Introduction to Mathematical Philosophy

Hannes Leitgeb, Stephan Hartmann

June 19, 2015

Contents

Week 2: Overview

Overview of Lecture 2: Truth

2.1 Introduction: Truth is another central topic of philosophy. Truth is a property that we can ascribe either to descriptive sentences or to what is expressed by descriptive sentence, that is, propositions. What a person believes to be true or believes to be false, depends of course on the person; but what is true or false is not relativized to persons in the same way.

2.2 What is Truth?: The obvious answer to the question 'What is truth?' is given by a definition of truth, just as the obvious answer to the question 'What is a grandfather?' is a definition of grandfatherhood. One traditional definition of truth is in terms of correspondence: a sentence is true if and only if it corresponds to reality (or to the facts). But the notion of correspondence is not clear enough, while we are searching for a definition of truth in familiar terms – terms that we do already understand.

2.3 The Truth Scheme I: Tarski was the first to state a precise definition of truth in familiar terms. According to Tarski, every such definition has to satisfy two desiderata: the definition must be formally correct and materially adequate. A definition of truth for a language L is defined to be materially adequate if it implies all truth equivalences for the language L, that is, all sentences of the form (T) 'A' is true if and only if A (where 'A' is to be replaced by an arbitrary descriptive sentence of L). This scheme (T) is called the 'truth scheme'. E.g.: under which conditions is the sentence 'Snow is white' true? Well, it is true if and only if snow is white. One should be able to derive this to be so from a definition of truth for a language L (when 'Snow is white' is a descriptive sentence of L).

2.4 The Truth Scheme II: Since 'Snow is white' is true if and only if snow is white, doesn't that show that we can always get rid of the truth predicate 'is true' if we like to? For we can always say 'Snow is white' directly rather than saying the longish 'The sentence 'Snow is white' is true'? However, we find that there are sentences from which one cannot in the same way eliminate the truth predicate: How could we say what is expressed by 'The last sentence asserted by Caesar is true' or 'Every sentence that is provable in mathematics is true' without using the truth predicate? That is why it is so useful to have the truth predicate in natural language. And it turns out that one cannot easily turn the truth scheme into our intended definition of truth either.

2.5 Defining the Language L_{simple}: We are ready to state a formally correct and materially adequate definition of truth in the Tarskian sense. We do so for a simple toy language which we call, accordingly, 'L_{simple}'. We determine the vocabulary and the grammatical rules of L_{simple} in precise terms; all of the sentences of L_{simple} are descriptive sentences that concern a couple of philosophers and a certain relationship that holds between some of them (the relation of one being a teacher of another).

2.6 Truth for L_{simple}: And then we state the first version of our definition of truth for L_{simple}. By applying the definition to the sentences of L_{simple} we can derive the truth equivalences for these sentences from our definition.

2.7 Recursive Definitions: There is one remaining issue: within some of the clauses in our definition of truth we have used the truth predicate again. Does this mean that the definition is viciously circular? Not really: we find that the definition is a so-called recursive one much in the same sense in which recursive definitions are understood in mathematics or computer science. By our definition, the truth conditions for a complex sentence are reduced to the truth conditions for simpler sentences; and the truth conditions for the simplest sentences are stated without invoking the truth predicate again. Nothing problematic follows from this. In the same way, also the set of sentences of our toy language L_{simple} can be defined recursively.

2.8 Explicit Definitions: It is even possible to turn the recursive definition of truth for L_{simple} into an explicit definition of truth for L_{simple} in which the truth predicate does not occur anymore in the defining clauses. The trick is to reformulate the defining clauses so that they talk about arbitrary sets of sentences. (The same can be done also for the recursive definition of sentencehood for L_{simple}.)

2.9 Theorems about Truth for L_{simple}: This second version of our definition of truth for the sentences of L_{simple} satisfies all of Tarski's desiderata: the definition is formally correct and materially adequate. And we define truth solely on the basis of terms that we do understand already: for we do understand expressions such as 'Socrates', 'is a teacher of', 'not', 'and', 'or', 'for all', 'set', and the like, and since we understand them, by our definition we also understand 'true'. Because the definitions of truth for L_{simple} and of 'sentence of L_{simple}' are formally precise, we are also able to prove now for sentences of L_{simple} that they have certain properties, which can be achieved in a manner similar to that in which mathematicians prove that all natural numbers have certain properties: by a principle of complete induction. For instance, one can prove by complete induction that all truth equivalences for sentences of L_{simple} can be derived from our definition of truth for L_{simple}, and we can prove, based on our definition of truth and complete induction, for every sentence of L_{simple} that either the sentence or its negation is true.

2.10 The Liar Paradox: However, as a famous argument from the philosophical tradition shows, it would be problematic to define truth in the same way for a language L that already includes the truth predicate for that very language L: that's the so-called Liar Paradox, which we state in terms of two premises and a conclusion.

2.11 Self-Referentiality and the Tarskian Hierarchy: The Liar Paradox involves a self-referential sentence: but the existence of such self-referential sentences is not by itself problematic. The real problem is to assume the truth equivalence to hold for a Liar sentence that says of itself that it is not true. The paradox does not affect our definition of truth for L_{simple} since the truth predicate for L_{simple} is not part of the vocabulary of L_{simple} (nor would all of the terms by which we define truth for L_{simple} be contained in the vocabulary of L_{simple}); for that reason one cannot derive the truth equivalence for any Liar-like sentence from our definition of truth for L_{simple}. More generally, one can build a hierarchy of object language, metalanguage, metametalanguage, and so on, such that in the metalanguage one can define truth for the object language, in the metametalanguage one can define truth for the metalanguage, and so on. But in none of these languages one could define truth for that very language itself, and none of the definitions of truth would be undermined by the Liar Paradox.

2.12 Conclusions: When we make a language sufficiently formal, we can state a recursive definition of truth for that language. By using the language of set theory, we can even turn such a recursive definition into an explicit one. The resulting definition of truth is formally correct, materially adequate, it only relies on terms that we already understand, and it avoids the Liar Paradox. Such a definition should thus count as a successful answer to the question 'What is truth?' (as being applied to the descriptive sentences of the language in question). And, on the basis of the definition, we can prove sentences to have certain truth-theoretic properties by means of a similar method by which mathematicians prove natural numbers to have certain mathematical properties.

Chapter 2

Week 2: Truth

2.1 Introduction (07:59)

Welcome to the second lecture of our Introduction to Mathematical Philosophy! In the first lecture, the calculus of real numbers and the basics of modern set theory guided us through the mazes of infinity. Today a mixture of formal languages, recursive definitions, set theory again, and the principle of complete induction will illuminate another central topic of philosophy: truth.

What is truth? Philosophers have been asking this question since the beginnings of philosophy. It's the question that Pilate poses to Jesus according to the Gospel of John. And even in ordinary life and in the sciences, where one is not typically asking that question, people are still presupposing some kind of answer to it. For the concept of truth gets applied a lot even outside of philosophy. More precisely: the term 'true', the linguistic expression, is used a lot in natural language discourse, even though people usually do not think much about its meaning. 'What you say is not true' says the policeman to the burglar: 'you were not at home that night, a neighbor saw you leave the house'. Or: 'We have excellent reasons to believe quantum mechanics to be true' says the physicist; 'it's been confirmed empirically a great number of times'. We regard knowably asserting something that is false, that is, something which is not true, as problematic, even morally problematic, at least in many contexts. And truth is usually regarded the aim of science; that is: the aim of science is to determine which sentences are true and which are not.

In all of these examples, truth and falsity have been ascribed to so-called descriptive or declarative sentences – sentences which describe something, in contrast with e.g. questions or imperatives – or, alternatively, truth and falsity are ascribed to what is expressed by descriptive sentences.

(Slide 1)

Truth as a property of

- descriptive sentences:

  The descriptive sentence ‘Snow is white’ is true.

Either we say: the descriptive sentence ‘snow is white’ is true, where by a sentence I mean here simply a sequence of symbols meaningfully put together.

Or we say: what the descriptive sentence ‘snow is white’ expresses, is true. What does the sentence ‘snow is white’ express? It expresses that snow is white. As philosophers say: the sentence ‘snow is white’ expresses the proposition that snow is white, where a proposition is not a string of symbols but rather what is expressed by, or said by, or conveyed by such a sequence of symbols.

(Slide 2)

Truth as a property of

- descriptive sentences:

  The descriptive sentence ‘Snow is white’ is true.

- propositions:

  What ‘snow is white’ expresses is true.

  That snow is white is true.

  The proposition that snow is white is true

In the next lecture we will introduce a more precise way of thinking about propositions in this sense of the word; but in the present lecture, sentences will be more important than propositions.

Remark on propositions: Indeed, the first aim of Lecture 3 will be to introduce you to a set-theoretic way of conceiving of propositions – propositions as sets of possible worlds.

So truth and falsity are typically predicated either of descriptive sentences or of the things that are expressed by such sentences: propositions. Sometimes, however, we use the term ‘true’ also more liberally: For instance, we might say that a certain story is true. But a story is just a sequence of sentences, and presumably we mean then that all the sentences in the story are true. Or we say that a certain belief is true: I believe that snow is white, and that belief is true. But every belief has a proposition as its content – I believe that snow is white: the proposition that snow is white is the content of my belief – and when we say that the belief is true we really mean that its content, a proposition, is true.

But shouldn't the term 'true', the truth predicate, be relativized to persons somehow? After all, doesn't everyone have his or her truth? Say, you and I trust different weather forecasts: it is true for me that it will be raining at a particular time and place, while it is true for you that it will be sunny then and there. Or we might have different tastes: it is true for me that the Beatles were the greatest pop/rock band of all times, while it is true for you that the Rolling Stones were the greatest. And so on.

At least in the case of the weather forecast, that's not a good way of expressing oneself. I believe it to be true that it will be raining at a particular time and place, while you believe something else: you believe it to be true that it will be sunny then and there. But one should not mix up believing to be true, holding to be true, taking to be true, accepting to be true with being true. Say, it is sunny at that particular time and place: then you were right; it is true that it is sunny at that particular time and place. I believed it to rain, and hence I believed something to be true which turned out to be false.

(Slide 3)

- I believe A, and A is true.
- I believe A, and A is false.
- I do not believe A, and A is true.
- I do not believe A, and A is false.

I believe A to be true, and A is true; I believe A to be true, maybe even for very good reasons, but A is false; I do not believe A to be true, and A is true; or I do not believe A to be true, and A is not true. Having a belief in A is logically independent of the truth of A – at least for most sentences A. Furthermore, take a typical everyday sentence such as 'Peter is at home right now'.

(Slide 4)

- 'Peter is at home right now' is true

  or 'Peter is not at home right now' is true.

Either he is at home, or he is not; in the first case the sentence is true, in the second case the sentence is false and hence its negation is true: 'Peter is not at home right now' is true. So for every such sentence, either the sentence is true or the negation, the corresponding negative sentence is true.

But it might well be the case that I neither believe that Peter is at home right now, nor do I believe the negation: I might simply suspend judgment on the matter, e.g. because I do not have good evidence either way.

(Slide 5)

- 'Peter is at home right now' is true

  or 'Peter is not at home right now' is true.

- But it is possible that:

  Neither 'Peter is at home right now' is believed by me

  nor 'Peter is not at home right now' is believed by me.

So belief allows for suspension of judgment, while truth does not. Belief is a super-interesting concept, which, by the way, will be the topic of the next lecture. But let us not confuse it with truth. While belief is indeed relative to persons – your beliefs won't always coincide with mine – that does not mean that truth is relative to persons. We may always say 'true' or 'false', without relativizing these terms to people.

The case of taste, which I had also mentioned before, is a bit different: while we all think that there are facts of the matter as far as the weather is concerned – if you are in doubt, come to Munich and get wet – it is much less clear that there are corresponding facts of the matter in the case of aesthetic judgements, or, moral judgements, or, for that matter, any kind of value judgement. It is simply not clear that the sentences that express value judgements are really descriptive sentences, whether there is something at all that they describe: an aesthetic or moral reality. And if they don't, then applying the truth predicate to them does not make much sense in the first place. It would be like saying of a question that it is true, or of an imperative statement that it is true, which would be awkward. Sentences expressing value judgements are another really interesting philosophical topic, but again I won't deal with them today. Within the context of this lecture, I will make sure that we will only consider applications of the truth predicate to sentences which are unproblematically descriptive.

Remark on moral and aesthetic judgements: Are moral statements true or false? Or are they more like imperatives, which are neither true nor false? And how about aesthetic judgements? If you are interested in questions like that, you can find more about them at:
http://plato.stanford.edu/entries/moral-cognitivism/
and
http://plato.stanford.edu/entries/aesthetic-judgment/.

