

# Summary of AK2030 Theory and Methodology of Science

Yashar Honarmandi  
yasharh@kth.se

November 9, 2020

## **Abstract**

This is a summary of AK2030 Theory and Methodology of Science.

This course, rather than teaching any particular aspect of science, will focus on general methodological and epistemological issues when approaching problems in science.

# Contents

1	General Philosophy of Science	1
2	Methodology	8

# 1 General Philosophy of Science

**Definitions** Definitions explain the meaning of words. They consist of a definiendum (what is to be defined), a definiens (that which defines the definiendum) and a defining connective (which describes how the two are connected. Words such as “is” as well as logical comparisons are in this category). In addition, there may be delimiters which explain that the definiendum is an aspect of something else. For instance, the definition “A system is spherically symmetric if it looks the same from all directions,” “spherically symmetric” is the definiendum and “A system is” is the delimiter.

**Lexical Definitions** A lexical definition is a definition based on a common understanding of a word or phrase.

**Stipulative Definitions** A stipulative definition is a definition based on reasons and arguments for choosing it in a particular way.

**Ambiguity** Ambiguity is the presence of two possible meanings of the same word.

**Vagueness** Vagueness is the unclarity of the limits of what is within the scope of a definition.

**Knowledge** Knowledge is true, justified belief. These three conditions are necessary, but not always sufficient. We will nevertheless proceed with this definition.

This definition is compatible with uncertainty, and recognizing uncertain knowledge is in fact vital in science. Naturally there are degrees of uncertainty.

In science, the justification of belief is central in the social aspects of its practice. An ideally justified belief requires consideration of all relevant reasons for that belief, and science is publically defended exactly to adhere to this ideal.

**Definitions Pertaining to Phenomena** A phenomenon  $X$  is

- predicted if one can know at what time  $X$  will occur.
- explained if the causes of  $X$  are known.
- designable if it is known to satisfy certain properties.

**Instrumentalism and Realism** Do scientific theories constitute claims that are true or false? Realists say yes, whereas instrumentalists say that theory only orders observations and do not in themselves contain truth.

**Inference** Inference is the logical process of going from a set of facts and/or evidence to some conclusion.

**Inductive and Deductive Reasoning** Inductive reasoning is the process of going from a set of observations to a general claim. Such reasoning is an attempt at extending our knowledge beyond what can be directly observed. Deductive reasoning is the process of explicating or extracting knowledge about something particular from previously possessed knowledge.

Inductive reasoning is fallible - even the best of efforts at induction might still overlook something for which one has no evidence. Deductive reasoning, on the other hand, is infallible - deduction necessitates that you have sufficient knowledge, and proper reasoning based on sufficient knowledge preserves the truth of your knowledge.

**Examples of Inference** Some examples of ways to infer follow. A few ways of inductive inference are

- Direct inference, where a property is extrapolated from a finite set of observations.
- Projection, where a sequence or trend is used to extrapolate a future observation.
- Generalization, where a hypothesis is formed based on a finite set of observations.

A few ways of deductive inference are

- Modus ponens, in which a set of assumptions  $A$  is combined with the idea that if  $A$  is true then their implications  $B$  are true to infer that  $B$  is true.
- Modus tollens, in which a conditional  $A \rightarrow B$  is combined with the observation that  $B$  is false to infer that  $A$  is false.

**Hume's Problem** As there exists an infinite set of inference rules, in particular inductive inference rules, hence the conclusions of any particular inference are only justified if the choice of inference rule is justified. Hume's problem discusses the justification of inference rules.

Hume's argument was the following:

1. Assume that inference is either deductive or inductive.
2. Assume that in order to justify an inference rule, it has to be inferred from some set of principles.
3. Assume that inference rules cannot be deduced, as past and previous inference need not be connected,
4. Thus, inference rules are arrived at inductively.

The implication of the latter statement is that there must be a set of inference rules allowing you to infer the validity of the rule you want use. But this set of inference rules is not necessarily valid in itself. The result is a so-called infinite regress of the same problem of justification, which according to Hume could only be resolved if no inference rule is justified.

**Solving Hume's Problem** Hume's problem would imply that the practice of science is itself irrational. This would be a complete disaster, and must therefore be resolved.

One way to solve this came from Karl Popper, who denied the premise that science rests on using inference rules. This will be discussed later.

Another solution is to deny the impact of Hume's argument by studying what offers justification, in the hope that this is not destroyed by the infinite regress. Foundationalism asserts that there exists a foundation upon which justification is built. Coherentism is an alternative, which desires for claims to be justified from some coherent system containing the set of previously accepted claims. In this view, inference rules are good if they are coherent with the rest of scientific knowledge. The result is that inference rules are merely manifestations of coherence and a way to connect different practices in a coherent manner.

