesting.
The Aztecs ritual's of renewal increased in frequency over the course of time.	The Aztecs' rituals of renewal increased in frequency over the course of time.

11. Words easily confused

"Effect" is most often a noun (the effect), and "affect" is almost always a verb.

Other pairs commonly confused: "lead"/"led" and "accept"/"except."

Check a glossary of usage to find the right choice.

Confused word	How to fix it
The recession had a negative **affect** on sales.	The recession had a negative **effect** on sales. (or) The recession **affected** sales negatively.
The laboratory instructor chose not to offer detailed **advise**.	The laboratory instructor chose not to offer detailed **advice**.

12. Misspellings

Spelling errors are usually perceived as a reflection of the writer's careless attitude toward the whole project.

Don't allow your hard work to be marred in this way!

In addition to comprehensive dictionaries, you may want to use electronic spell checks, spelling dictionaries, and lists of frequently misspelled words found in handbooks.

Before You Proofread

Be sure you've revised the larger aspects of your text. Don't make corrections at the sentence and word level if you still need to work on the focus, organization, and development of the whole paper, of sections, or of paragraphs.
Set your text aside for a while (15 minutes, a day, a week) between writing and proofing. Some distance from the text will help you see mistakes more easily.
Eliminate unnecessary words before looking for mistakes. See the Writing Center handout How to Write Clear, Concise, Direct Sentences.

Know what to look for. From the comments of your professors or a Writing Center instructor on past papers, make a list of mistakes you need to watch for.

When You Proofread

Work from a printout, not the computer screen. (But see below for computer functions that can help you find some kinds of mistakes.)
Read out loud. This is especially helpful for spotting run-on sentences, but you'll also hear other problems that you may not see when reading silently.

Use a blank sheet of paper to cover up the lines below the one you're reading. This technique keeps you from skipping ahead of possible mistakes.
Use the search function of the computer to find mistakes you're likely to make. Search for "it," for instance, if you confuse "its" and "it's;" for "-ing" if dangling modifiers are a problem; for opening parentheses or quote marks if you tend to leave out the closing ones.
If you tend to make many mistakes, check separately for each kind of error, moving from the most to the least important, and following whatever technique works best for you to identify that kind of mistake. For instance, read through once (backwards, sentence by sentence) to check for fragments; read through again (forward) to be sure subjects and verbs agree, and again (perhaps using a computer search for "this," "it," and "they") to trace pronouns to antecedents.
End with a spelling check, using a computer spelling checker or reading backwards word by word. But remember that a spelling checker won't catch mistakes with homonyms (e.g., "they're," "their," "there") or certain typos (like "he" for "the").

Subject – Verb Agreement

When to check for subject-verb agreement?

Watch subject-verb agreement:	Reason	Examples
When the subject follows the verb	When the subject follows the verb (especially in sentences beginning with the expletives "there is" or "there are"), special care is needed to determine the subject and to make certain that the verb agrees with it.	On the wall were several posters. There are many possible candidates. There is only one good candidate.
When the expletive "it" is the subject	The expletive "it" is always followed by a singular verb.	It is my car which stalls. It is their cars which stall.
When words like "each" are the subject	When used as subjects, words such as each, either, neither another anyone, anybody, anything someone, somebody, something one, everyone everybody, everything no one, nobody, nothing take singular verbs. Do not be confused by prepositional phrases which come between a subject and its verb. They do not change the number of the subject.	Each takes her turn at rowing. Neither likes the friends of the other. Everyone in the fraternity has his own set of prejudices. Each of the rowers takes her turn at rowing. Every one of the fraternity members has his own set of prejudices. top

When words like "none" are the subject	Other words such as none, any, all more, most, some may take either singular or plural verbs, depending on the context.	Some of the dollar was spent. Some of the dollars were spent. [Note: here the prepositional phrase does affect the subject. It tells you whether you are talking about a part of one thing (singular) or about a number of things (plural).]
When the subjects are joined by "and"	Subjects joined by "and" take plural verbs. Be aware: phrases such as "in addition to," "as well as," and "along with" do not mean the same thing as "and." When inserted between the subject and the verb, these phrases do not change the number of the subject.	Both Tom and Jane have English 167 papers due on Tuesday. Tom, as well as Jane, has an English 207 paper due Tuesday.
When singular subjects are joined by words like "or"	Singular subjects joined by "or," "nor," "either . . . or," or "neither. . . nor" take a singular verb.	Either the man or his wife knows the truth of the matter. Neither money nor power was important any longer.
When one subject is singular and one plural	If one subject is singular and one is plural, the verb agrees with the nearer subject.	Neither the television nor the radios work. Neither the radios nor the television works. top
When a linking verb is used	A linking verb ("is," "are," "was," "were," "seem" and others) agrees with its subject, not its complement.	Joe's favorite dessert is blueberry muffins. Blueberry muffins are Joe's favorite dessert.
When a collective noun is used	When regarded as a unit, collective nouns, as well as noun phrases denoting quantity, take singular verbs.	The whole family is active. (Family is a collective noun regarded as a unit.) The family have met their various obligations. (The individuals of the family are regarded separately.) A thousand bushels is a good yield. (a quantity or unit) A thousand bushels were crated. (individual bushels)