(On a personal note: The Beatles were the greatest pop/rock band of all times, and the Rolling Stones were not; fact of the matter! But that won't matter for now.)

2.2 What is Truth? (04:29)

Back to our initial question: What is truth? What is a true descriptive sentence or a true proposition?

When we ask a question like this, then before trying to answer it, it is helpful to pause for a second and to think about what kind of answer we would find acceptable at all.

Suppose that I already understand the terms '(biological) father of' and '(biological) mother of', but I do not know what a (biological) grandfather is. So I ask you: What is grandfatherhood? What is a grandfather? Ideally, in order to answer my question, you would state a definition of the term 'grandfather of' on the basis of terms that I do already understand.

For instance, you might say:

(Slide 6)

Grandfatherhood:

For all x, for all z: x is a grandfather of z if and only if

(i) there is a y, such that x is the father of y and y is the father of z,

or

(ii) there is a y, such that x is the father of y and y is the mother of z.

(i) and (ii) correspond to two different cases: according to the first case, x is a grandfather of z if x is the father of someone who is the father of z; according to the second case, x is a grandfather of z if x is the father of someone who is the mother of z. Grandfatherhood consists in either of these cases being satisfied.

As long as I understand the terms 'father of' and 'mother of', as well as all the logical symbols involved, such as 'there is', 'and', and the like, this will count as a successful answer to my question. Once I am told this definition, I do know what a grandfather is. Of course, this does not mean that I would know now of every two people in the world whether the one is a grandfather of the other – in order to know that, I would have to acquire a lot of empirical knowledge: I would need to travel, check birth certificates, ask around, make genetic tests, and so on. Not typically philosophical work. But that had not been the question anyway; the question wasn't 'Who is grandfather of whom?' The question was: what is a grandfather? And by determining under what conditions someone counts as a grandfather of someone else, the definition above made the meaning of the term 'grandfather of' perfectly clear to me.

Unlike truth, grandfatherhood is not a concept that is particularly interesting to philosophers. But we might still expect a similar kind of answer to the question ‘What is truth?’ as to the question ‘What is grandfatherhood?’. That is: ideally, one would state a definition of the term ‘true’, of the truth predicate, on the basis of other terms that we do already understand.

The qualification ‘that we do already understand’ is quite crucial here. For instance, assume that someone proposes the following answer to our main question: What is truth? It is correspondence with reality:

(Slide 7)

A descriptive sentence or a proposition is true if and only if

it corresponds to reality. [???]

There is some philosophical pedigree to this answer, though it is not so easy to tell who exactly in the philosophical tradition held this view; in any case, its guiding idea does look quite appealing: The descriptive sentence ‘snow is white’ is true. Why? It describes reality as it is; well for clean snow anyway. The sentence corresponds to reality. The only problem is that the meaning of the term ‘corresponds to’ is not clear enough: what does it mean for a sentence to correspond to reality? In what respect does it correspond? If it does, to what thing exactly does the sentence correspond? To reality as a whole, or to a part of reality, and if so to what part? If anything, we seem to understand the term ‘true’ better than the term ‘corresponds to’, which is why it does not seem to be particularly helpful to define truth in terms of correspondence. Can we define truth in terms of more familiar expressions, expressions which we do understand well already?

Quiz 08:
Please check for yourself: If someone asked you ‘What is truth?’ or ‘What do you mean by ‘true’?’, what would you be inclined to answer?

Remark on truth and correspondence: More on the traditional “Correspondence Theory of Truth” can be found at
http://plato.stanford.edu/entries/truth-correspondence/.

2.3 The Truth Scheme I (09:05)

The first person who actually stated such a precise definition of truth in familiar terms was the Polish mathematician and philosopher Alfred Tarski, who introduced his theory of truth in the 1930s and 1940s.

Remark on Tarski: If you want to know more about Alfred Tarski, please check out http://plato.stanford.edu/entries/tarski/.
The central works of Tarski on truth are:

Tarski, A., "Der Wahrheitsbegriff in den formalisierten Sprachen", *Studia Philosophica* 1 (1935), 261-405.
Tarski, A., "The Semantic Conception of Truth and the Foundations of Semantics", *Philosophy and Phenomenological Research* 4 (1944): 341-376.
Tarski, A., "Truth and Proof", *Scientic American* 220/6 (June 1969), 63-77.
Tarski, A., "The Concept of Truth in Formalized Languages", translation of Tarski 1935 by J.H. Woodger, in: Tarski 1983, pp. 152-278.
Tarski, A., *Logic, Semantics, Metamathematics*, second edition, ed. by J. Corcoran, Indianapolis: Hackett, 1983.

Tarski's goal was not to come up with a list of all truths and a list of all falsehoods, just as it had not been the goal in the grandfatherhood case before to come up with a list of who is a grandfather of whom: ultimately, it is the job of mathematicians to find out which mathematical sentences are true, it is the task of physicists to discover which physical sentences are true, it is the task of historians to determine which sentences about human history are true; and so on; philosophers should not aim to be the better mathematicians, physicists, historians, or the like. Instead the goal of a philosophical theory of truth is to clarify the meaning of the term 'true' by determining under which conditions a sentence counts as true. Whether these conditions actually apply in the case of some particular sentence may be a mathematical question, or a physical question, or a historical question, or the like, depending on what the sentence is like.

I will soon sketch a Tarskian definition of truth. But before I do so, let us pause again for a second and consider what exactly we should expect of such a definition. How do we know whether Tarski puts forward a plausible proposal for such a definition?

Tarski himself suggests two desiderata that a satisfactory definition of truth ought to satisfy: it should be formally correct and materially adequate.

(Slide 9)

In order to be satisfactory, a definition of truth has to be *formally correct* and *materially adequate.*

By 'formally correct' Tarski means that the definition should be precise, free from contradictions, and it should have the right form: the same form as other familiar definitions in philosophy, mathematics, or science, a form such as, e.g., the one of our definition of 'grandfather of' from before.

Remark on definitions: There is a whole logical-philosophical theory of definitions in general, for which formal methods play an important role again. If you want to read more about it, please have a look at
http://plato.stanford.edu/entries/definitions/.

By 'materially adequate' Tarski has in mind the following.

Consider this:

(Slide 10)

(1) 'Snow is white' is true if and only if snow is white.

This is an equivalence sentence, a type of sentence which we discussed briefly already in the first lecture. Such a sentence expresses that what is left of the 'if and only if' is equivalent to what is right of it. In the present case, the left-hand side of the equivalence sentence (1) is the sentence (2) 'Snow is white' is true.

(Slide 11)

Left-hand side of (1):

(2) 'Snow is white' is true.

This sentence (2) talks about another sentence, the sentence 'Snow is white', and it says of that sentence that it is true. And that very sentence 'Snow is white' is precisely the right-hand side of the equivalence (1):

(Slide 12)

Right-hand side of (1):

(3) Snow is white.

So the left-hand side of (1), which is sentence (2), says about the right-hand side of sentence (1), which is sentence (3), that it is true. In more schematic terms: the equivalence sentence (1) is of the form

(Slide 13)

Sentence (1) is of the form:

(T) 'A' is true if and only if A.

In sentence (1) the sentence 'Snow is white' occupies or replaces the place holder 'A' in our scheme (T); (1) is a particular instance of the general scheme (T). In other words: from

(T) we can generate sentence (1), and many other equivalences, by replacing the symbol 'A' by concrete sentences. Other such instances of (T) would be:

(Slide 14)

Other instances of (T):

'Tarski is a philosopher' is true if and only if Tarski is a philosopher.

'Munich is in Germany' is true if and only if Munich is in Germany.

'2+2=4' is true if and only if 2+2=4.

⋮

In each of these equivalences, the left-hand side ascribes truth to the right-hand side, where the left-hand side manages to talk about the sentence on the right by enclosing that very sentence within quotation marks. Using quotations marks is simply one convenient way of talking about linguistic objects. By means of quotation marks, the left-hand side of (1) mentions the sentence 'Snow is white', that is, sentence (3), while on the right-hand side of (1) that very sentence is used, that is, the sentence itself is stated in its normal sense, without any quotation marks.

(Slide 15)

(1) 'Snow is white' is true if and only if snow is white.

Left-hand side of (1):

(2) 'Snow is white' is true.

Right-hand side of (1):

(3) Snow is white.

Sentence (1) is of the form:

(T) 'A' is true if and only if A.

O.k. Now here is a question: Do you find sentence (1) plausible at all? Of course you do! Of course, the sentence 'Snow is white' is true if and only if snow is white. What else? Trivially so, as one might want to say. And similarly for other instances of (T). Let us call (T) the truth scheme. And let us call each of its instances a truth equivalence. Then this is what Tarski demands when he says that a definition of truth should be materially adequate:

(Slide 16)

Material Adequacy:

> A definition of truth for the descriptive sentences of a language L is materially adequate if and only if the definition implies all truth equivalences, that is, all instances of the truth scheme
>
> (T) 'A' is true if and only if A
>
> in which the place-holder 'A' is replaced by an arbitrary descriptive sentence in the language L.

Remark on material adequacy: When we say with Tarski that "A definition of truth for the descriptive sentences of a language L is materially adequate if and only if the definition implies...",
then the meaning of the phrase "the definition implies..." is left a bit vague. It can be made precise by replacing it by: "the definition (taken together with a sufficiently strong theory of syntax concerning the language L) logically implies..." and then clarifying both "sufficiently strong theory of syntax concerning the language L" and "logically implies". But for our present purposes it will be fine to leave things at the mere "the definition implies...".

For instance, if the language L in question is a part of English that includes the sentences 'Snow is white', 'Tarski is a philosopher', 'Munich is in Germany', and '$2 + 2 = 4$' as descriptive sentences, then a materially adequate definition of truth for L will entail all the corresponding truth equivalences for these sentences:

'Snow is white' is true if and only if snow is white. 'Tarski is a philosopher' is true if and only if Tarski is a philosopher. 'Munich is in Germany' is true if and only if Munich is in Germany. '$2 + 2 = 4$' is true if and only if $2 + 2 = 4$.

And so on.

And apart from being formally correct, Tarski expects any satisfactory definition of truth to be materially adequate in this sense. That is sort of a minimal constraint really: if the definition were not even able to derive all of these quite trivial instances of the truth scheme, then the definition would probably not be a good one; it would fail to support some sentences involving the truth predicate that we find obviously acceptable, indeed, acceptable on purely conceptual grounds, namely, all the relevant truth equivalences.

If a definition of truth is not materially adequate, then things are even worse: for the truth equivalences seem to capture what was right and understandable about the vague and unclear "truth is correspondence to reality" idea that I had mentioned before. For instance: while the left-hand side of

(Slide 17)

(1) 'Snow is white' is true if and only if snow is white.

talks about the truth of the sentence 'Snow is white', the right-hand side of (1) talks about snow and its being white. The left-hand side concerns the truth of a sentence, while the right-hand side concerns reality. And the vague and unclear notion of correspondence is replaced by the pretty precise and clear logical notion of 'if and only if'. So one might say that any definition of truth that fails to satisfy material adequacy fails to capture an important aspect of the, at least initially, attractive "truth is correspondence to reality" idea – perhaps that aspect of it that one can make good sense of.

Although the desideratum of material adequacy originates with Tarski, there is also more traditional support for it: Aristotle states in his Metaphysics that

(Slide 18)

Aristotle, *Metaphysics*:

> "to say of what is that it is, or of what is not that it is not, is true"

But that is reasonably similar to what the truth scheme expresses: for the 'if and only if' in

(Slide 19)

(T) 'A' is true if and only if A.

really expresses that either both sides of the equivalence are the case, or both are not the case. That is, the truth scheme says nothing else than

(Slide 20)

(A and 'A' is true) or (not A and 'A' is not true).

And this is not far at all from Aristotle's dictum: What is true?

"to say of what is (A) that it is ('A' is true), or of what is not (not A) that it is not ('A' is not true)"

Quiz 09:
What is the truth equivalence for the sentence 'The Munich Center for Mathematical Philosophy is part of LMU Munich'?
Solution

2.4 The Truth Scheme II (07:39)

At the same time, one should not overestimate the importance of the truth scheme: First of all, our acceptance of the truth scheme and its instances should not lure us into the trap of thinking that the truth predicate is really ultimately redundant, that we could always eliminate it in favour of a sentence that does not include the truth predicate. At first glance one might think: Since

'Snow is white' is true if and only if snow is white,

instead of saying to you

'Snow is white' is true

I might just as well say from the start

Snow is white.

