Introduction to Logic
UCL 2007/08

Nils Kurbis

Contents

Lecture 1. Introduction and Basic Notions of Logic

Most people will remember the scene from Monty Python's 'Holy Grail' in which Sir Bedevere explains to the people of his village how to tell whether someone is a witch or not. Sir Bedeveres argument is hilariously irrational. Put into a more tractable form than how it is originally presented, part of it runs somewhat like this:

(1) Witches burn.
(2) Whatever is made of wood, burns.
―――――――――――――――――――――――
(3) Therefore, witches are made of wood.

This is evidently a silly argument. Although it may well be true that whatever is made of wood burns, surely things made of other stuff also burn, say paper. So witches need not be made of wood, if they burn: they could also be made of paper. The conclusion *does not follow* from the premises: we can accept the premises of the argument (1) and (2) without accepting its conclusion (3).

Contrast Sir Bedevere's argument with the following one:

(1) Every cat is a mammal.
(2) No mammal is a fish.
―――――――――――――――――
(3) Therefore, no cat is a fish.

This is a much better argument. In this argument the conclusion *follows from* the premises: it cannot possibly be the case that both premises are true, but the conclusion false!

In studying logic, we study what it is that arguments of the latter kind have in common, and what distinguishes them from arguments of the kind Sir Bedevere puts forward. We will work with a somewhat idealised notion of an argument. For our purposes, the following definition will do:

DEFINITION. An *argument* is any set of declarative sentences, one of which is designated as the *conclusion* of the argument, the others being its *premises*.

We can write down arguments as lists of sentence, where the last one is separated by a line from the others, thereby distinguishing the premises from the conclusion. In the examples above, I have also numbered premises and conclusions, which helps referring to them in later discussion, but that is not required by the definition.

Our somewhat idealised definition has the effect that for our purposes we will call things 'arguments' which in ordinary life I suspect we wouldn't be easily convinced to class as such. For instance, the following is an argument:

(i) John likes a drink.

(ii) Therefore, Mary plays the piano.

Our definition of what an argument is ignores that in ordinary life we would require premises and conclusions to be somehow connected with or relevant to each other, not just random collections of sentences one of which is arbitrarily designated as a conclusion and the others as premises. This will have consequences in the development of logic that will probably strike you as rather counterintuitive. You'll see yourself what they are once we come across them!

Our definition of an argument is deliberately broad. It may well be true that some things we call arguments are not the kind of thing normally put forward as arguments in ordinary life. But because our definition is so broad, at least we are making sure that we do not *exclude* any arguments of the kind that use *declarative sentences* as premises and conclusions.

Thus we exclude from being premises and conclusions sentences which are not declarative, i.e. we are only dealing with sentences of the kind used in the examples above, but not, for instance, with

'Can I eat your dessert?'

or

'Ouch!'

or

'Don't hit your little sister!'

or

'Please go away.'

This excludes arguments (in the ordinary sense of that word), if there are any, that involve commands, requests, questions or exclamations. There might be examples involving commands or requests. Consider the following, due to Arthur Prior:

(1)	If God exists, go to church.
(2)	Don't go to Church!
(3)	Therefore, God does not exist.

Should we assimilate it to the other examples of arguments discussed so far? Prior's example tries to establish a certain conclusion, and in that sense it is similar to the other examples. However, it does so in a way which is rather different from what goes on in the other examples: in the latter, the premises are put forward as reporting facts, whereas here in the second premise we find a command or a request to take up a certain course of action. This points to a distinction which suffices to motivate excluding commands and requests from being premises and conclusions of arguments.

Another reason for excluding requests, commands, questions, exclamations etc. from being premises and conclusions of arguments is that restricting consideration to declarative sentence, we are considering only sentences that can be said to be *true* or *false*. For brevity's sake, from now on, when I talk about sentences, I mean declarative sentences. We call 'True' and 'False' the *truth-values* of sentence and abbreviate them by **T** and **F**. We assume that every sentence has either the truth-value **T** or the truth-value **F**, i.e. we assume every declarative sentence to be either true or false.

We observed at the beginning a crucial difference between Sir Bedevere's argument and the second argument about cats and fish. The former had premises which may very well be true, but a conclusion which may very well be false. Contrary to that, the second argument was of a kind such that it could not possibly be the case that the premises are true and the conclusion false: the conclusion *followed from* the premises. We use this observation to *define* a notion characterising arguments of second kind:

DEFINITION. An argument is *deductively valid* if and only if it is not possible for the premises to be true and the conclusion false.

We can also say that valid arguments are *truth-preserving*: if the premises of a valid argument are true, its conclusion must also be true. Arguments of Sir Bedevere's kind lack this property. Hence they are not deductively valid. They are *deductively invalid*:

DEFINITION. An argument is *deductively invalid* if and only it is not deductively valid.

It follows from the two definitions that an argument is deductively invalid if and only if it is possible for the premises to be true *and* the conclusion false.

Arguments thus come in two kinds: the valid ones and the invalid ones. The valid ones come in three kinds:

(i) True premises and a true conclusion, for instance:

(1)	Every cat is a mammal.
(2)	No mammal is a fish.
(3)	No cat is a fish.

(ii) At least one false premise and a true conclusion, for instance:

(1)	Every fish is human.
(2)	Socrates is a fish.
(3)	Socrates is human.

(iii) At least one false premise and a false conclusion, for instance:

(1)	London is either in France or in England.
(2)	London is not in England.
(3)	London is in France.

Obviously there can't be any valid arguments with true premises and a false conclusion, as then the argument would be invalid.

If an argument fails to convince us of the truth of its conclusion – assuming of course that we are not stubborn, or refuse to be rational, or can't follow the argument because it's too complicated, or its too dull to be followed – this may be because the argument is not valid, as was the case with Sir Bedevere's argument. But as the valid arguments come in the three kinds, there is another option for when we need not accept the conclusion of an argument: it might be that, although the argument is valid, one of the premises is false. Arguments which are valid and have true premises are particularly good. This is because if you accept the truth of its premises, such an argument should lead you to accept its conclusion. We call these arguments *deductively sound*:

DEFINITION. An argument is *deductively sound* if and only if it is deductively valid and its premises are true.

Hence argument (i) above is sound, whereas arguments (ii) and (iii) are *unsound*:

DEFINITION. An argument is *deductively unsound* if and only if it is not deductively sound.

It follows from the definition that an argument is deductively unsound if and only if it is either invalid or one of its premises is false, i.e. the unsound ones are those we called unconvincing at the beginning of this paragraph.

Notice that whether or not an argument is convincing depends, to some extend, at least, on who considers it. Whether or not an argument is unsound, by contrast, has no such subjective aspect: it is sound or unsound, no matter what you think about the argument. We might say that with our definition of unsoundness we have captured the objective core of our intuitions about unconvincing arguments.

Lecture 2. More Basic Notions

I have mentioned in the last lecture that our very wide definition of what we count as an argument for the purposes of logic – i.e. an argument is any set of declarative sentences one of which is designated as the conclusion of the argument, the others being its premises – has some counter-intuitive consequences. One such consequence is that, e.g.

(i)	John likes a drink.
(ii)	Therefore, Mary plays the piano.

is an argument according to our definition. But at least by our definition of validity, it is a bad argument, because it's invalid: it can have a true premise and a false conclusion. Later in this lecture we will encounter arguments which may appear rather random to intuition, but which our definition of validity will determine as good, valid arguments.

But first let's have a quick look at a certain kind of argument that is very often used in science and also in everyday life. Although these very often are good arguments in the sense that they convince people of their conclusions on the basis of their premises, we will not attempt to characterise in any detail what it is that makes these arguments good ones. Take, for instance, the following example. During an outbreak of cholera in Soho in 1854, the physician John Snow observed that people who contracted the disease had drunk water from a public water pump in Broadstreet. He concluded that the well was the source of the epidemic. His argument, thus, went somewhat like this:

Many people drunk water from the well and contracted cholera.
___
Therefore, the well is the source of the cholera outbreak.

This was as very good argument, so good indeed that the local authorities found it convincing enough to have the pump handle removed to prevent people from getting water from the well. But the conclusion doesn't follow from the premise with the kind of logical necessity that is characteristic of our definition of a valid argument: it could have been the case that all these people contracted cholera somehow else else and only by chance had they all drunk water from the well. Nonetheless, the conclusion that the well was contaminated is strongly supported by the premise. We can say that although the argument is not deductively valid, it is *inductively strong*. In this course, we are not going to study inductive arguments. We only concentrate on deductive arguments.[1]

Before continuing with the more counterintuitive consequences of our definitions of 'argument' and 'validity', we need some more terminology. Consider the following sentences:

(i) Alice is in Paris or in Rome.
(ii) Alice is not in Paris.
(iii) Alice is not in Rome.

Notice that we are now not considering an argument: none of the three sentences is singled out as the conclusion. We simply consider the three sentence as a collection. Such a collection is called a *set*: we do not care about the order or the number of times a item occurs in the collection; the numbering is merely added to facilitate referring to each sentence in the next paragraph.

There is something quite special about this set of sentences: they cannot possibly be all true together. For suppose (i) and (ii) are both true—then (iii) cannot possibly be also true; suppose (ii) and (iii) are both true—then (i) cannot possibly be true; suppose (i) and (iii) are both true; then (ii) cannot possibly be true!

Logicians have found such collections of sentences so interesting that they gave them a name: a set of sentence such that it is not possible for all its members to be true is called an *inconsistent* set. One of the reasons why these sets are quite important is that they are closely connected to the notion of validity. Take an inconsistent set. Its members cannot all be true together. Now take away some one sentence from the set. If the rest of the sentences in the set are all true, this sentence we have taken away must be false. In other words, the falsity of this sentence follows from the truth of all the others,

[1] The distinction is, I presume, the reason why Bergmann *et al.* chose the rather clumsy phrase 'deductively valid' instead of simply 'valid': the point is to remind us that there are very good arguments which we exclude from consideration, i.e. inductive ones.

i.e. denying the truth of the sentence and using this as the conclusion of an argument which has the other sentences in the set as premises produces a valid argument.

It goes without saying that not all sets of sentences are inconsistent. If a set is not inconsistent, then its members can all be true together. The 'opposite' of inconsistency is consistency:

DEFINITION. A set of sentences is *logically consistent* if and only it is possible for all the members of that set to be true.

We can now define inconsistency as that which is not consistent:

DEFINITION. A set of sentences is *logically inconsistent* if and only it is not consistent.

Notice that consistency and inconsistency are properties of *sets of sentences*, not of sentences.

We now have material at hand to discuss the first counterintuitive consequence of our definitions of 'argument' and 'validity'. Take a set of sentences that is inconsistent. By definition, the sentences in it cannot possibly be all true together. But now suppose we take an inconsistent set of sentences as the premises of an argument and chose an arbitrary new sentence as the conclusion of the argument. By definition, an argument is deductively valid if and only if it is not possible for the premises to be true and the conclusion false. But if the premises cannot all be true, as the set is inconsistent, this means that an argument the premises of which are inconsistent is already valid, no matter what has been chosen as the conclusion. This is irrelevant, because whatever the conclusion, it can never be the case that the premises of the argument are all true *and* the conclusion false, simply because the premises cannot be all true in the first place! Thus the following, for instance, is a valid argument:

(i)	Alice is in Paris or in Rome.
(ii)	Alice is not in Paris.
(iii)	Alice is not in Rome.
(iv)	Therefore, moon is made of cheese.

If that sounds very counterintuitive, it maybe a consolation to notice that at least the argument cannot be sound, because its premises cannot be all true. In other words, there is no way of establishing the conclusion on the basis of the premises. So maybe the fact that this is a valid argument is not so bad.

As mentioned, inconsistency is a property of sets of sentences, not of sentences. (A set could of course consist of only one sentence, but in general people take a set of one sentence to be different from the sentence itself and in any case it is useful to make this distinction here.) There is, however, a notion that applies to sentences, but not sets of sentences, which is closely connected to inconsistency. Consider the following sentence:

> It is raining and it is not raining.

This sentence cannot possibly be true. The weather sometimes is pretty bad, but it is never so bad that it is both, raining and not raining. We say that this sentence is *logically false*:

DEFINITION. A sentence is *logically false* if and only it is not possible for the sentence to be true.

So just as an inconsistent set of sentences is one the members of which cannot possibly be all true together, a logically false sentence is a sentence taken on its own which cannot possibly be true. Of course, a set consisting only of a single logically false sentence is logically inconsistent, as its members (this one logically false sentence) cannot possibly all be true, and conversely, a logically inconsistent set consisting of only one sentence has as its member a logically false sentence.

Now consider this sentence:

> It is raining or it is not raining.

This sentence cannot possibly be false. Whatever the weather is like, it is always either raining or not raining. We call such sentences *logically true*:

DEFINITION. A sentence is *logically true* if and only if it is not possible for the sentence to be false.

Notice that there is no term for sets of sentences that stand in a relation to logically true sentences analogous to the one in which inconsistent sets of sentences stand to logically false sentences.

Not all sentences are either logically true or logically false. There is a third option. There are sentences which are neither logically true nor logically false, for instance 'It is not raining'. These sentences are called *logically indeterminate*:

DEFINITION. A sentence is *logically indeterminate* if and only it is neither logically true nor logically false.

Again, there is no terminology to characterise sets which corresponds to the logical indeterminacy of sentences.

We can now discuss another counterintuitive consequence of our definitions of 'argument' and 'validity'. Consider any logically true sentence. Suppose you use it as the conclusion of an argument and take some arbitrary sentences as its premises. This argument is valid according to our definition! This is so because, as the conclusion is true anyway, it cannot possibly be the case that the premises are true *and* the conclusion false! So, for instance, this is a valid argument:

(1) Alice is in Paris, in Rome or in Madrid.
(2) Alice is in Madrid.
(3) Alice is not in Rome.
_____
(4) Therefore, it is raining or it is not raining.

If Alice is indeed in Madrid, this argument will even be sound. Contrary to the counterintuitive argument with the premises forming an inconsistent set, this time the terminology we have at hand does not allow us to say anything negative about the argument. We just have to live with the fact that it is valid and possibly sound, even though intuitively it looks like a pretty bad argument! Notice however that valid arguments, we said, are *truth preserving*: if the premises are true, so is the conclusion, and even if this kind of argument sounds somewhat strange, at least it has this property: it will never lead us astray: it'll never lead from true premises to a false conclusion.

There is one last definition we need to mention today. Consider the following pairs of sentences:

(1.1) John loves Mary.
(1.2) Mary is loved by John.

(2.1) Sometimes it is not raining.
(2.2) It is not raining all the time.

(3.1) London is to the north of Edinburgh.
(3.2) Edinburgh is to the south of London.

The sentences of each pair say the same thing. As you may have noticed by now, our definitions do not actually care an awful lot about what has been said by a sentence, or what they mean, or what they express. In logic, all we really care about is whether sentences are true or false, i.e. all we care about are the truth-values of sentences, or rather, not even that is the main concern, what truth-values sentences in fact have, but only what truth-values they *could* have. 'To say the same thing' therefore isn't a notion congenial to the way we are doing logic here, but we can capture some of this notion by noticing that, if two sentences 'say the same thing', then at least they never have different truth-values. Thus for our purposes all we need to care about 'saying the same thing' is captured by the following definition:

DEFINITION. The members of a pair of sentences are *logically equivalent* if and only it is not possible for one of the sentences to be true while the other sentence is false.

This is the last definition for this lecture.

To recapitulate, we defined notions relating to *sentences*, to *pairs of sentences*, to *sets of sentences* and to *arguments*. Make sure you keep these apart. When reading or using terminology, ask yourself what it applies to. The notions of *logical truth*, *logical falsity* and *logical indeterminacy* relate to *sentences*; the notion of *logical equivalence* relates to *pairs of sentences*; the notions of *consistency* and *inconsistency* relate to *sets of sentences*; the notions of *validity*, *invalidity*, *soundness* and *unsoundness* relate to arguments.

Getting acquainted with this terminology is a good exercise for using philosophical terminology in general correctly. Many essays in my experience suffer from incorrect use of terminology and thus are strictly speaking full of category mistakes—in other words false or even nonsensical claims. For instance, sentence are not valid. Only arguments are. There may of course be some intelligible way of speaking on which sentences may be called valid, but that is not the way of speaking that we have decided to follow with our definition. Similarly, arguments are not logically false. Only sentences are.

Again, sentences are not inconsistent. Only sets of sentences are.

The definitions we have given refer to some notion of possibility, e.g. it is not possible for the sentence to be true, or it is not possible for all the sentences in a set to be true together, or it is not possible for the premises to be true and the conclusion false. You might now ask yourself: how do we determine whether something like that is or is not possible? Consider the following argument:

(1)	Every beautiful person is a girl.
(2)	No boy is a beautiful person.
(3)	No boy is a girl.

This argument has true premises and a true conclusion. But is it valid? It seems to be the case that it is not possible for the premises to be true and the conclusion to be false—indeed, the conclusion, it seems, could not possibly be false!

The reason why we are inclined to say that (3) cannot possibly be false is that 'boy' and 'girl' have a certain meaning. I have already mentioned that logic as we are studying it in this course in general abstracts from the meanings of sentences or from what has been said. The reason for this is that *content* is always something connected to *particular* sentences. But in logic we are not primarily interested in particular cases of arguments; we are interested in what different arguments have in common, what *in general* makes arguments good or bad. Specific examples matter only in as far as they exemplify a general case. Our definitions aim at generality. We are not concerned with particular arguments, but rather with the *general forms* of arguments. We are not interested in talking about anything in particular, say boys and girls, but rather we are interested in whatever form arguments about whatever can have. Logic, it is sometimes said, has no particular subject matter, or maybe better, it has all possible subject matters as its subject matter.

When we use a notion of possibility in the definitions, we should keep in mind the aim with which the definitions are put forward, i.e. generalisation. So we won't be satisfied with calling the above argument valid. But we must have some criteria for when something is possible according to our logic and when it is not. The way to achieve this is to eliminate those features of the argument which determine its subject matter, i.e. boys, girls and beautiful people. So rather than having specific predicates like 'boy', 'girl' and 'beautiful person' in our argument, we could replace them by placeholders A, B and C:

(1)	Every A is a B.
(2)	No C is an A.
(3)	No C is a B.

Now we are not talking about anything in particular anymore. We are only considering the *form* of the argument, not its content. This puts us in a position to recognise that arguments of this form are not all valid. For instance, replace A with 'bird', B with 'animal', C with 'dog':

(1)	Every bird is an animal.
(2)	No dog is a bird.
(3)	No dog is an animal.

This argument has the same form as the boy-girl argument, but it has true premises and a false conclusion, hence is not valid. And so, because we are interested in the general case rather than the particular one, we put the former in the same category as the latter and call it invalid too. Contrast this case with the following form arguments can take:

(1)	Every A is a B.
(2)	No C is a B.
(3)	No C is a A.

No matter what you put for A, B and C, the argument is going to be valid. No argument of this form can be such that the premises are true and the conclusion false!

The method for determining whether something is possible or not according to our definitions in a way that takes into account the generality we aim at is *formalisation*: we replace those parts of sentences that determine a certain subject matter by place-holders or *variables* that stand for nothing in particular. In a way they stand for everything we can talk about: we can replace anything we like for A, B and C (within the bounds of grammar, of course). Formalisation isolates the general forms of arguments. It is often said that logical inferences are valid *by their form* and logic is said to be a *formal* science. We are doing *formal logic*.

We have seen that the definitions we have given so far have a drawback: because they refer to an intuitive notion of possibility, they are not quite suitable for the generality we require. We made this notion of possibility more suitable by explaining it in terms of formalisation. In fact, we will *redefine* the notions defined in terms of possibility in terms more suited to the project of formal logic, once we have introduced the methods of formalisation, which is the topic of the next lectures.

Lecture 3. Formalisation: Conjunction, Disjunction, Negation

At the end of the last lecture I explained how formalisation allows us to explicate the intuitive notion of possibility used in the definitions introduced so far. The objective is to abstract from content as far as possible and to isolate the *form* of arguments. We wish our notion of validity to be understood in such a way that all arguments *of a certain form* are valid, i.e. truth-preserving. In this lecture we will start introducing methods of formalisation. This involves introducing the use of symbols. Formalisation is to a certain extend a means of saving ink: using symbols enables us to express things in a shorter way than if sentences of ordinary English were used. Everything we say using symbols could be said using ordinary English, possibly at the cost of some prolixity. But in the second part of these lectures the latest you will realise that symbols are actually to some extend clearer than ordinary English, at least once you're used to them. Symbols help making distinctions that in ordinary language aren't easily available, and if made are clumsy and often quite hard to comprehend. Comparing discussions in modern logic books with ancient or mediaeval ones suffices to prove this point ... This is not to say that the meanings of sentences expressed using symbols are always immediate. Sometimes we need to do some thinking and reflect on the sentence to grasp it. But there is never any doubt what truth-value a sentence expressed in the formal language has, and we will introduce methods for determining truth-values in precise ways which always lead to a unique result. There are no ambiguities in the formal language.

Thus formalisation has three points: a) to provide a shorthand for sentence in ordinary language, b) more specifically a shorthand which is unambiguous, and c) a shorthand that exhibits the general forms of sentences and arguments.

In case you were wondering whether there is a difference between 'formal' and 'symbolic', I'll use these terms interchangeably. 'Formal', 'formalisation' and 'formal language' seems to have some more currency in contemporary literature, whereas 'symbolic', 'symbolisation' and 'symbolic language' seem to be more often used in older texts. But both are equally good ways of speaking. The first alternative seems to stress that in logic we are not interested in saying anything in particular: we focus on the form sentences that say something can have, not on their content. The second alternative seems to stress that nonetheless the point of using symbols is that they stand for or represent expressions having a certain content and can be used to say something.

In this and the next lectures, we will set up a formal language which serves the purposes isolated above. But before we introduce the first symbols of our formal language, we need to apply the motto of logicians *divide et impera* – divide and conquer! – and make some distinctions which allow us to exclude certain expressions of natural language from consideration in the formal language. Formalisation will proceed in two stages. At the first stage, we provide a symbol for an expression of ordinary English only if it is of such a kind that it turns sentences into sentences. We call these expressions *sentential connectives*. For instance, 'and' is such an expression. It forms, e.g., the *compound* sentence 'Mary is playing the piano and John goes to the pub' from the two *simple* ones 'John goes to the pub' and 'Mary is playing the piano'. 'It is not the case that', 'or' and 'if-then' have similar properties. This is to be understood as excluding 'Some' and 'All', because these are best understood as occurring in the context 'Some A are B' or 'All A are B', and thus form sentences from two predicates. Of course you might say that 'All', e.g., forms a sentence from a sentence, for instance 'All men are mortal' from 'Men are mortal'. But notice that 'Men are mortal' means the same as 'All men are mortal', but (*) 'All all men are mortal' doesn't mean the same as 'All men are mortal' but is, in fact, meaningless. Thus 'all' cannot just form sentences from sentences: To exclude (*) from being a legitimate construction, the structure of the sentence would have to be taken into account. The sentence that is used to form another sentence by appending 'all' must have a rather specific form, namely 'A are B'. Thus if 'all' and 'some' were connectives that form sentences from sentences, we have to take into account the subsentential structure of the sentences. Thus it is natural to say right from the start that 'all' and 'some' really should be treated as 'all ... are ...' and 'some ... are ... ', which are expressions forming sentences from two predicates.

Thus at the first stage of formalisation, we consider only expressions which form sentences from sentences, no matter what their substructure. Excluding 'some' and 'all' from symbolisation might come as a bit of a surprise, because this excludes us from formalising the syllogisms used as examples in the last lectures. But as mentioned, formalisation has two stages. 'Some' and 'all' come at the second stage. At the first stage, we only introduce the language and symbols of *sentential logic*, where we only consider whole sentences and sentences formed from them, but not subsentential structures. At the second stage, we take into account the structure of expression and subsentential parts of sentence, in particular those parts for which A and B stand in syllogisms, i.e. predicates. The second stage of formalisation is thus called *predicate logic*.

We haven't yet excluded enough to proceed to formalisation. You'll remember that we have introduced the two truth-values **T** and **F**, although we have not made much explicit use of them yet. We will use them at almost every step from now on, and in particular to provide a definition of what kinds of expressions are symbolised in the formal language we are about to construct. Contrast the following two sentences:

(i) John believes that the earth is flat.
(ii) It is not the case that the earth is flat.

Both sentences contain the sub-sentence 'the earth is flat': we can view both sentences as being constructed from this sentence by writing 'John believes that' and 'It is not the case that', resp., in front of it. There is a significant difference between these two examples. We know that 'the earth is flat' is false. This allows us to conclude that 'It is not the case that the earth is flat.' is true. But it does not allow us to conclude anything about the truth-value of 'John believes that the earth is flat.'. The formal language will only have symbols for the former kind of expression. We call them the *truth-functional* connectives:

DEFINITION. A sentential connective is used *truth-functionally* if and only if it is used to generate a compound sentence from one or more sentences in such a way that the truth-value of the generated compound is wholly determined by the truth-values of those one or more sentences from which the compound is generated, no matter what those truth-values may be.

This definition will become a lot clearer when we've given examples of the kind of connectives that are truth-functional and are graced with symbols in the language of sentential logic. So let's introduce the first one, which is *conjunction*.

By conjunction, we'll mean the word 'and'. Instead of using the English word, we will represent it by the symbol &. And instead of writing actual sentences to its left and right, we write capital letters like, A, B, C. We stipulate which ordinary language sentences are supposed to be represented by these capital letters. E.g., we stipulate that 'John goes to the pub.' is represented or, indeed, abbreviated by simply writing J and 'Mary is playing the piano.' is abbreviated by M. We can draw up a table where we record which sentence is abbreviated by which capital letter:

Letter:	Sentence it abbreviates:
M	Mary is playing the piano.
J	John goes to the pub.

A table like this one provides what is called an *interpretation* of the formal language: M and J by themselves of course mean nothing – they are not sentences of English (nor, indeed, any other language, I presume), they are merely letters – but we can of course stipulate that for the moment they are to stand for sentences of English that say something. We can now represent the sentence

Mary is playing the piano and John goes to the pub.

by the following sentence of the language of sentential logic:

$M \& J$

Suppose we extend our interpretation of capital letters of the language of sentential logic by adding the two following rows to the table:

Letter:	Sentence it abbreviates:
M	Mary is playing the piano.
J	John goes to the pub.
K	Kate is a brunette.
L	Luke prefers blondes.

We can now symbolise

Kate is a brunette and Luke prefers blondes.

by the following sentence of the symbolic language:

$K\&L$

These admittedly trivial examples illustrate how symbolisation exhibits the common form of sentences. The two sentences 'Mary is playing the piano and John goes to the pub.' and 'Kate is a brunette and Luke prefers blondes.' have a common structure, namely being constructed from two simpler sentences by means of conjunction. Here is a slightly more exciting example. Consider the following extension of the interpretation of the symbolic language:

Letter:	Sentence it abbreviates:
M	Mary is playing the piano.
J	John goes to the pub.
K	Kate is a brunette.
L	Luke prefers blondes.
N	Nancy is playing the piano.

We can now symbolise this sentence:

Mary and Nancy are playing the piano.

by the following sentence of the symbolic language:

$M\&N$

It is less obvious that 'Mary and Nancy are playing the piano.' shares a common structure with 'Mary is playing the piano and John goes to the pub.'

and 'Kate is a brunette and Luke prefers blondes.' than it is that the latter two share this structure. Nonetheless, nothing spectacular or unforeseen has happened yet.

Conjunction is often expressed by other phrases in ordinary English, e.g. by 'although', 'but', 'however'. We also symbolise all these phrases by &. So, for instance, 'Kate is a brunette, but Luke prefers blondes.' is symbolised as $K\&L$, 'Although Mary is playing the piano, John goes to the pub.' as $M\&J$, etc.. Here we have some more telling cases of common structure: whatever it is that 'although' or 'but' or 'however' express that distinguishes their use in ordinary English from that of 'and' is nothing that need concern us for the purposes of logic. Whatever this difference, whether you use 'and' or 'but' makes no difference to the *truth* of what you have said: the truth-values of 'Although Mary is playing the piano, John goes to the pub.' is the same as the truth-value of 'Mary is playing the piano and John goes to the pub.'. Although it may sometimes be more *appropriate* to use one rather than the other, and your utterance could be criticised for inappropriateness, it is quite plausible to assume that you wouldn't be criticised for having said something *wrong*.

In the last paragraph, we talked about the structure of sentences. Very abstractly speaking, the structure of the example sentences is the following: first comes a sentence, then the symbol & and then another sentence. When we talk about the structure of sentences, rather than using sentences to say things, we use *boldface* letters **P**, **Q**, **R** instead of normal faced ones. So we can now say that $M\&N$, $K\&L$ and $M\&J$ have the common structure **P**&**Q**. We call sentences of this structure *conjunctions* and **P** and **Q** are their *conjuncts*. Notice that $M\&N$, $K\&L$ and $M\&J$ say something, namely what is given by their interpretation, but **P**&**Q** says nothing, as it is not a sentence, but only the general form of a number of sentences.

There is something more interesting about the common form that these sentences have than just the fact that they have it. What we are in the process of doing is to construct a formal or symbolic language in which we can abbreviate or represent sentences of ordinary English. Thus we need to say something about how the new expressions are to be used, or what their meanings are. As we are in the process of constructing the language of sentential logic, we won't say anything about the meanings of capital letters, except what is said by their interpretation; we are not concerned with subsentential structures. You'll remember that we decided that only truth-functional sentential connectives may be in the language of sentential logic.

The point here was that if a truth-functional connective is used to construct a compound sentence from simpler ones, then the truth-value of the compound sentence is determined completely by the truth-value of the simpler sentences. In other words, given the truth-values of the simple sentences, we should be able to *calculate* the truth-value of the compound sentence. How this is to be done is reasonably simple in the case of conjunction. Suppose both simple sentences left and right of & have the truth-value **T**. Then the conjunction of both should have the same truth-value: if L and M are true, so is $L\&M$. On the other hand, you won't say something true if you assert $L\&M$, but L is false. Similarly if M is false. So if either or both of the two simple sentences has the truth-value **F**, $L\&M$ should also have the truth-value **F**. We can summarise this reasoning in the following table:

P	**Q**	**P & Q**
T	**T**	**T**
T	**F**	**F**
F	**T**	**F**
F	**F**	**F**

Such a table is called a *truth-table*: it specifies the truth-value of compound sentences on the basis of the truth-value of its parts. Notice that it takes into account all possible combinations of the two truth-values **T** and **F**. So whatever happens we have specified the truth-value of a sentence of the form **P&Q** on the basis of the truth-values of its conjuncts. So whenever we are given a sentence in the symbolic language, e.g. $L\&M$, we know how to calculate its truth-value on the basis of the truth-values of L and M.