		top
When a relative pronoun is used as a subject of an adjective clause	A relative pronoun ("who," "which," or "that") used as a subject of an adjective clause takes either a singular or plural verb in order to agree with its antecedent.	A vegetable that contains DDT can be harmful. (Adjective clause modifying the singular noun "vegetable.") Vegetables that contain DDT can be harmful. (Adjective clause modifying the plural noun "vegetables.") Mary is one of the students who have done honor to the college. (Adjective clause modifies the plural noun "students." "Students" is the antecedent of "who." In the above sentence Mary is just one of the students. So at least two students have done honor to the college. Compare that to: Mary is the only one of our students who has achieved national recognition. In this case, "one," not "students," is the antecedent of "who." Compare to the sentence above: Of all our students, Mary is the only one who has achieved national recognition.

Using Conjunctive Adverbs

Use conjunctive adverbs (or sentence adverbs) to:

indicate a connection between two independent clauses in one sentence
link the ideas in two or more sentences
show relationships between ideas within an independent clause.

Here are some examples of conjunctive adverbs:

also	indeed	then	hence
however	similarly	furthermore	nevertheless
otherwise	finally	moreover	thus
consequently	likewise	therefore	nonetheless

How to punctuate conjunctive adverbs

Guideline	Examples
When a conjunctive adverb connects two independent clauses in one sentence, it is preceded by a semicolon and followed by a comma.	Tuition increases, say officials, are driven by the universities' costs; consequently, tuition income typically covers less than 50% of college budgets.
If a conjunctive adverb is used in any other position in a sentence, it is set off by commas.	Nonetheless, some colleges are making efforts to trim budgets and pass along the savings. Secretary Bennett, however, maintains that more federal aid would only encourage universities to count on the government to meet any increases they might impose.

Using Coordinating Conjunctions

Coordinating conjunctions join grammatically similar elements (two nouns, two verbs, two modifiers, two independent clauses).

and	nor	but	yet
or	so	for	

Use coordinating conjunctions to indicate that the elements they join are equal in importance and in structure.

How to punctuate coordinating conjunctions

Guideline	Example
When a coordinating conjunction joins two independent clauses, a comma is used before the coordinating conjunction (unless the two independent clauses are very short). Conjunctions that are not followed by non-essential elements should never be followed by commas.	Perhaps no budget is without some fat, but university officials argue that their unique function requires special standards of evaluation.
When either independent clause in a compound sentence contains a comma to set off introductory or	The figures at elite universities, particularly, are enough to cause sticker shock; yet the current

non-essential elements, a reader may be confused by a comma before a coordinating conjunction. In this case, a semicolon may replace the comma.	increases at many schools are the lowest in a decade.

When NOT to punctuate coordinating conjunctions

Guideline	Examples
If a sentence begins with a coordinating conjunction, it is **not** followed by a comma.	Yet the typical tenured professor's salary of $43,500 still represents 10% less buying power than the equivalent salary in 1970.
Commas are **not** used between two verbs, two subjects, two complements, or two objects joined by a coordinating conjunction.	That confuses most analogies between universities and profit-making enterprises. [compound object of preposition] Endowments and gifts make up the rest. [compound subject] Georgetown, for example, has eliminated one-third of its graduate programs in the past five years and recently decided to close its dental school. [compound verb]

When to use a Semicolon

Semicolons help you connect closely related ideas when a style mark stronger than a comma is needed. By using semicolons effectively, you can make your writing sound more sophisticated.

Use a Semicolon to . . .	Reason	Example
Link two independent clauses	To connect closely related ideas	Some people write with a word processor**;** others write with a pen or pencil.
Link clauses connected by conjunctive adverbs or transitional phrases	To connect closely related ideas	But however they choose to write, people are allowed to make their own decisions**;** as a result, many people swear by their writing methods.
Link lists where the items contain commas	To avoid confusion between list items	There are basically two ways to write: with a pen or pencil, which is inexpensive and easily accessible**;** or by computer and printer, which is more expensive but quick and neat.
Link lengthy clauses or clauses with commas	To avoid confusion between clauses	Some people write with a word processor, typewriter, or a computer**;** but others, for different reasons, choose to write with a pen or pencil.

Rules for Using Semicolons

Rule	Example
A semicolon is most commonly used to link (in a single sentence) two independent clauses that are closely related in thought. When a semicolon is used to join two or more ideas (parts) in a sentence, those ideas are then given equal position or rank.	Some people write with a word processor**;** others write with a pen or pencil.
Use a semicolon between two independent clauses that are connected by conjunctive adverbs or transitional phrases.	But however they choose to write, people are allowed to make their own decisions**;** as a result, many people swear by their writing methods.
Use a semicolon between items in a list or series if any of the items contain commas.	There are basically two ways to write: with a pen or pencil, which is inexpensive and easily accessible**;** or by computer and printer, which is more expensive but quick and neat.
Use a semicolon between independent clauses joined by a coordinating conjunction if the clauses are already punctuated with commas or if the clauses are lengthy.	Some people write with a word processor, typewriter, or a computer**;** but others, for different reasons, choose to write with a pen or pencil.