Up to equivalence, there is no difference between my two assertions, and hence I can always eliminate the truth predicate if I like. The concept of truth is superfluous really.

Wrong! The reason is that this recipe of elimination works only for assertions of sentences of the form

'A' is true.

But take either of the following two sentences:

(Slide 21)

- The last sentence asserted by Caesar is true.
- Every sentence that is provable in mathematics is true.

I do not know which sentence exactly was the last one to be asserted by Caesar, let alone whether that last sentence is true; but I do still understand the sentence

- The last sentence asserted by Caesar is true.

And I cannot not replace that sentence by an equivalent sentence that does not include the truth predicate: the problem is that the last sentence asserted by Caesar is not stated within quotation marks here; and since I do not know what was Caesar's last sentence either, I cannot say that sentence directly, without invoking the truth predicate.

In the case of 'Every sentence that is provable in mathematics is true', the problem is that it ascribes truth to infinitely many sentences: while in the history of mankind human beings will actually prove only finitely many mathematical theorems, infinitely many mathematical theorems are *prov-able*, are such that they could be proven in principle from the axioms of mathematics by the rules of logic, if only we had the time, space, energy, and so on to do it. The sentence above says about all of these infinitely many *prov-able* theorems that they are true, and we cannot express that thought in any other way than by using the truth predicate. In particular, we cannot simply state all provable mathematical theorems directly, or otherwise we would have to formulate an infinitely long sentence of the form: mathematical theorem T_1 and mathematical theorem T_2 and mathematical theorem T_3 and... ad infinitum. The truth scheme does not tell us for either of our two examples how to reduce them to sentences that would not involve the truth predicate. So the power of the truth scheme is restricted.

Secondly, the truth scheme

(Slide 22)

(T) 'A' is true if and only if A.

is not by itself the answer that Tarski hopes to give to our main question "What is truth?". This is because the truth scheme is not itself the intended definition of the term 'true': as such it would not be formally correct, as it does not have the right form. It is but a scheme, a pattern; the 'A' in

'A' is true if and only if A

is but a slot or a blank: once the slot or blank is filled in, the result is a concrete sentence, but as the answer to our main question we want a concrete sentence of the form

(Slide 23)

For all x: x is true if and only if . . .

from the start, a sentence which defines 'true' much as we had defined 'grandfather of' before by

(Slide 24)

For all x, for all z: x is a grandfather of z if and only if . . .

You might think: The variables 'x' and 'z' in our definition of the term 'grandfather of' are also a little bit like slots or blanks: the only reason why we do not mind their presence is because they make perfect sense together with the two instances of 'for all' at the beginning.

Indeed, we could always, in principle, replace 'x' and 'z' by pronouns in natural language: we could say

(Slide 25)

Grandfatherhood:

For every two persons: *one* is a grandfather of *the other* if and only if

(i) there is a y, such that *the one* person is the father of y and y is the father of *the other* person,

or

(ii) there is a y, such that *the one* person is the father of y and y is the mother of *the other* person.

This would be harder to read but not really different in meaning. In a similar way, we can also get rid of 'y' in this definition in favor of some standard pronoun in natural language, if we want to.

But now you might wonder: Couldn't we fill in the blank in the truth scheme accordingly by putting 'for all A' in front of it:

(Slide 26)

For all A: 'A' is true if and only if A.

Why not take that as a definition of truth?

Well in principle, this might be an option, but one would first need to explain how this goes exactly and what exactly is meant by it. For instance, let us try to express the new proposal in natural language again. If one does so, presumably, one ends up with something like this here:

(Slide 27)

For every sentence/proposition, *it* (in quotation marks?) is true if and only if *it*.

On the left-hand side of the 'if and only if' one tries to mention an arbitrary sentence or proposition, while on the right-hand-side one intends to use, that is, to state the same arbitrary sentence or proposition without quotation marks, and the idea would be to have one pronoun – one variable, as it were – which does both of these jobs simultaneously. The 'for all' at the beginning would have to reach into quotations marks on the left of 'if and only if', if it's for sentences at least, whereas to the right of 'if and only if' it would have to supply the pronoun 'it' with the range of sentences or propositions that the 'it' substitutes for.

Turning something like that into a definition of truth may well be possible, but at least at first glance it is not clear how one should understand a sentence such as the above, what one would be allowed to infer from it, and so on. It is simply not a familiar sentence of English, nor of any of the more standard languages in science or philosophy, and it would need some work to turn it into one.

That is why Tarski does not regard the truth scheme as a definition of truth itself, nor does he try to turn it into a definition of truth directly. Instead he suggests a more roundabout way of defining truth. However, he does use the truth scheme as a measure of success: only if the truth equivalences that one can generate from the truth scheme are derivable from the official definition of truth, the definition will be materially adequate; and only if the definition is materially adequate, it will be satisfactory.

Remark on the truth scheme: There is actually a modern theory of truth which does take the truth scheme to be central, and which does not demand truth to be defined explicitly in the way that Tarski suggested: the so-called Deflationary Theory of Truth. For more on it, see
http://plato.stanford.edu/entries/truth-deflationary/

Quiz 10:
Is it possible to replace the sentence 'For every sentence x, either x is true or the negation of x is true' by a sentence which does *not* include the truth predicate, but which says the same as the given sentence?
Solution

2.5 Defining the Language L_{simple} (06:29)

So how does Tarski define truth?

You might have noticed that in our explanation of material adequacy,

(Slide 28)

Material Adequacy:

> A definition of truth for the descriptive sentences of a language L is materially adequate if and only if the definition implies all truth equivalences, that is, all instances of the truth scheme
>
> (T) 'A' is true if and only if A
>
> in which the place-holder 'A' is replaced by an arbitrary descriptive sentence in the

language L.

we were dealing with a definition of truth for the descriptive sentences of some given language L: not a definition of truth for all sentences of all languages whatsoever. This is at the same time convenient and important. We will see later why it is important, but for the moment it is simply convenient because it will allow me to state a Tarskian definition of truth just for a little toy language, which I can determine to be so that things remain simple and readily comprehensible. For instance, I will set up that toy language in the way that it contains only descriptive sentences: for that reason I will be able to simply say 'sentence' instead of the longish 'descriptive sentence', as there will be no other types of sentence around anyway.

Alright. The goal will be to define truth for a simple toy language L_{simple}. Let me first describe what L_{simple} is like: The basic expressions of that language are just a couple of names for philosophers, some predicates, that is, some terms that express properties or relations, and some logical symbols. Here is the complete vocabulary of L_{simple}:

(Slide 29)

Vocabulary of L_{simple}*:*

- Names: 'Socrates', 'Plato', 'Aristotle', 'Tarski'.
- Predicates: 'is a philosopher', 'is a teacher of'.
- Logical Symbols: 'it is not the case that' ('not'), 'and', 'or'.

Now for the grammar of our little toy language, which I will keep very simple again, much simpler than the grammar of English. These are the grammatical rules of L_{simple}:

(Slide 30)

Grammar of L_{simple}*:*

- If we put a name before 'is a philosopher' we get a sentence of L_{simple}.

For example:

(Slide 31)

'Socrates is a philosopher' is a sentence of L_{simple}.

Similarly,

(Slide 32)

Grammar of L_{simple}*:*

- If we put a name before 'is a teacher of' and another one after it, we get a sentence of L_{simple}.

For example:

(Slide 33)

‘Plato is a teacher of Aristotle’ is a sentence of L_{simple}.

But note that, for the same reason, also

(Slide 34)

‘Aristotle is a teacher of Plato’ is a sentence of L_{simple}.

Just as for English in general, we will be able to formulate various true sentences but also various false sentences in our toy language. The grammar of a language makes sure that we can only formulate meaningful sentences, but it does not itself know about truth or falsity as yet. Accordingly,

(Slide 35)

‘Socrates is a teacher of Socrates’ is a sentence of L_{simple}.

And so on.

Next step:

(Slide 36)

Grammar of L_{simple}:

- If we put ‘it is not the case that’ (‘not’) in front of a sentence of L_{simple}, we get a sentence of L_{simple}.

E.g., we have determined already that ‘Aristotle is a teacher of Plato’ is a sentence of L_{simple}. Therefore, also

(Slide 37)

‘It is not the case that Aristotle is a teacher of Plato’ is a sentence of L_{simple}.

Or shorter:

(Slide 38)

‘Aristotle is not a teacher of Plato’ is a sentence of L_{simple}.

For the same reason also

(Slide 39)

‘It is not the case that it is not the case that Aristotle is a teacher of Plato’ is a sentence of L_{simple}.

You might think: oh, that is the same sentence as ‘Aristotle is a teacher of Plato’ again, since the two negations, the two occurrences of ‘not’, cancel each other out. But let us be precise: we are still dealing with two distinct sentences here. For instance, ‘It is not the case that it is not the case that Aristotle is a teacher of Plato’ is longer than, consists of more signs than, the sentence ‘Aristotle is a teacher of Plato’. Intuitively, the one sentence is true if and only of the other one is, the one sentence says pretty much what the other sentence says, but they are still different sentences, they constitute different ways of saying the same.

Finally:

(Slide 40)

Grammar of L_{simple}:

- For every two sentences of L_{simple}, if we put an ‘and’ or an ‘or’ between them, then we get sentences of L_{simple}.

For example:

Since ‘Socrates is a philosopher’ is a sentence of L_{simple}, and also ‘Aristotle is not a teacher of Plato’ is a sentence of L_{simple}, it follows that

(Slide 41)

‘Socrates is a philosopher and Aristotle is not a teacher of Plato’ is a sentence of L_{simple}.

And since that is a sentence of L_{simple}, and ‘Plato is a teacher of Aristotle’ is a sentence of L_{simple}, also

(Slide 42)

‘(Socrates is a philosopher and Aristotle is not a teacher of Plato) or Plato is a teacher of Aristotle’ is a sentence of L_{simple}.

You see that I have put parentheses around

‘Socrates is a philosopher and Aristotle is not a teacher of Plato’

here: the reason is that I wanted to make clear that the ‘or’ connects that sentence and

‘Plato is a teacher of Aristotle’.

For instance,

(Slide 44)

‘Socrates is a philosopher and (Aristotle is not a teacher of Plato or Plato is a teacher of Aristotle)’ is a sentence of L_{simple}, too, but a different one: in it the 'or' connects 'Aristotle is not a teacher of Plato' and 'Plato is a teacher of Aristotle'. If one wanted to make the grammar of our toy language perfectly precise, then one should involve parentheses in the grammatical rules from the start, but let us be relaxed about this: we can always include parentheses when this is necessary to avoid misunderstandings.

That completes the grammatical rules of our language L_{simple}:

Grammar of L_{simple}:

- If we put a name before 'is a philosopher' we get a sentence of L_{simple}.
- If we put a name before 'is a teacher of' and another one after it, we get a sentence of L_{simple}.
- If we put 'it is not the case that' ('not') in front of a sentence of L_{simple}, we get a sentence of L_{simple}.
- For every two sentences of L_{simple}, if we put an 'and' or an 'or' between them, then we get sentences of L_{simple}.

All, and only what is a sentence according to these rules shall count as a sentence of L_{simple}.

Some bits of terminology (but it's not very important): 'Not'-sentences are also called 'negation sentences', 'and'-sentences are also called 'conjunction sentences', 'or'-sentences are also called 'disjunction sentences'.
E.g., 'Socrates is a philosopher and (Aristotle is not a teacher of Plato or Plato is a teacher of Aristotle)' is a conjunction sentence, as the 'and' is the dominating logical connective. Picture a reversed tree that starts with the 'and' as its top node, where there is a left branch downwards that leads from it to 'Socrates is a philosopher', and where there is a right branch downwards that leads from the top 'and' node to an 'or' node that corresponds to 'Aristotle is not a teacher of Plato or Plato is a teacher of Aristotle'. From that 'or' node there is again a left branch downwards that leads to a 'not' node that corresponds to 'Aristotle is not a teacher of Plato', and there is a right branch that leads from the 'or' node downwards to 'Plato is a teacher of Aristotle'. Finally, the 'not' node on the left leads downwards to 'Aristotle is a teacher of Plato'. This little tree depicts the syntactical structure of the sentence 'Socrates is a philosopher and (Aristotle is not a teacher of Plato or Plato is a teacher of Aristotle)'. The sentence is a conjunction sentence since the top node of its tree is an 'and' node.