So far we cannot yet say very exciting things in our language. We need some more connectives. When introducing the notion of truth-functional sentential connectives, I mentioned as an example of such a beast the connective 'It is not the case that'. We represent this phrase by ∼. It stretches intuition a little to call this expression a 'connective' as it doesn't connect anything. But it forms sentences from sentences, and that was our definition of what a 'connective' is, so we should not be led astray by intuitions; what matters is the definition we have given. The example I gave was:

It is not the case that the earth is flat.

If we abbreviate 'The earth is flat.' by E, we can symbolise this sentence by

$\sim E$

We call the symbol $\sim$ *negation* and $\sim E$ *the negation of* E.

In ordinary English, negation is often expressed by phrases rather different from 'It is not the case that'. More often we simply use 'not', as in 'Mary is not a brunette.' or 'John does not go to the pub.'. Furthermore, negation is often attached to *predicates* rather than sentences and expressed by prefixes like 'un-', 'non-', 'im-' and 'in-'. For instance, to deny that Aristotle was married, we could use the sentence 'Aristotle was unmarried.', to deny that something is plausible, we say that it is implausible. This observation allows us to say some more interesting things about structure than was the case with conjunction. 'Aristotle is unmarried.', 'Logic is indispensable.', 'Orange juice is non-alcoholic.', 'It is not the case that the earth is flat.' all have the same structure. Using the device of boldface letters, we can write this common structure as $\sim \mathbf{P}$. The next table gives an interpretation of capital letters of the formal language using ordinary English sentences, which we can then translate using negation:

Letter:	Sentence it abbreviates:
E	The earth is flat.
M	Mary is playing the piano.
A	Aristotle is married.
D	Logic is dispensable.
O	Orange juice is alcoholic.

Using $\sim$, these are translated as the following:

English	Symbols
It is not the case that the earth is flat.	$\sim E$
Mary is not playing the piano.	$\sim M$
Aristotle is unmarried.	$\sim A$
Logic is indispensible.	$\sim L$
Orange juice is non-alcoholic.	$\sim O$

Thus although in ordinary English we use different phrases to express negation, in the symbolic language the examples all turn out to be of the same form.

If a sentence has the truth-value **T**, i.e. it is true, then its negation should have truth-value **F**, i.e. it is false. Conversely, if a sentence has the truth-value **F**, i.e. it is false, then its negation has the truth-value **T**, i.e. it is true. This gives the following truth-table for negation:

P	**∼ P**
T	**F**
F	**T**

It might be worth noting here that in principle we need not add any other sentential connective to the language of sentential logic. Anything that can be expressed at all by using truth-functional sentential connectives can be expressed by using only conjunction and negation. But it makes things a little easier to introduce more symbols into the language!

We can now symbolise some more complex sentences. Using the interpretation given earlier, here are some examples:

Ordinary English	Symbols
Kate is a brunette and Luke does not prefer blondes.	$K\& \sim L$
Neither Nancy nor Mary are playing the piano.	$\sim N\& \sim M$
Although Mary is playing the piano, John does not go to the pub.	$M\& \sim J$
Nancy and Mary are not both playing the piano.	$\sim (N\&M)$

Notice the use of parentheses in the third example: the negation applies to the conjunction of N and M: we use parentheses to 'separate' the different parts of the sentence. If someone asserts that Nancy and Mary are not both playing the piano, then he asserts that one of N and M is false, i.e. they are not both true, which is to say that the conjunction of N and M is false, i.e. the negation of the conjunction is true. The use of parentheses should not present any difficulties for the moment. We will describe it precisely in the next but one lecture.

The last thing to do for this lecture is to introduce one further connective, which allows us to formalise a sentence like

Mary is not playing the piano or John goes to the pub.

We use this symbol to symbolise 'or': ∨. So, using once more the interpretation provided earlier, the sentence is formalised as:

$\sim M \vee J$

We call a sentence of the form $\mathbf{P}\vee\mathbf{Q}$ a *disjunction* and **P** and **Q** its *disjuncts*. We have to give a truth-table for disjunction. In the two cases we have discussed so far, the truth-tables were hardly controversial and unexciting. We now face the problem that 'or' in ordinary language is used ambiguously. When you go to a conference, you often get a choice between tea or coffee to go with your breakfast. 'Or' here is meant to *exclude* the option that you take both. We call this use of 'or' the *exclusive* use. But there is also another use. Suppose a friend tells you on Thursday that he intends to go to the cinema or to go to the Opera on Friday. If, on Saturday, he tells you that he's actually managed to do both, he hasn't said anything incorrect on Thursday. This use of 'or' is called the *inclusive* use. When we use ∨, we use it in the inclusive sense. We want the example of John and Mary to work in such a way that, given the disjunction is true, if Mary is playing the piano, John goes to the pub. However, if the disjunction is true, we include the case where John goes to the pub independently of what Mary is doing. The only case where the disjunction is false, then, is the case where John is not in the pub, even though Mary is playing the piano. These considerations give the following truth-table for disjunction:

P	Q	P ∨ Q
T	T	T
T	F	T
F	T	T
F	F	F

The truth-table for ∨ eliminates all ambiguities that are present in ordinary English 'or'. But notice that of course all these ambiguities can be avoided in ordinary English at the cost of more words: e.g. the inclusive use is often expressed by using 'and/or', the exclusive sense could be expressed by using 'Either A or B, but not both'. The exclusive sense of 'or' can be expressed in the formal language by using negation, conjunction and disjunction. You can try to find out how to do this, if you like.

Here are some examples of ordinary English sentences symbolised using ∨ and ∼:

Ordinary English	Symbols
Mary or Nancy is playing the piano.	$M \vee N$
Kate is not a brunette or Luke does not prefer blondes.	$\sim K \vee \sim L$
Mary and Nancy are not both playing the piano.	$\sim M \vee \sim N$
Neither Nancy nor Mary is playing the piano.	$\sim (N \vee M)$

Notice that we also formalised the last two sentences using conjunction and negation. We have two options for translating these sentences into the formal language. We can check that both both symbolisations 'say the same thing' in our sense of the word, i.e. they are logically equivalent as defined in the last lecture, by constructing a complex truth-table for each expression. We can construct complex truth-tables in stages. Take, for instance, $\sim \mathbf{P} \& \sim \mathbf{Q}$ First, we write down the different combinations of **T**s and **F**s that can be the truth-values of **P** and **Q**. Then, using the truth-table for negation, we calculate the truth-tables for $\sim \mathbf{P}$ and $\sim \mathbf{Q}$, and finally we apply the truth-table for conjunction to calculate the truth-table for $\sim \mathbf{P} \& \sim \mathbf{Q}$:

P	**Q**	$\sim \mathbf{P}$	$\sim \mathbf{Q}$	$\sim \mathbf{P} \& \sim \mathbf{Q}$
T	**T**	**F**	**F**	**F**
T	**F**	**F**	**T**	**F**
F	**T**	**T**	**F**	**F**
F	**F**	**T**	**T**	**T**

Let's also calculate the truth-table for $\sim (\mathbf{P} \vee \mathbf{Q})$. Here we first calculate the truth-table for $(\mathbf{P} \vee \mathbf{Q})$ by applying the truth-table for disjunction, and then the truth-table for $\sim (\mathbf{P} \vee \mathbf{Q})$ by applying the truth-table for negation:

P	**Q**	$\mathbf{P} \vee \mathbf{Q}$	$\sim (\mathbf{P} \vee \mathbf{Q})$
T	**T**	**T**	**F**
T	**F**	**T**	**F**
F	**T**	**T**	**F**
F	**F**	**F**	**T**

Both truth-tables have the same last row, the same combination of **T**s and **F**s. This means that any sentence of the forms $\sim \mathbf{P} \& \sim \mathbf{Q}$ and $\sim (\mathbf{P} \vee \mathbf{Q})$ have the same truth-value under any conditions. Thus they are logically equivalent and 'mean the same thing' for the purposes of logic. The same holds for sentences of the forms $\sim \mathbf{P} \vee \sim \mathbf{Q}$ and $\sim (\mathbf{P} \& \mathbf{Q})$, as you can check as a homework.

Lecture 4. Formalisation: Material Implication, Biconditional

Last lecture we introduced the first symbols of our formal language: conjunction &, negation $\sim$ and disjunction $\vee$. The use of each of these symbols is characterised by a *truth-table*, which specifies under which conditions sentences of the form $\mathbf{P}\&\mathbf{Q}$, $\mathbf{P} \vee \mathbf{Q}$ and $\sim \mathbf{P}$ are true or false. Truth-tables give a method for *calculating* the truth-values of compound sentences on the basis of the truth-values of the simple sentences they are built up from. In calculating compound sentences with more than one sentential connective, we often have to use *parentheses* to indicate the order in which the calculation is done. For instance, there is a difference in meaning between the following two sentences:

(i) Either Kate is not a brunette or John does not prefer blondes.
(ii) It is not the case that either Kate is a brunette or John does not prefer blondes.

In the first example, we only say something false in case Kate is a brunette and John prefers blondes. In the second example, we deny that a disjunction is true, i.e. we deny that at least one of the disjuncts is true, which is to say we assert that both disjuncts are false: we only say something true in case Kate is not a brunette and John prefers blondes. That's a rather important difference.

This difference is reflected in the formal language by the use of parentheses. The way parentheses are placed to structure sentences of the formal language pretty much mirrors the grammatical structure of the English sentences: the first sentence starts with 'either' followed by two clauses connected by 'or' in each of which a negation occurs; the second sentence starts with 'It is not the case that', which is followed by two clauses connected by 'or' in one of which a negation occurs. Using the interpretation of the last lecture, the two sentences are symbolised in the following way:

(i') $\sim K \lor \sim J$
(ii') $\sim (K \lor \sim J)$

Notice how the structure of the symbolic sentences mirrors the structure of the English sentences.

That both sentences of the formal language must have a different meaning can be seen by constructing their truth-tables. The construction of each truth-table proceeds differently. In the first case, we first calculate the truth-tables of $\sim J$ and $\sim K$ by applying the truth-table for negation to the simple sentences J and K, and then we calculate the disjunction $\sim K \lor \sim J$ by applying the truth-table for disjunction to $\sim J$ and $\sim K$:

K	J	$\sim K$	$\sim J$	$\sim K \lor \sim J$
T	**T**	**F**	**F**	**F**
T	**F**	**F**	**T**	**T**
F	**T**	**T**	**F**	**T**
F	**F**	**T**	**T**	**T**

The calculation of the truth-table for the second sentence proceeds differently: first we calculate the truth-table of $\sim J$ by applying the truth-table for negation to the simple sentence J, then we calculate the disjunction $K \lor \sim J$ by applying the truth-table for disjunction to K and $\sim J$, and last we calculate the truth-table of $\sim (K \lor \sim J)$ by applying the truth-table for negation to $K \lor \sim J$:

K	J	$\sim J$	$K \lor \sim J$	$\sim (K \lor \sim J)$
T	**T**	**F**	**T**	**F**
T	**F**	**T**	**T**	**F**
F	**T**	**F**	**F**	**T**
F	**F**	**T**	**T**	**F**

Observe that the two truth-tables have a different combination of **T**s and **F**s in their last row. That means that $\sim K \lor \sim J$ and $\sim (K \lor \sim J)$ are true and false under different conditions: for instance, if Kate is not a brunette and John does not prefer blondes (which is the case covered by the last line of each truth-table), $\sim K \vee \sim J$ is true and $\sim (K \vee \sim J)$ is false. Thus both sentences cannot say the same thing: they have different meanings and are not logically equivalent.

Notice incidentally that dropping 'either' from (ii) to form

(iii) It is not the case that Kate is a brunette or John does not prefer blondes.

gives a sentence which is ambiguous and can mean either (i) or (ii). No such ambiguity exists in the sentences of the formal language.

At the end of the last lecture we also discussed an example of two different sentences in the formal language which do say the same thing (in the sense explained). There are two ways of formalising the sentence

Neither Mary nor Nancy is playing the piano.

The two options are:

(i) $\sim M \& \sim N$
(ii) $\sim (M \vee N)$

Both sentences of the formal language say the same thing as they have the same truth-table. This is the truth-table for $\sim M \& \sim N$:

M	N	$\sim M$	$\sim N$	$\sim M \& \sim N$
T	**T**	**F**	**F**	**F**
T	**F**	**F**	**T**	**F**
F	**T**	**T**	**F**	**F**
F	**F**	**T**	**T**	**T**

And this is the truth-table for $\sim (M \vee N)$:

M	N	$M \vee N$	$\sim (M \vee N)$
T	**T**	**T**	**F**
T	**F**	**T**	**F**
F	**T**	**T**	**F**
F	**F**	**F**	**T**

Notice that here we have calculated the truth-tables for the *meaningful sentences* $\sim M \& \sim N$ and $\sim (M \vee N)$ which, given our interpretation of N and M as 'Nancy is playing the piano.' and 'Mary is playing the piano.', both formalise the English sentence 'Neither Mary nor Nancy is playing the piano.'. At the end of the last lecture, we calculated truth-tables which used boldfaced letters. This means we calculated not the truth-tables of specific sentences that mean something, but rather we showed that any two sentences *exhibiting specific forms* have the same truth-tables. They are *logically equivalent*, which, as we said in lecture 2, is our way of capturing for the formal language what the phrase 'to say the same thing' means when applied to sentences of ordinary English.

We also came across another example of a sentence of English which could be formalised in two different ways:

Mary and Nancy are not both playing the piano.

The two options are:

(i) $\sim M \vee \sim N$
(ii) $\sim (M \& N)$

Notice the difference between this pair and the pair we just discussed.
In this case, too, both sentences of the formal language say the same thing as they have the same truth-table. This is the truth-table for $\sim M \vee \sim N$:

M	N	$\sim M$	$\sim N$	$\sim M \vee \sim N$
T	**T**	**F**	**F**	**F**
T	**F**	**F**	**T**	**T**
F	**T**	**T**	**F**	**T**
F	**F**	**T**	**T**	**T**

And this is the truth-table for $\sim (M \& N)$:

M	N	$M \& N$	$\sim (M \& N)$
T	**T**	**T**	**F**
T	**F**	**F**	**T**
F	**T**	**F**	**T**
F	**F**	**F**	**T**

As in the previous example, both formal sentences have the same truth-table and are thus logically equivalent.

The two logical equivalences we have now recorded are so prominent in logic that they have a name. They are called *DeMorgan's Laws*.

DEMORGAN'S LAWS. $\sim (M\&N)$ and $\sim M \vee \sim N$ are *logically equivalent*, and so are $\sim M \& \sim N$ and $\sim (M \vee N)$.

We now introduce a further connective into our formal language. It is the *conditional*. It allows us to formalise sentences like

If Kate is a brunette, then Luke does not prefer blondes.

We use the symbol $\supset$, often called the 'horseshoe', to represent 'If-then...'. Thus the example is formalised as:

$K \supset \sim J$

A sentence of the form $\mathbf{P} \supset \mathbf{Q}$ is called a *conditional*, where $\mathbf{P}$ is called the *antecedent* and $\mathbf{Q}$ is called the *consequent* of the conditional.

'If-then...' is very closely connected to logical consequence. We often express arguments by using 'If-then...', e.g. 'If Alice is either in Paris or in Rome, but not in Paris, then she must be in Rome'. We need to take this into account when constructing a suitable truth-table for $\supset$. Appealing to the definition of valid and invalid arguments, the connection between sentences using 'If-then...' and arguments suggests, first, that if the conditional is true and the antecedent of the conditional is also true, then the consequent should be true, and, secondly, that if the antecedent is true and the consequent is false, the conditional should be false. These considerations give the first two lines of the truth-table for $\supset$:

P	**Q**	**P $\supset$ Q**
T	**T**	**T**
T	**F**	**F**
F	**T**	
F	**F**	

What about lines three and four? We have four options: (i) both are **F**, (ii) line three is **T** and line four is **F**, and (iii) line three is **F** and line four is **T**, (iv) both are **T**. Let's consider every option one by one.

Option (i). Both remaining lines are **F**.

Then the truth-table for $\supset$ would look like this:

P	**Q**	**P ⊃ Q**
T	**T**	**T**
T	**F**	**F**
F	**T**	**F**
F	**F**	**F**

This is just the truth-table for conjunction. But we certainly don't want 'If-then...' to mean the same as 'and'. Hence this truth-table is not suitable for $\supset$, if it is to mean 'If-then'.

Option (ii). Line three is **T** and line four is **F**.

Then the truth-table for $\supset$ would look like this:

P	**Q**	**P ⊃ Q**
T	**T**	**T**
T	**F**	**F**
F	**T**	**T**
F	**F**	**F**

The row underneath **P** $\supset$ **Q** has exactly the same combination of **T**s and **F**s in it as does **Q** on its own in the second row of the table. In other words, if this was the truth-table for **P** $\supset$ **Q**, then it would say the same thing as plain **Q**, which certainly doesn't match 'If-then...' in English. Hence this isn't an option for a truth-table for $\supset$ either.

Option (iii). Line three is **F** and line four is **T**.

Then the truth-table for $\supset$ would look like this:

P	**Q**	**P ⊃ Q**
T	**T**	**T**
T	**F**	**F**
F	**T**	**F**
F	**F**	**T**

If this was the truth-table for $\supset$, then $A \supset B$ is true just in case A and B have the same truth-value, i.e. $A \supset B$ would say that A and B are true and false under the same circumstances. But 'If-then...' really should be weaker than that: we want to say that if A is true, so is B, but if B is true, 'If A then B' need not say anything about A. This reasoning shows that the correct truth-table for $\supset$ is the remaining option (iv).

Option (iv). Both remaining lines are **T**.

Then the truth-table for $\supset$ looks like this:

P	**Q**	**P $\supset$ Q**
T	**T**	**T**
T	**F**	**F**
F	**T**	**T**
F	**F**	**T**

This is the only option we have for a truth-table for a formal analogue of 'If-then...' in the framework we have chosen.

The truth-table for $\supset$ has certain counterintuitive consequences of the kind we have already encountered when discussing the notion of validity. For instance, consider the following conditional:

If it is raining and it is not raining, then the moon is made of cheese.

Writing R for 'It is raining' and M for 'The moon is made of cheese', we can symbolise this sentence as

$$(R\& \sim R) \supset M$$

Calculating its truth-table gives the following:

R	M	$\sim R$	$(R\& \sim R)$	$(R\& \sim R) \supset M$
T	**T**	**F**	**F**	**T**
T	**F**	**F**	**F**	**T**
F	**T**	**T**	**F**	**T**
F	**F**	**T**	**F**	**T**

The sentence is a logical truth! It cannot possibly be false as the last row of the truth-table contains only **T**s.This mirrors exactly the counter-intuitive fact that an argument is logically valid according to our definition if its premises are inconsistent. Thus at least no essentially new counter-intuitive consequences ensue from the truth-table for the conditional. To the contrary, because such sentences as the above come out as logical laws, we can say that the conditional adequately captures the notion of validity defined earlier.

Here are two further counterintuitive consequences of the truth-table for $\supset$. As a matter of fact, the moon is not made of cheese. Now consider the following sentence:

If the moon is made of cheese, then it is raining.

Translated into the formal language, this is:

$M \supset R$

The last two lines of the truth-table for $\supset$, those where the antecedent is false, both have a **T** underneath $\mathbf{P} \supset \mathbf{Q}$, i.e. in these cases the conditional is true. Hence no matter what the weather, the truth-table for $\supset$ determines that this sentence is true, given that as a matter of fact the moon is not made of cheese!

Now consider the contraposition of the sentence (the contraposition of $\mathbf{P} \supset \mathbf{Q}$ is $\sim \mathbf{Q} \supset \sim \mathbf{P}$):

If it is not raining, then the moon is not made of cheese.

Its symbolisation in the formal language is:

$\sim R \supset \sim M$

Lines one and three of the truth-table for $\supset$, those were the consequent is true, both have a **T** underneath $\mathbf{P} \supset \mathbf{Q}$, i.e. in these cases the conditional is true. Hence no matter what the weather, the truth-table for $\supset$ determines that this sentence is true, given that as a matter of fact the moon is not made of cheese!

These counterintuitive truth-conditions for conditionals in the formal language are consequences of our decision to accept only two truth-values and only truth-functional connectives. There just isn't another connective that avoids these consequences in the framework we have chosen. Intuition may be stretched if we translate sentences of the formal language using $\supset$ back into English sentences using 'It-then'. However, intuition is not stretched to much if we keep in mind the decisions we have made at the beginning: a consequences of these decisions is that we only have a very weak conditional. In fact, any sentence $\mathbf{P} \supset \mathbf{Q}$ is logically equivalent to $\sim \mathbf{P} \vee \mathbf{Q}$ and $\sim (\mathbf{P} \& \sim \mathbf{Q})$, as you can check by constructing truth-tables for each sentence. In our formal language, $M \supset R$, e.g., says no more nor less than $\sim M \vee R$ and $\sim (M \& \sim R)$. And intuition is stretched a lot less (if at all) when it is noted that these are both true, as the moon is not made of cheese. It is worth keeping this in mind when you are inclined to think that the conditional of our formal language is somehow defective because of such counterintuitive examples. In any case, the conditional of the formal language has a perfectly good and precise meaning, as determined by its truth-table. We thus need not explain the meaning of $\supset$ with reference to the conditional in ordinary English at all. We can be content with saying that $\supset$ is as close as we can get in the formal language to defining a connective that has some of the characteristics of 'If-then' in ordinary English.

Lecture 5. The Syntax of Sentential Logic

The third option for a truth-table for the conditional we discussed in the last lecture (and rejected as unsuitable for a conditional) actually gives the truth-table for a useful other connective. We said that this truth-table is not appropriate for a conditional, because it is somewhat too strong: it gives a sentence which is true just in case both component sentences have the same truth-value. Let's use a new symbol to designate the connective with this truth-table:

P	**Q**	**P ≡ Q**
T	**T**	**T**
T	**F**	**F**
F	**T**	**F**
F	**F**	**T**

Then $A \equiv B$ says that if A is true, then B is true, and if B is true, A is true; in other words, A if and only if B (or A iff B for short). $\equiv$ is therefore called the *biconditional*, and $A \equiv B$ is logically equivalent to $(A \supset B)\&(B \supset A)$, as you can verify yourself by calculating its truth-table. The biconditional can be used to express logical equivalences in our language, e.g. those stated in DeMorgan's Laws: $(\sim A \vee \sim B) \equiv \sim (A\&B)$ and $(\sim A\& \sim B) \equiv \sim (A \vee B)$. If two sentences are logically equivalent, then the biconditional having them to the left and right of $\equiv$ is logically true.

In lecture 3 I mentioned that we build up our formal language in two stages. First we consider only *sentential connectives*, then in the second stage we add the *quantifiers* 'some' and 'all'. We have now introduced all the symbols of the first stage of our symbolic language. They are $\&, \sim, \vee, \supset$ and $\equiv$. We now need to discuss the 'grammar' of the language in some detail.

The grammar of a formal language is called its *syntax*. It makes precise how to construct sentences of the formal language and how parentheses are to be used as punctuation marks.

You will remember that there is a difference between boldfaced letters **P**, **Q**, **R** and normal-faced ones A, B, C etc.. We use the latter to abbreviate sentences of English and by giving an interpretation of them. The former we use to talk *about* such sentences. A, B, C etc. are sentences of our symbolic language, but **P**, **Q**, **R** are not. We say that **P**, **Q**, **R** are *metavariables* which *range over* expressions of the symbolic language. When we talk *about* a language, we do so in the *metalanguage*. By contrast, we call the symbolic language where A, B, C etc. are used to stand for sentences of English the *object language*. We also give the object language a short label: SL.

We can now give a precise definition of how sentences of SL are to be constructed. The expressions of the language are the *sentence letters* A, B, C etc. and the *truth-functional connectives* $\&, \sim, \vee, \supset$ and $\equiv$, and there are *parentheses* (and) as punctuation marks. The definition of 'sentence of SL' is going to be what is called an *inductive definition*:[2] we first lay down what the simplest sentences are and then give a method for how to construct more complex sentences from less complex ones. Any sentence of the language is constructed in a step-by-step way from sentence letters, connectives and parentheses according to the following method:

1. Every sentence letter is a sentence.
2. If **P** is a sentence, then $\sim \mathbf{P}$ is a sentence.
3. If **P** and **Q** are sentences, then $(\mathbf{P}\&\mathbf{Q})$ is a sentence.
4. If **P** and **Q** are sentences, then $(\mathbf{P} \vee \mathbf{Q})$ is a sentence.
5. If **P** and **Q** are sentences, then $(\mathbf{P} \supset \mathbf{Q})$ is a sentence.
6. If **P** and **Q** are sentences, then $(\mathbf{P} \equiv \mathbf{Q})$ is a sentence.
7. Nothing is a sentence unless it can be formed by repeated application of clauses 1-6.

Notice the parentheses around the sentence in clauses 3-6. Strictly speaking, any sentence of SL must have a pair of outer parentheses. We adopt the convention that we may drop these.

[2]'Inductive' as used here is not to be confused with 'inductive' as discussed in lecture 2, which concerned generalisations of statements on the basis of empirical evidence. The method of induction used in the definition of what a sentence of SL is also called 'mathematical induction'. Empirical generalisations lack the kind of logical necessity we are interested in in logic, mathematical induction doesn't.

Here are some examples of sentences of SL:

$((A\&B) \vee (A\&C))$
$(A \supset (B \supset A))$
$((\sim A\& \sim B) \equiv\sim (A \vee B))$
$\sim (A \supset B)$
$(A\& \sim A)$
$(\sim A \vee A)$

Here are some examples of strings of symbols that are not sentences:

$\sim A\&B \vee C \supset A$
$A \sim B$
$\sim AB \supset (C \equiv C)$
$A\&B\&C$
$A \vee (B\&C) \equiv (A \vee B)\&(A \vee C)$

We already discussed an example where parentheses were needed to translate sentences of English into the formal language: in lecture 4, $\sim K\vee \sim J$ and $\sim (K\vee \sim J)$ were used to abbreviate different sentences of English. Both have different truth-tables and thus mean different things. Without parentheses, ambiguities would arise. Consider, for instance, $A\vee B\&C$. This is not a sentence of SL. The reason is that it would be an ambiguous sentence: if you wanted to calculate its truth-table, you wouldn't know whether to use the one for $\vee$ first and then the one for &, or whether to proceed conversely. Parentheses indicate the order in which truth-table calculation proceeds and thereby ambiguity is avoided.

We now introduce some syntactical terminology that allows us to describe sentences of SL. First, a sentence of SL consisting only of a sentence letter is an *atomic sentence*. The next three clauses define what the *main connective* and the *immediate sentential components* of a sentence are:

1. If **P** is an atomic sentence, **P** contains no connectives and hence does not have a main connective. **P** has no immediate sentential components.
2. If **P** is of the form $\sim \mathbf{Q}$, where **Q** is a sentence, then the main connective of **P** is the tilde that occurs before **Q**, and **Q** is the immediate sentential component of **P**.
3. If **P** is of the form $\mathbf{Q}\&\mathbf{R}$, $\mathbf{Q} \vee \mathbf{R}$, $\mathbf{Q} \supset \mathbf{R}$, or $\mathbf{Q} \equiv \mathbf{R}$, where **Q** and **R** are sentences, then the main connective of **P** is the connective that occurs between **Q** and **R**, and **Q** and **R** are the immediate sentential components of **P**.

Here are some examples:

Sentence:	Main Con.:	Imm. Sent. Comp.:
$((A\&B) \vee (A\&C))$	$\vee$	$(A\&B)$, $(A\&C)$
$(A \supset (B \supset A))$	$\supset$	A, $(B \supset A)$
$((\sim A\& \sim B) \equiv \sim (A \vee B))$	$\equiv$	$(\sim A\& \sim B)$, $\sim (A \vee B)$
$\sim (A \supset B)$	$\sim$	$(A \supset B)$
$(A\& \sim A)$	$\&$	A, $\sim A$
$(\sim A \vee A)$	$\vee$	A, $\sim A$

The *sentential components* of a sentence are the sentence itself, its immediate sentential components, and the sentential components of its immediate sentential components. The *atomic components* of a sentence are the sentential components which are atomic sentences. So, using again the sentences of the last example, this table gives their sentential components:

Sentence:	Sentential Components:
$((A\&B) \vee (A\&C))$	$((A\&B) \vee (A\&C))$, $(A\&B)$, $(A\&C)$, A, B, C
$(A \supset (B \supset A))$	$(A \supset (B \supset A))$, $(B \supset A)$, A, B
$((\sim A\& \sim B) \equiv\sim (A \vee B))$	$((\sim A\& \sim B) \equiv\sim (A \vee B))$, $(\sim A\& \sim B)$, $\sim (A \vee B)$, $\sim A$, $\sim B$, $A \vee B$, A, B
$\sim (A \supset B)$	$\sim (A \supset B)$, $(A \supset B)$, A, B
$(A\& \sim A)$	$(A\& \sim A)$, $\sim A$, A
$(\sim A \vee A)$	$(\sim A \vee A)$, $\sim A$, A

Lecture 6. The Semantics of SL: Truth and Falsity

At the end of the second lecture we discussed what it means that something is possible according to the notion of possibility used in the definitions of logical validity, logical truth, logical falsehood etc.. I gave an example of an argument where it was not obvious whether it is possible for the premises to be true and the conclusion false:

(1)	Every beautiful person is a girl.
(2)	No boy is a beautiful person.
(3)	No boy is a girl.

We explicated the intuitive notion of possibility used in the definitions by appealing to the notion of *formalisation*. We said that possibility as used in the definitions should be understood with reference to the *form* of sentences and arguments: if it is not possible for the premises of an argument to be true and its conclusion to be false, then we should understand this as meaning that no argument *of this form* can have true premises and a false conclusion.

Our intuitive notion of possibility has, I suppose, little or nothing to do with formalisation. Thus the observations at the end of the second lecture call for new definitions of the basic logical concepts using the tools of formalisation rather than the intuitive notion of possibility. We introduced methods of formalisation in the third and fourth lectures, i.e. we introduced symbols and their truth-tables, and we defined the syntax of the language SL, which we use for formalisation. Now that we have these tools, in the next two lectures we will give new definitions of the basic logical concepts which avoid the intuitive notion of possibility altogether.

Notice the dialectics: we start with an intuitive notion of possibility in terms of which we define basic logical concepts. We then discover that these intuitive definitions do not quite settle in every case whether an argument is valid or not. In particular, this is due to a conflict that is brought out by the example above: on the one hand, it seems as if it is not possible for a boy to be a girl, given what we mean by 'boy' and 'girl', so the argument should be valid on the grounds alone that that its conclusion cannot be false; on the other hand, however, logic should have a kind of generality that is independent of particular subject matters, like boys and girls. We resolved the conflict by explicating the notion of possibility in terms of the *form* of an argument. This effectively allows us to avoid the notion of possibility altogether in the definitions of basic logical concepts. We won't follow the course of defining the basic logical notions of in terms of the forms of sentences and arguments, though, but introducing the methods of formalisation nonetheless gives us the tools to formulate new definitions which make no use of the notion of possibility and therefore do not rely on intuitions anymore. It remains to explicate in some more detail what is meant by the form of sentences and arguments, and how it is determined whether all arguments of a certain form are valid or not.[3]

Before we redefine the basic logical concepts, we need to introduce some more terminology and logical tools. Last lecture we discussed the *syntax* of the formal language SL and gave a precise definition of what counts as a sentence of SL. In this and the next lecture we will discuss what is called the *semantics* of SL in some detail. Semantics concerns the truth-conditions and truth-values of sentences. Just as we have first introduced the formal language on an intuitive basis and then gave precise definitions in lecture five, we have first introduced semantic notions intuitively via the truth-tables and now proceed to more precision.