**The Hypothetico-Deductive Method** The hypothetico-deductive method consists of

1. Formulating a hypothesis.
2. Deducing consequences of the hypothesis.
3. Test whether the consequences are true.
4. If the consequences are false, reject the hypothesis deductively. Otherwise, strengthen your confidence in the hypothesis inductively.

There are some requirements on the involved parts. A good hypothesis must be

- Either true or false.
- Not a tautology, i.e. something that is necessarily true or false by definition.
- Contain some generalization or discussion of the unobservable.

A good set of consequences must be

- Observable.
- Follow from valid deduction.
- Relevant for the hypothesis. This part informs what kinds of experiment you want to do. For instance, if the hypothesis is  $A$  implies  $B$ , then you would need to find cases where  $A$  is true and  $B$  is false at the same time to reject the hypothesis. In other words, studying cases where  $A$  is false or  $B$  is true does not help.

**Falsifiability and Falsification** A theory or hypothesis is falsifiable if it has implications the truth or falsehood of which can be observed. The act of observing an implication to be false is called falsification.

**Popper's Falsificationism** Popper attempted to circumvent Hume's problem by rejecting the ability of science to strengthen hypotheses, leaving it only with the ability to deductively falsify them. While this seems to solve the problem, some critiques of this solution are:

- It provides no way to give hypotheses confidence, in contrast to what seems reasonable and what is scientific practice.
- Along with an observation comes auxiliary hypotheses about the validity of the experiment. A particular falsification can only falsify the conjunction of the fundamental hypothesis and the auxiliary hypotheses, limiting our ability to perform modus tollens on the fundamental hypothesis. This is called the Duhem-Quine thesis.
- Hypotheses may be modified ad hoc solely to fit particular observations. Popper countered this by claiming that modifications that reduce falsifiability are ad hoc and should not be used, in contrast to other modifications, which might be necessary.

**Confidence** How can our confidence in a hypothesis be strengthened, and why does observations of consequences of a hypothesis increase our confidence in the hypothesis itself?

One answer might be found in the compatibility of the hypothesis and its consequences. However, relevance is also needed - otherwise, many observations might seemingly verify far-reaching conclusions. The issue of underdetermination is also prominent, as multiple hypotheses might be compatible with the same observations.

A better requirement is that the conclusions be unlikely if the hypothesis is false. Eliminating hypotheses based on this is called severe testing. While this is an improvement, one must first of all avoid double-counting evidence - if you use some piece of evidence to construct a hypothesis, you cannot use that evidence to verify your hypothesis. In addition, if there is low confidence in the hypothesis to begin with, severe testing will not sufficiently strengthen it.

**Frequentism** The frequentist view of probability is that it is the frequency with which some event occurs. As hypotheses are not events, it follows from this view that probabilities cannot be meaningfully assigned to hypotheses, limiting the possibility of quantitatively strengthening hypotheses.

**Observation** Loosely speaking, observation is the event of experience through the senses. It is key to all scientific practice. Observations can be divided into categories:

- Direct observation, which is unaided sensory experience.
- Aided direct observation, which uses instruments to amplify the senses.
- Indirect observation, in which an event is not directly observed and the event is experienced through its effect on its surroundings.

**Empiricism** Empiricism is the thesis that sensory knowledge is the ultimate basis for knowledge.

**Logical Empiricism and its Refutation** Logical empiricism is the position that theory and experiment can be separated, that is that theory does not inform experiment. If this were true, then any scientific theory would rest on a solid and independent foundation of experimental observation.

This position has been refuted. Its refutation is as follows: First of all, indirect observation always depends on theory, as theory dictates how to interpret how the indirect observations pertain to the underlying phenomena. Next, aided direct observation sometimes requires theory as instruments may distort the phenomena they are used to observe, and separating between instrumental effects and actual phenomena requires theory. Thus theory and observation cannot be separated.

**The Problem of Nomic Measurement** The circular relation between theory and observation is formulated in the problem of nomic measurement. Its statement is as follows:

- Assume that you want to measure property  $X$ , which is not directly observable.
- Assume that  $X$  may be inferred from  $Y$ , which is directly observable.
- For the inference to be performed, one needs some relation  $X = f(Y)$ . This relation cannot be obtained experimentally obtained as it would require  $X$  and  $Y$  to be known at the same time, violating the assumption.
- As  $f$  cannot be identified from measurement alone, it must come from theory. Thus theory and experiment are inextricably connected.