Quiz 11:
By applying the grammatical rules of L_{simple}, show that 'Socrates is a teacher of Tarski or it is not the case that Socrates is a teacher of Tarski' is a sentence of L_{simple}.
Solution

2.6 Truth for L_{simple} (09:42)

Let us now assume that we understand already all of the basic expressions in the vocabulary of L_{simple}: 'Socrates', 'Plato', 'Aristotle', 'Tarski', 'is a philosopher', 'is a teacher of', 'not', 'and', 'or'. Additionally, let us assume that we understand the terms 'name (in the vocabulary of L_{simple})', 'predicate (in the vocabulary of L_{simple})', 'sentence of L_{simple}', and the logical symbols 'for all' (with the relevant variables), 'if-then', and 'if and only if'. In fact, we do not really need to assume that we understand all of these expressions because we already understand them quite well. This being in place, we can now define truth for L_{simple} in a Tarskian manner.

Here is the definition of truth for L_{simple}, well, the first version of it:

(Slide 45)

Truth for L_{simple} (first version):

For all sentences x of L_{simple}:

- if x is the result of putting together the name 'Socrates' with the predicate 'is a philosopher', then

  x is true if and only if Socrates is a philosopher;

- if x is the result of putting together the name 'Plato' with the predicate 'is a philosopher', then

  x is true if and only if Plato is a philosopher;

- if x is the result of putting together the name 'Aristotle' with the predicate 'is a philosopher', then

  x is true if and only if Aristotle is a philosopher;

- if x is the result of putting together the name 'Tarski' with the predicate 'is a philosopher', then

  x is true if and only if Tarski is a philosopher;

(Slide 46)

Truth for L_{simple} (first version): [CONTINUED]

- if x is the result of putting together the name ‘Socrates’ with the predicate ‘is a teacher of’ and with the name ‘Socrates’, then

  x is true if and only if Socrates is a teacher of Socrates;

- if x is the result of putting together the name ‘Socrates’ with the predicate ‘is a teacher of’ and with the name ‘Plato’, then

  x is true if and only if Socrates is a teacher of Plato;

  $\vdots$

  [4 names times 4 names = 16 cases in total for ‘is a teacher of’]

(Slide 47)

Truth for L_{simple} (first version): [CONTINUED]

- (case: $x =$ ‘not’ $+\, y$)

  if there is a sentence y of L_{simple}, such that x is the result of putting together the logical symbol ‘it is not the case that’ with y, then

  x is true if and only if y is not true;

(Slide 48)

(Equivalently:

if x is the result of putting together the logical symbol ‘it is not the case that’ with a sentence y of L_{simple}, then

x is true if and only if y is not true.)

(Equivalently, and most precisely:

for all sentences y of L_{simple}, if x is the result of putting together the logical symbol ‘it is not the case that’ with y, then

x is true if and only if y is not true.)

Furthermore:

(Slide 49)

Truth for L_{simple} (first version): [CONTINUED]

- (case: $x = y +$ 'and' $+ z$)

  if there is a sentence y of L_{simple} and a sentence z of L_{simple}, such that x is the result of putting together y with the logical symbol 'and' and with z, then

  x is true if and only if y is true and z is true;

(Slide 50)

(Equivalently:

if x is the result of putting together a sentence y of L_{simple} with the logical symbol 'and' and with a sentence z of L_{simple}, then

x is true if and only if y is true and z is true.)

(Equivalently, and most precisely:

for all sentences y of L_{simple}, for all sentences z of L_{simple}, if x is the result of putting together y with the logical symbol 'and' and with z, then

x is true if and only if y is true and z is true.)

Finally:

(Slide 51)

Truth for L_{simple} (first version): [CONTINUED]

- (case: $x = y +$ 'or' $+ z$)

  if there is a sentence y of L_{simple} and a sentence z of L_{simple}, such that x is the result of putting together y with the logical symbol 'or' and with z, then

  x is true if and only if y is true or z is true;

(Slide 52)

> (Equivalently:
>
> if x is the result of putting together a sentence y of L_{simple} with the logical symbol 'or' and with a sentence z of L_{simple}, then
>
> x is true if and only if y is true or z is true.)
>
> (Equivalently, and most precisely:
>
> for all sentences y of L_{simple}, for all sentences z of L_{simple}, if x is the result of putting together y with the logical symbol 'or' and with z, then
>
> x is true if and only if y is true or z is true.)

Let us first see how this definition is meant to work.

For example: We know from the rules of grammar for L_{simple} that the sentence

(Slide 53)

'Socrates is a philosopher' is a sentence of L_{simple}.

So the sentence 'Socrates is a philosopher' is one of the xs that our definition is concerned with. So whatever follows the 'For all sentences x of L_{simple}' in our definition applies also to the sentence 'Socrates is a philosopher'. The only if-then sentence, the only case in our definition that is relevant for that sentence is the first one:

(Slide 54)

- if x is the result of putting together the name 'Socrates' with the predicate 'is a philosopher', then

  x is true if and only if Socrates is a philosopher;

for all the other cases in our definition deal with other sentences of L_{simple}.

'Socrates is a philosopher' is indeed the result of putting together the name 'Socrates' with the predicate 'is a philosopher', it is an x of the required form, indeed the only such x, and the definition of truth tells us now that this very sentence is true if and only if Socrates is a philosopher. In other words, we can derive from our definition of truth together with the fact that 'Socrates is a philosopher' is a sentence of L_{simple} that is the result of putting together the name 'Socrates' with the predicate 'is a philosopher':

(Slide 55)

‘Socrates is a philosopher’ is true if and only if Socrates is a philosopher.

That is exactly the kind of truth equivalence that we need to derive from our definition in order to prove the definition materially adequate in the sense of Tarski.

In the same way, by applying other if-then statements, other cases, of our definition, we can conclude:

(Slide 56)

‘Plato is a teacher of Aristotle’ is true if and only if Plato is a teacher of Aristotle.

‘Aristotle is a teacher of Plato’ is true if and only if Aristotle is a teacher of Plato.

‘Socrates is a teacher of Socrates’ is true if and only if Socrates is a teacher of Socrates.

Now how about the negation sentence ‘It is not the case that Aristotle is a teacher of Plato’, which is also a sentence of L_{simple} according to the grammatical rules of that language:

(Slide 57)

‘It is not the case that Aristotle is a teacher of Plato’ is a sentence of L_{simple}.

The relevant part of the definition above is this one:

(Slide 58)

- if there is a sentence y of L_{simple}, such that x is the result of putting together the logical symbol ‘it is not the case that’ with y, then

  x is true if and only if y is not true;

y: ‘Aristotle is a teacher of Plato’

x: ‘It is not the case that Aristotle is a teacher of Plato’

There is indeed a sentence y, such that the sentence ‘It is not the case that Aristotle is a teacher of Plato’ is the result of putting together the logical symbol ‘it is not the case that’ with y; in fact, the corresponding y is uniquely determined: it is the unnegated sentence ‘Aristotle is a teacher of Plato’.

What our definition tells us, therefore, is that the sentence

(Slide 59)

'It is not the case that Aristotle is a teacher of Plato' is true if and only if it is not the case that 'Aristotle is a teacher of Plato' is true.

But we already know from before under which conditions 'Aristotle is a teacher of Plato' is true: it is true if and only if Aristotle is a teacher of Plato.

(Slide 60)

'It is not the case that Aristotle is a teacher of Plato' is true if and only if it is not the case that Aristotle is a teacher of Plato.

So by substituting equivalents for equivalent statements, which is a logically valid inference even in the context of 'it is not the case that', we can derive:

'It is not the case that Aristotle is a teacher of Plato' is true if and only if it is not the case that Aristotle is a teacher of Plato.

But that is yet another truth equivalence: the truth equivalence for the negation sentence 'It is not the case that Aristotle is a teacher of Plato'.

In a similar way, we can also derive

(Slide 61)

'It is not the case that it is not the case that Aristotle is a teacher of Plato' is true if and only if it is not the case that it is not the case that Aristotle is a teacher of Plato.

from our definition.

Accordingly, the relevant part of our definition of truth as far as the sentence 'Socrates is a philosopher and it is not the case that Aristotle is a teacher of Plato' is concerned, is

(Slide 63)

- if there is a sentence y of L_{simple} and a sentence z of L_{simple}, such that x is the result of putting together y with the logical symbol 'and' and with z, then

  x is true if and only if y is true and z is true;

x: 'Socrates is a philosopher and it is not the case that Aristotle is a teacher of Plato'

y: 'Socrates is a philosopher'

z: 'it is not the case that Aristotle is a teacher of Plato'

And what the definition of truth tells therefore us is that

(Slide 64)

'Socrates is a philosopher and it is not the case that Aristotle is a teacher of Plato' is true if and only if 'Socrates is a philosopher' is true and 'it is not the case that Aristotle is a teacher of Plato' is true.

Once again, by substituting equivalents for each other, we can derive from this and from what we had derived before:

(Slide 65)

'Socrates is a philosopher and it is not the case that Aristotle is a teacher of Plato' is true if and only if Socrates is a philosopher and it is not the case that Aristotle is a teacher of Plato.

That is yet another truth equivalence for a sentence in L_{simple}. And so on and so forth.

Quiz 12:
Derive the following from our definition of truth: 'Socrates is a teacher of Tarski or it is not the case that Socrates is a teacher of Tarski' is true if and only if Socrates is a teacher of Tarski or it is not the case that Socrates is a teacher of Tarski.
Solution

Remark on truth: Now that you have caught a first glimpse of a Tarskian theory of truth, you might be interested in getting an overview of various different philosophical theories of truth, including Tarski's, and in comparing them with each other:
`http://plato.stanford.edu/entries/truth/`.

2.7 Recursive Definitions (10:07)

By the way, e.g., since we also know that Socrates is a philosopher, and Aristotle is not really a teacher of Plato, we also know from the truth equivalences that we have just derived that the sentence 'Socrates is a philosopher' is true, while the sentence 'Aristotle is a teacher of Plato' is not true.

(Slide 66)

'Socrates is a philosopher' is true.

'Aristotle is a teacher of Plato' is not true.

But that's not really the issue now. One might say that these latter bits of knowledge have two sources: a conceptual one, such as knowledge of the the truth equivalence "Socrates is a philosopher' is true if and only if Socrates is a philosopher'; but also an empirical source, such as knowledge of the historical fact that Socrates is a philosopher. These two sources of knowledge combined by a little bit of logical reasoning lead to our knowledge of, e.g., the sentence 'Socrates is a philosopher' being true. But the empirical part of that knowledge is not really philosophical knowledge: in this case, it would be historical knowledge.

Back to philosophy: By now it should not come as a big surprise that all truth equivalences for sentences in L_{simple} are derivable from the definition above, given the corresponding definition of 'sentence of L_{simple}' and some basic facts on how the sentences of L_{simple} are built up from the basic terms of L_{simple}. In other words:

(Slide 67)

The definition of truth for L_{simple} is materially adequate.

I will return to that point later.

We have stated a definition of truth that is materially adequate. All linguistic expressions in that definition, except maybe for the truth predicate, are understood. The idea is now that the meaning of 'true' gets pinned down by the definition, just as the meaning of 'grandfather of' was captured by our previous definition of that term. And that seems to be the case. Voila!

There is one slight problem though: In the case of the definition of grandfatherhood, the term 'grandfather of' did not occur on the right-hand side of the definition. And it better did not, or otherwise the definition of grandfatherhood would have been circular: we would have defined the term 'grandfather of' by means of, amongst others, the term 'grandfather of' again. However, in the case of our definition of truth from above the term 'true' does occur on the right-hand sides of some of the various parts of our definition. For instance, in the relevant case for negation sentences,

(Slide 68)

- if there is a sentence y of L_{simple}, such that x is the result of putting together the logical symbol 'it is not the case that' with y, then

  x is true if and only if y is not true;

'true' shows up to the left of 'if and only if' and to its right!

So something seems to be fishy here, right? As things stand, the definition does not seem to be formally correct.

Well: not so fast. There are circularities and circularities. For instance: mathematicians, and even more so computer scientists, are very much used to what they call recursive definitions. Here is an example: assume that addition, the summing up, of natural numbers is already well understood. But now you want explain to someone – a human being or a computer – what multiplication, what the product of two natural numbers is. Then one might turn to the following recursive definition of multiplication:

(Slide 69)

- (Multiplying 0 with m yields 0.)

  $mult(0, m) = 0.$

- (Multiplying $1 + n$ with m yields the sum of m and the result of multiplying n with m.)

  $mult(1 + n, m) = m + mult(n, m).$

This is a definition of multiplication in which the defined term 'mult' occurs on both sides of the second equation.

Nevertheless mathematicians and computer scientists do not worry about this. Why?