One of the most important features of the formal language is that its sentences and the truth-tables for its connectives are constructed in such a way that the truth-value of every sentence is uniquely determined by the truth-values of their atomic components. The truth-value of complex sentences can be calculated on the basis of the truth-values of the atomic ones. Thus, if a truth-value is assigned to each atomic sentence of the language of SL, the truth-values of all the sentences of SL are determined. We call such an assignment of truth-values to the atomic sentences of SL, unsurprisingly,

[3]The preliminary way in this objective has been approached in the second lecture could also be extended to notions relating to sentences: e.g., if no sentence of a certain form can be false, the sentence is logically valid. Question: can we extend it to cover also the notion of logical equivalence?

a *truth-value assignment*:

DEFINITION. A *truth-value assignment* is an assignment of truth-values (**T**s or **F**s) to the atomic sentences of SL.

Thus a truth-value assignment specifies for each atomic sentence of SL, and hence, by the truth-tables, for every sentence of SL, whether it is true or false.

Notice that we assume that the atomic sentences of SL are *independent* of each other. This means that the truth-value of one atomic sentence does not effect the truth-value of another atomic sentence. More precisely, for any two atomic sentences, we consider all four combinations of the two truth-values **T** and **F**. Thus we abstract a good deal from ordinary language. For instance, it is arguably impossible for '*a* is red' and *a* is green' to be both true: nothing can be both red and green. These two sentence also look very 'atomic': they are not composed of any other, simpler sentences. But with respect to the formal language, either we have to conclude that these are not after all atomic sentences, or that after all something can be both red and green, or, best of all options, we ignore that, given what 'red' and green' mean, nothing can be both red and green. The last option again makes use of the methodology we have been using since the first lecture: we are always interested in the more general case; we ignore the particulars and aim at generality. We don't care an awful lot about what sentences actually mean; we only care about the form of sentences. It may well be true that certain sentences which arguably are 'atomic' sentences – and notice that this is a very precise term: we would definitely translate '*a* is red' into SL by using a single sentence letter and giving it this interpretation – of ordinary language have a meaning such that not all four possibilities of combinations of truth-values are possible. Nonetheless, in ignoring this fact we do not *exclude* anything; if we assume that all four possibilities can obtain, we definitely take into account the three possibilities that ordinarily we think are all that is given. And in any case, ignoring that '*a* is red' and '*a* is green' cannot be both true together, given what they ordinarily are taken to mean, does not ignore anything we couldn't get into an argument after all by incorporating into the argument the information reflection on the meanings of 'red' and 'green' supplies. Suppose we have an argument in which both sentences figure, for instance:

(1)	a is red.
(2)	Therefore, a is not green.

Let's use the following interpretation of the language of SL:

Letter:	Sentence it abbreviates:
R	a is red.
G	a is green.

Then formalising the argument shows that we shouldn't count it as valid, given our explication of the notion of possibility in terms of formalisation, because there certainly are arguments of the following form which have a true premise but a false conclusion, i.e. are invalid:

(1)'	R
(2)'	$\sim G$

It is, however, very easy to make the argument valid by adding a further premise. The formalisation does not take into account the intended meaning of the sentence letters, but R and G should not be both true given their interpretation. We can of course add this information as a further premise. If R and G are not supposed to be both true, then $\sim (R\&G)$ is true. And if we add this as a premise, the argument becomes valid:

(x)	$\sim (R\&G)$
(1)'	R
(2)'	$\sim G$

To conclude this discussion we note that all the information that considerations of the meanings of sentences of English give us may be added as additional premises to arguments. That way we do justice to our intuitions concerning what is possible and what is not, given what sentences of ordinary English mean. It is, accordingly, not very problematic that the notion of a truth-value assignment assumes atomic sentences to be independent of each other, which deviates strongly from ordinary English.

We assume a truth-value assignment to assign a truth-value to *every* sentence letter of SL. SL has infinitely many sentence letters. So we cannot actually write a complete truth-value assignment down. However, if we consider a particular sentence and wish to calculate its truth-value, only a fragment of a truth-value assignment suffices, namely that fragment which assigns truth-values to the atomic components of the sentence. And this fragment is finite, as any sentence has only finitely many atomic components. Similarly, if we want to calculate the truth-table of a sentence, we only need to take into account every combination of **T**s and **F**s for those atomic sentences that actually occur in the sentence (as we have been doing in the last lectures). So, for instance, if there is only one atomic sentence in the sentence, the truth-table has only two rows, as there are only two combinations of **T**s and **F**s, and these are the only cases we need to consider:

A	$\sim A$	$A \vee \sim A$
T	**F**	**T**
F	**T**	**T**

Each row in the table corresponds to a fragment of several truth-value assignments. A truth-value assignment assigns a truth-value to *every* sentence letter of SL. But all that interests us in calculating truth-table for $A \vee \sim A$ is the part of them that determines the truth-value of A. Whatever a truth-value assignment does to the other sentence letters of SL, e.g. B, C, D etc., does not affect the truth-value of $A \vee \sim A$, as they are not atomic components of that sentence, so we need not take this information into account when calculating the truth-table.

If a sentence has two atomic components, the truth-table has four rows, e.g.:

A	B	$A \& B$	$\sim B$	$\sim (A \& B)$	$\sim (A \& B) \equiv \sim B$
T	**T**	**T**	**F**	**F**	**T**
T	**F**	**F**	**T**	**T**	**T**
F	**T**	**F**	**F**	**T**	**F**
F	**F**	**F**	**T**	**T**	**T**

Again, we need not take into account all the other assignments of truth-values to sentence letters that truth-value assignments make: all we are interested in are fragments of them, and there are only four different kinds of truth-value assignment that we need to consider: the first row stands for all those assignments assigning **T** to both A and B, the second stands for all those assigning **T** to A and **F** to B, the third stands for those assigning **F** to A and **T** to B and the fourth stands for all those assigning **F** to both A and B. Each row in fact corresponds to the set of truth-value assignments that agree on which truth-values they assign to the two sentence letters A and B. We will allow ourselves the terminological luxury of speaking of *the* truth-value assignment assigning, for instance, **T** to A and **F** to B, or corresponding to the third line of the truth-table, even though there is not one unique such truth-value assignment.

If a sentence has three atomic components, the truth-table has eight rows, e.g.:

A	B	C	$(A\&B)$	$(A\&B) \vee C$
T	**T**	**T**	**T**	**T**
T	**T**	**F**	**T**	**T**
T	**F**	**T**	**F**	**T**
T	**F**	**F**	**F**	**F**
F	**T**	**T**	**F**	**T**
F	**T**	**F**	**F**	**F**
F	**F**	**T**	**F**	**T**
F	**F**	**F**	**F**	**F**

This is already quite a large truth-table, and you won't ever see a larger one in these lectures! (Although you may in exercises ...)

And so on. In general, if a sentence has n atomic components, its truth-table has 2^n rows. This means that truth-tables very quickly get very large and unwieldy. But in many cases we are not actually interested in the whole truth-table. Rather, we are interested either in a specific line of the truth-table, or in whether a truth-table has a line of a specific form. Here are two definitions that capture what is going on at each line:

DEFINITION. A sentence is *true on a truth-value assignment* if and only if it has the truth-value **T** on the truth-value assignment.

DEFINITION. A sentence is *false on a truth-value assignment* if and only if it has the truth-value **F** on the truth-value assignment.

Now consider, for instance, the question whether $(A\&B) \vee C$ is equivalent to $A\&(B \vee C)$. We should expect that they are not. To show this, it suffices to find one truth-value assignment on which one of them is true and the other false. For instance, all the truth-value assignments sharing the fourth row of the truth-table for $(A\&B) \vee C$ will do, because on these assignments, $A\&(B \vee C)$ is false, while $(A\&B) \vee C$ is true, as the following calculation shows, where we only calculate the one row we are interested in now:

A	B	C	$B \vee C$	$A\&(B \vee C)$
F	**T**	**T**	**T**	**F**

Such a construction is called a *shortened truth-table.* This single line of a truth-table together with the 4th line of the truth-table for $(A\&B) \vee C$ shows that there are truth-value assignments on which one of the sentences is true and the other false, which shows that they are not equivalent. Hence to show this we do not need to calculate the complete truth-tables, because we need not consider *all* truth-value assignments, but only *some* of them which have the right properties.

We can now give a definition in terms of truth-value assignments that replaces the notion of logical equivalence:

DEFINITION. Sentences **P** and **Q** of SL are *truth-functionally equivalent* if and only if there is no truth-value assignment on which **P** and **Q** have different truth-values.

We also have the terminology at hand to redefine the notions of truth-functional truth, falsity and indeterminacy:

DEFINITION. A sentence **P** of SL is *truth-functionally true* if and only if **P** is true on every truth-value assignment.

For instance, $A \vee \sim A$ is truth-functionally true, as on every truth-value assignment it comes out as true, as its truth-table shows. Truth-functionally true sentences are also called *tautologies.*

DEFINITION. A sentence **P** of SL is *truth-functionally false* if and only if **P** is false on every truth-value assignment.

For instance, $A\& \sim A$ is truth-functionally false, as on every truth-value assignment it comes out as false, as its truth-table shows. Truth-functionally false sentences are also called *contradictions.*

DEFINITION. A sentence **P** of SL is *truth-functionally indeterminate* if and only if **P** is neither truth-functionally true nor truth-functionally false.

For instance, $(A\&B)\vee C$ is truth-functionally indeterminate, for, as its truth-table shows, it can be true as well as false on different truth-value assignments. Truth-functionally indeterminate sentences are also called *contingent* sentences.

In many cases we are not interested in the whole truth-table for a sentence. This is in particular so if they are very large. Also, mostly we are interested in whether a sentence of SL has one of the properties just defined. To find out whether, for instance, a sentence is truth-functionally true, we need not draw a whole truth-table. All we need to know is that there is no truth-value assignment on which the sentence is false. We can achieve this by testing whether we can find a truth-value assignment on which it is true. Take, for instance, $A \supset (B \supset A)$. To see whether there is a truth-value assignment on which it is false, we can start with assuming that calculating its truth-value for some truth-value assignment gives the truth-value **F**:

A	B	$B \supset A$	$A \supset (B \supset A)$
			F

The truth-table for $\supset$ tells us furthermore that if $A \supset (B \supset A)$ is false, then A must be true and $B \supset A$ must be false. We can enter this information in the table:

A	B	$B \supset A$	$A \supset (B \supset A)$
T		**F**	**F**

The truth-table for $\supset$ also tells us that if $B \supset A$ has the truth-value **F**, then B has the truth-value **T** and A has the truth-value **F**. But we said already that A must be **T** if $A \supset (B \supset A)$ is **F**! Hence our attempt to find a truth-value assignment on which $A \supset (B \supset A)$ is assigned the truth-value **F** failed: we reached a contradiction and cannot complete the shortened truth-table.

Similarly, if we want to know whether a sentence is truth-functionally false, it suffices to test whether there can be a truth-value assignment on which it gets the truth-value **T**. And if we want to find out whether a sentence is truth-functionally indeterminate, it suffices to find out whether is one truth-value assignment on which it is true and another one on which it is false.

It may be necessary to consider more than one line when constructing a shortened truth-table. Take, for instance, $\sim (A\& \sim B) \equiv (A \supset B)$, and assume we want to find out whether it is truth-functionally true. First, we assume that there is a truth-value assignment on which this sentence can be false:

A	B	$\sim B$	$A\& \sim B$	$A \supset B$	$\sim (A\& \sim B)$	$\sim (A\& \sim B) \equiv (A \supset B)$
						F

But then we have two ways of continuing, as the truth-table for $\equiv$ says that a sentence $\mathbf{P} \equiv \mathbf{Q}$ can be false on a truth-value assignment under two conditions: if **P** is true and **Q** is false, and if **P** is false and **Q** is true. So we need to add a further row to cover both cases:

A	B	$\sim B$	$A\& \sim B$	$A \supset B$	$\sim (A\& \sim B)$	$\sim (A\& \sim B) \equiv (A \supset B)$
				T	**F**	**F**
				F	**T**	**F**

We continue with each row separately. Applying the truth-table for negation gives adds the following information: if $\sim (A\& \sim B)$ is **T**, then $A\& \sim B$ is **F**, and if $\sim (A\& \sim B)$ is **F**, then $A\& \sim B$ is **T**:

A	B	$\sim B$	$A\& \sim B$	$A \supset B$	$\sim (A\& \sim B)$	$\sim (A\& \sim B) \equiv (A \supset B)$
			T	**T**	**F**	**F**
			F	**F**	**T**	**F**

Now we can use the information in the first row to give the truth-values of A and $\sim B$, because $A\& \sim B$ can only be **T** if they both are **T**:

A	B	$\sim B$	$A\& \sim B$	$A \supset B$	$\sim (A\& \sim B)$	$\sim (A\& \sim B) \equiv (A \supset B)$
T		**T**	**T**	**T**	**F**	**F**
			F	**F**	**T**	**F**

But now we run into trouble: if $\sim B$ is **F**, then B must be **T**, but if $A \supset B$ is **F**, then A must be **T** and B must be **F**. But B cannot be both **T** and **F**, so we cannot complete the first row of the shortened truth-table. Thus we've closed off one option for constructing a truth-value assignment on which $\sim (A\& \sim B) \equiv (A \supset B)$ is false.

Of course there is still the other option to be considered. In the second row, the **F** under $A \supset B$ gives us the information that A must be **T** and B must be **F**:

A	B	$\sim B$	$A\& \sim B$	$A \supset B$	$\sim (A\& \sim B)$	$\sim (A\& \sim B) \equiv (A \supset B)$
T		**T**	**T**	**T**	**F**	**F**
T	**F**		**F**	**F**	**T**	**F**

But if B is assigned the truth-value **F**, then $\sim B$ must be assigned the truth-value **T**, and as A is assigned the truth-value **T**, too, $A\& \sim B$ must be **T**. But we have already given it the truth-value **F**. Hence again we reach a contradiction can cannot complete the shortened truth-table. This has closed off the other option for $\sim (A\& \sim B) \equiv (A \supset B)$ to be false, and thus it follows that there cannot be a truth-value assignment on which this sentence is false, and hence it must be truth-functionally true.

You will have noticed that there were several options of how to continue entering information into the shortened truth-table. For instance, in the first row, instead of using the information that $A\& \sim B$ is **T**, we could have used the information that $A \supset B$ is **T**. However, using this information would have resulted in more cases to be considered, as there are of course three ways for $A \supset B$ to be true on a truth-value assignment. Similarly, in the second row, instead of using the information that $A \supset B$ is **F**, we could have used the information that $A\& \sim B$ is **F**. Again, this would have resulted in three new cases to be considered. In other words, the choice of which information to use and how to continue trying to complete the shortened truth-table was motivated by a *strategy* to create a shortened truth-table which is as short as possible. A general method for constructing shortened truth-tables economically is always to continue with those truth-values, if possible, which are the only one of their kind in the last row of a truth-table for a connective of SL, i.e. **F** in the case of sentences of the form $\mathbf{P} \vee \mathbf{Q}$ and $\mathbf{P} \supset \mathbf{Q}$, and **T** in the case of sentences of the form $\mathbf{P}\&\mathbf{Q}$.

Lecture 7. The Semantics of SL: Consistency and Validity

At the very beginning of this lecture series it has been claimed that logic is about arguments. But since then we haven't actually seen any arguments anymore. We'll do something about this very soon.Unfortunately, before we can do so we need to work towards finishing the replacement of intuitive logical concepts defined in terms of logical possibility with notions defined in terms of truth-value assignments. Luckily, though, we need only introduce one concept, and then we'll look at a real argument and use the insights from the discussion to define some more terms!

We have so far re-defined the notions of logical truth, falsity, indeterminacy and equivalence in terms of truth-value assignments, and we called the re-defined notions truth-functional truth, falsity, indeterminacy and equivalence, respectively. Next we'll have a look at the notions of logical consistency and inconsistency. Consider the set consisting of the following two sentences:

(1) Jo is a girl.
(2) Jo is a boy.

Is this set logically consistent or inconsistent? You should guess by now what I'm going to say on that issue. Of course, given the meanings of 'boy' and 'girl' in English, these two sentences cannot both be true together, and hence the set should count as logically inconsistent. However, and that's something you've heard a few times by now, logic should be general, rather than being tied to particular subject matters such as boys and girls. As we've seen in the last lecture, the concept of a truth-value assignment completely abstracts from the content of sentences, and in this way insures the kind of generality we are aiming for (recall my comments on formalisation in earlier lectures). So we need to re-define the notions of logical consistency and inconsistency in terms of truth-value assignments. The definitions shouldn't be altogether too surprising, so here they are:

DEFINITION. A set of sentences of SL is *truth-functionally consistent* if and only if there is a truth-value assignment on which all the members of the set are true.

DEFINITION. A set of sentences of SL is *truth-functionally inconsistent* if and only if it is not truth-functionally consistent.

Suppose we use the following interpretation:

Letter:	Sentence it abbreviates:
B	Jo is a boy.
G	Jo is a girl.

We can then see that there is no problem in assigning **T** to both B and G (forgetting, so to say, what the sentences they represent mean!), hence there is at least one truth-value assignment (and in fact infinitely many of them) on which both are true.

If, as a matter of fact, Jo cannot both be a boy and a girl, we can do justice to our intuitions by adding another sentence to the set which expresses this fact, namely $\sim (B \& G)$. Then the following set is truth-functionally inconsistent:

$$\{B, G, \sim (B \& G)\}$$

In case you are not familiar with this notation, it is common to write sets by listing the objects in the set and enclosing them in curly brackets, as in the example (of course, this can only be done if there are only finitely many objects in the set!).

And now it's time for an argument. Consider the following passage I came across in a book on the Black Death in Ireland:[4]

> If the causative pathogen was *Yersinia pestis*, then it was carried by the black rat, which was to be found primarily in large cities, especially ports, and which thrives in dry surroundings such as

[4]Maria Kelly: *The Great Dying: A History of the Black Death in Dublin*, 2002.

> grain depots and in the holds of ships. Because the black rat itself is very sedentary and rarely travels outside a limited radius, the bacillus is usually transmitted over wide areas by the passive transport of the rat and flea. Ships carrying grain or cloth brought infected rats into the ports of Europe, and infected fleas most likely travelled in the merchandise of merchants and mariners. Alternatively, if the epidemic was caused by some other pathogen, then it was most likely carried by merchants and travellers and transmitted by direct contact.

It is fairly obvious (making some non-controversial assumptions) that it follows from this passage that the causative pathogen of the plague was transmitted through transport by ship. This can easily be shown by applying methods of logic in analysing and formalising the passage. The first question to answer is how to translate this argument most economically into SL. We don't need to translate every piece of information we get from this passage with its own sentence letter (although of course we could proceed that way). It suffices to focus on the main strands of the argument. For instance, the information about the habitat of the black rat is not essential to the main point of the first sentence, which is that (1) if the causative pathogen was *Yersinia pestis*, then it was carried by the black rat. Of course, the information this passes over is supportive of a further point made in the next sentence. However, here too we need not pay too much attention to detail, because the main point is that (2) if the causative pathogen was carried by the black rat, then it was transmitted through transport by ship. The third sentence, too, gives some additional information, but again, we can leave it at noting that this information is used to support the truth of the two main points. Notice the overlap in vocabulary used to state (1) and (2). This will inform the use of sentences letters in formalisation as well as the further analysis of the argument. Using phrases we have already used in stating the main points of the first three sentences of the argument, as well as making some non-controversial assumptions about how people travelled to Ireland in prior to 1348, we can say that the fourth sentence states that (3) if the causative pathogen was not *Yersinia pestis*, then it was transmitted through transport by ship.

We have now isolated three premises that are stated in the argument. Our analysis motivates using the following interpretation for purposes of formalisation:

Letter:	Sentence it abbreviates:
A	The causative pathogen was *Yersinia pestis.*
B	The causative pathogen was carried by the black rat.
C	The causative pathogen was transmitted through transport by ship.

The three premises of the argument and their formalisations are then:

Formalisation:		Premise:
(1)	$A \supset B$	If the causative pathogen was *Yersinia pestis*, then it was carried by the black rat.
(2)	$B \supset C$	If the causative pathogen was carried by the black rat, then it was transmitted through transport by ship.
(3)	$\sim A \supset C$	If the causative pathogen was not *Yersinia pestis*, then it was transmitted through transport by ship.

The conclusion of the argument, we said, is that the causative pathogen of the plague was transmitted through transport by ship, i.e. plain C. How can we show that C follows from the premises $A \supset B$, $B \supset C$ and $\sim A \supset C$? Well, it would follow if it is not possible for the premises to be true and the conclusion false. So what we could try to do is to see whether we can construct something like a shortened truth-table, not with the purpose of determining the truth-value of only one sentence, but rather where we assume C to be **F** and $A \supset B$, $B \supset C$ and $\sim A \supset C$ to be all **T**:

A	B	C	$\sim A$	$A \supset B$	$B \supset C$	$\sim A \supset C$	C
				T	**T**	**T**	**F**

I use a double line to separate the atomic components of the premises and conclusion of the argument from its premises, and also to to separate the premises from the conclusion. For clarity's sake I have entered plain C twice in the table, but of course it would have sufficed to enter it only once.

From the assumption that C is **F** and that $B \supset C$ and $\sim A \supset C$ are both **T**, it follows that B and $\sim A$ must both be **F**, by the truth-table for $\supset$:

A	B	C	$\sim A$	$A \supset B$	$B \supset C$	$\sim A \supset C$	C
	F	**F**	**F**	**T**	**T**	**T**	**F**

But if $\sim A$ is **F**, then A must be **T**, and so, as B is **F**, it follows from the truth-table for $\supset$ that $A \supset B$ must be **F**, rather than **T**, as assumed. Thus we reach a contradiction and cannot complete the shortened truth-table. It follows that there is no truth-value assignment on which all premises of the argument, $A \supset B$, $B \supset C$ and $\sim A \supset C$, are true and the conclusions C is false. Hence the argument is valid.

Alternatively, we could of course have constructed complete truth-tables for the premises and conclusions, and then checked whether in each row where a **T** occurs underneath every premises, a **T** also occurs underneath the conclusion, but as they together have three atomic components, this would have resulted in a rather large object.

In the last paragraphs we have initially worked with the intuitive notion of validity defined in terms of possibility, and then worked our way towards a definition of validity in terms of truth-value assignments. The discussion should have indicated what the new definition:

DEFINITION. An argument of SL is *truth-functionally valid* if and only if there is no truth-value assignment on which all the premises are true and the conclusion is false.

DEFINITION. An argument of SL is *truth-functionally invalid* if and only if it is not truth-functionally valid.

What I called 'additional' or 'supportive' information in the analysis of the argument is important to establish that its three premises are in fact true. This is the reason why we don't have to take this information into account if we are just interested in the *validity* of the argument. But the information is, of course, vital to establish whether the argument is *sound*. As was the case with logical validity, an argument is truth-functional validity or invalid quite independently of whether its premises are true. Just as we defined the *deductive soundness* of an argument we could now, if we liked, define what it means that an argument is *truth-functionally sound*. This is a notion which seems to be uncongenial to Bergmann *et al.*, as it does not appear in the book. But anyway, the definition would lay down that an argument is *truth-functionally sound* if and only if it is truth-functionally valid and the premises are true on their intended interpretation.

The way truth-functional validity has been defined makes reference to arguments. Bergmann *et al.* also define a notion of *truth-functional entailment* in addition. Entailment holds between a sentence and a set of sentences of

SL. The definition is:

DEFINITION. A set Γ of sentences of SL *truth-functionally entails* a sentence $\mathbf{P}$ if and only if there is no truth-value assignment on which every member of Γ is true and $\mathbf{P}$ is false.

I am not entirely sure why Bergmann *et al.* define both these notions. Maybe they assume that arguments only have finitely many premises, although this does not follow from their definition, which is that an argument is a set of two or more sentences, one of which is designated as the conclusion and the others as the premise. When we consider entailment, no restriction on the size of Γ is made—although no such restriction has been made on premises of arguments either, as far as I can see! Anyway, we can now introduce some notation to abbreviate 'Γ entails $\mathbf{P}$'. We write this in symbols of the meta-language as $\Gamma \vDash \mathbf{P}$.

We can use this notation to state theorems about the semantics of SL. For instance, in lecture 2 we noted that if a set of sentence is inconsistent, then, as they cannot all be true together, removing one sentence from the set and using the denial of its truth as the conclusion of an argument the premises of which are all the other sentences of the set produces a valid argument. To express this using the symbols introduced in the last paragraph, denote the *union* of two sets Γ and Δ by $\Gamma \cup \Delta$, i.e. this is the set which contains all members of Γ as well as of Δ. Then what has just been said establishes the following:

THEOREM. If $\Gamma \cup \{\mathbf{P}\}$ is truth-functionally inconsistent, then $\Gamma \vDash \sim \mathbf{P}$.

The converse is true, too:

THEOREM. If $\Gamma \vDash \sim \mathbf{P}$, then $\Gamma \cup \{\mathbf{P}\}$ is truth-functionally inconsistent.

For, if Γ entails $\sim \mathbf{P}$, then whenever all sentences of Γ are assigned the truth-value $\mathbf{T}$, $\sim \mathbf{P}$ is also assigned the truth-value $\mathbf{T}$, and thus $\mathbf{P}$ must be assigned the truth-value $\mathbf{F}$, by the truth-table for negation. Thus whenever all of Γ are true, $\mathbf{P}$ is false, i.e. the sentences in the union $\Gamma \cup \{\mathbf{P}\}$ can never be all true together, hence this set is truth-functionally inconsistent. *Q.e.d.*

Lecture 8. Proof-Theory for SL: Conjunction, Implication

In the last lectures we discussed the semantics of SL, including definitions of the concepts of validity and entailment in terms of truth-value assignments and we have introduced ways of determining whether arguments are valid or sentences entailed by sets of sentences. In principle, we could leave matters there; in a sense, we have achieved what we wanted: we have devised methods for determining whether arguments are valid and invalid, which was one of the aims we formulated for the study of logic. But logic, as we have been doing it so far, always had two sides, a semantic one and a syntactic one, and so now that we have dealt with semantic side of arguments, we can expect this to be supplemented considerations belonging to the syntactic side. Furthermore, the tests for validity introduced in the last lecture always assumed that we had been given some argument first, i.e. we had a set of premises and a conclusion, and then we applied the shortened truth-table method or the complete truth-tables to figure out whether the argument is valid or not, i.e. whether there is a truth-value assignment making all the premises true and the conclusion false or not. But in many cases of reasoning in real life and philosophy we haven't actually been given an argument, but only a set of assumptions. In this and the next lecture we will introduce methods for deriving sentences from other sentences, which allow us to construct arguments from assumptions. We call such constructions *deductions* or *derivations*.

We proceed in the following way: for each connective of the language SL, we will give rules that specify what follows from a sentence with this connective as the main connective, and rules that specify under which conditions we can infer a sentence with this connective as the main connective. These rules are called, unsurprisingly, *rules of inference* or *derivation rules*. The first kind of rules are called *introduction rules*, the second kind *elimination* *rules*. The introduction rules specify how to introduce a formula with the main connective into the derivation on the basis of formulas that are already in the deduction; the elimination rules, however, do not in some sense 'get rid' of some formula: the formula is still in the deduction. The formula is 'eliminated' in the sense that we have inferred some other sentence from it: the eliminated formula is a preliminary conclusion on the way to the conclusion we actually aim to derive from the assumptions given and is 'eliminated' because we have shifted our attention from it to something else. This way we set up what is called a *natural deduction system*: the rules mirror more or less closely moves we naturally make when reasoning (up to a certain point). The system of derivation rules for the connectives $\sim$, $\supset$, $\vee$, $\&$ and $\equiv$ is called SD.

The rules belong to the realm of syntax, as they do not refer to truth-value assignments, but only to the structure of sentences of SL. Of course, to motivate why we use these rules rather than others we will refer to the intended interpretations of the logical connectives, and this may imply referring to their truth-tables. Once formulated, however, the rules themselves will not make any reference to truth-tables, and no reference to the intended meanings of the connectives, although the rules also succeed in conferring a certain meaning to the connectives, as they codify their use in deductive arguments, and this meaning will inevitably be very close to the meanings conferred on them by the truth-tables—indeed, it is arguable that the meanings conferred on connectives by truth-tables and the meaning conferred on them by the rules we are going to give are identical, at least assuming that every sentences is either true or false.[5]

We first need to address the question of how to structure derivations. We will do this by first specifying how derivations are started, and then each rule will tell us how to continue. You'll recognise the principle: that's how we proceeded in the definitions of what counts as a sentence of SL: we give a basic case, and then explain how to go on. To start a derivation, we list assumptions of the derivation one underneath the other, then draw a vertical line to their left and number each and write 'Assumption' to the right of each assumption, and finally close them off from what is going to be the rest of the derivation by drawing a horizontal line underneath them. So, if, for

[5]Why is this so? Well, assuming that the rules we are interested in are valid ones, i.e. always lead from true premises to true conclusions, we can read the rules off the truth-tables (by asking what rules the truth-tables make valid), and conversely, assuming that every sentences is either true or false we can read off the truth-tables from the rules.

instance, your assumptions are A, B and C, we start a derivation by writing:

1	A	Assumption
2	B	Assumption
3	C	Assumption

Derivations are then continued by adding rows generated by applications of rules of inference. Each new line is also numbered, and we specify how the line has been derived, by which rules applied to which lines.

We need some rules for the connectives. Let's start with the simplest one, conjunction. Obviousl a conjunction **P**&**Q** follows if both its conjuncts **P** and **Q** are given. This gives the introduction rule for conjunction. Conversely, from the conjunction **P**&**Q** both **P** and **Q** follow. This gives the elimination rules for conjunction. We can write the first rule, more congenial to the subject of formal logic, in the following way:

Conjunction Introduction (&I)

	P
	Q
▷	**P**&**Q**

This is to be read as meaning that, if amongst the assumptions of the deduction or amongst any sentence derived from them there are sentences **P** and **Q**, you may add a new row to the deduction in which you write **P**&**Q**. We will have to make this a little bit more precise when all the rules of SD have been formulated, but for the time being it is precise enough.