**Operationalization** Direct observation, while powerful in a sense, is generally error-prone and can often only give qualitative descriptions, hence indirect observational methods are usually preferred. The process of linking properties to directly observable effects is called operationalization of that property. The two are linked by some stable relation which guarantees that when the property is present, so is the observed effect. Using operationalization we may infer information about the property from observing the effect.

The quality criteria for an operationalization are

- That the underlying property be well-defined in order to allow for proper inference.
- That the relation between the property and the effect is valid.
- That the relation is sufficiently stable.
- That the effect is publically observable with sufficient precision.

**Operationalism** Observationalism is the position that everything is defined through the operations by which we observe or measure them. This is generally not a very helpful position for observationalization.

The criticisms of observationalism are

- Observationalism prohibits two differently performed measurements to pertain to the same underlying concept.
- Observationalism makes it impossible to criticize a measurement for not properly capturing an underlying concept.

**Measurement** The process of measurement consists of the following:

1. Define the concept that you would like to measure.
2. Operationalize the concept.
3. Specify a measure and define units of comparison.
4. Represent the results with numbers.

The requirements for a good measurement are:

- There must be a unit of comparison.
- The unit must be sufficiently stable.
- Anyone must have access to the same measure.

**Scales** The representation part of a measurement implies the need for scale. There are five kinds of scales:

1. Nominal scales, in which samples are given ID numbers.
2. Ordinal scales, in which objects are ordered according to some qualitative measure.
3. Interval scales, which allow the comparison of distances.
4. Ratio scales, which allow the comparison of both distances and ratios.
5. Absolute scales, which allow the comparison of absolute numbers, i.e. where the absolute numbers have a significance by themselves.

Different scales of the same type made to represent the same property must represent the same empirical structure. Each set of scales is defined by the set of allowed transformations, and combined with the previous this implies that what can be inferred from a measurement is limited by the allowed transformations. The allowed transformations are:

- Transformations that preserve uniqueness for nominal scales.
- Positive monotone transformations for ordinal scales.
- Positive linear transformations for interval scales.
- Positive scalar transformations for ratio scales.
- None for absolute scales.

**Measurement Error** The measurement error is defined as the difference between the measured and true value of some property. It comes in two forms: Systematic and random error.

Before introducing the two, we introduce the notions of precision and accuracy. Precision is the absence of variation, and accuracy is closeness to the true value. Systematic errors are thus inaccuracies, and random errors are imprecision.

**Reducing Error** Random error can be reduced by repeating measurements and performing new operationalization in such a way that you obtain more precise measurements.

Before discussing reducing systematic error, we introduce convergent and divergent validity. The former is that different ways to measure the same intended property should yield the same result, and the latter is that different ways to measure different intended properties should yield different results.

Now, to reduce systematic error a first approach is to identify the causes. The convergent and divergent validity of your results is one way to look. You can also work to reduce observer effects, as well as stabilize and standardize your measurements.

The control of error represents a set of auxiliary hypotheses which enter in the hypothetico-deductive method.

**Experiments** An experiment is a controlled observation in which the observer manipulates the real variables that are believed to influence the outcome, both for the purpose of intervention and control. Its purpose is to justify accepting or rejecting a hypothesis. Its characteristics are:

- manipulation.
- intervention with the independent variables.
- control of disturbing factors.
- observation.

**Mill's Method of Difference** Mill's method of difference describes the logic and procedure of experiments. It has the following steps:

1. Ask what causes a phenomenon  $E$ .
2. Conjecture that  $C$  is the cause.
3. Produce situations  $S_1$  and  $S_2$  in which neither  $C$  nor  $E$  occur and such that all relevant causal factors are the same.
4. Activate  $C$  in  $S_1$  only.
5. Observe  $E$  in  $S_1$  only.
6. Assert that something causes  $E$  in  $S_1$  and that nothing causes  $E$  in  $S_2$ .
7. As the two only differ by  $C$ , conclude that  $C$  causes  $E$ .

Baked into these steps are a number of assertions and ideas which might be difficult to realize - for instance, it might be hard to know what factors are causal. This necessitates good design.

**Randomization** Randomization is the process of dividing samples into treatment and control groups using randomness. Experiments set up this way are called randomized controlled trials.

Randomness is desired because it might eliminate selection bias, help convince others that your experiment is not rigged and blinds the identity of the treatment from both investigators and subjects. However, randomization is not the only way to achieve these goals.