Because multiplication of a number with m is reduced thereby to a multiplication of a smaller number with m. And the product of that smaller number with m is reduced to a multiplication of a yet smaller number with m. And so on, until the number that is to be multiplied with m becomes so small that it is 0 (see Figure 2.1): that is exactly where the initial part of the definition applies: $mult(0, m) = 0$ (see Figure 2.2).

$$\begin{aligned}
& mult(1+n, m) \\
= \; & m + mult(n, m) \\
= \; & m + m + mult(n-1, m) \\
= \; & m + m + m + mult(n-2, m) \\
= \; & m + m + m + \ldots + mult(1, m) \\
= \; & m + m + m + \ldots + m + mult(0, m)
\end{aligned}$$

Figure 2.1: Calculation (part 1)

$$
\begin{aligned}
& mult(1+n, m) \\
= \; & m + mult(n, m) \\
= \; & m + m + mult(n-1, m) \\
= \; & m + m + m + mult(n-2, m) \\
= \; & m + m + m + \ldots + mult(1, m) \\
= \; & m + m + m + \ldots + m + 0
\end{aligned}
$$

Figure 2.2: Calculation (part 2)

And in that part of the definition, the term '*mult*' does not occur on the right-hand side of the equation anymore. The whole sequence of reductions by definition bottoms out in something that can be understood without any understanding of '*mult*'. Such recursive definitions are circular, but they are not viciously circular: nothing bad follows from them. In fact, every good programmer knows how extremely useful recursive procedures can be, and all the standard programming languages allow for recursive procedures that, as it were, "call themselves".

Now, exactly the same holds for the definition of 'true' for sentences of L_{simple} above. By our definition, the truth procedure "calls itself".

The truth condition for the complex sentence

'Socrates is a philosopher and Aristotle is not a teacher of Plato'

is determined, by definition, from the truth conditions for the simpler sentences

'Socrates is a philosopher' and 'Aristotle is not a teacher of Plato',

and the truth condition for the latter of these is determined by the truth conditions for the yet smaller sentence

'Aristotle is a teacher of Plato'.

Finally, the truth conditions for the maximally simple sentences 'Socrates is a philosopher' and 'Aristotle is a teacher of Plato' are formulated without invoking the term 'true' anymore (see Figure 2.3):

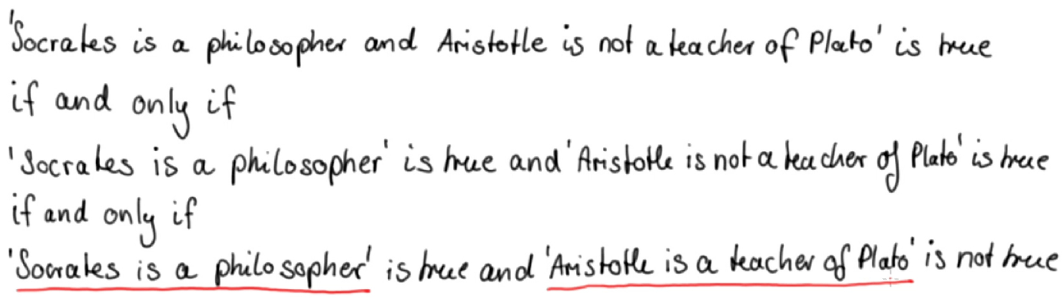

Figure 2.3: "Calculation" of truth conditions

(Slide 70)

- if x is the result of putting together the name 'Socrates' with the predicate 'is a philosopher', then

  x is true if and only if Socrates is a philosopher;

- if x is the result of putting together the name 'Aristotle' with the predicate 'is a teacher of' and with the name 'Plato', then

  x is true if and only if Aristotle is a teacher of Plato;

The only terms used on these right-hand sides are 'Socrates', 'is a philosopher', 'Aristotle', 'Plato', and 'is a teacher of'. The whole sequence of reduction by definition bottoms out in something that can be understood without any prior understanding of 'true'.

Our definition of truth from above is thus a recursive definition in the same sense as the definition of multiplication: it is a definition of truth for sentences in which the truth conditions for complex sentences are determined from the truth conditions for less complex sentences. It is circular, but it is not viciously circular.

In fact, if one thinks about it, already our definition of 'sentence of L_{simple}' was a recursive definition, though we did not emphasize that at the time: the simplest sentences were defined explicitly to be sentences of L_{simple}; and the complex sentences were defined to be sentences of L_{simple} if their parts were. In other words:

(Slide 71)

Sentencehood for L_{simple} (first version):

For all x:

- if x is the result of putting together the name 'Socrates' with the predicate 'is a philosopher', then x is a sentence of L_{simple};

  ⋮

- if there is a sentence y of L_{simple}, such that x is the result of putting together the logical symbol 'it is not the case that' with y, then x is a sentence of L_{simple};

(Slide 72)

> (Equivalently:
>
> if x is the result of putting together the logical symbol 'it is not the case that' with a sentence y of L_{simple}, then x is a sentence of L_{simple}.)
>
> (Equivalently, and most precisely:
>
> for all sentences y of L_{simple}, if x is the result of putting together the logical symbol 'it is not the case that' with y, then x is a sentence of L_{simple}.)
>
> ⋮

And so on. And nothing else is a sentence of L_{simple}.

Languages that are defined from a given vocabulary in such a precise recursive manner are called formal languages. So L_{simple} is a formal language, and both the syntax of L_{simple} – everything to do with vocabulary and grammar – and the semantics of L_{simple} – everything to do with truth conditions – are recursive in nature.

And this is not just a random coincidence or an artifact of the manner in which we set up our toy language L_{simple}: linguists and philosophers of language believe that pretty much the same is true of natural language in general. Indeed, if natural language were not recursive in this sense, it would be hard to explain how we are able to understand sentences which we have never seen or heard before. For example: I bet you have not ever seen the sentence

(Slide 73)

It is not the case that (Aristotle is a teacher of Tarski, and Tarski is a teacher of Aristotle).

before in your whole life. But still, by understanding its basic expressions, how these basic expressions are put together grammatically, and how the truth condition of a complex sentence like that depends on the truth conditions of its parts, you are able to understand the sentence. You understand the sentence says it is not the case that both of the embedded sentences are true; that at least one of them is false; that therefore Aristotle is not a teacher of Tarski, or Tarski is not a teacher of Aristotle, or neither is a teacher of the other. Whether the sentence is actually true or not, that is what it says. Without grammar and truth conditions being recursive, you would not have been able to figure that out for a sentence that you haven't ever seen before.

Remark on recursiveness and language: As philosophers say, the meaning of sentences in natural language is compositional. That is: the meaning of a sentence in natural language is determined by the meanings of its parts (and by the manner in which these parts are put together). More on this can be found at
<http://plato.stanford.edu/entries/compositionality/.

Quiz 13:
Find a recursive definition of the function f on numbers $1, 2, 3, \ldots$ so that $f(n) = 1 \cdot 2 \cdot \ldots \cdot n$.
(So $f(n)$ is such that it is the product of the numbers $1, 2, 3, \ldots, n$.)
Solution

2.8 Explicit Definitions (08:02)

Maybe you still worry about recursive definitions. Don't: there is in fact a way of turning the recursive definitions from above into definitions that are formally correct in the same sense in which the definition of grandfatherhood is: where any circularities, even harmless ones, are avoided. Tarski himself exploited this fact which had been known to philosophers and logicians at least since the wonderful work by the German mathematician and philosopher Gottlob Frege, who created what we now call modern formal logic at the end of the 19th and the beginning of the 20th century. Every recursive definition can be transformed into a standard definition without any kind of circularity.

Remark on Frege: If you want to know more about Gottlob Frege, please consult
http://plato.stanford.edu/entries/frege/.
(The other pioneer on recursive definitions and on how to turn them into explicit ones was Richard Dedekind, whom I had mentioned already in the first lecture.)

In the case of truth for L_{simple}, this is how it works:

(Slide 74)

Truth for L_{simple} (second version):

For all sentences x of L_{simple}:

x is true if and only if x is a member of all sets Y of sentences of L_{simple} for which the following holds: for all x',

- if x' is the result of putting together the name 'Socrates' with the predicate 'is a philosopher', then

  x' is a member of Y if and only if Socrates is a philosopher;

  $\vdots$

(See Figure 2.4 and Figure 2.5.)

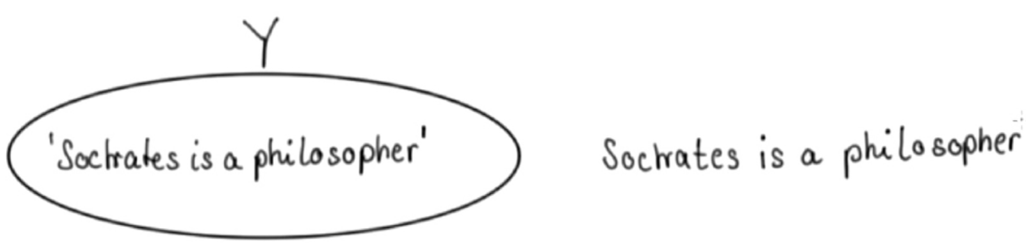

Figure 2.4: Illustration 1 for "truth for L_{simple}"

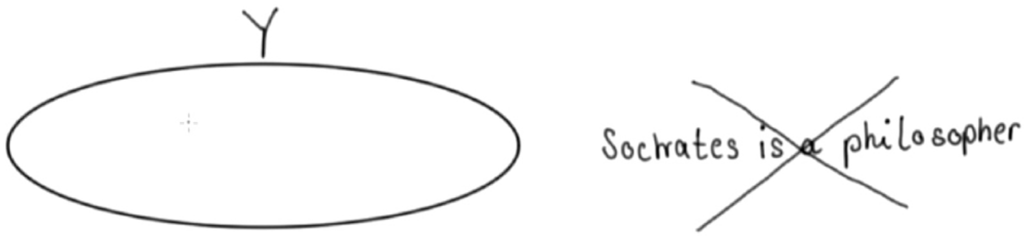

Figure 2.5: Illustration 2 for "truth for L_{simple}"

(Slide 75)

⋮

- if x' is the result of putting together the name 'Socrates' with the predicate 'is a teacher of' and with the name 'Socrates', then

  x' is a member of Y if and only if Socrates is a teacher of Socrates;

- if x' is the result of putting together the name 'Socrates' with the predicate 'is a teacher of' and with the name 'Plato', then

  x' is a member of Y if and only if Socrates is a teacher of Plato;

⋮

(Slide 76)

Truth for L_{simple} (second version): [CONTINUED]

- (case: $x' =$ 'not' $+\, y$)

  if there is a sentence y of L_{simple}, such that x' is the result of putting together the logical symbol 'it is not the case that' with y, then

  x' is a member of Y if and only if y is not a member of Y;

(See Figure 2.6 and Figure 2.7.)

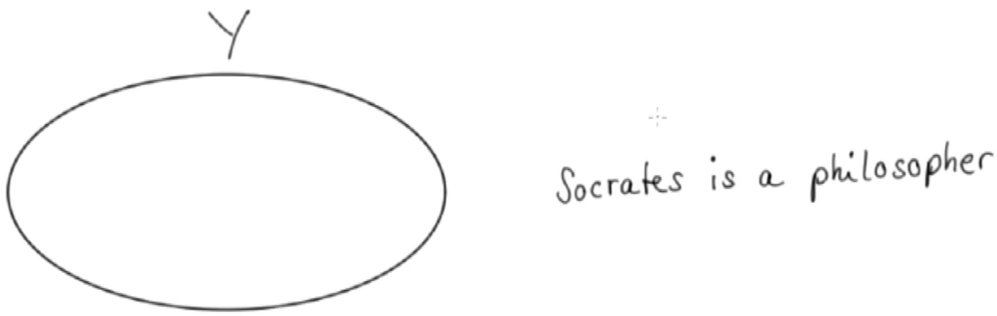

Figure 2.6: Illustration 3 for "truth for L_{simple}"

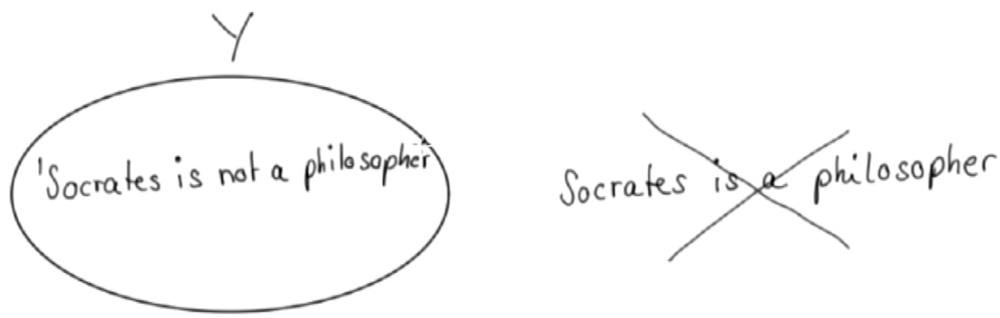

Figure 2.7: Illustration 4 for "truth for L_{simple}"

(Slide 77)

(Equivalently:

if x' is the result of putting together the logical symbol 'it is not the case that' with a sentence y of L_{simple}, then

x' is a member of Y if and only if y is not a member of Y.)