Furthermore, if on some line of the derivation **P**&**Q** occurs, we may add a new line, in which we can write **P** or **Q**, which we write in the following way:

Conjunction Elimination (&E)

	P&**Q**	or		**P**&**Q**
▷	**P**		▷	**Q**

These are all the rules for conjunction.

As Arthur Prior once observed, the best way to learn logic is to do logic, so let's do a simple derivation, using the rules for conjunction, to see how the system of natural deduction works. For instance, we can show that from $(A\&B)\&C$ we may infer $A\&(B\&C)$. Notice that the rules are to be understood as quite rigid: if you want to infer **P**&**Q**, you need in the derivation *first* a row with **P** in it and *in some later row* you need **Q**. This may sound somewhat pedantic, which of course it is, but it helps to remember that the rules are supposed to be the kind of thing that a computer could be programmed to use, and a computer has a rather limited ability to carry out commands and only does exactly what you tell it to do. The rules are intended to be of such a kind that no ingenuity is needed to apply them. Anyway, here is the deduction:

1	$(A\&B)\&C$	Assumption
2	$A\&B$	1 &E
3	A	2 &E
4	B	2 &E
5	C	1 &E
6	$B\&C$	4, 5 &I
7	$A\&(B\&C)$	3, 6 &I

That's a very simple and not very interesting derivation. We need some more rules for connectives to make derivations more exciting.

The next connective we'll add to SD is the conditional. It is obvious what its elimination rule should be: if you have $A \supset B$ on a line in a derivation, and you also have A somewhere, then you can derive B:

Conditional Elimination (⊃E)

	$\mathbf{P} \supset \mathbf{Q}$
	$\mathbf{P}$
▷	$\mathbf{Q}$

This rule is often called *modus ponens*.

The introduction rule for $\supset$ is more complicated to state than the ones we had so far, although it is not difficult to see under which conditions you can derive $A \supset B$: if you can show that, assuming A, B can be derived, then you should be allowed to derive $A \supset B$. Notice that under these circumstances, $A \supset B$ should be true independently of whether A is true or false, because we conclude that *if* A is true, so is B. The conclusion $A \supset B$ does not depend on the truth of A anymore: although the truth of B on its own is conditional upon the truth of A, the conditional $A \supset B$ is true unconditionally.

The way we have just stated informally the introduction rule for $\supset$ requires us to add new assumptions to derivations, which we call *auxiliary assumptions*, in contrast to the *primary assumptions*, which are assumptions taken from a set of sentences which we assume to be true. We do this by adding a new 'level' to the derivation, i.e. we allow there to be more than one vertical line to the left of formulas. At each step of a derivation, we can add new auxiliary assumptions, which open up new *subderivations* and add a new vertical line. For instance, after having derived a few sentences from our assumptions A, B and C, we may wish to add a new auxiliary assumption D:

1	A	Assumption
2	B	Assumption
3	C	Assumption
4	$A\&B$	1, 2 &I
5	$B\&C$	2, 3 &I
6	$\quad D$	Assumption

If this subdeduction then ends with a sentence $D\&A$, say, we can apply material conditional introduction to derive $D \supset (D\&A)$. Notice what we have done then: we have, under the assumption D, derived $D\&A$, and then conclude that $D \supset (D\&A)$, i.e that if D, then $(D\&A)$. This conclusion does not depend on the assumption D anymore: we have incorporated the assumption into the antecedent of a conditional, and due to the nature of the conditional, this allows us to 'close off' the subderivation; we do not need to make the auxiliary assumption D any more.

The rules for conditional introduction, then, is this:

Material Conditional Introduction (⊃I)

		P
		⋮
		Q
▷	**P** ⊃ **Q**	

We indicate that the subderivation has been closed off by returning to the level of the derivation at which the auxiliary assumption has been made.

Here is an example of a derivation using the rules for the material conditional. We show that from $A \supset B$ and $B \supset C$ we may derive $A \supset C$:

1	$A \supset B$		Assumption
2	$B \supset C$		Assumption
3		A	Assumption
4		B	1, 3 ⊃E
5		C	2, 4 ⊃E
6	$A \supset C$		3–5 ⊃I

Due to the framework in which derivations are carried out, we need to add a rule which is structural in character and has nothing to do with connectives of the formal language SL. The rule allows us to repeat a sentence, and is therefore called Repetition:

Repetition (R)

	P
	⋮
▷	**P**

This may look like a rather unnecessary rule, but it is indeed required at times. Remember that we apply the rules quite mechanically. So if we want to introduce a conjunction **P**&**Q**, we strictly speaking first have to have a row with **P** in it and then a later one with **Q** in it. Thus the rules is needed when deriving $B\&A$ from $A\&B$, if we first derive A and then B from the assumption:

1	$A\&B$	Assumption
2	A	1 &E
3	B	1 &E

Strictly speaking, we can now not apply &I, because that would only give us what we already had, namely $A\&B$. So we need repetition:

1	$A\&B$	Assumption
2	A	1 &E
3	B	1 &E
4	A	2 R
5	$B\&A$	3, 4 &I

Of course, had we derived first B and then A, we could have derived $B\&A$ from $A\&B$ without using repetition:

1	$A\&B$	Assumption
2	B	1 &E
3	A	1 &E
4	$B\&A$	2, 3 &I

In many cases, the rule of repetition is not one that in fact needs to be used – often there are ways of doing derivations which avoid the rule –, but it allows for a certain amount of flexibility once you have already started your deduction.

Repetition is, however, necessary in some cases. Consider, for instance, that $B \supset A$ should be derivable from A, as, if A is true, so is $B \supset A$. Here we need to apply repetition in order to get a derivation of A from B in the subderivation:

1	A	Assumption
2	B	Assumption
3	A	1 R
4	$B \supset A$	2–3 ⊃I

Without repetition the introduction rule for ⊃ would not allow us to infer $B \supset A$.

Maybe you are asking yourself now, but how do I prove $A \supset (B \supset A)$, which is a truth-functional truth? If a sentence is a truth-functional truth, then it is true independently of any assumptions we are making; no assumption is needed so ensure its truth. Notice that the vertical lines to the left of sentences indicate which assumptions are made to ensure that the sentences in each line below the assumptions are true. Thus, if a sentence is a logical truth, it should have a derivation which ends with a row that has a line to its left that is not begun by a set of assumptions, i.e. the deduction has only auxiliary assumptions, but no primary assumptions, and all the subdeductions begun by them have been closed off. As an example, here is a proof of $A \supset (B \supset A)$:

1	A	Assumption
2	B	Assumption
3	A	1 R
4	$B \supset A$	2–3 ⊃I
5	$A \supset (B \supset A)$	1–4 ⊃I

It is obvious that the leftmost vertical line in this deduction plays hardly any role at all. Alternatively, thus, we could adopt the convention that deductions of logical truths end in a row with a formula without a vertical line to its left, to indicate that the sentence is true unconditionally. And thus the proof of $A \supset (B \supset A)$ is the following:

1	$\mid A$	Assumption
2	$\mid \mid B$	Assumption
3	$\mid \mid A$	1 R
4	$\mid B \supset A$	2–3 ⊃I
5	$A \supset (B \supset A)$	1–4 ⊃I

A logical truth does not depend on anything for its truth, so no assumptions need to be made to ensure its truth, and as there is no vertical line to the left of $A \supset (B \supset A)$, as a matter of fact the assumptions A and B made in the course of the derivation are only auxiliary assumptions.

Here is a derivation using the rules for conjunction and for the conditional. We'll derive $(A \& B) \supset C$ from $A \supset (B \supset C)$:

1	$\mid A \supset (B \supset C)$	Assumption
2	$\mid \mid A \& B$	Assumption
3	$\mid \mid A$	2 &E
4	$\mid \mid B \supset C$	1, 3 ⊃E
5	$\mid \mid B$	2 & E
6	$\mid \mid C$	4, 5 ⊃E
7	$\mid (A \& B) \supset C$	2–6 ⊃I

The converse is also true, i.e from $A \& B \supset C$ we can derive $A \supset (B \supset C)$:

1	$A\&B \supset C$	Assumption
2	$\quad A$	Assumption
3	$\quad\quad B$	Assumption
4	$\quad\quad A\&B$	2, 3&I
5	$\quad\quad C$	1, 4 ⊃E
6	$\quad B \supset C$	3–5 ⊃I
7	$A \supset (B \supset C)$	2–6 ⊃I

Notice how here each auxiliary assumption gets its own subderivation. Although derivations can start with any number of assumptions (as long as that number is finite), subderivations always only start with one assumption. The reason for this is that we introduce auxiliary assumptions and subdeductions only to close them off by applications of rules like conditional introduction, and these rules dictate the structure of the beginnings of subdeductions. We'll introduce three more rules that allow to close off subdeductions in the next lectures.

A last example: from $A \supset B$, we can derive $A\&C \supset B\&C$

1	$A \supset B$	Assumption
2	$\quad A\&C$	Assumption
3	$\quad A$	2 &E
4	$\quad B$	1, 3 ⊃E
5	$\quad C$	2 &E
6	$\quad B\&C$	4, 5 &I
7	$A\&C \supset B\&C$	2–6 ⊃I

Lecture 9. Proof-Theory for SL: Negation

Last week we introduced how to do deductions. This is the most complicated thing we've done so far, so don't be disappointed if you don't quite understand how to do a deduction at first attempt. It really is the kind of thing you have to practice: you have to see some examples of deductions and do some yourself, and eventually you'll get the hang of it. One problem surely is that deductions look rather alien and it may not be completely lucid why they have the form they do. Here it helps to remember that we need *some* way of writing down deductions, some way of how to organise reasoning from assumptions to a conclusion in a systematic way. There are in fact many different ways of achieving this aim. There is no necessity that we do deductions in the way we do it here. Thus it is not as if there is always something to be understood for each detail of a deduction: certain aspects of deductions are simply chosen arbitrarily, and these are things you have to learn by heart, as it were, not something that has a deeper reason. It's a bit like grammar: there is no deeper, logical reason why the third person singular of many verbs of English is formed by adding an 's' at its end. That's just how we speak. Similarly, there is no deeper reason why we write sentences underneath each other next to vertical lines. That's just how we've decided to do deductions. What is important is this: first, we have rules of inferences that tell us precisely what can be done at each step in a deduction and how conclusions can be arrived at, and secondly, there is some way of organising deductions that shows which assumptions are made at which point in the deduction, with ways of separating subdeductions from the main deduction so that we can see which assumptions are made on the way to which conclusions.

Last week we introduced the rules for conjunction and the material conditional. This already allowed us to do a few deductions, but not yet very interesting ones. Today we'll add the rules for negation.

Suppose you reason from some assumptions $A_1 \dots A_n$ that, say, a philosopher makes. Now you ask yourself, is what the philosopher says consistent with another assumption, say B? For instance, you might think that B is an assumption well worth considering. Now suppose you reason from $A_1 \dots A_n$ together with the auxiliary assumption B and you arrive at two sentences, C and $\sim C$. What does that mean? The philosopher's assumptions $A_1 \dots A_n$ together with B entail two sentences that cannot be true together, i.e. the set $\{A_1 \dots A_n, B\}$ is inconsistent: it cannot be the case that all sentences in this set are true, because, if that were the case, C as well as $\sim C$ must be true, as each sentences is entailed by the set, which is impossible. Thus at least one of the sentences in the set $\{A_1 \dots A_n, B\}$ must be false. Then granting the philosopher that his assumptions $A_1 \dots A_n$ are true, it follows that B must be false, and so $\sim B$ must be true. Thus we can infer $\sim B$ from $A_1 \dots A_n$, i.e. what the philosopher holds true entails $\sim B$. Hence we have now shown that the philosopher is committed to $\sim B$, and that is of course independent of the auxiliary assumption B: we have shown that the auxiliary assumption that B together with the other assumptions $A_1 \dots A_n$ leads to a contradiction, so the philosopher cannot hold that B is true, and thus he must hold that B is false, i.e. $\sim B$ is true.

Summarising this reasoning in rule-form, we get the following:

Negation Introduction (∼I)

| | **P**
| | ---
| | **Q**
| | ∼ **Q**
▷ | ∼ **P**

Applying negation introduction allows us to close off a subderivation which was started with an auxiliary assumption which has been assumed for *reductio ad absurdum*, as this rule is often called. The conclusion drawn does not depend on the truth of the auxiliary assumption any more, because the assumption has been reduced to absurdity, and its negation is asserted instead.

Here is an easy deduction making use of this rule. Suppose you assume $A \supset B$ and $\sim B$. It should follow that $\sim A$, and we can show that it does with the following deduction:

1	$A \supset B$	Assumption
2	$\sim B$	Assumption
3	$\quad A$	Assumption
4	$\quad B$	1, 3 $\supset$E
5	$\quad \sim B$	2 R
6	$\sim A$	3–5 $\sim$I

The inference from $A \supset B$ and $\sim B$ to $\sim A$ is often called *modus tollendo tollens*.

Assume that you assume $\sim A \supset \sim B$ and B. Then it should follow that A is true. For, if A was false, i.e. $\sim A$ was true, then, by $\supset$E, $\sim B$ would also be true, but, as we assumed B to be true, we have arrived at a contradiction; thus the auxiliary assumption $\sim A$ cannot be true, i.e. must be false, and thus A is true. This exemplifies another rule for negation. If from an auxiliary assumption of the form $\sim \mathbf{P}$ we derive two sentences $\mathbf{Q}$ and $\sim \mathbf{Q}$, then we can infer $\mathbf{P}$ and close off the subderivation starting with $\sim \mathbf{P}$:

Negation Elimination ($\sim$E)

	$\quad \sim \mathbf{P}$
	$\quad \mathbf{Q}$
	$\quad \sim \mathbf{Q}$
$\triangleright$	$\mathbf{P}$

This rule is also called *reductio ad absurdum*: this time it is $\sim \mathbf{P}$ that has been reduced to absurdity, and therefore $\mathbf{P}$ must be true. In fact, the two rules are equivalent, if $\sim\sim A$ is equivalent to A, which it is as the semantics of SL shows. Replacing $\mathbf{P}$ by $\sim \mathbf{P}$ in $\sim$E in fact gives an application of $\sim$I with conclusion $\sim\sim \mathbf{P}$. Hence instead of using the present rule for negation elimination, we could use as an alternative rule the rule of double negation elimination:

<u>*Double Negation Elimination* ($\sim\sim$E)</u>

	$\sim\sim \mathbf{P}$
$\triangleright$	$\mathbf{P}$

Going back to our rule of negation elimination, we can show that from $\sim A \supset\sim B$ and B, A can be derived:

1	$\sim A \supset \sim B$	Assumption
2	B	Assumption
3	$\sim A$	Assumption
4	B	2 R
5	$\sim B$	1, 3 $\supset$E
6	A	3–5 $\sim$E

We could give an alternative deduction using negation introduction and double negation elimination: then instead of applying $\sim$E in line 5, we would apply $\sim$I and derive, not A, but $\sim\sim A$. If we apply double negation elimination next, we get the desired conclusion A, but in seven steps, rather than just six.

For the rest of this lecture we'll mainly practice doing deductions. Now that we have the rules for negation, conjunction and implication we can deduce some more interesting things. But first here is a shorter way of saying that from assumptions $A_1 \ldots A_n$ we can derive B: we write this as $A_1 \ldots A_n \vdash B$.

Example 1. $\sim A \vdash \sim (A \& B)$:

1	$\sim A$	Assumption
2	$\quad A \& B$	Assumption
3	$\quad A$	2 &E
4	$\quad \sim A$	1 R
5	$\sim (A \& B)$	2–4 $\sim$I

Example 2. $A \supset B \vdash \sim (A \& \sim B)$:

1	$A \supset B$	Assumption
2	$\quad A \& \sim B$	Assumption
3	$\quad A$	2 &E
4	$\quad B$	1, 3 $\supset$E
5	$\quad \sim B$	2 &E
6	$\sim (A \& \sim B)$	2–5 $\sim$I

Example 3. $A \supset B, A \supset C \vdash A \supset (B \& C)$:

1	$A \supset B$	Assumption
2	$A \supset C$	Assumption
3	$\quad A$	Assumption
4	$\quad B$	1, 3 $\supset$E
5	$\quad C$	2, 3 $\supset$E
6	$\quad B \& C$	4, 5 &I
7	$A \supset (B \& C)$	3–6 $\supset$I

Example 4. $\sim (A \& \sim B) \vdash A \supset B$:

1	$\sim (A \& \sim B)$	Assumption
2	A	Assumption
3	$\sim B$	Assumption
4	$A \& \sim B$	2, 3 &I
5	$\sim (A \& \sim B)$	1 R
6	B	3–5 ∼E
7	$A \supset B$	2–6 ⊃I

Example 5. $A \supset B, A \supset \sim B \vdash \sim A$:

1	$A \supset B$	Assumption
2	$A \supset \sim B$	Assumption
3	A	Assumption
4	B	1, 3 ⊃E
5	$\sim B$	2, 3 ⊃E
6	$\sim A$	3–5 ∼I

This logical inference is obviously very close to *reductio ad absurdum*. It shows that if a sentence implies another sentences and also its negation, then it must be false.

For a change, let's have a look at some philosophical arguments. *Reductio ad absurdum* is one of the most common principles of reasoning in philosophy. Consider, for instance, reasoning that leads to a paradox: a paradox often (although not always) has the form that certain assumptions lead to a contradiction. Here is an example involving two sentences p and s:

p: s is false.
s: p is true.

If p is true, s must be false, as that is what p says, but then p cannot be true, as otherwise what s says is true. If p is false, s must be true, but then p cannot be false, as that is what s says. Either way, we have a contradiction.

What *reductio ad absurdum* tells you now is that some of your assumptions must have been false. In some cases of paradoxes it is notoriously difficult to say what the wrong assumption was, and in some cases, indeed, such an investigation has driven rather deep research in mathematics. There are other examples of paradoxes where it is claimed to be straightforward what assumptions to reject. Consider, for instance, an argument that leads to Zeno's Paradox of the Arrow:

<u>Zenos Paradox of the Arrow:</u>

(1) Consider an arrow moving on a line from point a to point b.
(2) Then the arrow must pass each point of the line one by one: first it is at a, then at a_1, then at a_2 etc. etc. until it reaches b_{n-1}, b_n and then finally b.
(3) Let x be a point on the line.
(4) Assume the tip of the arrow is in point x.
(5) Then the arrow is not moving, because what is at one point is at rest.
(6) But then the arrow does not pass each point of the line one by one, as it is not moving.
(7) But (6) contradicts (2).
(8) So the arrow is not moving.

We can formalise this argument by letting A stand for 'The arrow moves' and B for 'The arrow passes each point one by one'. Zenos argument makes use of the principle of *reductio ad absurdum*. Steps (1)–(2) show that the assumption that (A) the arrow is moving, implies that (B) it passes each point one by one (so we could also conclude $A \supset B$). Steps (1)–(6) show that the assumption that (A) if the arrow is moving, also implies that ($\sim B$) it does not pass each point one by one (so we could conclude $A \supset\sim B$). Hence A entails two sentences which cannot be true together, and Zeno concludes by *reductio ad absurdum* that ($\sim A$) the arrow is not moving.

Now consider the following counter-argument to Zeno's *reductio ad absurdum* of the assumption that the arrow moves:

Against Zeno's Paradox of the Arrow:

(1) Assume Zeno is right and there is no movement.
(2) Nevertheless, the arrow changes its position from a to b.
(3) So there is movement, because that's just what we mean by movement.
(4) Hence even if there is no movement, there is movement.
(5) Therefore, there is movement.

Is this a good argument? Let A be 'There is movement'. Steps (1)–(4) show that if ($\sim A$) there is no movement, then (A) there is movement, i.e. $\sim A \supset A$. Let's see whether we can derive A from this assumption. To do so, we assume $\sim A \supset A$ and $\sim A$ and see whether we can derive a contradiction. If so, we can apply negation elimination and conclude A.

This does indeed work:

1	$\sim A \supset A$	Assumption
2	$\sim A$	Assumption
3	A	1, 2 ⊃E
4	$\sim A$	2 R
5	A	2–4 ∼E

The counter-argument to Zeno's Paradox is indeed valid. The deduction establishes a principle called *consequentia mirabilis*: if A is entailed by its own negation, then A must be true: $\sim A \supset A \vdash A$. I presume you can guess why this inference has been called 'the miraculous consequences'!

Lecture 10. Proof-Theory for SL: Disjunction, Basic Notions of Proof-Theory

So far we have introduced derivation rules for &, $\sim$ and $\supset$. We now need rules for $\vee$. The introduction rule for this symbol is straightforward: if A is true, so are $A \vee B$ and $B \vee A$, so if we have A somewhere in a derivation, we should be able to add $A \vee B$ or $B \vee A$. In other words, the introduction rule for $\vee$ is the following:

Disjunction Introduction ($\vee$I)

	P			**P**
		or		
$\triangleright$	$\mathbf{P} \vee \mathbf{Q}$		$\triangleright$	$\mathbf{Q} \vee \mathbf{P}$

Using this rule, we can show that $\sim (A \vee B) \vdash \sim A \& \sim B$, which is on half of one of DeMorgan's Laws. Let's first reflect on how to construct the deduction. We want to derive a *conjunction*. We can expect the last step of the deduction to be by &I, as both conjuncts must follow from the premises, if their conjunction does. So the first part of the strategy is for deriving $\sim A\$ \sim B$ from $\sim (A \vee B)$ is to derive both $\sim A$ and $\sim B$ as a preparation for applying &I. Furthermore, $\sim A$ and $\sim B$ are the *negations* of $\sim A$ and $\sim B$, resp., and we can expect these to be derivable by $\sim$I, as, if $\sim A$ follows from $\sim (A \vee B)$, assuming A must lead to a contradiction. More specifically, $A \vee B$ is then derivable, which contradicts $\sim (A \vee B)$. Similarly for B. Implementing this strategy gives the following deduction:

1	$\sim (A \vee B)$	Assumption
2	$\quad A$	Assumption
3	$\quad A \vee B$	2 ∨I
4	$\quad \sim (A \vee B)$	1 R
5	$\sim A$	2–4 ∼I
6	$\quad B$	Assumption
7	$\quad A \vee B$	6 ∨I
8	$\quad \sim (A \vee B)$	1 R
9	$\sim B$	6–8 ∼I
10	$\sim A \& \sim B$	5, 9 &I

Here is another example: $\sim (A \& B) \vdash \sim A \vee \sim B$. Let's again think about a strategy for construction the deduction. This time, we cannot expect the last step of the derivation to be by an introduction rule of the main connective of its conclusion, as neither $\sim A$ nor $\sim B$ should follow from $\sim (A \& B)$ alone. Reflection on the form of the inference rules available in SD shows that the only rule we can reasonably expect to end the deduction is ∼E.[6] Hence we assume, for *reductio ad absurdum*, $\sim (\sim A \& \sim B)$ and attempt to derive a contradiction as a preparation for applying ∼E. How are we going to get a contradiction? Well, neither assuming A nor assuming B on their own will suffice, because we neither have $\sim A$ nor $\sim B$, nor do A and B alone lend themselves to deriving sentences contradiction $\sim (A \& B)$ or $\sim (\sim A \vee \sim B)$, as inspection of the rules of SD shows. Howevere, if we assume *both* of them, clearly we can derive a sentence contradicting $\sim (A \& B)$, i.e. $A \& B$. This then allows us to reduce one of these two assumptions, say B, to absurdity, so that we can apply ∼I. This in turn allows us to derive a sentence contradicting $\sim (\sim A \vee \sim B)$, i.e. $\sim A \vee \sim B$, which allows us to reduce the auxiliary assumption A to absurdity, and again we can apply ∼I, this time with the conclusion $\sim A$. And this once more allows us to derive a sentence contradicting $\sim (\sim A \vee \sim B)$, namely, as before, $\sim A \vee \sim B$. And now we have

[6] Of course there can be other deductions not ending with this rule, but these won't be very straightforward ones.

almost derived the desired conclusion, because closed off the subdeductions beginning with the auxiliary assumptions A and B, and what is left is the subdeduction beginning with the auxiliary assumption $\sim(\sim A \vee \sim B)$. The latter has now also has been reduced to absurdity and hence we can apply ∼E, to derive the conclusion $\sim A \vee \sim B$. In other words, the derivation is the following:

1	$\sim(A \& B)$	Assumption
2	$\quad \sim(\sim A \vee \sim B)$	Assumption
3	$\quad\quad A$	Assumption
4	$\quad\quad\quad B$	Assumption
5	$\quad\quad\quad A \& B$	3, 4 &I
6	$\quad\quad\quad \sim(A \& B)$	1 R
7	$\quad\quad \sim B$	4–6 ∼I
8	$\quad\quad \sim A \vee \sim B$	7 ∨I
9	$\quad\quad \sim(\sim A \vee \sim B)$	2 R
10	$\quad \sim A$	3–9 ∼I
11	$\quad \sim A \vee \sim B$	10 ∨I
12	$\quad \sim(\sim A \vee \sim B)$	2 R
13	$\sim A \vee \sim B$	2–12 ∼E

This is one half of another one of DeMorgan's Laws.

What about the elimination rule for disjunction? Suppose you know that $A \vee B$ is true. Assume you can show that from A a sentence C follows, and also that the same sentence follows from B. Then C must be true, because either A is true or B is true: whatever is the case, C follows from each one. Hence we use the following rule as the elimination rule for ∨:

Disjunction Elimination (∨E)

| $\mathbf{P} \vee \mathbf{Q}$
| | $\mathbf{P}$
| |---
| |
| | $\mathbf{R}$
| | $\mathbf{Q}$
| |---
| |
| | $\mathbf{R}$
▷ | $\mathbf{R}$

In other words, if you have a sentence of the form $\mathbf{P} \vee \mathbf{Q}$ somewhere on the derivation, and now you make the auxiliary assumption $\mathbf{P}$ and show that from it a sentence $\mathbf{R}$ follows and you also show that the same sentence follows from the auxiliary assumption $\mathbf{Q}$, then you can close off both subderivations and infer $\mathbf{R}$, as then $\mathbf{R}$ follows from $\mathbf{P} \vee \mathbf{Q}$ alone. This rule is often called *proof by cases*: if you have derived $\mathbf{P} \vee \mathbf{Q}$, you know that either of two cases must hold, namely either $\mathbf{P}$ or $\mathbf{Q}$, although you may not know which; you then consider each case separately to see whether in each case something else follows; if that is the case, you know it must be true, because your considerations have exhausted all cases and each case entails the same sentence.

We can use disjunction elimination to prove the other halves of DeMorgan's Laws. First, $\sim A \vee \sim B \vdash \sim (A \& B)$. The strategy for constructing the derivation is the following. The premise is a disjunction. So we can expect we have to use disjunction elimination, because there is no other way in which we could use the premise to derive the conclusion (we can't, for instance, expect to use it as part of a *reductio ad absurdum*, as there is no straightforward way of deriving the negation of the premise). I.e. we need two subdeductions beginning with auxiliary assumptions $\sim A$ and $\sim B$, respectively. The conclusion is the negation of the conjunction $A \& B$. Now if in each subdeduction we begin a new subdeduction with $A \& B$ as the auxiliary assumption, we can easily derive sentences contradicting $\sim A$ and $\sim B$ respectively, i.e. A and B, by conjunction elimination, which then in turn allows us to close off the subdeductions beginning with $A \& B$. The result is that the two subdeductions we need as a preparation for disjunction elimination, the one beginning with $\sim A$ and the other one with $\sim B$, each end with an application of negation introduction, deriving the formula $\sim (A\&B)$. Hence the conditions for applying $\vee$E are fulfilled and we can close the subdeductions off. The derivation then looks like this:

1	$\sim A \vee \sim B$	Assumption
2	$\sim A$	Assumption
3	$A\&B$	Assumption
4	A	3 &E
5	$\sim A$	2 R
6	$\sim (A\&B)$	3–5 ∼I
7	$\sim B$	Assumption
8	$A\&B$	Assumption
9	B	8 & E
10	$\sim B$	7 R
11	$\sim (A\&B)$	8–10 ∼I
12	$\sim (A\&B)$	1, 2–6, 7–11 $\vee$E

Next comes the derivation showing that $\sim A\& \sim B \vdash \sim (A \vee B)$. This is going to be somewhat more complicated than the deductions we already had. We need to derive the negation of a disjunction. We can expect the last step to be by negation introduction, i.e. after assuming the primary assumption $\sim A\& \sim B$ we start a subdeduction beginning with the auxiliary assumption $A \vee B$. We now need to derive a contradiction in this subdeduction. We now have a conjunction and a disjunction as premises to work with. Notice that the conjunction conjoins the negations of the sentences that the disjunction disjoins. Hence contradictions are forthcoming, if in the next steps we begin new subdeductions with A and B as auxiliary assumptions, as preparations for applying $\vee$E. The aim would be to derive, from each, A and B, a sentence of the form $\mathbf{P}\& \sim \mathbf{P}$,[7] then to apply $\vee$E to close off the two subdeductions,

[7] You may have noticed that I have used the word 'contradiction' ambiguously since the last lecture. The term 'contradiction' was introduced as a shorter way of characterising a sentence is truth-functionally false. Now, in discussing the proof-theory of SL, I have used

and then to apply &E to this contradiction twice, to derive **P** and ∼**P** on their own, which then allows us to apply ∼E to close off the subdeduction beginning with $A \vee B$, to derive the desired conclusion $\sim (A \vee B)$. The problem is that in the two subdeductions beginning with A and B, we can so far only derive *different* contradictions, i.e. in the one beginning with A we get a contradiction by applying &E to $\sim A \& \sim B$, deriving $\sim A$ and then applying &I, and in the other one we can get a contradiction by deriving $\sim B$ and proceeding in the same fashion. But we need to derive *the same* contradiction in both subdeductions. To see how this can be done we need to digress a little. Suppose you have an inconsistent set of premises, e.g. A and $\sim A$. We said that the definitions of logical validity or truth-functional validity entail that a inconsistent sets entail everything. We should expect something similar to be the case in proof-theory, and we can indeed show that there is a deduction from a contradiction as assumptions to any sentence whatsoever as conclusion:

1	A	Assumption
2	$\sim A$	Assumption
3	$\quad \sim D$	Assumption
4	$\quad A$	1 R
5	$\quad \sim A$	2 R
6	D	3–5 ∼I

Clearly, nothing in this deduction hangs on the shapes of A and D. This observation helps the strategy for proving $\sim A \& \sim B \vdash \sim (A \vee B)$. For, instead of D we could use an arbitrary contradiction, say $C \& \sim C$, and then derive this contradiction in the two subdeductions beginning with A and B, respectively. This, then, to the same contradiction having been derived from

it one the one hand to characterise that which is needs to be derived for the conditions of applications of ∼I and ∼I to be given, i.e. two sentences, one of which is of the form **P** and the other of the form $\sim$ **P**. Contradictions, then, are the 'absurdities' referred when we called these rules versions of *reductio ad absurdum*. One the other hand, I have used the term to name sentences of the form $\mathbf{P} \& \sim \mathbf{P}$. This ambiguity is harmless, as we could formulate the negation rules in such a way that, instead requiring a contradiction of the first kind, they require one of the second kind.

both A and B, and so we can apply $\vee$E and finally, as this contradiction is also the conclusion of the application of this rule, apply negation introduction and close off the subdeduction beginning with $A \vee B$, which completes the derivation. In symbols, it looks like this:

1	$\sim A \& \sim B$	Assumption
2	$A \vee B$	Assumption
3	A	Assumption
4	$\sim (C \& \sim C)$	Assumption
5	A	3 R
6	$\sim A$	1 &E
7	$C \& \sim C$	4–6 ~E
8	B	Assumption
9	$\sim (C \& \sim C)$	Assumption
10	B	8 R
11	$\sim B$	1 &E
12	$C \& \sim C$	9–11 ~E
13	$C \& \sim C$	2, 3–7, 8–12 $\vee$E
14	C	13 &E
15	$\sim C$	13 &E
16	$\sim (A \vee B)$	2–15 ~I

There is slightly shorter deduction, which also makes use of some interesting possibilities available in construction deductions, namely that assuming A and then repeating it gives a deduction of A from A, which is required by an application of $\vee$E. Another feature to notice is that so far, applications of $\vee$E have exhibited a certain neat symmetry in the subdeductions begun with each disjunct. This need not always be the case, as the next deduction shows:

1	$\sim A \& \sim B$	Assumption
2	$A \vee B$	Assumption
3	A	Assumption
4	A	3 R
5	B	Assumption
6	$\sim A$	Assumption
7	B	5 R
8	$\sim B$	1 &E
9	A	6–8 ∼E
10	A	2, 3–4, 5–9 ∨E
11	$\sim A$	1 &E
12	$\sim (A \vee B)$	2–11 ∼I

The second deduction may be slightly easier, but the first one introduces a feature of SD that may not be altogether too obvious: when assuming formulas as auxiliary assumptions, we are not limited to assuming only such ones as somehow make use of the atomic components of the primary premises: we can assume any sentence we like, even completely random and disconnected ones. The important point is that assuming the sentences leads to the desired result and that the subdeductions begun with them as auxiliary assumptions are closed off in the course of the deduction.