Randomization does not help to equally distribute background factors between the treatment and control groups. Imbalances may instead be avoided by being checked and controlled for. One might also use so-called stratified randomization, where background factors are included in the randomization process.

Randomization has no bearing of the broadness of your sample, meaning that it adds nothing to the generalizability of your results to other populations than that which you have sampled.

**Repeatability, Reproducibility and Replicability** If an experiment is described such that it contains enough information that others can repeat it, the experiment is repeatable. If a competent repetition of the experiment yields the same result, the experiment is reproducible. If a competent independent experiment using independent data, methods and experimental infrastructure yields the same results, the experiment is said to be replicable. These categories will be important when looking for errors in experimental design.

**Design Errors** An erroneous design may manifest as:

- Failure to control for all relevant factors.
- Confirmation bias(/interpretation problem) in observations.
- The observer effect(/influence problem) spoiling your results.
- The placebo effect interfering.
- Selection bias in your treatment and control groups.

If an experiment sufficiently controls for such errors, conclusions drawn from the experiment are said to be internally valid.

To detect such errors, one might employ what knowledge one possesses, of either theory or experimental practice, to identify them, or one might investigate the repeatability, reproducibility and replicability of the experiment to find out where it might have gone wrong.



**Experimental Control** Experimental control consists of identifying relevant features that might interfere with your experiment and being able to influence these features such that alternative explanations can be ruled out.

A few methods for implementing experimental control are:

- Dividing your subjects into treatment and control groups.
- Keeping other factors constant by finding or creating situations with the same background.
- Eliminating background factors altogether by creating special circumstances. One way to do this is to blind your study. There is single blinding, where the subjects do not know whether they are in the treatment or control group, and double blinding, where neither the subject nor the experimenter knows to which group the subject belongs.
- Separate out external factors by some special experimental construction.

**Non-Experimental Procedures** I briefly mention some examples of non-theoretical activities in science that are not experiments. The first is observational studies, where you study subjects with no ability to manipulate, intervene or control the subjects. Two others are natural experiments and field experiments, in which you have no ability to manipulate or intervene, but by construction you may still control for some factors. A final one is simulation, where you have the ability to both manipulate, intervene and control, but only on a representation of your system.

**Models** Models are important tools in science used to help us describe the world. They have different aspects, and may be approached through any of them. These are:

- Representations, where the models attempt to represent some target in the real world to some extent. This is useful when studying the real-world target might be impossible or infeasible, morally or ethically prohibited or cognitively difficult.
- Idealizations, where the model contains only the relevant aspects of the real-world target. Models are generally analogies of the real-world target. These may be positive (containing similar aspects), negative (containing idealizations) or neutral (containing descriptions of things that cannot be known in the target).
- Purpose-dependent tools - some models describe certain properties better than others.
- Things to be manipulated. The kind of analogy provided by the model limits the available set of manipulations - if some property is idealized away in the model, then investigating that property in the model has no bearing on the target. Manipulations may also be used to reveal neutral analogies as either positive or negative.

**Models and Experiments** Models and experiments are similar in that they contain some variables, the manipulation of something and the observation of the effect. However, they differ in the type of errors that can be made. Experiments rely on internal validity, whereas this is less of a problem for models. By contrast, models have greater issues with justifying neutral analogies.

**Quality Criteria for Models** Models do not necessarily have quality criteria that apply equally well in all context, and in fact these often trade off with one another. Nevertheless, some criteria are:

- Similarity to the target with respect to relevant properties. This may even be termed accuracy.
- Robustness. A model is robust with respect to some condition if changing it does not change the result.
- Precision. A model is more precise than another if the specification of the former imply the specification of the latter.
- Simplicity. A model is simpler than another if it contains fewer variables, fewer parameters and fewer observations.
- Tractability. A model is tractable with respect to some principles if the results can be obtained by applying the set of principles to the model.
- Transparency. A model is transparent if the user is capable of understanding how the results are produced.

**Modelling Strategies** I here introduce two modelling strategies. Their approaches reflect how we learn from models.

The first is to consider models as mirrors of the world. In this view you desire your model to be precise, a feature that you typically gain at the cost of simplicity, tractability and transparency. You will still not gain enough precision to avoid issues of external validity, i.e. to guarantee that the model describes the target.

The other view is to view models as isolations of relevant features of a system. This view requires that the target be divisible, at least to some extent, in such a way. It is also difficult to validate, as the target also has interaction effects.

## 2 Methodology

Methodology is the study of the design of methods for gaining knowledge. It is different from general philosophy of science in that it does not contain, for instance, comparison, choice and justification of methods.