(Equivalently, and most precisely:

for all sentences y of L_{simple}, if x' is the result of putting together the logical symbol 'it is not the case that' with y, then

x' is a member of Y if and only if y is not a member of Y.)

(Slide 78)

Truth for L_{simple} (second version): [CONTINUED]

- (case: $x' = y +$ 'and' $+ z$)

  if there is a sentence y of L_{simple} and a sentence z of L_{simple}, such that x' is the result of putting together y with the logical symbol 'and' and with z, then

  x' is a member of Y if and only if y is a member of Y and z is a member of Y;

(See Figure 2.8 and Figure 2.9.)

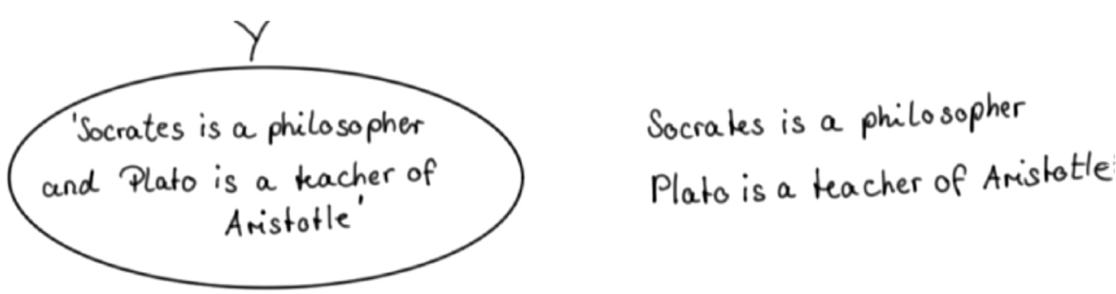

Figure 2.8: Illustration 5 for "truth for L_{simple}"

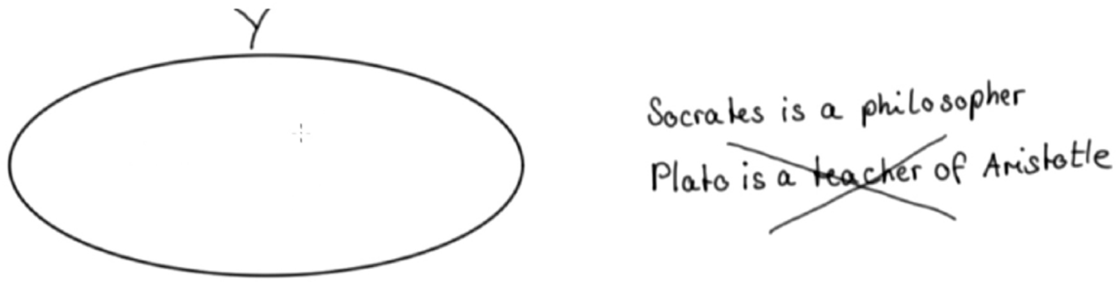

Figure 2.9: Illustration 6 for "truth for L_{simple}"

(Slide 79)

(Equivalently:

if x' is the result of putting together a sentence y of L_{simple} with the logical symbol 'and' and with a sentence z of L_{simple}, then

x' is a member of Y if and only if y is a member of Y and z is a member of Y.)

(Equivalently, and most precisely:

for all sentences y of L_{simple}, for all sentences z of L_{simple}, if x' is the result of putting together y with the logical symbol 'and' and with z, then

x' is a member of Y if and only if y is a member of Y and z is a member of Y.)

(Slide 80)

Truth for L_{simple} (second version): [CONTINUED]

- (case: $x' = y + \text{'or'} + z$)

  if there is a sentence y of L_{simple} and a sentence z of L_{simple}, such that x' is the result of putting together y with the logical symbol 'or' and with z, then

  x' is a member of Y if and only if y is a member of Y or z is a member of Y.

(See Figure 2.10, Figure 2.11, and Figure 2.12.)

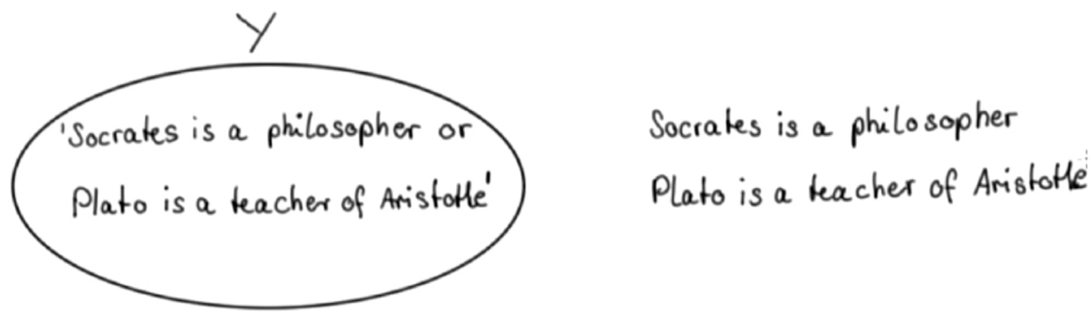

Figure 2.10: Illustration 7 for "truth for L_{simple}"

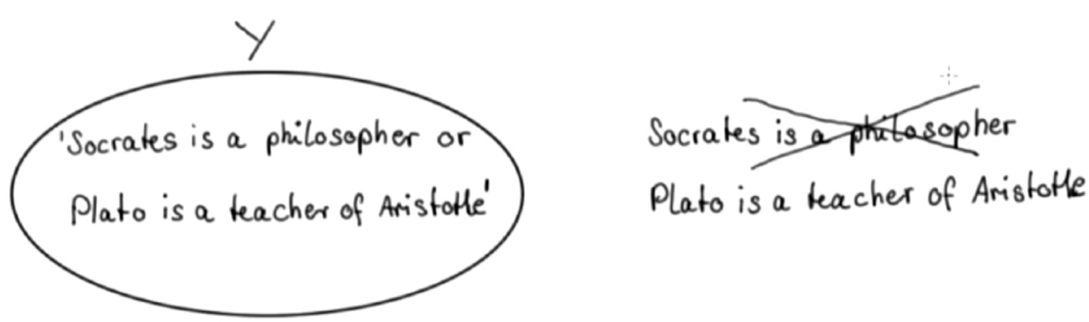

Figure 2.11: Illustration 8 for "truth for L_{simple}"

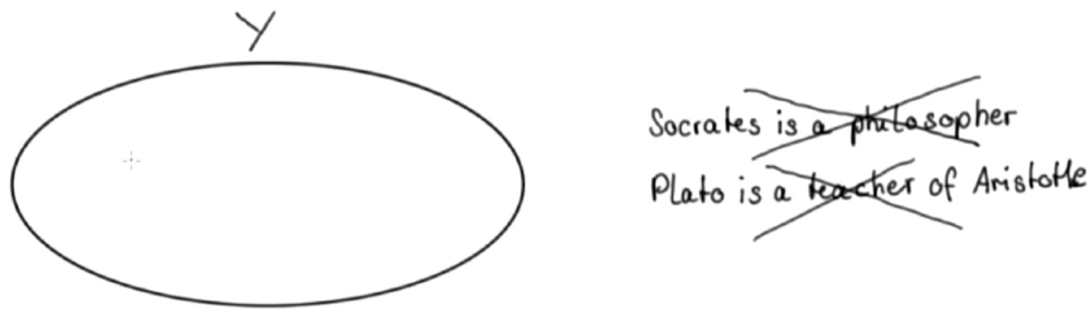

Figure 2.12: Illustration 9 for "truth for L_{simple}"

(Slide 81)

> (Equivalently:
>
> if x' is the result of putting together a sentence y of L_{simple} with the logical symbol 'or' and with a sentence z of L_{simple}, then
>
> x' is a member of Y if and only if y is a member of Y or z is a member of Y.)
>
> (Equivalently, and most precisely:
>
> for all sentences y of L_{simple}, for all sentences z of L_{simple}, if x' is the result of putting together y with the logical symbol 'or' and with z, then
>
> x' is a member of Y if and only if y is a member of Y or z is a member of Y.)

Basically, wherever we had said 'x is true' before, we now say 'x (or now: x') is a member of the set Y; whatever we demanded of truth before, we now demand of certain sets Y of sentences of L_{simple}; and we call a sentence of L_{simple} true if and only if it is a member of all such sets Y. By a set we mean again what had been explained already in the first lecture. Instead of talking of truth on the right-hand side of our definition, we now talk of all sets Y of sentences that obey the same conditions as truth. We say more or less what we had said before, but now we manage to say it without using the predicate 'true' itself on the right-hand side. In this way, we can state a definition of truth that is not just materially adequate but also formally correct: the truth predicate 'true' does not show up anymore on the right-hand side of the definition. Instead, we talk about sets of a certain kind there. The definition is of course a bit more complicated than the one of 'grandfather of' – it involves more cases and it contains expressions such as 'for all sets Y' – but other than that the two definitions are of the same kind and equally harmless. Our definition of truth for the sentences of L_{simple} ticks all boxes: it is free of any formal flaws, and it can be seen to still imply all truth equivalences, as intended.

The same method can also be applied to state a standard, non-recursive definition of 'sentence of L_{simple}':

(Slide 82)

Sentencehood for L_{simple} (second version):

For all x:

x is a sentence of L_{simple} if and only if x is a member of all sets Y for which the following holds: for all x',

- if x' is the result of putting together the name 'Socrates' with the predicate 'is a philosopher', then x' is a member of Y;

  $\vdots$

- if there is a member y of Y, such that x' is the result of putting together the logical symbol 'it is not the case that' with y, then x' is a member of Y;

(Slide 83)

> (Equivalently:
>
> if x' is the result of putting together the logical symbol 'it is not the case that' with a member y of Y, then x' is a member of Y.)
>
> (Equivalently, and most precisely:
>
> for all members y of Y, if x' is the result of putting together the logical symbol 'it is not the case that' with y, then x' is a member of Y.)
>
> $\vdots$

And so on.

In this case, wherever we had said 'x is a sentence of L_{simple}' before, we now say 'x (or now: x') is a member of the set Y'; whatever we demanded of sentencehood in L_{simple} before, we now demand of certain sets Y; and we call something a sentence of L_{simple} if and only if it is a member of all such sets Y.

In a sense, the recursive structure of our previous definitions of truth and of 'sentence of L_{simple}' is preserved by our new and improved definitions, as virtually all parts of our previous definitions survive in our new definitions if only in slightly reformulated form. In fact it should be easier to understand now why the circularity in our previous definitions was not really a problem: because it can be eliminated in favour of definitions that are formally correct in the traditional sense of the word. Also note that our formally correct definition of truth for sentences of L_{simple} builds on, or presupposes, our formally correct definition of what a sentence of L_{simple} is.

Quiz 14:
Let a function f be defined recursively on the numbers $0, 1, 2, \ldots$ as follows:
$f(0) = 1$. $f(n+1) = f(n) + 2$.
Can you find a way of defining the same function f explicitly, that is, a way of assigning the same value to '$f(n)$' but without mentioning in the definition the function f itself again?
Solution

2.9 Theorems about Truth for L_{simple} (13:44)

As far as the sentences in our toy language L_{simple} are concerned, we found what we were looking for: a formally correct and materially adequate definition of truth. Everyone who understands the everyday expressions 'Socrates', 'Plato', 'Aristotle', 'Tarski', 'is a philosopher', 'is a teacher of', the logical expressions 'not', 'and', 'or', 'for all' (with the relevant variables), 'if-then', and 'if and only if', who understands the mathematical term 'set', who understands in addition the syntactical expressions 'name (in the vocabulary of L_{simple})', 'predicate (in the vocabulary of L_{simple})', and 'sentence of L_{simple}', and who knows the relevant English grammar of our definition, can also understand the definition of truth for L_{simple} and hence understands truth for L_{simple}.