These deductions are already quite complicated. It lies in the nature of systems of natural deduction that, if they only have very few rules, some deductions are rather clumsy and long. That's why the book introduces another system, SD+, which has some more rules of inference, which serve as 'short cuts' and make deductions a rather easier business.

Now that we have introduced the rules of the system SD, we need some definitions of terminology that allows us to talk about derivations in SD. First, here is a notion that we could have introduced earlier:

DEFINITION. A sentence **P** occurring in a deduction is *in the scope of assumptions* $\mathbf{Q}_1 \ldots \mathbf{Q}_n$ if and only if the scope lines immediately to the left of each assumption (i.e. the ones lines beginning with these assumptions) are also to the left of **P**.

Scope lines are the vertical lines to the left of formulas in deductions, beginning either with the primary assumptions or auxiliary assumptions. We use this notion to define more precisely what being derivable in SD means:

DEFINITION. A sentence **P** is *derivable in SD* from a set of sentences Γ of SL if and only if there is a derivation in SD in which all the primary assumptions are members of Γ and **P** occurs in the scope only of those assumptions.

We have already introduced a notation for this, which is $\Gamma \vdash \mathbf{P}$. Notice that it is only required that *some* members of Γ occur as the primary assumptions of the derivation. Derivations always only have finitely many primary (and auxiliary) assumptions, because derivations are things that can be written down: infinitely many things cannot be written down. Nonetheless, a sentence **P** may be said to be derivable from an infinitely large set Γ, if a finite subset Δ of Γ suffices to derive **P**.

The last notion can be used to give a definition of validity somewhat different from the semantic definition using the notion of truth-value assignments. Notice that when we discussed the semantics of SL the primitive notion in terms of which all others were defined was the notion of a truth-value assignment. Now the primitive notion is that of a deduction.

DEFINITION. An argument of SL is *valid in SD* if and only of the conclusion of the argument is derivable in SD from the set consisting of the premises.

DEFINITION. An argument of SL is *invalid in SD* if and only if it is not valid in SD.

We now need a notion that corresponds to that of *truth-functional truth*. It is the following:

DEFINITION. A sentence **P** of SL is a *theorem in SD* if and only if **P** is derivable in SD from the empty set.

We already had an example of a logical truth in the first lecture on derivations. Here it is again:

1	A	Assumption
2	B	Assumption
3	A	1 R
4	$B \supset A$	2–3 ⊃I
5	$A \supset (B \supset A)$	1–4 ⊃I

$A \supset (B \supset A)$ is in the scope of no primary assumptions, therefore it has been derived 'from the empty set' of primary assumptions. We write this as $\vdash A \supset (B \supset A)$.

Here is another example of a theorem:

$\vdash (A \supset C) \supset ((B \supset C) \supset ((A \vee B) \supset C))$

1	$A \supset C$	Assumption
2	$B \supset C$	Assumption
3	$A \vee B$	Assumption
4	A	Assumption
5	C	1, 4 ⊃E
6	B	Assumption
7	C	2, 6 ⊃E
8	C	3, 4–5, 6–7 ∨E
9	$(A \vee B) \supset C$	3–8 ⊃I
10	$(B \supset C) \supset ((A \vee B) \supset C)$	2–9 ⊃I
11	$(A \supset C) \supset ((B \supset C) \supset ((A \vee B) \supset C))$	1–10 ∨I

Here is a less exciting definition, corresponding to truth-functional equivalence:

DEFINITION. Sentences **P** and **Q** of SL are *equivalent in SD* if and only if

$\mathbf{Q}$ is derivable in SD from $\{\mathbf{P}\}$ and $\mathbf{P}$ is derivable in SD from $\{\mathbf{Q}\}$.

Hence the deductions of DeMorgan's Laws have shown that $\sim (A \vee B)$ and $\sim A \& \sim B$ are equivalent in SD, and so are $\sim A \vee \sim B$ and $\sim (A \& B)$.

Finally, there is the notion of consistency defined in terms of derivations in the system of natural deduction:

DEFINITION. A set of sentences of SL is *inconsistent in SD* if and only if both a sentence $\mathbf{P}$ of SL and its negation $\sim \mathbf{P}$ are derivable in SD from Γ.

DEFINITION. A set of sentences of SL is *consistent in SD* if and only if it is not inconsistent in SD.

As already noted, an interesting fact about inconsistent sets is that every sentence is derivable from them.

Lecture 12. Predicate Logic and its Syntax: Syllogisms and Beyond

The formal language SL is a somewhat poor language. It does not enable us to express things we would express in ordinary English with sentences like 'Every man is mortal.', 'Something is rotten in the state of Denmark.', 'Nobody loves you when you're down and out.'. A consequence of this that the language SL is incapable of expressing such paradigms of reasoning as for instance the following two syllogisms:

(1) No men are islands.
(2) All men are mortal.
___
(3) Therefore, some mortals are not islands.

(4) No islands are man.
(5) All men are animals.
___
(6) Therefore, some animals are not islands.

We obviously need to do something about that, not the least in order to be in a position to test whether the two examples just given are valid or invalid arguments.[8]

Closely connected is another shortcoming of the language of SL. Consider

[8] The scholastics would have called the forms of these syllogisms 'Felapton' and 'Fesapo', respectively, and would indeed have considered them to be valid. Later, when we define a concept of validity appropriate to classify syllogisms, it will turn out that they are valid if (and only if) it is assumed that there are men in the first case and animals in the second: if a premise is added to the effect that there are men/animals, the argument is valid; if no such premise is added, the argument is invalid.

the following sentences:

(7) John loves Mary.
(8) John is fat.
(9) Mary loves Lucy.

In SL, we would formalise them with three different sentence letters, for instance L, F and P. But obviously, this glosses over some reasonably interesting facts about these three sentences: L, F and P stand, as atomic sentences, in no interesting relation whatsoever to each other. However, there are some interesting relations between the English sentences (7), (8) and (9): in (7) and (9) the words 'loves' and 'Mary' occur, in (7) and (8) 'John' occurs.

Why such occurrences of common phrases in different sentences is of interest becomes immediate when we consider the following two sentences:

(10) Mary is taller than John.
(11) John is taller than Lucy.

From this it follows that (13) Mary is taller than Lucy, because (12) whoever is taller than someone else, who in turn is taller than a third person, is also taller than the third. We should be able to formalise this inference, but this is not possible in SL. If we symbolise (10) as M, (11) as N and 'Mary is taller than Lucy' as O, the best we can do in SL is to express that if Mary is taller than John, and John is taller than Lucy, then Mary is taller than Lucy, which would be: (*) $(M \& N) \supset O$. But this covers only the particular case, where Mary is taller than John, John is taller than Lucy, and therefore Mary is taller than Lucy. Contrary to that, the property of tallness that makes this inference possible is perfectly general. For instance, if Lucy is taller than Paul, and Paul is taller than Peter, then Lucy is taller than Peter. SL would force us to formalise this using three different sentence letters, e.g. A, B and C, and then we would also need to add a premise corresponding to (*), i.e. $(A \& B) \supset C$, if we wish to formalise the inference as a valid one. And again, we have not succeeded in expressing what is common to M, N, O, A, B and C, namely that they relate people to each other in order of tallness, and in particular, we have no means to infer, from the sentences of SL, that Mary is taller than all of them. In short, what is missing in SL is not only that we cannot express what different atomic sentences have in common – what common subsentential components they have – but also that we haven't got a way of expressing *generality*, as in 'whoever is taller than someone else, who in turn is taller than a third person, is taller than the third'.

A formal language fulfilling these two *desiderata* – being able to express what is common to different atomic sentences and being able to express generality – will also allow us to formalise syllogisms and, indeed, a vast number of arguments and sentences not of the traditional syllogistic form as for instance the argument about Mary's tallness in the last paragraph. To set up a language in which this is possible is the topic of this lecture: it is the language PL of *predicate logic*.

The first *desideratum* is fulfilled if we introduce expressions into the formal language doing the work of *names* and *predicates* in ordinary language. We use *lower case letters* $a, b, c, \ldots$ (possibly with subscripts as in $a_1, a_2, a_3 \ldots b_1, b_2, b_3 \ldots c_1, c_2, c_3 \ldots$) to abbreviate names of English; we call these symbols also *names*, or, more technically, *individual constants*. Names or individual constants stand for *objects*. Predicates are expressions forming sentences out of names. They come in different kinds. For instance, there are *one-place predicates*. These are expressions that form a sentence out of *one* name. For instance, in ordinary English 'is fat' is a one-place predicate. If attached to the name 'Jones', it forms the sentences 'Jones is fat'. Next, there are *two-place predicates*. For instance, in ordinary English 'loves' is a two-place predicate. If put between the names 'John' and 'Mary' (in this order) it forms the sentences 'John loves Mary'. Then there are *three-place predicates*. They form sentences out of three names, as, for instance, 'is between ... and ...' does out of 'Paris', 'London' and 'Berlin' (in this order): the result is the sentence 'Paris is between London and Berlin'. And so on. In the formal language, we use upper case letters F, G, H etc. to stand for predicates, and add primes $'$ to indicate how many places they have: H', G', F' for one-place predicates, H'', F'', G'' for two-place predicates, and so on for n-place predicates $F'''^{\cdots}$, $G'''^{\cdots}$, $H'''^{\cdots}$, where H, G and F are followed by n primes. In practice, when it is clear from the context how many places a predicate has, we omit the primes: they only serve a technical purpose in the definition of a formula of PL, but are not needed if we know what the interpretation of a predicate letter is. As for sentence letters of SL, we can use subscripts to distinguish predicate letters of PL.

To construct sentences consisting of predicates and names in the formal language we deviate somewhat from ordinary English, where names either come before the predicate (if it's a one-place one) or to the left and right (if it's a two-place one) or at various other places (if it's a three-or-more-place one). In PL, we write individual constants always *after* predicate letters.

For instance, we can now formalise some of the sentences we have discussed earlier: let F stand for 'is fat', L for 'loves' and T for 'is taller than'; furthermore, let l stand for 'Lucy', m for 'Mary', j for 'John'; then (7) becomes Ljm, (8) Fj, (9) Lml, (10) Tmj and (11) becomes Tjl.

Two things are worth noting. First, 'Lucy loves Mary' of course becomes Llm and 'John is taller than Mary' is Tjm. This means that we need to pay attention to the *order* in which names occur in sentences. Secondly, the example of the three-place predicate 'is between ... and ...' indicated that even in English sometimes we need to mark where names are supposed to go if we are to form a sentence out of the predicate. In a sense, predicates have a number of 'gaps' that are filled by names in forming sentences, and if we have a predicate with more than one gap, we also need to say something about which name fills which gap. Both questions, where and how many gaps a predicate has and which name goes into which gap when constructing a sentence, is solved by the use of *variables*. To mark the gaps in a predicate, we use w, x, y, z, with subscripts where necessary, as in 'x is fat', 'x loves y', 'x is between y and z'. Sentences are then formed by replacing the variables with names, and instead of referring to the order of the names, we can indicate which variable is replaced by which name and in that way ensure that we are actually forming the sentences we intended to form, i.e. 'John loves Mary.' rather than 'Mary loves John.', and 'Paris is between London and Berlin.' rather than, e.g., 'London is between Berlin and Paris.'. Correspondingly, in PL we abbreviate predicates of ordinary English by expressions of the kind Fx, Hxy, $Gxyz$ and so on.

I shall use 'predicate' ambiguously to stand for expressions like F, G, H on their own (possibly with primes or subscripts), but also for these followed by the appropriate number of variables. If disambiguation is needed, I shall call G, F and H *predicate letters*. Expressions of the kind $F'x$, $H''xy$, $G'''xyz$ are also called *atomic formulas* of PL, where this term also covers expressions resulting from replacing one or more variable in these expressions by names. If all variables in an atomic formula are replaced by names, we get an *atomic sentence*. The latter are either true or false, depending on whether objects named by the names have the right properties or stand in the right relation to each other.

To fulfil the second *desideratum*, we introduce expressions expressing *generality* into the language PL, i.e. expressions which are roughly the formal analogues for 'some' and 'all' and the like in English. I mentioned in the third lecture that 'some' and 'all' are best understood as always occurring in the contexts ‘some ... are ...’ and ‘all ... are ...’, i.e. they are expressions forming sentences out of two predicates, for instance ‘some men are mortal’ and ‘all men are mortal’ from ‘men’ and ‘mortal’. ‘Some’ and ‘all’ are of course also used in other contexts, and there are other expressions which can do the same job as ‘some’ and ‘all’. For instance, ‘every’ is often used instead of ‘all’, as in ‘Every man is mortal’, and it is also used in contexts where only one predicate is used, as in, e.g., ‘Everything is miserable.’. Similarly, instead of ‘some’ we can use, for instance, ‘there is’ as in ‘There is beer in the fridge’, and we can use this phrase too for sentences formed out of one predicate instead of two, as in, e.g., ‘Something is wrong here’. And most obviously, no one would demand ‘some’ to occur only in the contexts ‘some ... are ...’ as of course ‘are’ can often be replaced by ‘is’ as in ‘Some man is an island’. In setting up our formal language of predicate logic, we shall leave behind these intricacies of English grammar. We shall introduce two formal expressions, $\forall$ and $\exists$, called the *universal quantifier* and the *existential quantifier*, respectively, which we will let do the work of expressions of ordinary English like ‘every’, ‘all’, ‘some’, ‘there is’ and whatever else there might be—as far as possible, that is.

Quantifiers are expressions that turn predicates into sentences.[9] But before we describe the grammar of quantifiers of PL, let’s have a closer look at how variables work. An expression like ‘x loves y’ is almost like a sentence. In a sense it is halfway between a predicate and a sentence: taken on its own it asserts nothing, because of the ‘gap’ represented by the variables, but if we used x and y as names of objects and declared them to stand, for instance, for Lucy and Mary, then ‘x loves y would make an assertion which is either true or false. Something similar is true for pronouns and demonstratives in English. For instance, if it is specified which person ‘she’ refers to something can be asserted by ‘She is pretty.’, namely that the person refers to has this property, but if it is not specified who ‘she’ refers to, the sentence is as gappy as a mere predicate, as nothing has been asserted to have the property. Similarly, ‘This is a nightmare.’ says something only if some object is referred to by ‘this’, but if that is not the case, nothing has been asserted. Now notice that a sentence like ‘All men are mortal’ can be rephrased so that pronouns occur in it, for instance ‘Take anything you like: if it is a man, then it is mortal’ or ‘It is true of anything whatsoever that if it is a man, then it is

[9] Which of course means that predicates are just as much expressions turning quantifiers into sentences as they are expressions turning names into sentences.

mortal'. In PL, a variable like x also works pretty much in the same way as 'it' does in these sentences.[10] Replacing the pronoun 'it' by the variable x in 'if it is a man, then it is mortal' gives a construction of the kind I have called "halfway between a sentence and a predicate": 'if x is a man, then x is mortal'. Using Mx to stand for 'x is a man' and Dx to stand for 'x is mortal', we can formalise it as $Mx \supset Dx$.

In formalising the sentences occurring in the syllogisms with which the lecture began, we want to say that this complex predicate expressed by $Mx \supset Dx$ – being mortal if a man – is true of everything, i.e. whatever you might chose to refer to by x, $Mx \supset Dx$ is true, i.e., whatever x might be, $Mx \supset Dx$, or for any x, $Mx \supset Dx$. We use the universal quantifier followed by the variable x, $\forall x$ to stand for 'whatever x might be' or 'for any x' or 'every x is such that'. Then the whole thing becomes: $\forall x(Mx \supset Dx)$. This, then, is the formalisation of 'All men are mortal' in PL.

Now consider 'Some mortals are not islands'. First, we can reformulate this as 'There is something such that it is a man and it is not an island.' or 'At least one thing is such that it is a man and not an island.'. Replacing pronouns by variables in 'it is a man and it is not an island.', we get 'x is a man and x is not an island.'. Using Ix to stand for 'x is an island', we can formalise the latter as $Mx\& \sim Ix$. The whole sentence 'Some mortals are not islands.' can then be rephrased as meaning something like this: 'There is an x such that x is mortal and x is not an island.'. We use the existential quantifier followed by x to stand for 'There is some x is such that', so that the whole sentence is formalised as $\exists x(Mx\& \sim Ix)$.

In formalising sentences of English in the language PL, we record which predicates and names of PL are used to stand for which predicates and names of English in a *symbolisation key* or an *interpretation*, as we've called it in the lectures on sentential logic. A symbolisation key will also specify what is called a *universe of discourse*: it's the collection of things we are talking about in the moment, for instance, physical objects: the universe of discourse contains the objects the variables of PL (or pronouns of English) stand for. As examples, let's formalise the syllogisms from the beginning of the lecture.

Symbolisation Key:

[10] We could also use variables like demonstratives, but this is somewhat non-standard.

UD:	Physical objects
Mx:	x is a man
Dx:	x is mortal
Ix:	x is an island
Ax:	x is an animal

First syllogism:

(1') $\sim \exists x(Mx \& Ix)$
(2') $\forall x(Mx \supset Dx)$

(3') $\exists x(Mx \& \sim Ix)$

Second syllogism:

(4') $\sim \exists x(Ix \& Mx)$
(5') $\forall x(Mx \supset Ax)$

(6') $\exists x(Ax \& \sim Ix)$

As another example, let's formalise the argument about Mary being taller than other people. First, here it is again in English:

(10) Mary is taller than John.
(11) John is taller than Lucy.
(12) Whoever is taller than someone who is taller than a third person, is taller than this third person.
---
(13) Mary is taller than Lucy.

It is a remarkable fact about the Aristotelian logic that was predominant until the late nineteenth century that it was unable to analyse this rather straightforward argument in logical terms. The argument does not have the form of a syllogism, and everything not of this form was outside the scope of Aristotelean formal logic. In PL, we can easily formalise the argument. Here is the interpretation of symbols in terms of English:

Symbolisation Key:

UD:	Persons
Txy:	x is taller than y
m:	Mary
j:	John
l:	Lucy

We can reformulate the third premise using variables, thus: 'for any x, any y, any z, if x is taller than y, then, if y is taller than z, x is taller than z'. And now the argument is formalised thus:

(10') Tmj
(11') Tjl
(12') $\forall x \forall y \forall x((Txy \& Tyz) \supset Txz)$
___
(13') Tml

As was done for the case of SL, we can give a precise, inductive definition of what counts as a formula of the language PL—the *formal syntax* of PL. First, we need to say what the *vocabulary* of PL is:

Sentence Letters:	$A, A_1, A_2 \ldots B, B_1, B_2 \ldots Z, Z_1, Z_2$
Predicate Letters:	$A', A'_1, A'_2 \ldots A'', A''_1, A''_2 \ldots A''', A'''_1, A'''_2 \ldots B', B'_1, B'_2 \ldots B'', B''_1, B''_2 \ldots B''', B'''_1, B'''_2 \ldots Z', Z'_1, Z'_2 \ldots Z'', Z''_1, Z''_2, Z''', Z'''_1, Z'''_2$
Individual Terms:	
a) Individual Constants:	$a, b, c \ldots, a_1, a_2 \ldots, b_1, b_2 \ldots, c_1, c_2 \ldots$
b) Individual Variables:	$w, x, y, z, w_1, x_1, y_1, z_1, w_2, x_2, y_2, z_2 \ldots$
Connectives:	$\sim, \&, \vee, \supset, \equiv$
Quantifier Symbols:	$\forall, \exists$
Punctuation Marks:	$(,)$

Next, we to define most basis kinds of formulas of PL:

DEFINITION. An *atomic formula* of PL is an expression of PL which is either a sentence letter or an n-place predicate followed by n individual terms.

One more definition:

DEFINITION. An **x**-*quantifier* is a quantifier symbol followed by the variable **x**.

Finally, this is the definition of 'Formula of PL':

1. Every atomic formula is a formula of PL.
2. If $\mathbf{P}$ is a formula of PL, so is $\sim \mathbf{P}$.
3. If $\mathbf{P}$ and $\mathbf{Q}$ are formulas of PL, so are $(\mathbf{P} \& \mathbf{Q})$, $(\mathbf{P} \vee \mathbf{Q})$, $(\mathbf{P} \supset \mathbf{Q})$, $(\mathbf{P} \equiv \mathbf{Q})$.
4. If $\mathbf{P}$ is a formula of PL that contains at least one occurrence of $\mathbf{x}$ and no **x**-quantifier, then $(\forall \mathbf{x})\mathbf{P}$ and $(\exists \mathbf{x})\mathbf{P}$ are formulas of PL.
5. Nothing else is a formula of PL.

Lecture 13. Basic Syntactic Notions of Predicate Logic

Just as we defined notions to characterise sentences of SL and their parts – i.e. sentential components and main connective –, we now need to define corresponding notions for formulas of PL. First, we define a notion somewhat broader than 'truth-functional connective':

DEFINITION. A *logical operator* is an expression of PL which is either a quantifier or a connective.

Next comes the definition of concepts relating to formulas and their parts, namely *immediate subformula*, *subformula* and *main logical operator*:

1. If **P** is an atomic formula of PL, then **P** contains no logical operator, and hence no main logical operator, and **P** is the only subformula of **P**.
2. If **P** is a formula of PL of the form $\sim \mathbf{Q}$, then the tilde $\sim$ preceding **Q** is the main operator of **P**, and **Q** is the immediate subformula of **P**.
3. If **P** is a formula of PL the form $(\mathbf{Q} \vee \mathbf{R})$, $(\mathbf{Q} \& \mathbf{R})$, $(\mathbf{Q} \supset \mathbf{R})$ or $(\mathbf{Q} \equiv \mathbf{R})$, then the connective between **Q** and **R** is the main logical connective of **P**, and its immediate subformulas are **Q** and **R**.
4. If **P** is a formula of PL of the form $\forall \mathbf{x}\mathbf{Q}$ or $\exists \mathbf{x}\mathbf{Q}$, then the quantifier that occurs before **Q** is the main logical operator of **P**, and **Q** is the immediate subformula of **P**.
5. If **P** is a formula of PL, then every subformula (immediate or not) of a subformula of **P** is a subformula of **P**, and **P** is a subformula of itself.

Quantifiers range over fixed parts of formulas. Which one is captured by the following definition, which makes use of the concepts just defined:

DEFINITION. The *scope of a quantifier* in a formula **P** of PL is the subformula **Q** of **P** of which that quantifier is the main logical operator.

Example: $\forall x(Fx \supset \exists y(Gy \& \forall z(Hz \supset Rxyz)))$

Quantifier	Scope
$\forall x$	$(Fx \supset \exists y(Gy \& \forall z(Hz \supset Rxyz)))$
$\exists y$	$(Gy \& \forall z(Hz \supset Rxyz))$
$\forall z$	$(Hz \supset Rxyz)$

We also need some terminology to distinguish two kinds of variables:

DEFINITION. An occurrence of a variable **x** in a formula **P** of PL is *bound* if it is within the scope of an **x**-quantifier.

DEFINITION. An occurrence of a variable **x** in a formula **P** of PL is *free* if it is not bound.

Example: $\forall xFxy$: x is a bound variable, y is a free variable

If a formula contains open variables, it does not really say anything; as remarked in the last lecture, such a formula is a bit like 'She is pretty', where we aren't told who she is. Contrary to that, if a formula contains no free variables, but all are bound, we always know what has been said, given we know the interpretation of the predicates and names. In other words, we always have a complete sentence:

DEFINITION. A formula **P** is a *sentence* of PL if and only if no occurrence of a variable in **P** is free.

Examples: $\forall x \exists yFxy$ is a sentence. $\forall xFxy$ is not.

Finally, one more definition. Suppose you know that every man is mortal.

Then it follows that Socrates is mortal, if he is a man. We can express this as saying that, if s stands for Socrates, and $\forall x(Mx \supset Dx)$ formalises 'All men are mortal', then we get a true sentence by dropping the quantifier and replacing the variable x of $Mx \supset Dx$ by the name s, the result being $Ms \supset Ds$. The resulting sentence a *substitution instance* of the quantified sentence. In the following definition, $\mathbf{Q}(\mathbf{a}/\mathbf{x})$ means that $\mathbf{x}$ is replaced by $\mathbf{a}$ in $\mathbf{Q}$. Notice that the concept is defined only for *sentences*, not formulas in general.

DEFINITION. If $\mathbf{P}$ is a sentence of PL of the form $\forall\mathbf{x}\mathbf{Q}$ or $\exists\mathbf{x}\mathbf{Q}$, and $\mathbf{a}$ is an individual constant, then $\mathbf{Q}(\mathbf{a}/\mathbf{x})$ is a *substitution instance* of $\mathbf{P}$. The constant $\mathbf{a}$ is the *instantiating constant*.

Example:

Sentence	Substitution Instance	Instantiating Constant
$\forall x \forall y Fxy$	$\forall y Fay$	a
$\forall y Fay$	Fab	b
$\forall x \forall z \exists y Bxyz$	$\forall z \exists y Bayz$	a

For the rest of the lecture, we'll go through some examples of formalisations of sentences of English into PL. We've done some syllogisms already last lecture, so now we'll do some examples not of the syllogistic form. Consider the following pair of sentences:

1. Everyone loves someone.
2. Someone is loved by everyone.

The pair seems to exhibit the same grammatical phenomenon of turning the active into the passive voice that also characterises the following two sentences

3. Peter loves Mary.
4. Mary is loved by Peter.

3. and 4. say exactly the same thing. But 1. and 2. don't.[11] Here is another

[11]It is sometimes claimed – in logic books and elsewhere – that some people commit what is called the 'quantifier-shift fallacy', which is the fallacy that assumes that 1. and 2. are logically equivalent. It can be rather difficult to see how anyone could ever commit this fallacy, if both sentences are juxtaposed in the way they are here. What we can conclude,

pair of sentences:

5. Something causes everything.
6. Everything is caused by something.

If 1. and 2. did say the same thing, then 5. and 6. would also say the same thing, as they, too, exhibit the structure exhibited by 3. and 4., where the active is turned into the passive voice. But 5. certainly says something different from what 6. says, because 6. may well be true while 5. may well be false. 5. is true just in case there is at least one event which caused all other events, which of course entails that this event causes itself, as 'everything' means 'everything' rather than 'everything else'. It is a very strong claim that something like that exists, and indeed, if nothing can cause itself, it must be false. It is much easier to imagine 6. to be true; indeed, it is a very plausible thing to say, if every event can be traced back to its causes. Hence we cannot conclude from the fact that 3. and 4. say the same thing and that 1. and 2., and 3. and 4., share the same 'surface grammar', that the latter pairs also say the same thing. This difference is reflected in the formal language by a difference in the order in which quantifiers occur.

1. and 2. say something like the following:

1.' For every x, there is some y such that x loves y.
2.' There is some y such that for every x, x loves y.

5. and 6. can be rephrased as:

5.' There is some x such that for every y, x causes y.
6.' For every y, there is an x such that x causes y.

Using Lxy to stand for 'x loves y' and Cxy for 'x causes y', formalisation yields:

---

however, is that, as 3. and 4. are logically equivalent, 1. and 2. would also be logically equivalent if quantifiers worked like names do: 'everyone' names everyone, and 'someone' names someone, but noone in particular (an *individuum vagum*, as we might call it in scholastic mode). 4. and 5. say the same thing, once in the active once in the passive voice, and if one thought that quantifiers are names (their surface grammar is, after all, very similar to that of names), one should conclude that 1. and 2. say the same thing, once in the passive, once in the active voice. But they don't. Hence quantifiers are not names.

1.” $\forall x \exists y Lxy$
2.” $\exists y \forall x Lxy$
5.” $\exists x \forall y Cxy$
6.” $\forall y \exists x Cxy$

In contrast, letting p stand for ‘Peter’ and m for ‘Mary’, there is no difference in the formalisations of 3. and 4.: both are formalised as Lpm.

Notice that the *order* in which the variables are bound by the quantifiers is also important. It makes a difference whether we first bind x by the universal quantifier and then y by the existential quantifiers, or first y by the existential and then x by the universal quantifier. If, however, the two variables are bound by the same quantifier, there is no difference in meaning; e.g. $\forall x \forall y Cxy$ and $\forall y \forall x Cxy$ say the same thing.

Given a two-place predicate, there are six different formulas we can build from it and the quantifiers. The following table lists them, with English translations:

1.	$\forall x \forall y Cxy$	Everything causes everything.
2.	$\forall x \exists y Cxy$	Everything causes something.
3.	$\forall y \exists x Cxy$	Everything is caused by something.
4.	$\exists y \forall x Cxy$	Something is caused by everything.
5.	$\exists x \forall y Cxy$	Something causes everything.
6.	$\exists x \exists y Cxy$	Something causes something.