Common Mistakes to Avoid

Avoid using a comma when a semicolon is needed:

Incorrect Comma Use	Why It's Wrong	Correct Semicolon Use
The cow is brown, it is also old.	Both parts of the sentence are independent clauses, and commas should **not** be used to connect independent clauses if there is no coordinating conjunction. This mistake is known as a **comma splice.**	The cow is brown**;** it is also old.
I like cows, however, I hate the way they smell.	The conjunctive adverb **however** signals a connection between two independent clauses, and commas should **not** be used to connect independent clauses if there is no coordinating conjunction.	I like cows**;** however, I hate the way they smell
I like cows: they give us milk, which tastes good, they give us beef, which also tastes good, and they give us leather, which is used for shoes and coats.	It's unclear what the three list items are, since the items are separated by commas.	I like cows: they give us milk, which tastes good**;** they give us beef, which also tastes good**;** and they give us leather, which is used for shoes and coats.
Cows, though their bovine majesty has been on the wane in recent millenia, are still one of the great species of this planet, domesticated, yet proud, they ruminate silently as we humans pass	It's unclear where the first independent clause ends and the second independent clause begins.	Cows, though their bovine majesty has been on the wane in recent millenia, are still one of the great species of this planet**;** domesticated, yet proud, they ruminate silently as we humans pass tumultuously

tumultuously by.		by.

Avoid using a semicolon when a comma is needed:

Incorrect Semicolon Use	Why It's Wrong	Correct Comma Use
The cow is brown; but not old.	The coordinating conjunction **but** doesn't require a semicolon, since the second part of the sentence isn't an independent clause.	The cow is brown**,** but not old.
Because cows smell; they offend me.	The first part is not an independent clause, so no semicolon is required.	Because cows smell**,** they offend me.

Using Commas

Use the guidelines below to learn how to use commas effectively in your writing.

Using commas to punctuate restrictive and non-restrictive modifiers

Use commas to set off **non-restrictive** modifiers.

Do not use commas to set off **restrictive** modifiers.

Non-restrictive vs. restrictive modifiers

non-restrictive modifier adds information that is not essential to our understanding of the sentence; if we remove it from the sentence, the basic meaning of the sentence does not change.

A **restrictive** modifier identifies, or limits the reference of, the noun it modifies.

The table below illustrates these definitions.

Type of modifier	Example	Explanation
Non-restrictive	The statue of his mother by Joseph Smith, dated 1894, sold for over a million dollars.	The date of Joseph Smith's statue can be removed from the sentence without altering its meaning: The statue of his mother by Joseph Smith sold for over a million dollars.
Restrictive	The painting dated 1894 is a forgery; the one dated 1892 is genuine.	The phrases "dated 1894" and "dated 1892" cannot be detached from the sentence without making the meaning unclear: The painting [which one?] is a forgery; the one [which one?] is genuine.

Further Examples

Note the distinction and the different punctuation in the following pairs of sentences:

Non-restrictive	Restrictive
William Carlos Williams, the poet, was also a farmer.	The poet William Carlos Williams was also a farmer.
John, who has been drinking, should not drive.	People who have been drinking should not drive.
Many Americans travel to Mexico, where Laetrile is legal and readily available.	Many Americans travel to countries where Laetrile is legal and readily available.
In spring, when the water is high, the lake surges over the rocks.	At times when the water is high the lake surges over the rocks.
The waiters, dressed in their white jackets, are already arranging the chairs on the sidewalk.	The waiters dressed in white jackets serve in the main dining room; those in red serve in the coffee shop.

Dashes

Dashes, when used sparingly and correctly, can be used to make your writing sound more sophisticated.

Use dashes to . . .	Reason	Example
Indicate sudden changes in tone or thought within a sentence	To emphasize the contradiction between ideas	There is an illness in many foreign services--the people in them are only good at following instructions.* I am under the impression that she has no instructions at all--and doesn't need any.* The exuberant--I should say lunatic--quality of his ravings electrified the crowd. *
Set off some sentence elements	To insert parenthetical commentary while emphasizing their importance (Parentheses tend to diminish the importance of what's enclosed in them)	Over a candlelit dinner last month at Spaso House, the ambassadorial residence in Moscow, Robert Strauss and his wife Helen listened as two Senators--Republican Robert C. Smith of New Hampshire and Democrat John Kerry of Massachusetts--agreed that the way to bring American audiences "out of their chairs" these days was simply to say, in Smith's words, "We won the cold war, and we're not going to send one dime in aid to Russia."* Strauss favors--as does, sotto voce, the Administration--early admission of Russia to the International Monetary Fund.*
Create emphasis	To connect ideas strongly to each other	To feed, clothe, and find shelter for the needy--these are real achievements.