While L_{simple} has been really simple, Tarski actually provided similar kinds of definitions of truth for much more complex languages: languages with much greater vocabulary, vocabularies that might include the usual names and predicates from, say, mathematics, physics, history, and more; languages with more logical symbols, such as 'if-then', 'if and only if', 'for all', 'there is', and more; languages with more complex grammar; languages with sentences that speak about infinitely many objects, such as the natural numbers, or all sets of natural numbers, and so forth. While, accordingly, the definitions of truth for these languages are more complex than our definition of truth for L_{simple}, they are equally doable, and the construction of the corresponding definition is based on similar methods: in particular, stating a recursive definition of truth, and then turning that recursive definition into a standard, so-called explicit, non-recursive one.

I will not state any of these more complex definitions of truth for more complex languages here. Instead let me show you what further conclusions we can draw from our definition of truth for L_{simple}.

First of all, L_{simple} is not that simple: its set of sentences is already an infinite set. This is because by applying the logical symbols over and over again to sentences in L_{simple}, one gets longer and longer sentences of L_{simple}, and there is no end to this. Indeed, one can prove that the set of sentences of L_{simple} is of the same size as the set of natural numbers, in the sense explained in the last lecture: there is a pairing off between the members of the two sets. So our definition defines truth for infinitely many sentences in one fell swoop, which is not that bad.

Secondly, once one has stated the definitions of 'sentence of L_{simple}' and truth for L_{simple} as precisely as it was done above, it becomes possible to prove general claims about truth for L_{simple} by means of similar methods as those by which mathematicians prove statements about all natural numbers.

Mathematicians prove that all natural numbers have a certain property P by first showing that

(Slide 84/1)

- 0 has that property P;

  and then by proving that

  (Slide 84/2)

- for all natural numbers n: if n has the property P, then also $n+1$ has that property P.

By the so-called principle of complete induction over natural numbers – which has nothing to do with inductive or probabilistic reasoning – it follows from this that all natural numbers have property P.

(Slide 85)

- 0 has that property P;
- for all natural numbers n: if n has the property P, then also $n+1$ has that property P.

From this, by *complete induction over natural numbers*, it follows:

- All natural numbers n have property P.

For 0 has P, and if 0 has P also 1 has the property, and if 1 has it also 2, and so on, throughout all the natural numbers.

Remark on complete induction over natural numbers: This is the standard method of proof that mathematicians use in order to prove that all natural numbers have a certain property. You can find it explained in almost every introductory textbook on mathematics. Sometimes the term 'complete' is used to denote a closely related (and equivalent) method of proving that all natural numbers have a certain property P: First show that 0 has that property P. And then prove for all natural numbers n: if all the numbers $0, ..., n$ have the property P, then also $n+1$ has that property P. So here 'if all the numbers $0, ..., n$ have the property P' replaces 'if n has the property P' from before.
However, in our lectures, I do not use the term 'complete' in this way: by saying 'complete induction' I merely want to signal that I do not mean anything like inductive reasoning (or probabilistic reasoning) here which will become a topic from the third lecture when we will turn to the topic of subjective probability.

Similarly, in our case, one proves first of the simplest sentences, like 'Socrates is a philosopher', that they have a property P:

(Slide 86/1)

- 'Socrates is a philosopher' has property P.

  $\vdots$

Then one proves:

(Slide 86/2)

- For all sentences x of L_{simple}: if x has the property P, then also the result of putting together the logical symbol 'it is not the case that' with x has property P.
- For all sentences x of L_{simple}, for all sentences y of L_{simple}: if x has the property P and also y has the property P, then the sentence of L_{simple} that results from putting together x with 'and' and y has the property P.
- For all sentences x of L_{simple}, for all sentences y of L_{simple}: if x has the property P and also y has the property P, then the sentence of L_{simple} that results from putting together x with 'or' and y has the property P.

By the principle of complete induction over sentences in L_{simple} it follows from this that all sentences of L_{simple} have property P.

For the simplest sentences have P, and if they have P then also the sentences that can be built by applying a logical symbol to them have P, and if those have P then also the sentences that can be built by applying a logical sign to them have property P, and so on, throughout all the sentences of L_{simple}. We know that we reach all sentences of L_{simple} in this way, by the definition of L_{simple} that was stated above.

So this principle of complete induction over sentences or formulas of a formal language L really says: if (i) property P holds for the simplest sentences of L, and (ii) if property P is preserved when sentences that have P are put together and another sentence is built from them according to the grammatical rules of L, then it follows from this: (iii) all sentences in L have property P. If you want to know more about this version of induction, then really any logic textbook or any logic lecture notes will do in which so-called soundness and completeness theorems are proven: one needs induction in order to prove these theorems, which is why the authors of these books will normally explain complete induction over sentences/formulas before they turn to the proofs of these theorems.

What justification do we have for these principles of complete induction? Either one simply takes them for granted, as axioms, as it were – and no mathematician would complain about that – or, even better, one derives them from set theory, which is possible. As mentioned in the first lecture, set theory forms the foundation of modern mathematics, and the principles of induction above can be shown to follow from it.

Here are two examples of theorems that one can prove by the method of complete induction over sentences in L_{simple}:

(Slide 88/1)

Theorem:

For all sentences x of L_{simple}:

the result of putting together a quotation mark, the sentence x itself, another quotation mark, 'if and only if', and the sentence x again, follows from the definition of truth for L_{simple}

(if taken together with the definition of 'sentence of L_{simple}' and basic facts concerning the grammar of these sentences).

In other words: every truth equivalence for a sentence of L_{simple} follows from the definition of truth for L_{simple} – the definition is provably materially adequate. Intuitively, this was already clear from our examples from before, but it is one thing to be intuitively clear, and another thing to actually prove things.

(Slide 88/2)

E.g., the following sentence follows from the definition of truth for L_{simple}:

$\underbrace{\text{‘Socrates is a philosopher’}}_{x}$ is true if and only if $\underbrace{\text{Socrates is a philosopher}}_{x}$.

Second Theorem:

(Slide 89/1)

Theorem:

For all sentences x of L_{simple}:

x is true or the result of putting together ‘it is not the case that’ with x is true.

In other words: for every sentence of L_{simple}, either the sentence is true or its negation is true.

(Slide 89/2)

E.g.:

‘$\underbrace{\text{Socrates is a philosopher}}_{x}$’ is true or

‘it is not the case that $\underbrace{\text{Socrates is a philosopher}}_{x}$’ is true.

The last theorem shows that our definition of truth is not just useful in the sense of determining the meaning of ‘true’ in a formally correct and materially adequate manner; it is also useful in the sense that one can derive general laws of truth from it, such as: for each of the infinitely many sentences of L_{simple}, either the sentence is true or its negation is. I had mentioned that little law of truth already at the beginning of this lecture when I distinguished truth from belief, but it is one thing to mention it and another thing to actually prove it in a manner that is just as precise as the mathematics of natural numbers.

Let me illustrate the method of proof by induction over sentences of L_{simple} by sketching the proof of the second theorem; the first theorem I will simply take for granted now.

(Slide 90)

Proof of second theorem (sketch):

- Property P:

  x is true or the result of putting together 'it is not the case that' with x is true.

First we prove that the simplest sentences of L_{simple} have property P.

For instance:

(Slide 91/1)

By logic:

- Socrates is a philosopher or it is not the case that Socrates is a philosopher.

By our definition of truth for L_{simple}:

- 'Socrates is a philosopher' is true if and only if Socrates is a philosopher.
- 'It is not the case that Socrates is a philosopher' is true if and only if it is not the case that Socrates is a philosopher.

From this it follows:

- 'Socrates is a philosopher' is true or 'It is not the case that Socrates is a philosopher' is true.

That is:

- 'Socrates is a philosopher' does have property P. ✓

By substituting equivalent sentences for each other, we get from Socrates being a philosopher or Socrates not being a philosopher:

'Socrates is a philosopher' is true or 'It is not the case that Socrates is a philosopher' is true.

Hence, the sentence 'Socrates is a philosopher' does have property P.

Analogously, for the other maximally simple sentences of L_{simple}.

(Slide 92)

Next we want to prove:

- For all sentences x of L_{simple}:

  if x has the property P, then also the result of putting together 'it is not the case that' with x has property P.

In order to be able to state this part of the proof more succinctly, I will abbreviate the result of putting together 'it is not the case that' with x by: $\neg x$.

(Slide 93)

Next we want to prove:

- For all sentences x of L_{simple}:

  if x has the property P, then also $\underbrace{\neg x}_{\text{negation of } x}$ has property P.

(Slide 94/1)

Assume x has property P: (i) x is true or (ii) $\neg x$ is true.

We need to show now that in this case also $\neg x$ has property P: $\neg x$ is true or $\neg\neg x$ is true.

By assumption, x is true or $\neg x$ is true.

(Slide 94/2)

Assume the second to be the case: (ii) $\neg x$ is true.

Then we are done, for that means:

(Slide 94/3)

By logic:

- $\neg x$ is true or $\neg\neg x$ is true.

  $\neg x$ has property P. ✓

Now assume the other possible case, the first case: x is true.

(Slide 95/1)

Assume x has property P: (i) x is true or (ii) $\neg x$ is true.

Assume the other possible case: (i) x is true.

By logic:

- x is not not true.
- It is not the case that it is not the case that x is true.
- It is not the case that x is not true.

(The latter three expressions are equivalent. In any case, whatever the x, the truth of x is doubly negated here.)

(Slide 95/2)

By our definition of truth for L_{simple}:

- x is not true if and only if $\neg x$ is true.

By substituting equivalents for equivalents in the context of 'it is not the case that' it follows:

(Slide 95/3)

By logic:

- It is not the case that $\neg x$ is true.

Here the truth of the negation of x is negated once only.

But now we can apply the definition of truth again: whatever the x, it holds: it is not the case that x is true if and only if $\neg x$ is true.

(Slide 96/1)

[CONTINUED]

We had:

- It is not the case that $\neg x$ is true.

By our definition of truth for L_{simple}:

- It is not the case that x is true if and only if $\neg x$ is true.

This also applies to an x that is itself a negation sentence, something of the form $\neg x$ (see Figure 2.13).

[CONTINUED]

We had:

- It is not the case that $\neg x$ is true.

By our definition of truth for L_{simple}:

- It is not the case that $\neg x$ is true if and only if $\neg\neg x$ is true.

By logic:

- $\neg\neg x$ is true.

By logic again:

- $\neg x$ is true or $\neg\neg x$ is true.

  $\neg x$ has property P. ✓

Figure 2.13: $\neg x$

We can get rid of the 'it is not the case that' and plug in another $\neg$ in front of $\neg x$ as compensation (see Figure 2.14).

[CONTINUED]

We had:

- It is not the case that $\neg x$ is true.

By our definition of truth for L_{simple}:

- It is not the case that $\neg x$ is true if and only if $\neg\neg x$ is true.

By logic:

- $\neg\neg x$ is true.

By logic again:

- $\neg x$ is true or $\neg\neg x$ is true.

  $\neg x$ has property P. ✓

Figure 2.14: Plugging in another '$\neg$' in front of $\neg x$

In other words: by substituting equivalents for equivalents again it follows that

(Slide 96/2)

By logic:

- $\neg\neg x$ is true.

That is: the double negation of x is true.

But that implies:

(Slide 96/3)

By logic again:

- $\neg x$ is true or $\neg\neg x$ is true.

  $\neg x$ has property P. ✓

In either case we derived what we needed to derive. So under the assumption x has property P, that is, x is true or $\neg x$ is true, it does indeed follow that also $\neg x$ has property P, that is, $\neg x$ is true or $\neg\neg x$ is true.

The two remaining cases of sentences of L_{simple} connected by 'and' and sentences of L_{simple} connected by 'or' are similar to the case of negation sentences; one only needs to distinguish four cases now instead of two, but that is all.

So we are ready to apply the law of complete induction over sentences in L_{simple}:

(Slide 97)

From this, by *complete induction over sentences in L_{simple}*, it follows:

- All sentences of L_{simple} have property P.

That is:

- For all sentences x of L_{simple}: x is true or $\neg x$ is true. ✓

We have derived a general law of truth from our definition of truth, the definition of 'sentence of L_{simple}', some rules of logic, and the method of proof by complete induction – an almost arithmetical proof in the service of a philosophical account of truth!