Notice how binding the second variable of Cxy first gives sentences best translated into the passive voice. Instead of $\forall y \exists x Cxy$ and $\exists y \forall x Cxy$ we could also change the order of the variables following C in 3. and 4. and write $\forall x \exists y Cyx$ and $\exists x \forall y Cyx$. Thus the passive voice can either be formed by changing the order of the variables following directly after the quantifier symbols, or alternatively changing the order of the variables following the two-place relation.

Lecture 14. Identity

Translating English into symbols is often more or less a matter of doing it word by word: words of English are replaced by expressions of PL, so that the structure of English sentences is often very similar to the structure of sentences of PL. But it is not always the case that we can just translate a word of English corresponding to a logical operator of PL as that symbol. Consider, for instance, ‘Polar bears and grizzly bears are dangerous’. We cannot, letting Px stand for ‘x is a polar bear’, Gx for ‘x is a grizzly bear’ and Dx for ‘x is dangerous’, formalise this as $\forall x((Gx \& Px) \supset Dx)$. For then we have said that things which are *both* polar bears and grizzly bears are dangerous. Obviously, this is not what is meant. Rather, what is meant is $\forall x((Gx \vee Px) \supset Dx)$, i.e. anything which is either a polar bear or a grizzly bear is dangerous. Or consider ‘Someone who fails logic deserves pity.’. Letting Fx stand for ‘x fails logic’ and Dx for ‘x deserves pity’, neither of the following two formalisations will do. A first attempt might be: $\exists x(Fx \supset Dx)$. But this is too weak: it merely asserts that at least one person has the property of deserving-pity-if-failing-logic. A second attempt might be: $\exists x(Fx \& Dx)$. This is too strong: it asserts that there actually is someone who fails logic and deserves pity. What is meant by ‘Someone who fails logic deserves pity’ is of course that whoever fails logic deserves pity. So translating it into PL will not make use of the existential quantifier at all, but rather of the universal one, i.e. $\forall x(Fx \supset Dx)$.

It is instructive to take formulas of PL and consider under which conditions they are true. Let’s use the following interpretation:

UD:	physical things
Ax	x is an animal
Ux	x is a unicorn
Ex	x is equine

Consider the following sentences:

(1) $\forall x(Ux \supset (Ex\&Ax))$
(2) $\forall x(Ux \supset\sim (Ex\&Ax))$
(3) $\exists xUx\&\exists yUy$
(4) $\exists xUx\& \sim \exists yUy$

(1) can be translated into English as 'Whatever is a unicorn is an equine animal', or 'Every unicorn is an equine animal', or, more concisely, 'Unicorns are equine animals'. But is this sentences true? The use of expressions of ordinary English corresponding to the universal quantifier like 'every', 'everything', 'everyone', 'whoever', 'whatever' etc. often are used to say not only that everything of one kind also has another property, but also suggest that something exists which has both properties. For instance, if I assert that all my children have the flu, then I am most certainly taken to have children. If I have no children we'd say that I shouldn't assert that all my children have the flu. Thus 'Unicorns are equine animals' may be taken to imply that there are unicorns, but as there aren't, it should be false. However, it nonetheless seems true that unicorns are equine animals, by considering the kind of creature they are supposed to be, whether there are any or not. So from that perspective, (1) should be true. Thus ordinary English seems undecided when it comes to the question whether the translation of (1) into it is true or false. In a sense, of course, this is an ambiguity. PL is different. One reason for its existence is to provide a language in which to avoid inconsistency. Remember that a formula with the material conditional as main connective is true if its antecedent is false. Hence, whatever you might chose x to stand for, $(Ux \supset (Ex\&Ax))$ is true of it, i.e. that thing is an-equine-animal-if-a-unicorn—a rather clumsy property to have, but nonetheless one that everything indeed has, as nothing is a unicorn. Thus PL decides that (1) is true. But what about (2)? If there are no unicorns, again the antecedent $(Ux \supset\sim (Ex\&Ax))$ is true of whatever x may stand for, and hence PL decides that (2) is true, too. This may sound counterintuitive. But that feeling is dispelled if we remember *why* according to PL both are true. The reason is once more the material conditional, as with so many other counterintuitive results that we encountered during these lectures. In PL, (1) and (2) are not contradictories: they can both be true together.[12]

How to translate (3) into English? You might be tempted to say that it

[12] However, they cannot be both false together.

means 'There are two unicorns': after all, what you are saying is that there is an x which is a unicorn and a y which is also a unicorn, hence if both x and y are unicorns, there must be at least two of them. Similarly, you might be tempted to say that (4) means that there us only one unicorn, because one half of it says that x is a unicorn, but the other half says that there is no y which also is a unicorn. But this is a *non sequitur*. Suppose there is only one unicorn. Now in that case 'There is a unicorn' is true. This is exactly what $\exists xUx$ says, and so does $\exists yUy$. But then both conjuncts of (3) are true, and consequently the conjunction of them must be true. Thus (3) in fact says exactly the same as each of its conjuncts taken on its own. So let's have a look at (4). The first conjunct definitely says that there is a unicorn. And the second conjunct denies this: it says that there is no unicorn. Hence (4) is a contradiction. It can never be true. To conclude, in PL, the shape of bound variables plays no role in a sentence over and above the structure it gives to which variables are bound by which quantifiers in the sentence: two sentences with exactly the same structure of bound variables, which differ only in the shape of the variables, are equivalent.

How do we express that, for instance, there is exactly one unicorn, or that there are at most or least two? So far, we cannot express this. But consider how you would express such a sentence in a version of English closer to PL. 'There is exactly one unicorn' means something like this: 'There is a unicorn, and if you come across another one, it was the one you've already counted'. Similarly, 'There are at most two unicorns' means something like 'There is a unicorn and possibly another one, but if you come a across a third one, it was one you've already counted'. And 'There are at least two unicorns' means 'There is a unicorn, and at least one other, different one'. Now of course logic does not care an awful lot about what we can count and what we came across, so this choice of words is somewhat uncongenial to PL. However, what these considerations show is that to translate 'at least', 'at most', or 'exactly' followed by a number we need some way of expressing sameness and difference in the language. To do so we introduce a new two-place predicate to express sameness. Negating a sentence expressing sameness of two objects we get one expressing difference.

Thus we extend the language of PL by adding the two-place predicate $=$ for *identity*. We'll form a formula from it by writing terms to its left and right, e.g. $a = b$. Notice that this is all we need to do. We need not change any definitions about the formulas of the language. The definitions of *atomic formula*, *formula* etc. can stay what they are. We'll call the language PL extended by the symbol = PLE.

Let's see how to express 'There is exactly one unicorn' in PLE. The approximation of the last but one paragraph can be improved on by excising the phrases connected to counting and coming across, for instance thus: 'There is a unicorn, and whatever else is a unicorn is identical to this one'. In symbols: $\exists x(Ux \& \forall y(Uy \supset y = x))$. This sentence can only be true if there is exactly one unicorn, because if there were two, then the second conjunct of the conjunction following the existential quantifier would require that it is identical with x, as $(Uy \supset y = x)$ is true of it as the y. Notice what the last sentence says: if you chose arbitrarily an x and a y, then, if y is a unicorn, x is the same as y. Suppose x is also a unicorn. Then there can be at most one unicorn, if this sentence is true of all y and x: $\forall x \forall y((Ux \& Uy) \supset y = x)$. Notice that there may be no unicorn at all and this sentence still be true.

'There are at most two unicorns' then means something like this: 'If there are two things which are unicorns, then, if there is a third one, it is identical to one of the two'. In symbols: $\forall x \forall y(Ux \& Uy \& \forall z(Uz \supset (z = x \vee z = y)))$.

'There are at least two unicorns' means something like 'There are two unicorns, and they both are different', i.e. $\exists x \exists y((Ux \& Uy) \& \sim x = y)$.

'There are exactly two unicorns' says that there are at least two and at most two unicorns: $\exists x \exists y(((Ux \& Uy) \& \sim x = y) \& \forall z(Uz \supset (z = x \vee z = y)))$.

Let's sum this up in a table:

(i)	There is at least one F.
(i')	$\exists x Fx$
(ii)	There is at most one F.
(ii')	$\forall x \forall y((Fx \& Fy) \supset x = y)$
(iii)	There is exactly one F.
(iii')	$\exists x(Fx \& \forall y(Fy \supset x = y))$
(iv)	There are at least two Fs.
(iv')	$\exists x \exists y((Fx \& Fy) \& \sim x = y)$
(v)	There are at most two Fs.
(v')	$\forall x \forall y \forall z(((Fx \& Fy) \& Fz) \supset (z = x \vee z = y))$
(vi)	There are exactly two Fs.
(vi')	$\exists x \exists y(((Fx \& Fy) \& \sim x = y) \& \forall z(Fz \supset (z = x \vee z = y)))$

I suspect you can see the pattern emerging.

Identity allows us to solve a logical puzzle. Consider the sentence

(5) The present King of France is bald.

Is this sentence true or false? Well, if true, there should be a present King of France, as the sentence asserts of him that he is bald. But there is no such person. Hence the sentence must be false. But if it is false, doesn't that mean that the following is true:

(6) The present King of France is not bald.

But now it still looks as if there must be a King of France, as (6) asserts of him that he is not bald, and so the sentence cannot be false either. How to solve this dilemma?

Russell argued that (5) says something like this: there is someone who is the present King of France, and there is exactly one such person, and he is bald. The definite article 'the' suggests that the King of France exists and that there is a unique object of this kind. The phrase 'the present King of France' is called a *definite description*. Letting Fx stand for 'x is present King of France' and Bx for 'x is bald', we can analyse (5) as

(5') $\exists x((Fx \& \forall y(Fy \supset x = y)) \& Bx)$

(6) then says that there is exactly one King of France and he is not bald. In other words:

(6') $\exists x((Fx \& \forall y(Fy \supset x = y)) \& \sim Bx)$

We can now see that (6') is not the negation of (5'), and hence there is no contradiction in saying that they are both false. Both entail that there is a unique person who is the present King of France, but as there is no such person, both sentences can be false without contradiction, as one is not the negation of the other.

We can express that there is a unique F somewhat more concisely as $\exists x \forall y(Fy \equiv x = y)$. We can prove that this sentence is true if and only if there is exactly one F. a) If this sentence is true, there must be at least one F. The sentence says that for some x, everything which is F is identical to it. Obviously, then, these x must be F, as they are self-identical. Suppose, furthermore, that there are two things which are F. As everything which is an F is identical to some x, they both are identical to x, and hence are identical to each other, given the properties of identity. Hence if $\exists x \forall y(Fy \equiv x = y)$ is true, there is exactly one F. b) Conversely, assume that there is exactly one F. Then clearly everything which is F is identical to it, so that, calling the thing which is F x, $\forall y(Fy \equiv x = y)$ is true of it. Thus obviously there is something such that $\forall y(Fy \equiv x = y)$ is true of it, hence $\exists x \forall y(Fy \equiv x = y)$. Finally, putting a) and b) together shows that $\exists x \forall y(Fy \equiv x = y)$ is true if and only if there is exactly one F.

We can extend the language of PLE even further by adding symbols for *functions*. Functions are expressions which, when appended to a name, form a complex name of an object. For instance, 'the father of' is a function, and when appended to 'John', as in 'the father of John', can be used to refer to whoever is John's father. Functions are mostly used in mathematics, so let's have a look at an example involving the functions addition and multiplication.

Here is a famous conjecture, made by the mathematician Christian Goldbach in the 18th century, and which is still an unsolved problem of mathematics:

(a) Every even number greater than 2 is the sum of two primes.

Using variables, we can rephrase (1) as:

(b) For any number x, if x is even and greater than 2, then there are two numbers y and z such that x and y are prime numbers and the sum of y and z is equal to x.

It is not difficult to formalise (b) using the following interpretation:

UD: natural numbers
Ex: x is even
$x > y$: x is greater than y
Px: x is prime
$x + y$: the sum of x and y

The formalisation of (1) then is:

(c) $\forall x((Ex \& x > 2) \supset \exists y \exists z((Py \& Pz) \& y + z = x)))$

This formalisation has assumed that the predicates 'is even' and 'is prime' are primitive predicates. But of course we can analyse 'is even' in terms of identity and the function of multiplication. 'x is even' means:

(i) There is a number w such that 2 times w is equal to x.

Thus we add to the interpretation:

$x \times y$: x times y

'x is even' can now be formalised as:

(i') $\exists w(2 \times w = x)$

Replacing Ex by (2') in (1") gives:

(d) $\forall x((\exists w(w \times 2 = x) \& x > 2) \supset \exists y \exists z((Py \& Pz) \& y + z = x))$

Finally, x is a prime number' means:

(ii) x is different from 1 and if it is the product of two numbers, one of them must be equal to 1

In other words:

(ii') x is different from 1 and whenever two numbers y and z are such that y times z is equal to x, then either y is equal to 1 or z is equal to 1

Hence we can express 'x is a prime number' in symbols in the following way:

(ii") $x \neq 1 \& \forall y \forall z(x = y \times z \supset (y = 1 \vee z = 1))$

Replacing (ii") in (d) gives the formalisation of (a) using only the functions addition and multiplication, and the relation 'is larger than' as non-logical, mathematical primitives:

(e) $\forall x((\exists w(w \times 2 = x) \& x > 2) \supset \exists y \exists z(((y \neq 1 \& \forall v_1 \forall v_2(y = v_1 \times v_2 \supset (v_1 = 1 \vee v_2 = 1))) \& (z \neq 1 \& \forall v_1 \forall v_2(z = v_1 \times v_2 \supset (v_1 = 1 \vee v_2 = 1)))) \& y + z = x))$

Lecture 15. Formalisation and Axiomatisation

The formal language PLE allows us to formalise general properties of relations. Take, for instance, tallness and use the set of people as the universe of discourse. Let Txy stand for 'x is taller than y' or, equivalently, 'y is smaller than x'. Obviously, if x is taller than y, then, if y is taller than z, x is taller than z, or in symbols:

(1) $\forall x \forall y \forall z(Txy \supset (Tyz \supset Txz))$

Furthermore, nothing is taller than itself:

(2) $\sim \exists x Txx$

Another property of tallness is this one: if x is taller than y, then y is not taller than x:

(3) $\forall x \forall y(Txy \supset \sim Tyx)$

Given the way the world is, there are things which are incomparable with respect to the relation 'being taller than': if two presons are equally tall, neither is taller than the other. Hence the following is not true:

(4) $\forall x \forall y(Txy \vee Tyx)$

In English: of any two persons, one is smaller than the other. What is true is the negation of (4), $\sim \forall x \forall y(Txy \vee Tyx)$, or equivalently:

(5) $\exists x \exists y(\sim Txy \& \sim Tyx)$

There are two persons neither of which is taller than the other.

Notice the difference between (1)-(3) and (4)/(5). (1)-(3) are true as a consequence of the kind of concept 'being taller than' is. (4) is true as a matter of empirical accident: as a matter of fact, some people have the same height, but of course the world might have been such that all people have different heights. So (4)/(5) is not a 'conceptual truth' about the tallness, but one we know to be true given the properties of the things in the universe of discourse. Correspondingly, obviously, if (4)/(5) were in fact true. (1)-(3) are what could be called *axioms* for tallness. Appealing to an informal notion of logical truth, we might even say that (1)-(3) are logically true, given the interpretation of Txy as 'x is taller than y'. Whatever is the case with the people in the universe of discourse, (1)-(3) must be true if Txy is to mean 'x is taller than y'.

Given that some people are of equal height, the following two sentences are also false:

(6) $\exists x \forall y(x \neq y \supset Txy)$
(7) $\exists y \forall x(x \neq y \supset Txy)$

(5) says that someone is taller than everyone else and (8) that someone is smaller than everyone else. Both are false, because, as some people are of equal height, i.e. are incomparable with respect to 'being taller than', that there are no smallest or tallest people.

Notice that, given the interpretation of Txy as 'x is taller than y', (6) and (7) entail that there is *exactly one* tallest/smallest person. For suppose that two different people, call them a and b, are such that everyone is smaller than them, i.e. $\forall y(a \neq y \supset Tay)$ and $\forall y(b \neq y \supset Tby)$. As *ex hypothesi* a and b are different, it follows from these two sentences that Tab and Tba. But this contradicts (2), which entails that it can never be the case that of two things the first is taller than the second and conversely, the second taller than the first. Hence, a and b cannot be different. The argument for the case of the smallest person is exactly parallel. Hence, as (2) is an axiom for tallness, we need not express 'There is exactly one tallest person' and 'There is exactly one smallest person' as the more complex:

(8) $\exists x(\forall y(x \neq y \supset Txy) \& \forall z(\forall w(z \neq w \supset Tzw) \supset z = x))$
(9) $\exists x(\forall y(x \neq y \supset Tyx) \& \forall z(\forall w(z \neq w \supset Twz) \supset z = x))$

(8) and (9) have the structure of 'There is exactly one F' introduced in the last lecture, where F is replaced by the rather complex properties $\forall y(x \neq y \supset Txy)$ and $\forall y(x \neq y \supset Tyx)$, respectively, which are the properties of being-taller-than-everyone-else and being-smaller-than-everyone-else, respectively.

As there are only finitely many people, it is not the case that there always is someone who is taller or smaller than a given person, and so the following two are false: 'everyone is taller than someone' and 'everyone is smaller than someone':

(10) $\forall x \exists y Txy$
(11) $\forall y \exists x Txy$

The falsity of (10) and (11) entails that (12) and (13) are true:

(12) $\exists x \forall y \sim Txy$
(13) $\exists y \forall x \sim Txy$

(12) says that someone is not taller than everyone, (13) that someone is not smaller than everyone. This is quite trivially true, as I, for instance, am neither smaller nor taller than everyone, as I'm neither smaller nor taller than myself (or everyone else who is as tall/small as me). We can express something more substantial by noting that the ordering of people by the relation of tallness has beginning and endpoints, which is expressed by the following:

(14) $\exists x \forall y (\sim Txy)$
(15) $\exists y \forall x (\sim Txy)$

(14) says there is someone such that he is not taller than everyone: this is true for someone who has the height of the tallest people: he's taller than some and equally tall as others, but the ordering does not go further than him, so there is no one taller than him. Similarly, (15) says that there is someone such that he is not smaller than everyone: this is true for someone who has the height of the smallest people: he is smaller than some, and equally small as others, but the ordering does not go lower than him, so there is no one smaller than him.

Of course tallness is not the only relation which satisfies conditions (1) to (3): as already mentioned, T could also be interpreted as 'smaller than'. Another options might be 'is more difficult to understand than', 'is more expensive than' and $<$ and $>$ of mathematics.

Here is some terminology categorising relations. Relations satisfying condition (1) are called *transitive* relations, those satisfying condition (2) *asymmetric*, and those satisfying condition (3) *irreflexive*.

Let's contrast 'being taller than' with a closely related relation: 'x is as tall as or taller than y', formalised as Oxy. This relation is also transitive. But it is not irreflexive. To the contrary, everything is as tall or taller than itself:

(16) $\forall x \forall y \forall z(Ox \supset (Oy \supset Oz))$
(17) $\forall x Oxx$

Furthermore, Oxy is also not assymetric. If x is as tall as y, then both, x is as tall or taller than y, and hence y is as tall or taller than x. Nonetheless, the following does not hold:

(18) $\forall x \forall y(Oxy \supset Oyx)$

If x is taller than y, then Oxy is true, but not conversely: Oyx is false.

Relations satisfying condition (17) are called *symmetric*. An example of a symmetric relation is 'x is as tall as y'. Formalised as Axy, it has the following properties:

(19) $\forall x Axx$
(20) $\forall x \forall y(Axy \supset Ayx)$
(21) $\forall x \forall y \forall y(Axy \supset (Ayz \supset Axz))$

Thus Axy is reflexive, symmetric and transitive. Such a relation is called an *equivalence relation*. Another example of an equivalence relation is 'x is equal to y'.

Identity shares some properties with 'being as tall as': it too is an equivalence relation. But it also satisfies another logical principle, from which, in fact, all these properties except reflexivity follow: if two things are identical, then whatever is true of the one is true of the other, or:

(22) $\forall x \forall y(x = y \supset (Fx \supset Fy))$

This is often called the *indiscernibility of identicals*. Notice that the converse principle 'If everything that is true of one thing is also true of another thing, then they are identical', often called the *identity of indiscernibles*, is not expressible in the language of PLE! We cannot quantify over properties.

It might be worth contrasting the relations of identity and difference. They differ in some of their properties but share others. Let's write Dxy for 'x and y are different'. Difference is irreflexive: nothing is different from itself. It is symmetric: if x is different from y, then y is different from x. It is not transitive: if x is different from y, and y is different from z, then it does not follow that x is different from y.

Let's have a look at another example of a relation, possibly more interesting than 'taller than'. As the universe of discourse we'll use the set of moments of time in the past. The relation we'll consider in is 'x is earlier than y', which we'll formalise as Exy. This relation is also transitive and irreflexive: no moment of the past is earlier than itself, and if one moment of earlier than another, which is earlier than a third, it is earlier than the third. Contrary to tallness, the relation 'earlier than' does not 'branch'. The past goes in a straight line: any two different moments in time are comparable to each other in the order imposed on them by 'earlier than'. In case of tallness, it was possible that two people were equally tall, which entailed they were incomparable in terms of the relation of 'being taller than'. In the case of 'earlier than', as there is only one past, there are no two distinct moments of the past which are incomparable. In other words, if two moments of time are distinct, then either one is earlier than the other, or the latter earlier than the first. Also, if x is earlier than y, then y is not earlier than x, for any two moments of time. This gives the following axioms for Exy:

(23) $\forall x \forall y \forall z(Exy \supset (Eyz \supset Exz))$
(24) $\forall x \forall y(x \neq y \supset (Exy \vee Eyx))$
(25) $\forall x \forall y(Exy \supset \sim Eyx)$

From (25) it follows that 'earlier than' is irreflexive, i.e. $\sim \exists x Exx$: for assume that some moment of time, call it a, is earlier than itself, i.e. Eaa. Then by (25) a is also not earlier than itself, i.e. $\sim Eaa$. That's a contradiction, and hence no moment of time can be earlier than itself. (24) is equivalent to $\forall x \forall y(x = y \vee (Exy \vee Eyx))$, i.e. of any two moments of time, either they are the same or one is earlier than the other. Together with (25) it follows that for any two moments of time x and y exactly one of $x = y$, Exy and Eyx holds.

The relation 'earlier than' is somewhat more difficult than 'taller than'. For instance, we do not know whether time has a beginning or not. Be that as it may, we can formalise each claim. Suppose that there is a beginning of time. Then the following axiom is also true of Exy:

(26) $\exists x \forall y (x \neq y \supset Exy)$

(26) says that there is a moment in the past which is earlier than every other moment. Together with the other properties of Exy this entails that there is exactly one such moment: the first moment of time.

If instead we assume time not to have a beginning, we use this axiom:

(27) $\forall x \exists y Eyx$

If (27) is assumed to be true, it ensures that for every moment of the past, there is an earlier one.

Another open question is whether between any two moments of the past, where one is earlier than the other, there always is another one, in other words:

(28) $\forall x \forall y (Exy \supset \exists z (Exz \& Ezy))$

If we consider the present moment of time to be one of the past moments, then our universe of discourse has a last moment:

(29) $\exists x \forall y (y \neq x \supset Eyx)$

Of course this might be true even if the present moment is not one of the past moments, namely if time is discrete and consists of moments following each between which there are no other moments. In this case (28) would of course be false. If (28) is true of the earlier-than relation, and we don't consider the present moment as part of the past moments, then it is rather implausible that there is a last point in the sequence of past moments, i.e. (29) should be false. If there is neither a last nor a first moment of the past, then we can express this slightly more shortly as:

(30) $\forall y \exists x \exists z (Exy \& Eyz)$

Relations satisfying (19)-(21) are called *linear orderings*. Relations satisfying (24) are called *dense*, and those satisfying (26) *without end-points*.

Now consider the relation 'is earlier than or simultaneous with', the universe of discourse remaining the same, formalised as Sxy. This relation is transitive and reflexive, but neither symmetric nor asymmetric. In other words, we have (31) and (32) as axioms, but neither (33) nor (34):

(31) $\forall x \forall y \forall z(Sxy \supset (Syz \supset Sxz)$
(32) $\forall x Sxx$
(33) $\forall x \forall y(Sxy \supset Syx))$
(34) $\forall x \forall y(Sxy \supset \sim Syx))$

'Earlier than or simultaneous with' satisfies a property which is weaker than asymmetry, namely that of any two *different* moments in the past, either the first is earlier than or simultaneous to the the second or conversely:

(35) $\forall x \forall y(x \neq y \supset (Sxy \supset \sim Syx))$

Sxy also has another interesting property. As there is only one past, which is composed of moments of time, if x is earlier than or simultaneous with y and y is earlier than or simultaneous with x, then x must be the same moment as y. In other words, Sxy satisfies the following property:

(36) $\forall x \forall y((Sxy \& Syx) \supset x = y)$

Relations satisfying (36) are called *anti-symmetric*. If a relation satisfies (31), (32) and (36) it is called a *partial ordering*.

Here is a summary of properties of relations we have discussed in this lecture:

	R is ...	iff ...
1	*reflexive*	$\forall x Rxx$
2	*irreflexive*	$\sim \exists x Rxx$
3	*symmetric*	$\forall x \forall y(Rxy \supset Ryx)$
4	*asymmetric*	$\forall x \forall y(Rxy \supset \sim Ryx)$
5	*anti-symmetric*	$\forall x \forall y((Rxy \& Ryx) \supset x = y)$
6	*transitive*	$\forall x \forall y \forall z(Rxy \supset (Ryz \supset Rxz))$
7	*dense*	$\forall x \forall y(Rxy \supset \exists z(Rxz \& Rzy))$
8	*without end-points*	$\forall y \exists x \exists z(Rxy \& Ryz)$
9	*weakly connected*	$\forall x \forall y(x \neq y \supset (Rxy \vee Ryx))$
10	*connected*	$\forall x \forall y(Rxy \vee Ryx)$

Here are some common terms to categorise relations satisfying a number of these properties:

Relations satisfying ...	are called...
1, 3, 6	*equivalence relations*
1, 5, 6	*partial orderings*
4, 6, 9	*linear orderings*
1, 5, 6, 10	*simple orderings*

Lecture 16. The Semantics of PLE: Truth and Interpretation

Truth-tables provided a semantics for the language SL of sentential logic: they enabled us to determine whether a sentence is true or false given the truth-values of its atomic components. We now need a semantics for the language PL of predicate logic. In a sense, we have been doing semantics all along, disguised in the symbolisation keys. As a preparatory step towards a mathematically more rigourous treatment of the semantics of predicate logic it is sufficient to notice that symbolisation keys can provide us with means of determining the truth-values of atomic formulas, if we take into account what we know about how the world is like. This is why it is legitimate to talk about an *interpretation* when talking about a symbolisation key; the symbolisation key gives the meanings of the symbols of PLE, which in turn provides us with a means of determining the truth-conditions of sentences of PLE. Consider, for instance, the following key:

UD:	places
Ixy:	x is in y
Cx:	x is a city
Sx:	x is a country
l:	London
p:	Paris
f:	France
e:	England

It provides an interpretation of formulas of PLE in the sense that we can determine whether they are true or false given what we know about places. Consider, for instance:

(1) $Cl\&Ile$
(2) Ilf
(3) $\sim Ipe$
(4) $\sim (Ipe \vee Ipf)$
(5) $Ipl \supset Ipe$
(6) $Ile \supset Ipl$

(1) is true, because London is a city in England; (2) is false, because London is not in France; (3) is true, because Paris is not in England; (4) is false, because Paris is France; (5) is true, because Paris is not in London; (6) is false, for the same reason and because London is in England.

We can also determine the truth-values of sentences involving quantifiers:

(7) $\forall x(Cx \supset Ixe)$
(8) $\forall x(Sx \supset \exists y(Cy\&Ixy))$
(9) $\forall x\forall y(Ixy \supset Cx\&Sy))$

(7) is false because not every city is in England; (8) is true because there's a city in every country; (9) is false, because it's not true that if one place is in another, the larger one is a country and the smaller one a city: for instance, Hyde Park is in London, but neither is London a country nor Hyde Park a city.

Semantics can help us finding out whether we have translated a sentence of English correctly into symbols, given an interpretation. Consider the following two sentence. More often than never they are both offered as formalisations of the same sentence of English, namely 'Every city is in some country':

(10) $\forall x(Cx \supset \exists y(Sy\&Ixy))$
(11) $\forall x\exists y((Cx\&Sy) \supset Ixy)$

(10) is true, because every city is in some country, or maybe more precisely: for every city, there is a country in which it is. (11) is hard to translate into fluent English: it says something that might be put like this: any city is in something if that thing is a country. Let's have a closer look at this example. What would have to be the case with x and y so that $(Cx\&Sy) \supset Ixy$ is true of them? It would suffice that either x is not a city or y is not a country. For instance, let x be London and y be Paris, then $Cx\&Sy$ is false, hence $(Cx\&Sy) \supset Ixy$ is true. For it to be false, we need an x which is a city and a y which is a country, where x is not in y: for instance, let x be London and

y be France. Still, if x refers to London, $\exists y((Cx\&Sy) \supset Ixy)$ is nonetheless true, because, as noted, Paris is something that makes $(Cx\&Sy) \supset Ixy$ true, if y refers to it. The next question is whether x can be anything we like and $\exists y((Cx\&Sy) \supset Ixy)$ still be true. The answer is yes. We'll prove it in two steps. First, if x is not a city, $\exists y((Cx\&Sy) \supset Ixy)$ is trivially true, because Cx is then false, and y can be whatever we like without there being a chance of making $(Cx\&Sy) \supset Ixy$ false, hence $\exists y((Cx\&Sy) \supset Ixy)$ is also true. Hence, for every non-city x, there is something, y, that makes $(Cx\&Sy) \supset Ixy$ true. Secondly, if x is a city, we can also find a suitable y which makes $(Cx\&Sy) \supset Ixy$ true. Take, for instance, something that is not a country. So $\exists y((Cx\&Sy) \supset Ixy)$ is true also for any city x. Thus it is true whether x is a city or not, and as everything either is a city or not, $\exists y((Cx\&Sy) \supset Ixy)$ is true for anything whatsoever, hence $\forall x\exists y((Cx\&Sy) \supset Ixy)$ is true.

Thus both, (10) and (11) are true. But the reasoning showing that (11) is true also shows that it is true for the wrong reasons: it is somewhat too easy to make true, which makes (11) a rather uninteresting claim. (10) on the other hand is more substantial: it is true because, given any city, we can find a country in which it lies. This becomes blatant if we interchange the interpretations of Cx and Sx, so that Cx means 'x is a country' and Sx means 'x is a city'. This change in interpretation does not change the truth-value of (11): it is still true, for reasons similar to why (11) was true on the original interpretation. But (10) now is false, because it now says that every country is in some city. Hence (10) and (11) can have different truth-values on different interpretations, and that means they cannot be used as formalisations of the same sentence of English.