**Choice of Method** To answer the question of what methods should be chosen, three approaches are typically found:

- The conventional approach, in which you choose the same methods as your teachers or peers.
- The outcome-oriented approach, in which you choose the method that gives the best results.
- The reason-based approach, in which you choose the method for which you have the overall best reasons. Reasons here means considerations of the method with respect to a set of goals that you want to achieve.

Their advantages and disadvantages may be summarized as follows:

Approach	Advantages	Disadvantages
Conventional	<ul style="list-style-type: none"><li>• Makes choices easy</li></ul>	<ul style="list-style-type: none"><li>• Leaves you less open to correcting methodological mistakes and being critical of your own results</li><li>• Makes it hard to collaborate with others due to inflexibility</li></ul>
Outcome-oriented	<ul style="list-style-type: none"><li>• Nada</li></ul>	<ul style="list-style-type: none"><li>• Is too vague - what are the best results, for instance, and who judges this?</li><li>• Science often involves long-term planning, so the optimal choice of method might be hard to know a priori</li></ul>
Reason-based	<ul style="list-style-type: none"><li>• Flexible</li></ul>	<ul style="list-style-type: none"><li>• Nada</li></ul>

Table 1: Advantages and disadvantages of approaches to method choice.

**Statistics** Statistics is a set of mathematical methods for dealing with quantitative data. It is divided into descriptive and inferential statistics.

**Lying With Statistics** The practice of lying with statistics involves taking good data and using proper statistical methods to produce false or misleading claims.

**Methodology and Statistics** The methodological aspects of statistics entail choosing the right statistical tool for some particular task, justifying the choice and choosing a proper representation of the data.

**Why Use Statistical Inference?** Statistical inference might be useful for hypothesis testing if:

- The hypothesis has stochastic implications.
- One wants to quantify error or confidence.

Note that the statistical treatment generally weakens individual accounts of confirmations and falsification if the hypothesis is deterministic.

**Fisher's Significance Testing** Fisher's significance testing is a framework for rejecting hypotheses. Its steps are as follows:

1. Specify a hypothesis  $H$ .
2. Devise an experiment to test  $H$  and specify the possible outcomes of the experiments.
3. Determine the distribution of the test statistics.
4. Observe the outcome of the experiment.
5. Calculate the  $p$ -value, defined as the probability of observing a result as least as extreme as the observed one given  $H$ .
6. If  $p$  is smaller than some threshold value, reject  $H$ .

**$p$ -Value Abuse** Fisher's significance testing leaves room for exploitation. A few examples of how this can be done are:

- Modifying the hypothesis so as to mislead readers about what you have shown.
- Change the experiment or sample size until you get the desired result.
- Change how outcomes are partitioned.
- Use other sample distributions.
- Change the threshold.

**Neyman-Pearson Testing** Fisher's framework only leaves room for considering one hypothesis at a time. The Neyman-Pearson framework allows you to test two mutually exclusive and jointly exhaustive hypotheses. Labelling one of them as  $H_0$ , the outcome is designated according to the table below.

	$H_0$ is true	$H_0$ is false
$H_0$ is accepted	Correct	Type-II error
$H_0$ is rejected	Type-I error	Correct

Table 2: Designation of outcomes in Neyman-Pearson testing.

The typical approach is to set the desired type-I error rate to some acceptable value and perform power analysis on your test given this. The power of the test is defined as the probability of rejecting  $H_0$  given that it is true, and depends on the set type-I error rate, the magnitude of the effect of interest and the sample size. Once that is done, you compare the obtained  $p$ -value to the set type-I error rate and use that to accept or reject  $H_0$ .

**Bayesian Statistics** The Bayesian approach to statistics provides another way of using statistics for testing a hypothesis. The steps are:

1. Formulate a set of competing hypothesis.
2. Determine prior probabilities of each hypothesis being true.
3. Collect data not used for informing your assignment of probabilities.
4. Determine the probability of the data given the hypothesis.
5. Compute the probability of the hypothesis given the data using Bayes' theorem.
6. Update the prior probabilities using the above.

Some issues with this approach include:

- The issue of assign prior probabilities. This may be solved by approaching the issue in a subjectivist manner, where one realizes that as long as the priors are not 0 or 1 everyone will agree in the end, or an objectivist approach, where one starts the process by dividing one's belief equally between the hypotheses.
- The issue of old evidence, which may not be reused.
- The issue of uncertain evidence, as the Bayesian approach builds on letting the probability of the evidence occurring going to 1.