Logic is not concerned an awful lot with what the world is like. Consequently, we don't have to give interpretations in such a way that the truth-values of formulas are based on what is actually the case. Consider the example we discussed a little while ago about Mary, John and Lucy and their respective tallness. We don't have to have some real persons in mind to determine the truth-values of sentences in which the names 'Mary', 'Lucy' and 'John' occur. It is sufficient to specify a possible situation. For instance, let Mary be taller than Lucy, and let John be as tall as Lucy. Here is the symbolisation key:

UD: Mary, John, Lucy
Txy: x is taller than y
m: Mary
j: John
l: Lucy

Now take the following sentences:

(12) Tml
(13) Tjl
(14) $\forall xTmx$
(15) $\forall x(x \neq m \supset Tmx)$
(16) $\exists xTjx$
(17) $\exists xTxj$

In the situation described, (12) is true, because Mary is taller than Lucy; (13) is false, because John is as tall as Lucy; (14) is false, because Mary is not taller than herself, i.e. she is not taller than everyone; (15), however, is true, because Mary is taller than everyone else who is not Mary; (16) is false, because no one is smaller than John; (17) is true, because someone, i.e. Mary, is taller than John.

Suppose we stipulate alternatively that Lucy is taller than John, and John is taller than Mary, for instance. Then the truth-values of the sentences (12), (15), (16) and (17) change, the others remaining the same.

Predicate letters of PL can be interpreted as we like: the symbols of the language taken by themselves are meaningless. An interpretation gives them meaning. If we ask ourselves whether we can find an interpretation of a sentence on which it is true, we are considering what the sentences of PL could mean, and here we are completely free concerning what we let predicate letters (and names) stand for. However, when giving a semantics for sentences of PLE, we need of course interpret the symbol $=$ as identity. This is the only predicate letter of the language which has a fixed interpretation. This has some consequences worth noting. First of all, any interpretation must be such that any statement of the form $\mathbf{a} = \mathbf{a}$ is true. Secondly, if $\mathbf{a} = \mathbf{b}$ is interpreted as true, i.e. **a** and **b** stand for the same object, then, if something is true of one, it must be true of the other. In other words, if $\mathbf{a} = \mathbf{b}$ is true, then, if **Pa**, the sentence resulting from replacing **a** by **b** (in some, but not necessarily all places), i.e. **Pb**, must also be true. If an interpretation does not satisfy these conditions, then we have not succeeded in giving the symbol

= its intended interpretation as identity.

We are now in a position to define notions of logical truth, falsity, indeterminacy, etc.. We have noted that some sentences of PLE can come out as true on one interpretation, but false on others. This is of course exactly parallel to the case of SL, where some sentences may come out as true or false on different truth-value assignment. In PLE, the work of a truth-value assignment is done by the more complex one of an interpretation. An interpretation is, informally, a device that specifies, for each name what it refers to, for each predicate which objects it is true of, for each relation which objects stand in it to each other, and for each sentence letter whether it is true or false. For the time being, this suffices; we shall have a closer look at interpretation in the next lecture. We can now start the defining. First, a formally precise notion of logical truth for predicate logic:

DEFINITION. A sentence **P** of PLE is *quantificationally true* if and only if **P** is true on every interpretation.

Secondly, here is the notion capturing the informal notion of logical falsehood:

DEFINITION. A sentence **P** of PLE is *quantificationally false* if and only if **P** is false on every interpretation.

Finally, we need a term for sentences which are neither of the two:

DEFINITION. A sentence **P** of PLE is *quantificationally indeterminate* if and only if **P** is neither quantificationally true nor quantificationally false.

All the sentences given earlier in this lecture are examples of sentences which are quantificationally indeterminate. We already encountered an example of a sentence that is quantificationally false, or a contradiction:

(18) $\exists x Fx \& \sim \exists y Fy$

Of course, all sentences of SL which are truth-functionally false are also quantificationally false in PLE. Here are two examples of quantificationally true sentences:

(19) $Fa \supset \exists xFx$
(20) $\forall xFx \supset Fa$

(19) says that if a is F, then there is something which is F. Whatever property F might express and whatever object a might refer to, this must be true. (20) says that if everything has the property F, then a particular object named by 'a' also has it. Again this must be true whatever F and a are.

Next, we need a notion expressing that two formulas say the same thing. This is captured by the following definition:

DEFINITION. Sentences **P** and **Q** of PLE are *quantificationally equivalent* if and only if there is no interpretation on which **P** and **Q** have different truth-values.

For instance, the following pairs of sentences exhibit this property:

(21.a) $\forall xFx$
(21.b) $\sim \exists x \sim Fx$

(22.a) $\forall x(Fx \supset\sim Gx)$
(22.b) $\sim \exists x(Fx \& Gx)$

(21.a) says that everything is F, (21.b) that nothing fails to be F: both always have the same truth-value, no matter what F is. (22.a) and (22.b) both say that no F is a G, and thus, too, are quantificationally equivalent.

The last definitions for today capture the notions of consistency and inconsistency for PLE:

DEFINITION. A set of sentences of PLE is *quantificationally consistent* if and only if there is at least one interpretation on which all the members of the set are true.

DEFINITION. A set of sentences of PLE is *quantificationally inconsistent* if and only if the set is not quantificationally consistent.

Of course we'll also need notions of validity and invalidity, but we'll leave that for the next lecture.

Lecture 17. The Semantics of PLE: Validity and Satisfaction

To assess whether arguments formalised in predicate logic are valid or invalid, we need definitions of these notions. Here they are:

DEFINITION. An argument of PLE is *quantificationally valid* if and only if there is no interpretation on which every premise is true and the conclusion is false.

DEFINITION. An argument of PLE is *quantificationally invalid* if and only if the argument is not quantificationally valid.

Arguments have only finitely many premises. So these definitions do not cover the case where we want to draw inferences from infinitely many sentences. The following definition caters for the more general case:

DEFINITION. A set Γ of sentences of PLE *quantificationally entails* a sentence $\mathbf{P}$ of PLE if and only if there is no interpretation on which every member of Γ is true and $\mathbf{P}$ is false.

Notice that these definitions only cover the cases where the premises and conclusions are *sentences* (this follows from the definition of an argument). In PLE, however, there is also the wider category of *formulas*, which are like sentences only that they may contain free variables. To interpret formulas, we need to assign objects to the variables. The notion of a *variable assignment* is also needed to give recursive truth-conditions for formulas of the form $\exists\mathbf{x}\mathbf{Q}$ and $\forall\mathbf{x}\mathbf{Q}$. In sentential logic, the truth-tables provided us with a means of calculating the truth-value of a sentence on the basis of the truth-values of its atomic components. We want to do something similar for sentences of the forms $\exists\mathbf{x}\mathbf{Q}$ and $\forall\mathbf{x}\mathbf{Q}$ on the basis of their component formula, i.e. $\mathbf{Q}$, which has $\mathbf{x}$ free. To do so, we need something like a notion of truth that is applicable to formulas with free variables. As will become apparent during the discussion, we won't call this notion 'truth', but satisfaction, which may avoid some possible confusions. Once this is done, we can redefine the notions of truth and falsity on an interpretation in terms of satisfaction, and extend the definition of entailment to cover also the cases where the premises and conclusions are from the wider class of formulas. We could also extend the notion of an argument so as to allow for free variables in the premises or the conclusion.

First, we need to be a little more specific about what an interpretation as a device allowing us to determine truth-conditions of sentences of PLE does to predicate letters. We said that, for instance, Fa is true on an interpretation, just in case the object referred to by a has the property F on the interpretation, and similarly, Rab is true on an interpretation just in case the objects referred to be a and b stand in relation R to each other on the interpretation, and also that $Babc$ is true of a, b and c just in case a, b and c stand in the relation B to each other on the interpretation. Thus we can say that an interpretation assigns to a one-place predicate letter F the individuals having the property that F is interpreted as expressing, it assigns to a two-place letter R the pairs of objects standing in the relation to each other that R is interpreted as expressing, to a three-place predicate letter B the triples standing in the relation to each other that B is interpreted as expressing, and so on, four-place predicate letters are assigned quadruples of objects of the universe of discourse, five-place predicate letters quintuples, six-place letters sextuples, etc.. We can generalise this: an interpretation of the language of PLE assigns to each n-place predicate letter *n-tuples* of objects of the universe of discourse. n-tuples are written in the following way: $\langle o_1 \dots o_n \rangle$. And furthermore an interpretation assigns to each constant an object and to each sentence letter a truth-value.

Here is an example. Take the following symbolisation key:

UC:	England, France, Spain, Germany and their capitals
a:	London
b:	Paris
c:	Berlin
d:	Madrid
e:	England
f:	France
g:	Germany
h:	Spain
Fx:	x is a city
Gxy:	x is capital of y
$Hxyz$:	y is between x and z
$Iwxyz$:	w is as far away from x as y is from z

We can summarise what is the case in the world in a table like the following:

Predicate	True of
x is a city	London, Paris, Berlin, Madrid
x is capital of y	London-England, Paris-France, Berlin-Germany, Madrid-Spain
y is between x and z	Berlin-Paris-Madrid, Madrid-Paris-Berlin Germany-France-England England-France-Germany Germany-France-Spain Spain-France-Germany
w is as far away from x as y is from z	London-Berlin-Paris-Madrid, Berlin-London-Paris-Madrid, Berlin-London-Madrid-Paris, London-Berlin-Madrid-Paris, Madrid-Paris-Berlin-London, Paris-Madrid-Berlin-London, Paris-Madrid-London-Berlin, Madrid-Paris-London-Berlin

We can represent more formally which n-tuples of objects are assigned to predicate letters of the language PLE by an interpretation:

Predicate	assigned by interpretation
F	a, b, c, d
G	$\langle a, e\rangle, \langle b, f\rangle, \langle c, g\rangle, \langle d, h\rangle$
H	$\langle c, b, d\rangle, \langle d, b, c\rangle, \langle g, f, e\rangle, \langle e, f, g\rangle, \langle g, f, h\rangle, \langle h, f, g\rangle$
I	$\langle a, c, b, d\rangle, \langle c, a, b, d\rangle, \langle c, a, d, b\rangle, \langle a, c, d, b\rangle, \langle d, b, c, a\rangle,$ $\langle b, d, c, a\rangle, \langle b, d, a, c\rangle, \langle d, b, a, c\rangle$

The table shows more clearly what an interpretation does: not only does it give meaning to the otherwise meaningless strings of symbols that are the sentences of PLE: crucially, it does so by specifying which objects in the universe of discourse have the properties or stand in the relations expressed by the predicate letters. The interpretation thereby determines which atomic sentences of PLE are true on it. Notice that we do not also need to say explicitly which things fail to have a property or don't stand in a relation to each other. What isn't specified in the list doesn't hold. The list specifies both, what is and what is not the case, by specifying what is the case.

Before going on to discussing variable assignments, let's see how we can use the interpretation to assign truth-values to sentences and to determine whether arguments are valid. For instance, consider the following sentences:

(1) $\forall x(Fx \supset \exists y Gxy)$
(2) $\forall y \forall y \forall z(Hxyz \supset Fy)$
(3) $\forall x \forall y(Ixdya \supset (x = b \& y = c))$

(1) says that every city is a capital. That's true on our interpretation, because the universe of discourse contains only capital cities. (2) says that everything that is between two things is a city. That's false, because France is between two things, but it is not a city. (3) is also true, because Paris and Berlin are the only things in the universe of discourse that make $Ixdya$ true, where x is assigned Paris and y Berlin.

The truth-values an interpretation allows us to assign to sentences of PLE of course then also allow us to determine whether arguments are invalid. Consider the following, quite obviously silly, argument:

(1) $\forall x(Fx \supset \exists yGxy)$
(2) Fa
―――――――――――
(3) Gaf

This argument is invalid, because, on our interpretation, all the premises are true but the conclusion is false.

It is rather more difficult to determine whether an argument is valid: how do we find out that an argument never has true premises and a false conclusion for *any* interpretation of the symbols occurring in them? One way is of course to prove that this holds: assuming there to be an interpretation on which the premises are true and the conclusion false should lead to a contradiction if the argument is valid. Another way will be provided by the system of natural deduction for PLE to be introduced in the next lectures.

Variable assignments do what their name says: they assign objects in the universe of discourse to the free variables of a formula. They do so relative to an interpretation. A variable assignment 'completes' an interpretation in case formulas contain free variables. We say that variable assignments *satisfy* formulas on interpretations, a notion we are going to define by induction in a few paragraphs. Informally, where $\mathbf{t_1 \ldots t_n}$ are *terms*, i.e. either variables or individual constants, what is meant is that if the objects assigned to the individual constants amongst $\mathbf{t_1 \ldots t_n}$ by the interpretation i and to the variables amongst $\mathbf{t_1 \ldots t_n}$ by the variable assignment v form an n-tuple assigned to the predicate letter $\mathbf{Q}$ by the interpretation, then v satisfies $\mathbf{Qt_1 \ldots t_n}$ on i. On the basis of this concept we can give a formally precise definition of truth on an interpretation for the whole language PLE, encompassing sentences as well as formulas.

The reason for this intermediate step via the notion of satisfaction is the following. Variables can stand for anything in the universe of discourse of the interpretation. Hence what has been said by a formula containing free variables depends on which variable assignment we are using. In a sense, a variable assignment is like a demonstrative gesture that fixes the reference of a demonstrative on an occasion of its use: a variable assignment fixes the reference of free variables used on a specific occasion.

Satisfaction is a notion relative to a variable assignment *and* an interpretation. We could of course have called the concept of satisfaction 'truth on an interpretation for a variable assignment', or something along those lines, but this might have been confusing. We want our notion of *truth on an interpretation* to be independent of this variety: it is a notion which makes no reference to a specific variable assignment any more. It is a notion which is not relative to variable assignments, in the sense that different variable assignment do not change whether a sentence is true on an interpretation *simpliciter*.

Let's write $v(\mathbf{x})$ to refer to the object the variable assignment v assigns to the variable **x** and $i(\mathcal{E})$ to refer to whatever the interpretation i assigns to the expression $\mathcal{E}$; i.e. if $\mathcal{E}$ is a constant, $i(\mathcal{E})$ is an object of the universe of discourse, if $\mathcal{E}$ is an n-place predicate letter, $i(\mathcal{E})$ is a set of n-tuples $\langle o_1 \ldots o_n \rangle$ of objects of the universe of discourse, and if $\mathcal{E}$ is a sentence letter, then $i(\mathcal{E})$ is one of the truth-values **T** and **F**. When giving truth-conditions of arbitrary formulas in the language, we'll need to consider what both, v and i do to *terms* of the language, where a term is either a constant or a variable. Where **t** is a term, let $i/v(\mathbf{t})$ stand for the object i assigns to **t** if it is a name or the object v assigns to **t** if it is a variable. Finally, $v[\mathbf{x}/\mathbf{o}]$ is a variable assignment which is just like v, except that it assigns the object **o** from the universe of discourse to the variable **x**.

We can now define what it means that a variable assignment satisfies a formula on an interpretation:

1. If **P** is a sentence letter, then a variable assignment v satisfies **P** on interpretation i if and only if $i(\mathbf{P}) = \mathbf{T}$.
2. If **P** is an atomic formula of the form $\mathbf{Qt_1 \ldots t_n}$, then v satisfies **P** on i if and only if $\langle i/v(\mathbf{t_1}) \ldots i/v(\mathbf{t_n}) \rangle$ is one of $i(\mathbf{Q})$.
3. If **P** is of the form $\sim \mathbf{Q}$, then v satisfies **P** on i if and only if v does not satisfy **Q** on i.
4. If **P** is of the form **P**&**Q**, then v satisfies **P** on i if and only if v satisfies **P** and **Q**.

5.-7. Similarly for $\mathbf{P} \supset \mathbf{Q}$ etc..

8. If **P** is of the form $\forall\mathbf{xQ}$, then v satisfies **P** on i if and only if for every member **o** of the universe of discourse of i, $v[\mathbf{x}/\mathbf{o}]$ satisfies **Q** on i.
9. If **P** is of the form $\exists\mathbf{xQ}$, then v satisfies **P** on i if and only if for some member **o** of the universe of discourse of i, $v[\mathbf{x}/\mathbf{o}]$ satisfies **Q** on i.

Notice that v only plays a role in clauses 2., 8. and 9., i.e. only where free variables may be involved.

We can now define truth on an interpretation in terms of satisfaction:

DEFINITION. A sentence **P** of PLE is *true on an interpretation* i if and only if every variable assignment v for i satisfies **P** on i.

DEFINITION. A sentence **P** of PLE is *false on an interpretation* i if and only if no variable assignment v for i satisfies **P** on i.

These definitions hold only for sentences. The reason for the restriction is that we want to ensure that whatever it is that truth and falsity on an interpretation apply to, it should be such that everything of this kind is either true or false.

The point of introducing the notion of satisfaction is not just to enable us to give a definition of truth on an interpretation. Probably more importantly, we can now give a definition of logical entailment for formulas as premises and conclusions:

DEFINITION. A set of formulas Γ of PLE entails a formula **P** of PLE if and only if for every interpretation i and every variable assignment v, if v satisfies all members of Γ on i, then v satisfies **P** on i.

Should we extend the notion of an argument so that premises and conclusions can be from the wider class of formulas, not just sentences, we need to give new definitions of the notions of validity and invalidity of arguments. As it is fairly obvious what they would be, I won't give them here.

Lecture 18. Proof-Theory for PLE: Universal and Existential Quantification

The last thing we need to do to complete our treatment of logic is to extend the system of natural deduction to contain also rules for the quantifiers $\forall$ and $\exists$. Let's start with $\forall$. It is not difficult to see what a suitable elimination rule for $\forall$ is: if everything has a certain property, then a particular thing has it. Recall the notation $\mathbf{P}(\mathbf{a}/\mathbf{x})$: it denotes the formula which results from $\mathbf{P}$ by substituting all occurrences of $\mathbf{x}$ by $\mathbf{a}$. $\mathbf{P}(\mathbf{a}/\mathbf{x})$ is called a *substitution instance* of $\forall\mathbf{x}\mathbf{P}$. Thus, the elimination rules for $\forall$ allows us to infer any substitution instance $\mathbf{P}(\mathbf{a}/\mathbf{x})$ of $\forall\mathbf{x}\mathbf{P}$ from that very formula. In other words:

Universal Elimination ($\forall$E)

	$\forall\mathbf{x}\mathbf{P}$
$\triangleright$	$\mathbf{P}(\mathbf{a}/\mathbf{x})$

We can use this rule to derive the logical law that says that if everything has a property F, then a particular thing, say a, also has this property:

$\vdash \forall xFx \supset Fa$

1	$\quad \forall x Fx$	Assumption
2	$\quad Fa$	1 ∀E
3	$\forall x Fx \supset Fa$	(1)-(2) ⊃I

Another example, if every F is a G, then, if a is F, it is also G:

$\forall x(Fx \supset Gx), Fa \vdash Ga$

1	$\forall x(Fx \supset Gx)$	Assumption
2	Fa	Assumption
3	$Fa \supset Ga$	(1) ∀E
4	Ga	(2), (3) ⊃E

Furthermore, if everything which is $\sim F$ is also G, then, if a does not have property G, it follows that a is F:

$\forall x(\sim Fx \supset Gx), \sim Ga \vdash Fa$

1	$\forall x(\sim Fx \supset Gx)$	Assumption
2	$\sim Ga$	Assumption
3	$\sim Fa \supset Ga$	(1) ∀E
4	$\quad \sim Fa$	Assumption
5	$\quad Ga$	(3), (4) ⊃E
6	$\quad \sim Ga$	(2) R
7	Fa	(3)-(6) ∼E

The elimination rule for the universal quantifier is straightforward, so let's move on to the next rule.

The introduction rule for the universal quantifier is somewhat more complicated. Assume that, for some arbitrarily chosen individual a, Fa has been derived. Then everything should have the property F, because, as a was arbitrary, we could have chosen any other object instead of it. What does it mean that a is arbitrary? It means that first of all, that a does not occur in the primary assumptions or premises of the argument: otherwise we would make specific assumptions about a, as a would name a specific thing one of the premises is about: in that case a could hardly be called arbitrary. Furthermore, if a is arbitrary, we should not make any other specific assumptions about a either, i.e. a should not occur in an auxiliary assumption either. If we let logic alone decide what properties a has, then we can call a arbitrary. Here is an example. If everything which is an F also is a G, then we should be able to infer that if everything is F, everything is G. In other words, we want $\forall x(Fx \supset Gx) \vdash \forall xFx \supset \forall xGx$. This should be the beginning of the deduction:

1	$\forall x(Fx \supset Gx)$	Assumption
2	$Fa \supset Ga$	(1) $\forall$E
3	$\quad \forall xFx$	Assumption
4	$\quad Fa$	(3) $\forall$E
5	$\quad Ga$	(2), (4) $\supset$E

Now have a look at a. It does not occur in any primary assumption of the deduction. It also does not occur in an auxiliary assumption in the subdeduction ending with Ga in line (5). In fact, it has been arbitrarily chosen in steps (2) and (4) to instantiate $\forall x(Fx \supset Gx)$ and $\forall xFx$. We could just as well have chosen any other lower case letter. Thus Ga holds for arbitrary names. Hence we should be able to apply the introduction rule for the universal quantifier and infer $\forall xGx$, and finally, the introduction rule for $\supset$ to derive the desired conclusion:

$\forall x(Fx \supset Gx) \vdash \forall xFx \supset \forall xGx$

1	$\forall x(Fx \supset Gx)$	Assumption
2	$Fa \supset Ga$	(1) ∀E
3	$\quad \forall xFx$	Assumption
4	$\quad Fa$	(3) ∀E
5	$\quad Ga$	(2), (4) ⊃E
6	$\quad \forall xGx$	(5) ∀I
7	$\forall xFx \supset \forall xGx$	(3)-(6) ⊃I

In general, the introduction rule for the universal quantifier is this one:

Universal Introduction (∀I)

$\mathbf{P}(\mathbf{a}/\mathbf{x})$

▷ $\forall\mathbf{x}\mathbf{P}$

provided that a is arbitrary and we also need to make sure that all occurrences of **a** in $\mathbf{P}(\mathbf{a}/\mathbf{x})$ have been replaced by **x** to form $\forall\mathbf{x}\mathbf{P}$. In other words, the following two conditions have to be fulfilled:

(i) **a** does not occur in an undischarged assumption
(ii) **a** does not occur in $\forall\mathbf{x}\mathbf{P}$

The second condition is needed to exclude applications of the rule as in the following failed deduction:

1	$\forall xIxx$	Assumption
2	Iaa	(1) ∀E
3	$\forall xIxa$	(2) ∀I

The last step in this 'deduction' is incorrect. Otherwise, interpreting Ixy as 'x is identical to y', it would follow, as everything is identical to itself ($\forall xIxx$), that everything is identical to a certain object a, i.e. there could only be one object. Certainly that is not true, so the inference of $\forall xIxa$ from $\forall xIxx$ is invalid. We block it in the natural deduction calculus by requiring that *all* occurrences of **a** in $\mathbf{P}(\mathbf{a}/\mathbf{x})$ are replaced by **x**: this requirement is violated in the attempted deduction.

Here is another example of something we can prove using the rules of inference for the universal quantification:

$\vdash \forall x(Fx\&Gx) \equiv (\forall xFx\&\forall xGx)$

1	$\forall x(Fx\&Gx)$	Assumption
2	$Fa\&Ga$	(1) ∀E
3	Fa	(2) &E
4	$\forall xFx$	(3) ∀I
5	Ga	(2) &E
6	$\forall xGx$	(5) ∀I
7	$\forall xFx\&\forall xGx$	(4), (6) &I
8	$\forall xFx\&\forall xGx$	Assumption
9	$\forall xFx$	(8) &E
10	Fa	(9) ∀E
11	$\forall xGx$	(8) &E
12	Ga	(11) ∀E
13	$Fa\&Ga$	(10), (12) &I
14	$\forall x(Fx\&Gx)$	(13) ∀I
15	$\forall x(Fx\&Gx) \equiv (\forall xFx\&\forall xGx)$	(1)-(7), (8)-(14) ≡I

This is obviously true: if everything is both, F and G, then everything is F and everything is G, and conversely.

Now for the rules for the existential quantifier. This time it is the introduction rule which is straightforward. If a particular thing, say a, has the property F, then something has the property F:

Existential Introduction (∃I)

$$\begin{array}{l|l} & \mathbf{P(a/x)} \\ & \\ \triangleright & \exists \mathbf{xP} \end{array}$$

We can use this rule to prove that if everything has the property F, then something does:

$\vdash \forall xFx \supset \exists xFx$:

1	$\forall xFx$	Assumption
2	Fa	1 ∀E
3	$\exists xFx$	(2) ∃I
4	$\forall xFx \supset \exists xFx$	(1)-(3) ⊃I

We can also prove an important relation between existential and universal quantification:

$\sim \forall x \sim Fx \vdash \exists xFx$

1	$\sim \forall x \sim Fx$	Assumption
2	$\sim \exists xFx$	Assumption
3	Fa	Assumption
4	$\exists xFx$	(3) ∃I
5	$\sim \exists xFx$	(2) R
6	$\sim Fa$	(3)-(5) ∼I
7	$\forall x \sim Fx$	(6) ∀I
8	$\sim \forall x \sim Fx$	(1) R
9	$\exists xFx$	(2)-(8) ∼E

To prove the converse – $\exists xFx \vdash \sim \forall x \sim Fx$ – we need the elimination rule for the existential quantifier, which we'll introduce now.

The elimination rule for the existential quantifier is more complicated. Suppose you have deduced that a certain formula, say B, follows from an assumption in which a name occurs, say Fa. Then Fa entails B. Now consider the weaker assumption, that there is an x which is F, $\exists xFx$. If a was arbitrarily chosen, then we can infer that $\exists xFx$ also entails B, because in that case a can do the job of an instance of F, of which at least one must exist, if $\exists xFx$ is true. Again, it is essential that a is arbitrary. $\exists xFx$ asserts the existence of some instances of F, but nothing about any instances in particular. All we know is that there are some things which are F; whatever they are, let's call one of them a: we assume nothing over and above about a than that it is an instance of F-ness. This means that we cannot assume a to have any other properties besides F-ness, except what follows logically from the other assumption we have made, in which a, then, must not occur. Furthermore, a should also not occur in B: if it does not, then we ensure that the conclusion we draw is independent of which things have the property F. Here is a slightly different way at looking at the rule. Suppose you derive a sentence $\exists xFx$. Thus you know that something must be F. Give one of the things which are F an arbitrary name, i.e. one not occurring undischarged premises, for instance a, and start a new subdeduction beginning with Fa as auxiliary assumption. Notice that the inference from $\exists xFx$ to Fa is of course invalid: that something is F does not entail that a specific a is F, as $\exists xFx$ can of course be interpreted as true, while Fa is interpreted as false. But we are not inferring Fa from $\exists xFx$: what we do is to start a new subdeduction beginning with Fa as an auxiliary assumption. We then attempt to infer a formula, say B, which does not contain a, i.e. it is independent of our choice of a. Thus on every interpretation on which Fa is true, B is true. But, as a was arbitrary and does not occur in B, we can now assert that B must also follow from $\exists xFx$: under the assumption that an arbitrary thing is F, we have derived a formula which is independent of what that thing is, hence it follows from the weaker assumption that $\exists xFx$.

The rule for existential quantifier elimination, then, is this one:

Existential Elimination (∃E)

	$\exists \mathbf{xPx}$
	$\mathbf{P(a/x)}$
	⋮
	$\mathbf{Q}$
▷	$\mathbf{Q}$

The following three conditions need to be fulfilled:

(i) **a** does not occur in an undischarged assumption
(ii) **a** does not occur in **Q**
(iii) **a** does not occur in $\exists \mathbf{xPx}$

In other words, when we come across a formula of the form $\exists \mathbf{xPx}$ in a deduction, we add a subdeduction beginning with the auxiliary assumption $\mathbf{P(a/x)}$, where **a** is a constant not occuring in $\exists \mathbf{xP}$ and any undischarged assumption, i.e. it is arbitrary, and then aim to derive a sentence **Q** which also does not contain **a**. Then we can close off the subdeduction and assert **Q** as following from $\exists \mathbf{xPx}$.

Condition (iii) is not met in the following failed deduction:

1	$\exists x Rxa$	Assumption
2	Raa	Assumption
3	$\exists x Rxx$	(2) ∃I
4	$\exists x Rxx$	(1), (2)-(3) ∃E

That the inference of $\exists Rxx$ from $\exists x Rxa$ is invalid is easily seen if we interpret a as referring to the number 1 and Rxy as the relation 'x is less than y': then it is true that there is something which is less than 1, i.e. 0, but not true that something is less than itself. Condition (iii) ensures that such invalid inferences are excluded from being inferable in the system of natural deduction.

Here are some examples of deductions using the elimination rule for the existential quantifier.

$\forall x(Gx \supset Fx), \exists xFx \vdash \exists xGx$

1	$\forall x(Gx \supset Fx)$	Assumption
2	$\exists xGx$	Assumption
3	$Ga \supset Fa$	(1) ∀E
4	$\quad Ga$	Assumption
5	$\quad Fa$	(3), (4) ⊃E
6	$\quad \exists xFx$	(5) ∃I
7	$\exists xFx$	(2), (4)-(6) ∃E

$\exists xFx \vdash \sim \forall x \sim Fx$

1	$\exists xFx$	Assumption
2	$\quad Fa$	Assumption
3	$\quad\quad \forall x \sim Fx$	Assumption
4	$\quad\quad Fa$	(2) R
5	$\quad\quad \sim Fa$	(3) ∀E
6	$\quad \sim \forall x \sim Fx$	(3)-(5) ∼I
7	$\sim \forall x \sim Fx$	(1), (2)-(6) ∃E

$\vdash \exists x(Fx \vee Gx) \equiv (\exists xFx \vee \exists xGx)$

1	$\exists x(Fx \vee Gx)$	Assumption
2	$Fa \vee Ga$	Assumption
3	Fa	Assumption
4	$\exists xFx$	(3) ∃I
5	$\exists xFx \vee \exists xGx$	(4) ∨I
6	Ga	Assumption
7	$\exists xGx$	(6) ∃I
8	$\exists xFx \vee \exists xGx$	(7) ∨I
9	$\exists xFx \vee \exists xGx$	(2), (3)-(5), (6)-(8) ∨E
10	$\exists xFx \vee \exists xGx$	(1), (2)-(9) ∃E
11	$\exists xFx \vee \exists xGx$	Assumption
12	$\exists xFx$	Assumption
13	Fa	Assumption
14	$Fa \vee Ga$	(13) ∨I
15	$\exists x(Fx \vee Gx)$	(14) ∃I
16	$\exists x(Fx \vee Gx)$	(12), (13)-(15) ∃E
17	$\exists xGx$	Assumption
18	Ga	Assumption
19	$Fa \vee Ga$	(18) ∨I
20	$\exists x(Fx \vee Gx)$	(19) ∃I
21	$\exists x(Fx \vee Gx)$	(17), (18)-(20) ∃E
22	$\exists x(Fx \vee Gx)$	(11), (12)-(16), (17)-(21) ∨E
23	$\exists x(Fx \vee Gx) \equiv (\exists xFx \vee \exists xGx)$	(1)-(10), (11)-(22) ≡I

Lecture 19. Proof-Theory for PLE: Identity, Functions and Basis Notions of Proof-Theory

To complete the system of natural deduction for predicate logic, we need to add rules for identity. The introduction rule for identity used by Bergmann *et al.* is somewhat non-standard, as it contains not only identity but also universal quantification. The rule allows for the derivation of the sentence 'Everything is self-identical' at any step in a deduction:

Identity Introduction (=I)

$\triangleright \mid \forall \mathbf{x}(\mathbf{x} = \mathbf{x})$

Alternatively, we could use the rule that allows to infer at any step in a deduction any sentence asserting, of an arbitrary object, that it is self-identical:

Alternative Identity Introduction (=I)

$\triangleright \mid \mathbf{a} = \mathbf{a}$

Both rules obviously amount to the same thing: it is an obvious consequence of Bergmann *et al.* original rule that it allows us to infer any sentence of the form $\mathbf{a} = \mathbf{a}$ at any step in a deduction: this follows by an application $\forall$E after $\forall \mathbf{x}(\mathbf{x} = \mathbf{x})$ has been introduced by an application of Bergmann *et al.* identity introduction. Conversely, the alternative identity introduction rule allows us to derive a sentence of the form $\forall \mathbf{x}(\mathbf{x} = \mathbf{x})$ at any step in the deduction: this follows by $\forall$I, applied after an application of alternative identity introduction, where $\mathbf{a}$ has been chosen such as not to occur in any undischarged assumption, which, of course, is always possible, as there is an unlimited supply of names in the language.[13]

The elimination rule for identity is the more substantial one. It is a version of Leibniz Law: if a and b are the same, then, if a has property P, so does b. In rule form:

Identity Elimination (=E)

	$\mathbf{a = b}$
	$\mathbf{P}$
▷	$\mathbf{P(a//b)}$

Here $\mathbf{P(a//b)}$ means that $\mathbf{b}$ has been replaced by $\mathbf{a}$ at zero or more places in $\mathbf{P}$.

Let's use the rules for identity to prove some things about it. For instance, if a has the property F, and b has a property G which excludes it from having F, then a cannot be the same as b:

$\vdash Fa, Gb, \forall x(Gx \supset\sim Fx) \vdash\sim a = b$

1	Fa	Assumption
2	Gb	Assumption
3	$\forall x(Gx \supset\sim Fx)$	Assumption
4	$Gb \supset\sim Fb$	(3) ∀E
5	$\sim Fb$	(2), (4) ⊃E
6	$\quad a = b$	Assumption
7	$\quad Fb$	(1), (6) =E
8	$\quad \sim Fb$	(5) R
9	$\sim a = b$	(6)-(8) ∼I

[13]Recall the discussion of arbitrarily chosen objects in connection to universal quantifier introduction.

An example would be if a is red all lover and b is blue all over: if something is blue all over it is not red all over, hence a and b must be different.

We can also show that identity has some important properties. For instance, we can prove that it is transitive:

$\vdash \forall x \forall y \forall z(x = y \supset (y = z \supset x = z))$

1	$a = b$	Assumption
2	$b = c$	Assumption
3	$a = c$	(1), (2) =E
4	$b = c \supset a = c$	(2)-(3) ⊃I
5	$a = b \supset (b = c \supset a = c)$	(1)-(4) ⊃I
6	$\forall z(a = b \supset (b = z \supset a = z))$	(5) ∀I
7	$\forall y \forall z(a = y \supset (y = z \supset a = z))$	(6) ∀I
8	$\forall x \forall y \forall z(x = y \supset (y = z \supset x = z))$	(7) ∀I

Identity is also symmetric:

$\vdash \forall x \forall y(x = y \supset y = x)$

1	$a = b$	Assumption
2	$\forall x(x = x)$	=I
3	$a = a$	(2) ∀E
4	$b = a$	(1), (3) =E
5	$a = b \supset b = a$	(1)-(4) ⊃I
6	$\forall y(a = y \supset y = a)$	(5) ∀I
7	$\forall x \forall y(x = y \supset y = x)$	

Obviously, identity is reflexive, which follows immediately by (=I).

Once we have identity, it makes a lot of sense to extend the language to contain function symbols. A function is an expression which, similar to a predicate, takes a fixed number of names and forms a new expression out of them. The difference is that functions don't form sentences out of names, but rather other, more complex names. Take, for instance, 'father of'. This is a function: when appended to the name 'Anna', say, we get 'father of Anna' which gives us a means of referring to Anna's father. Functions correlate things with other things; for instance, the function 'father of' correlate's Anna with her father. Another example is the addition functions in mathematics: + correlates two objects with their sum, for instance it correlates 7 and 5 with the number 12. Functions, just like predicates come with different numbers of places: 'father of' is a one-place function, + is a two-place function, and so on. Generalising, n-place functions are correlations of n objects, or alternatively n-tuples of objects, from the universe of discourse with objects from the universe of discourse. Crucially, an n-place function always correlates n objects with *exaclty* one object. Otherwise we could not use functions to construct complex names of objects: names in the formal language are such that they always name exactly one thing. So, for instance, although 'sister of' may sound on the face of it like a function, given the obvious grammatical similarity with 'father of', it is not a function if the universe of discourse contains, for instance, me: I have two sisters, so 'sister of Nils' cannot be used to name a unique object. We also require functions to correlate any tuple of the right number of objects to an object. Thus, if the universe of discourse also contains people who don't have sisters, for instance my friend Daniel, again 'sister of' cannot be a function in our sense, because 'sister of Daniel' names no one, as Daniel doesn't have a sister. On the other hand, if the universe of discourse contains only people with exactly one sister, then 'sister of' can be introduced into our language as function. Notice that this means that mathematical functions like addition are only functions if either a) the universe of discourse is restricted to the right kinds of objects for pairs of which sums are defined, e.g., numbers, or b) we need to decide the question what, for instance, 'Mary+John' is supposed to refer to: this could be done by laying down that it refers to some arbitrarily chosen object of the universe of discourse, say Lucy.

We call that which is correlated by a function its *argument*, if we consider it to be an n-tuple of object, or its *arguments* in the plural, if we consider it to be a plurality of objects,[14], and that which the argument/s is/are cor-

[14]n-tuples of objects are, although in some sense made up of n objects, once more individual things.

related to the *value* of the function for the argument/s. Thus there are two conditions that have to be fulfilled for something to be a function:

1. An n-place function must correlate one and only one value with each n-tuple of arguments.
2. Arguments and values must be from the same universe of discourse.

These conditions ensure that the deviant situations discussed earlier won't arise.

We call complex names formed by means of functions *terms*. We count amongst them also the individual constants and individual variables. Here is the definition:

1. Individual variables are terms.
2. Individual constants are terms.
3. If $\mathbf{f}$ is an n-place function and $\mathbf{t_1} \ldots \mathbf{t_n}$ are terms, then $\mathbf{ft_1} \ldots \mathbf{t_n}$ is a term.
4. Nothing else is a term.

We call individual constants and variables *simple terms*, all others *complex terms*. A term (simple or complex) containing a variable is called *open*, otherwise it is *closed*.

We have formulated the rules for the quantifiers in the natural deduction system only with reference to individual constants, i.e. one kind of simple terms. But of course, if everything is F, then also every object referred to by a complex term has the property F. Similarly, if the object named by a complex terms has property F, then something has property F. And, furthermore, if two objects named by (simple or complex) terms are identical, then, if one of them has the property F, so does the other. Finally, if we use the alternative rule for identity introduction, anything referred to be a complex terms is self-identical. In other words, in ∀E, ∃I, =E and =I we can use complex terms where in the statement of the rules we have used the individual constants **a** and **b**. We require these terms to be closed. However, in ∀I and ∃E we can't use a complex term to replace the arbitrary name **a**, because a complex term would not refer to an arbitrary object: for instance, if the term is fa, then we would be referring to an object which is one of those objects which are values of the function f, and thus we would be referring to an object in a specific region of the universe of discourse. In the case of ∀I, if we have derived Ffa, where a does not occur in any undischarged assumption, we have not succeeded in showing that an arbitrary object has the property F, but only that arbitrary objects that are values of the function f have the property F: obviously this is not good enough to justify the inference that everything has the property F. Similarly, in the case of ∃E, by making the auxiliary assumption Ffa, we do not merely assume that an object has the property F, but that one which is value of the function f has F: but obviously, all that $\exists xFx$ reports is that some things have the property F, but not that some things which are value of f have F, which is rather more specific.

It remains to redefine the basic notions of proof-theory which we have introduced for the system SD of which the present system PDE is an extension so that they also apply to the latter. This is straightforward: we only need to replace 'SL' by 'PLE' and 'SD' by 'PDE'. Nonetheless, I'll give the definitions again here:

DEFINITIONS:

A sentence **P** is *derivable in PDE* from a set of sentences Γ of PLE if and only if there is a derivation in PDE in which all the primary assumptions are members of Γ and **P** occurs in the scope only of those assumptions.

An argument of PLE is *valid in PDE* if and only of the conclusion of the argument is derivable in PDE from the set consisting of the premises.

An argument of PLE is *invalid in PDE* if and only if it is not valid in PDE.

A sentence **P** of PLE is a *theorem in PDE* if and only if **P** is derivable in PDE from the empty set.

Sentences **P** and **Q** of SL are *equivalent in PDE* if and only if **Q** is derivable in PDE from {**P**} and **P** is derivable in PDE from {**Q**}.

A set of sentences of PLE is *inconsistent in PDE* if and only if both a sentence **P** of PLE and its negation ∼ **P** are derivable in PDE from Γ.

A set of sentences of SL is *consistent in PDE* if and only if it is not inconsistent in PDE.

Definitions

Lecture 1

1. An *argument* is any set of declarative sentences, one of which is designated as the *conclusion* of the argument, the others being its *premises*.

2a. An argument is *deductively valid* if and only if it is not possible for the premises to be true and the conclusion false.

2b. An argument is *deductively invalid* if and only it is not deductively valid.

3a. An argument is *deductively sound* if and only if it is deductively valid and its premises are true.

3b. An argument is *deductively unsound* if and only if it is not deductively sound.

Lecture 2

4a. An set of sentences is *logically consistent* if and only it is possible for all the members of that set to be true.

4b. An set of sentences is *logically inconsistent* if and only it is not consistent.

5a. A sentence is *logically false* if and only it is not possible for the sentence to be true.

5b. A sentence is *logically true* if and only it is not possible for the sentence to be false.

5c. A sentence is *logically indeterminate* if and only it is neither logically true nor logically false.

6. The members of a pair of sentences are *logically equivalent* if and only it is not possible for one of the sentences to be true while the other sentence is false.

Part II. Formal Definitions: Sentential Logic

Lecture 3

7. A sentential connective is used *truth-functionally* if and only if it is used to generate a compound sentence form one or more sentences in such a way that the truth-value of the generated compound is wholly determined by the truth-values of those one or more sentences from which the compound is generated, no matter what those truth-values may be.

8. Truth-tables for Conjunction, Negation and Disjunction:

P	**Q**	**P & Q**
T	**T**	**T**
T	**F**	**F**
F	**T**	**F**
F	**F**	**F**

P	**∼ P**
T	**F**
F	**T**

P	**Q**	**P ∨ Q**
T	**T**	**T**
T	**F**	**T**
F	**T**	**T**
F	**F**	**F**

Lecture 4

9. DeMorgan's Laws: $\sim (M \& N)$ and $\sim M \vee \sim N$ are *logically equivalent*, and so are $\sim M \& \sim N$ and $\sim (M \vee N)$.

10. Truth-table for the Material Conditional:

P	**Q**	**P ⊃ Q**
T	**T**	**T**
T	**F**	**F**
F	**T**	**T**
F	**F**	**T**

Lecture 5

11. Truth-table for the Material Biconditional:

P	**Q**	**P ≡ Q**
T	**T**	**T**
T	**F**	**F**
F	**T**	**F**
F	**F**	**T**

The Syntax of SL

12. The expressions of SL:

Sentence letters:	A, B, C etc.
Truth-functional Connectives:	&, ∼, ∨, ⊃, ≡
Parentheses:	(,)

13. Definition of *sentence of* SL:

1. Every sentence letter is a sentence.
2. If **P** is a sentence, then ∼ **P** is a sentence.
3. If **P** and **Q** are sentences, then (**P**&**Q**) is a sentence.
4. If **P** and **Q** are sentences, then (**P** ∨ **Q**) is a sentence.
5. If **P** and **Q** are sentences, then (**P** ⊃ **Q**) is a sentence.
6. If **P** and **Q** are sentences, then (**P** ≡ **Q**) is a sentence.
7. Nothing is a sentence unless if can be formed by repeated application of clauses 1-6.

14. A sentence consisting only of a sentence letter is an *atomic sentence*.

15. The *main connective* and the *immediate sentential components* of sentences:

1. If **P** is an atomic sentence, **P** contains no connectives and hence does not have a main connective. **P** has no immediate sentential components.

2. If **P** is of the form $\sim$ **Q**, where **Q** is a sentence, then the main connective of **P** is the tilde that occurs before **Q**, and **Q** is the immediate sentential component of **P**.

3. If **P** is of the form **Q**&**R**, **Q** $\vee$ **R**, **Q** $\supset$ **R**, or **Q** $\equiv$ **R**, where **Q** and **R** are sentences, then the main connective of **P** is the connective that occurs between **Q** and **R**, and **Q** and **R** are the immediate sentential components of **P**.

16. The *sentential components* of a sentence are the sentence itself, its immediate sentential components, and the sentential components of its immediate sentential components.

17. The *atomic components* of a sentence are the sentential components which are atomic sentences.

The Semantics of SL

Lecture 6

18. A *truth-value assignment* is an assignment of truth-values (**T**s or **F**s) to the atomic sentences of SL.

19a. A sentence is *true on a truth-value assignment* if and only if it has the truth-value **T** on the truth-value assignment.

19b. A sentence is *false on a truth-value assignment* if and only if it has the truth-value **F** on the truth-value assignment.

20a. A sentence **P** of SL is *truth-functionally true* (or a *tautology*) if and only if **P** is true on every truth-value assignment.

20b. A sentence **P** of SL is *truth-functionally false* (or a *contradiction*) if and only if **P** is false on every truth-value assignment.

20c. A sentence **P** of SL is *truth-functionally indeterminate* (or *contingent*) if and only if **P** is neither truth-functionally true nor truth-functionally false.

21. Sentences **P** and **Q** of SL are *truth-functionally equivalent* if and only if there is no truth-value assignment on which **P** and **Q** have different truth-values.

Lecture 7

22a. A set of sentences of SL is *truth-functionally consistent* if and only if there is a truth-value assignment on which all the members of the set are true.

22b. A set of sentences of SL is *truth-functionally inconsistent* if and only if it is not truth-functionally consistent.

23a An argument of SL is *truth-functionally valid* if and only if there is no truth-value assignment on which all the premises are true and the conclusion is false.

23b. An argument of SL is *truth-functionally invalid* if and only if it is not truth-functionally valid.

24. A set Γ of sentences of SL *truth-functionally entails* a sentence **P** (in

symbols: $\Gamma \vDash \mathbf{P}$) if and only if there is no truth-value assignment on which every member of Γ is true and $\mathbf{P}$ is false.

The Natural Deduction System SD

Lecture 8

25. Reiteration:

$$\begin{array}{r|l} & \mathbf{P} \\ & \\ \rhd & \mathbf{P} \end{array}$$

26. Conjunction

a. Introduction (&I):

$$\begin{array}{r|l} & \mathbf{P} \\ & \mathbf{Q} \\ \rhd & \mathbf{P}\&\mathbf{Q} \end{array}$$

b. Elimination (&E):

$$\begin{array}{r|l} & \mathbf{P}\&\mathbf{Q} \\ \rhd & \mathbf{P} \end{array} \quad \text{or} \quad \begin{array}{r|l} & \mathbf{P}\&\mathbf{Q} \\ \rhd & \mathbf{Q} \end{array}$$

27. Material Conditional

a. Introduction (⊃I):

$$\begin{array}{r|l} & \begin{array}{|l} \mathbf{P} \\ \hline \\ \mathbf{Q} \end{array} \\ \rhd & \mathbf{P} \supset \mathbf{Q} \end{array}$$

b. Elimination (⊃E):

$$\begin{array}{r|l} & \mathbf{P} \supset \mathbf{Q} \\ & \mathbf{P} \\ \rhd & \mathbf{Q} \end{array}$$

Lecture 9

28. Negation

a. Introduction (∼I):

```
  | | P
  | |---
  | | Q
  | | ∼ Q
▷ | ∼ P
```

b. Elimination (∼E):

```
  | | ∼ P
  | |---
  | | Q
  | | ∼ Q
▷ | P
```

Lecture 10

29. Disjunction

a. Introduction (∨I):

```
  | P                  | P
           or
▷ | P ∨ Q            ▷ | Q ∨ P
```

b. Elimination (∨E):

```
  | P ∨ Q
  | | P
  | |---
  | | R
  | | Q
  | |---
  | | R
▷ | R
```

30. A sentence **P** in a deduction is *in the scope of assumptions* $\mathbf{Q}_1 \ldots \mathbf{Q}_n$ if

and only if the scope lines immediately to the left of each assumption (i.e. the ones lines beginning with these assumptions) are also to the left of **P**.

31. A sentence **P** is *derivable in SD* from a set of sentences Γ of SL if and only if there is a derivation in SD in which all the primary assumptions are members of Γ and **P** occurs in the scope only of those assumptions.

32a. An argument of SL is *valid in SD* if and only of the conclusion of the argument is derivable in SD from the set consisting of the premises.

32b. An argument of SL is *invalid in SD* if and only if it is not valid in SD.

33. A sentence **P** of SL is a *theorem in SD* if and only if **P** is derivable in SD from the empty set.

34. Sentences **P** and **Q** of SL are *equivalent in SD* if and only if **Q** is derivable in SD from $\{\mathbf{P}\}$ and **P** is derivable in SD from $\{\mathbf{Q}\}$.

35a. A set of sentences of SL is *inconsistent in SD* if and only if both a sentence **P** of SL and its negation $\sim \mathbf{P}$ are derivable in SD from Γ.

35b. A set of sentences of SL is *consistent in SD* if and only if it is not inconsistent in SD.

Part III. Formal Definitions: Predicate Logic

Lecture 12

36. The *vocabulary* of PL:

Sentence Letters:	$A, A_1, A_2 \ldots B, B_1, B_2 \ldots Z, Z_1, Z_2$
Predicate Letters:	A', A'_1, A'_2 ... A'', A''_1, A''_2 ... A''', A'''_1, A'''_2 ... B', B'_1, B'_2 ... B'', B''_1, B''_2 ... B''', B'''_1, B'''_2 ... Z', Z'_1, Z'_2 ... Z'', Z''_1, Z''_2, Z''', Z'''_1, Z'''_2
Individual Terms:	
a) Individual Constants:	$a, b \ldots, a_1, a_2 \ldots, b_1, b_2 \ldots$
b) Individual Variables:	$w, x, y, z, w_1, x_1, y_1, z_1, w_2, x_2, y_2, z_2 \ldots$
Connectives:	$\sim, \&, \vee, \supset, \equiv$
Quantifier Symbols:	$\forall, \exists$
Punctuation Marks:	(,)

37. An *atomic formula* of PL is an expression of PL which is either a sentence letter or an n-place predicate followed by n individual terms.

38. An **x**-*quantifier* is a quantifier symbol followed by the variable **x**.

39. The definition of 'Formula of PL':

1. Every atomic formula is a formula of PL.
2. If **P** is a formula of PL, so is $\sim$ **P**.
3. If **P** and **Q** are formulas of PL, so are $(\mathbf{P}\&\mathbf{Q})$, $(\mathbf{P} \vee \mathbf{Q})$, $(\mathbf{P} \supset \mathbf{Q})$, $(\mathbf{P} \equiv \mathbf{Q})$.
4. If **P** is a formula of PL that contains at least one occurrence of **x** and no **x**-quantifier, then $(\forall \mathbf{x})\mathbf{P}$ and $(\exists \mathbf{x})\mathbf{P}$ are formulas of PL.
5. Nothing else is a formula of PL.

Lecture 13

40. A *logical operator* is an expression of PL which is either a quantifier or a connective.

41. The *immediate subformula*, *subformula* and *main logical operator* of a formula of PL:

1. If **P** is an atomic formula of PL, then **P** contains no logical operator, and hence no main logical operator, and **P** is the only subformula of **P**.
2. If **P** is a formula of PL of the form $\sim$ **Q**, then the tilde $\sim$ preceding **Q** is the main operator of **P** and **Q** is the immediate subformula of **P**.
3. If **P** is a formula of PL the form $(\mathbf{Q} \vee \mathbf{R})$, $(\mathbf{Q}\&\mathbf{R})$, $(\mathbf{Q} \subset \mathbf{R})$ or $(\mathbf{Q} \equiv \mathbf{R})$, then the connective between **Q** and **R** is the main logical connective of **P**, and its immediate subformulas are **Q** and **R**.
4. If **P** is a formula of PL of the form $\exists \mathbf{x}\mathbf{Q}$ or $\forall \mathbf{x}\mathbf{Q}$, then the quantifier that occurs before **Q** is the main logical operator of **P**, and **Q** is the immediate subformula of **P**.
5. If **P** is a formula of PL, then every subformula (immediate or not) of a subformula of **P** is a subformula of **P**, and **P** is a subformula of itself.

42. The *scope of a quantifier* in a formula **P** of PL is the subformula **Q** of **P** of which that quantifier is the main logical operator.

43a. An occurrence of a variable **x** in a formula **P** of PL is *bound* if it is within the scope of an **x**-quantifier.

43b. An occurrence of a variable **x** in a formula **P** of PL is *free* if it is not bound.

44. A formula **P** is a *sentence* of PL if and only if no occurrence of a variable in **P** is free.

45. If **P** is a sentence of PL of the form $\forall \mathbf{xQ}$ or $\exists \mathbf{xQ}$, and **a** is an individual constant, then $\mathbf{Q}(\mathbf{a}/\mathbf{x})$ is a *substitution instance* of **P**. The constant **a** is the *instantiating constant*.

46. The language PLE is constructed by adding to Definition 36 a clause specifying that $=$ is a two-place predicate for identity, and to 37 that if **a** and **b** are names, then $\mathbf{a} = \mathbf{b}$ is an atomic sentence. All other definitions 38-45 then hold for PLE if 'PL' is replaced by 'PLE'.

Lecture 16

47. An interpretation specifies, for each name what it refers to, for each predicate which objects it is true of, for each relation which objects stand in it to each other, and for each sentence letter whether it is true or false.

48a. A sentence **P** of PLE is *quantificationally true* if and only if **P** is true on every interpretation.

48b. A sentence **P** of PLE is *quantificationally false* if and only if **P** is false on every interpretation.

48c. A sentence **P** of PLE is *quantificationally indeterminate* if and only if **P** is neither quantificationally true nor quantificationally false.

49. Sentences **P** and **Q** of PLE are *quantificationally equivalent* if and only if there is no interpretation on which **P** and **Q** have different truth-values.

50a. A set of sentences of PLE is *quantificationally consistent* if and only if there is at least one interpretation on which all the members of the set are true.

50b. A set of sentences of PLE is *quantificationally inconsistent* if and only if the set is not quantificationally consistent.

Lecture 17

51a. An argument of PLE is *quantificationally valid* if and only if there is no interpretation on which every premise is true and the conclusion is false.

51b. An argument of PLE is *quantificationally invalid* if and only if the argument is not quantificationally valid.

52. A set Γ of sentences of PLE *quantificationally entails* a sentence **P** of PLE if and only if there is no interpretation on which every member of Γ is true and **P** is false.

53. Notation (not used in the Bergmann *et al.*):

$v(\mathbf{x})$	is the object the variable assignment v assigns to the variable $\mathbf{x}$.
$i(\mathcal{E})$	is whatever the interpretation i assigns to the expression $\mathcal{E}$ of PLE: If $\mathcal{E}$ is a constant, $i(\mathcal{E})$ is an object of the universe of discourse, if $\mathcal{E}$ is an n-place predicate letter, $i(\mathcal{E})$ is a set of n-tuples $\langle o_1 \ldots o_n \rangle$ of objects of the universe of discourse, and if $\mathcal{E}$ is a sentence letter, then $i(\mathcal{E})$ is one of the truth-values $\mathbf{T}$ and $\mathbf{F}$.
$\mathbf{t}$	stands for terms, i.e. constants or variables.
$i/v(\mathbf{t})$	is the object i assigns to $\mathbf{t}$ if it is a name, or the object v assigns to $\mathbf{t}$ if it is a variable.
$v[\mathbf{x}/\mathbf{o}]$	is a variable assignment just like v, except that it assigns the object $\mathbf{o}$ from the universe of discourse to the variable $\mathbf{x}$.

54. Satisfaction:

1. If $\mathbf{P}$ is a sentence letter, then a variable assignment v satisfies $\mathbf{P}$ on interpretation i if and only if $i(\mathbf{P}) = \mathbf{T}$.
2. If $\mathbf{P}$ is an atomic formula of the form $\mathbf{Qt_1 \ldots t_n}$, then v satisfies $\mathbf{P}$ on i if and only if $\langle i/v(\mathbf{t_1}) \ldots i/v(\mathbf{t_n}) \rangle$ is one of $i(\mathbf{Q})$.
3. If $\mathbf{P}$ is of the form $\sim \mathbf{Q}$, then v satisfies $\mathbf{P}$ on i if and only if v does not satisfy $\mathbf{Q}$ on i.
4. If $\mathbf{P}$ is of the form $\mathbf{P}\&\mathbf{Q}$, then v satisfies $\mathbf{P}$ on i if and only if v satisfies $\mathbf{P}$ and $\mathbf{Q}$.

5.-7. Similarly for $\mathbf{P} \supset \mathbf{Q}$ etc..

8. If $\mathbf{P}$ is of the form $\forall\mathbf{x}\mathbf{Q}$, then v satisfies $\mathbf{P}$ on i if and only if for every member $\mathbf{o}$ of the universe of discourse of i, $v[\mathbf{x}/\mathbf{o}]$ satisfies $\mathbf{Q}$ on i.
9. If $\mathbf{P}$ is of the form $\exists\mathbf{x}\mathbf{Q}$, then v satisfies $\mathbf{P}$ on i if and only if for some member $\mathbf{o}$ of the universe of discourse of i, $v[\mathbf{x}/\mathbf{o}]$ satisfies $\mathbf{Q}$ on i.

55a. A sentence $\mathbf{P}$ of PLE is *true on an interpretation* i if and only if every variable assignment v for i satisfies $\mathbf{P}$ on i.

55b. A sentence **P** of PLE is *false on an interpretation* i if and only if no variable assignment v for i satisfies **P** on i.

56. *Alternative Definition of Entailment applicable to Formulas:* A set of formulas Γ of PLE entails a formula **P** of PLE if and only if for every interpretation i and every variable assignment v, if v satisfies all members of Γ on i, then v satisfies **P** on i.

The Natural Deduction System PDE

Lecture 18

57. Universal Quantification

a. Introduction (∀I):

	$\mathbf{P}(\mathbf{a}/\mathbf{x})$
▷	$\forall\mathbf{xP}$

b. Elimination (∀E):

	$\forall\mathbf{xP}$
▷	$\mathbf{P}(\mathbf{a}/\mathbf{x})$

provided that, for (∀I), the following two conditions are fulfilled:

(i) **a** does not occur in an undischarged assumption
(ii) **a** does not occur in $\forall\mathbf{xP}$

58. Existential Quantification

a. Introduction (∃I):

$$\begin{array}{r|l} & \mathbf{P}(\mathbf{a}/\mathbf{x}) \\ & \\ \triangleright & \exists \mathbf{x}\mathbf{P} \end{array}$$

b. Elimination (∃E):

$$\begin{array}{r|l} & \exists \mathbf{x}\mathbf{P}\mathbf{x} \\ & \begin{array}{|l} \mathbf{P}(\mathbf{a}/\mathbf{x}) \\ \hline \\ \mathbf{Q} \end{array} \\ \triangleright & \mathbf{Q} \end{array}$$

provided that, for (∃E), the following three conditions are fulfilled:

(i) **a** does not occur in an undischarged assumption
(ii) **a** does not occur in **Q**
(iii) **a** does not occur in ∃**xPx**

Lecture 19

59. Identity

a. Introduction (=I):

$$\begin{array}{r|l} \triangleright & \forall \mathbf{x}(\mathbf{x} = \mathbf{x}) \end{array}$$

b. Elimination (=E):

$$\begin{array}{r|l} & \mathbf{a} = \mathbf{b} \\ & \mathbf{P} \\ \triangleright & \mathbf{P}(\mathbf{a}\ /\ /\ \mathbf{b}) \end{array}$$

60. Conditions on functions:

1. An n-place function must correlate one and only one value with each n-tuple of arguments.
2. Arguments and values must be from the same universe of discourse.

61. Definition of 'Term':

1. Individual variables are terms.
2. Individual constants are terms.
3. If **f** is an n-place function and $\mathbf{t_1}\ldots\mathbf{t_n}$ are terms, then $\mathbf{f}\mathbf{t_1}\ldots\mathbf{t_n}$ is a term.
4. Nothing else is a term.

62. Individual constants and variables are *simple terms*, all others *complex terms*. A term (simple or complex) containing a variable is called *open*, otherwise it is *closed*.

63. In ∀E, ∃I, =E and =I complex terms can be used where in the statement of the rules the individual constants **a** and **b** occur.

64. A sentence **P** is *derivable in PDE* from a set of sentences Γ of PLE if and only if there is a derivation in PDE in which all the primary assumptions are members of Γ and **P** occurs in the scope only of those assumptions.

65a. An argument of PLE is *valid in PDE* if and only of the conclusion of the argument is derivable in PDE from the set consisting of the premises.

65b. An argument of PLE is *invalid in PDE* if and only if it is not valid in PDE.

62. A sentence **P** of PLE is a *theorem in PDE* if and only if **P** is derivable in PDE from the empty set.

63. Sentences **P** and **Q** of SL are *equivalent in PDE* if and only if **Q** is derivable in PDE from {**P**} and **P** is derivable in PDE from {**Q**}.

64a. A set of sentences of PLE is *inconsistent in PDE* if and only if both a sentence $\mathbf{P}$ of PLE and its negation $\sim \mathbf{P}$ are derivable in PDE from Γ.

64b. A set of sentences of SL is *consistent in PDE* if and only if it is not inconsistent in PDE.