Quiz 15:
In a previous quiz, a function f was defined recursively on the numbers $0, 1, 2, ...$ as follows: $f(0) = 1$. $f(n + 1) = f(n) + 2$.
And then we determined the following equivalent explicit definition of f:
$f(n) = 2n + 1$.
But back then we did not actually *prove* the two definitions to be equivalent. Now please do prove by complete induction over natural numbers that for all n: if $f(0) = 1$ and for all m, $f(m + 1) = f(m) + 2$, then it holds that $f(n) = 2n + 1$. So the property P is in this case: n has P if and only if, if $f(0) = 1$ and for all m, $f(m + 1) = f(m) + 2$, then $f(n) = 2n + 1$.
Solution

Quiz 16:
Fill in the missing part of the inductive proof of the theorem in the lecture as far as 'and'-sentences are concerned. That is, prove:
if (y is true or $\neg y$ is true) and (z is true or $\neg z$ is true), then $y\&z$ is true or $\neg(y\&z)$ is true.
(I write '$y\&z$' here as a shorthand for: the result of putting y together with 'and' and with z. By '$\neg$' I mean negation ('not').)
Solution

2.10 The Liar Paradox (07:59)

I have said before that much in the same way in which we were able to define truth for L_{simple} it is possible to define truth for much more expressive languages, languages with much more interesting vocabularies and grammar.

Now here is a thought: since we have convinced ourselves that the term 'true' is perfectly meaningful and unproblematic, why not add that term to the vocabulary of L_{simple}, extend the grammatical rules accordingly, and then define truth for this extended language? After all: we do not just think that Plato is a teacher of Aristotle, and hence, by our definition of truth, that the sentence 'Plato is a teacher of Aristotle' is true, we also think that for pretty much the same reason the sentence "Plato is a teacher of Aristotle' is true' is true. While in everyday life and the sciences we might hardly ever be interested in a sentence like that, it is perfectly natural for a philosopher to be interested in truths about truths: for philosophers do reflect on everything really! And take the languages for which we might want to define truth to their extreme: consider a full natural language, such as English. We are of course interested in the notion of truth for English as a whole; but English does contain the term 'true' already in its vocabulary. So a definition of truth for English as a whole should allow us to derive sentences that talk about the truth of sentences that talk about the truth of sentences again, and the like.

But there is danger just around the corner. This has been known for a long time: the ancient Greek philosophers Eubulides and Epimenides presented versions of this threat in their work, and a hint of this even found its way into a letter of St Paul in the New Testament.

Remark on Eubulides and Epimenides: A bit more on them and on their work can be found at
http://plato.stanford.edu/entries/dialectical-school/,
http://mathworld.wolfram.com/EubulidesParadox.html,
http://mathworld.wolfram.com/EpimenidesParadox.html.

Let me state the worry in my own terms. But before I do so, take a good look at this sentence:

(Slide 98)

($*$) The sentence in Lecture 2 that is introduced by a star symbol is not true.

O.k.? Don't think about it too much. For the moment, let me just assure you that this sentence will be the only one in this Lecture 2 to be introduced by a star symbol.

Now consider the following argument:

(Slide 99)

> (P1) The sentence in Lecture 2 that is introduced by a star symbol is identical to the sentence
>
> 'The sentence in Lecture 2 that is introduced by a star symbol is not true'.

This is simply an empirical truth: it just happens to be the case that the very sentence in Lecture 2 that is introduced by a star symbol is the sentence that starts with 'The sentence in Lecture 2...' and which ends with '... is not true.' If things had been different, maybe, the sentence in Lecture 2 introduced by a star symbol would have been a different one: maybe 'Snow is white'. But as it happens, the sentence in Lecture 2 that is introduced by a star symbol is just the one mentioned by P1.

The second premise of the argument is an instance of our, by now very familiar, truth scheme:

(Slide 100)

> (P2) 'The sentence in Lecture 2 that is introduced by a star symbol is not true' is true if and only if
>
> the sentence in Lecture 2 that is introduced by a star symbol is not true.

Once again we take the truth scheme and we instantiate the place-holder 'A' by a sentence: in this case, we generate the truth equivalence for the sentence 'The sentence in Lecture 2 that is introduced by a star symbol is not true'. P2 is plausible, simply because all truth equivalences are plausible.

Finally, here is the conclusion of the argument:

(Slide 101)

> (C) 'The sentence in Lecture 2 that is introduced by a star symbol is not true' is true
>
> if and only if
>
> 'The sentence in Lecture 2 that is introduced by a star symbol is not true' is not true.

This conclusion is of the form 'A if and only if not A': but that is a logical contradiction; a sentence cannot be equivalent to its negation. For this would mean that either both sides are the case – A and not A – or neither of them is – neither A nor not A – neither of which is logically possible.

At the same time, the argument from P1 and P2 to C is logically valid, in the sense explained in the first lecture: For assume P1 and P2 to be true.

(Slide 102)

> (P1) The sentence in Lecture 2 that is introduced by a star symbol is identical to the sentence
>
> 'The sentence in Lecture 2 that is introduced by a star symbol is not true'.
>
> (P2) 'The sentence in Lecture 2 that is introduced by a star symbol is not true' is true if and only if
>
> the sentence in Lecture 2 that is introduced by a star symbol is not true.
>
> ---
>
> Premise 1 entails:
>
> 'The sentence in Lecture 2 that is introduced by a star symbol is not true' is true if and only if
>
> the sentence in Lecture 2 that is introduced by a star symbol is true.

This is because logic tells us that if something is identical to something else, then the one can be substituted for the other – this is a logically valid inference. According to P1, the sentence in Lecture 2 that is introduced by a star symbol is identical to the sentence

'The sentence in Lecture 2 that is introduced by a star symbol is not true.'

But that means the one is true if and only if the other is true. In other words:

'The sentence in Lecture 2 that is introduced by a star symbol is not true' is true if and only if the sentence in Lecture 2 that is introduced by a star symbol is true.

Additionally, Premise 2 was:

(Slide 103)

> (P1) The sentence in Lecture 2 that is introduced by a star symbol is identical to the sentence
>
> ‘The sentence in Lecture 2 that is introduced by a star symbol is not true’.
>
> (P2) ‘The sentence in Lecture 2 that is introduced by a star symbol is not true’ is true if and only if
>
> the sentence in Lecture 2 that is introduced by a star symbol is not true.
>
> ---
>
> Premise 1 entails:
>
> ‘The sentence in Lecture 2 that is introduced by a star symbol is not true’ is true if and only if
>
> the sentence in Lecture 2 that is introduced by a star symbol is true.
>
> Premise 2 says:
>
> ‘The sentence in Lecture 2 that is introduced by a star symbol is not true’ is true if and only if
>
> the sentence in Lecture 2 that is introduced by a star symbol is not true.

(See Figure 2.16.)

(P1) The sentence in Lecture 2 that is introduced by a star symbol is identical to the sentence
'The sentence in Lecture 2 that is introduced by a star symbol is not true'.

(P2) 'The sentence in Lecture 2 that is introduced by a star symbol is not true' is true if and only if
the sentence in Lecture 2 that is introduced by a star symbol is not true.

Premise 1 entails:
'The sentence in Lecture 2 that is introduced by a star symbol is not true' is true if and only if
the sentence in Lecture 2 that is introduced by a star symbol is true.

Premise 2 says:
'The sentence in Lecture 2 that is introduced by a star symbol is not true' is
if and only if
true if and only if
the sentence in Lecture 2 that is introduced by a star symbol is not true.

Figure 2.15: Relations

Putting these two equivalences together, by logical reasoning, yields the contradictory conclusion of our argument:

(Slide 104)

> (P1) The sentence in Lecture 2 that is introduced by a star symbol is identical to the sentence
>
> 'The sentence in Lecture 2 that is introduced by a star symbol is not true'.
>
> (P2) 'The sentence in Lecture 2 that is introduced by a star symbol is not true' is true if and only if
>
> the sentence in Lecture 2 that is introduced by a star symbol is not true.
>
> ---
>
> The sentence in Lecture 2 that is introduced by a star symbol is true
>
> if and only if
>
> The sentence in Lecture 2 that is introduced by a star symbol is not true.

Which we can rewrite, if we want to:

> (C) 'The sentence in Lecture 2 that is introduced by a star symbol is not true' is true
>
> if and only if
>
> 'The sentence in Lecture 2 that is introduced by a star symbol is not true' is not true.

So we are facing a paradox again, in the sense encountered already in the first lecture of this course: we are confronted with a logically valid argument the premises of which are plausible if taken by themselves but the conclusion of which is absurd.

What has gone wrong? Of course, the trouble is to do with the sentence that I, diabolically, showed you before I presented the paradoxical argument:

(Slide 105)

($*$) The sentence in Lecture 2 that is introduced by a star symbol is not true.

This sentence, the sentence after the star symbol, is talking about itself (see Figure 2.16): It says of itself that it is not true. It's like it is saying: I am not true. And then applying to that self-referential sentence something that is completely innocent in all "normal" cases – namely, the truth scheme, and the rules of logic – leads to disaster.

(∗) The sentence in Lecture 2 that is introduced by a star symbol is not true.

Figure 2.16: Self-referentiality of the (*)-sentence

Call the sentence after the star symbol 'a Liar sentence', since, talking loosely, it says of itself that it lies. Accordingly, the paradox itself is called the Liar paradox in the philosophical literature.

Remark on the Liar Paradox: Here is the fastest and simplest way of deriving a contradictory conclusion from P1 and P2:
The sentence in Lecture 2 that is introduced by a star symbol is true, if and only if, by P1, 'The sentence in Lecture 2 that is introduced by a star symbol is not true' is true, if and only if, by P2,
the sentence in Lecture 2 that is introduced by a star symbol is not true.
That is: A if and only if not A. (Where 'A' is short for 'The sentence in Lecture 2 that is introduced by a star symbol is true'.)
If you want to know more about the Liar Paradox, please take a look at
http://plato.stanford.edu/entries/liar-paradox/.

Quiz 17:
(1) Can you also derive a contradiction from the truth equivalences for these two sentences:
A: Sentence B is true.
B: Sentence A is not true.
(2) Can you also derive a contradiction from the truth equivalence for this sentence:
C: Sentence C is true.
Solution

2.11 Self-Referentiality and the Tarskian Hierarchy (12:00)

Your first, and very natural, reaction to the Liar paradox is probably: there shouldn't be sentences like that! There should not be Liar sentences. Unfortunately, it is hard to see how one could ban them from existence. Should we make it illegal to include sentences like the one after the star in one's lectures? Anyway: now it has happened, P1 is true, and there is nothing that we can do about it.

One can show that there actually are many further reasons why the existence of such self-referential sentences simply has to be acknowledged: for instance, if one does what cryptographers like to do, and what the Austrian mathematical logician Kurt Gödel famously did in the proof of his so-called Incompleteness Theorems – that is: coding sentences and other linguistic expressions by natural numbers in a way that can be programmed on a computer – then, as Gödel proved, there will always be sentences, in fact, infinitely many of them, which in some sense talk about themselves. Their existence follows simply from basic principles of arithmetic.

Remark on Gödel: More on Kurt Gödel, his two incompleteness theorems, and on self-referentiality by Gödel's diagonalization lemma can be found at:
http://plato.stanford.edu/entries/goedel/
and
http://plato.stanford.edu/entries/paradoxes-contemporary-logic/#ParDia.
But note that Gödel's results are mathematical theorems about provability which involve the construction of arithmetical sentences that express (in a sense that can be made mathematically precise) "their own unprovability"; they are *not* philosophical arguments about truth which involve the construction of semantic sentences that express "their own falsity". This being said, it is clear that Gödel's proof of his incompleteness theorems were inspired by the philosophical Liar paradox.

Finally, it is not the case either that self-referentiality just by itself is the problem in the paradox above.

Take this sentence:

(Slide 106)

(+) The sentence in Lecture 2 that is introduced by a plus sign consists of 66 signs.

If you count the signs of this plussed sentence – the sentence that starts with 'The sentence in...' and which ends with 'of 66 signs.' – you will find that it does consist of 66 signs (not counting spaces, but counting the full stop at the end). The sentence expresses an empirical truth: the sentence says about itself something that is true. It says this about a sentence, and it just happens that the sentence about which it says this is itself. No contradiction follows from the sentence; of course not: no contradiction could ever follow logically from a true sentence!

Remark on self-reference: You can read more on this remark at
http://plato.stanford.edu/entries/self-reference/.

So let us simply accept Premise 1. If it is not the bad guy, then there is just one other premise to blame: P2, which is an instance of the truth scheme:

(Slide 107)

(P2) 'The sentence in Lecture 2 that is introduced by a star symbol is not true' is true if and only if

the sentence in Lecture 2 that is introduced by a star symbol is not true.

(T) 'A' is true if and only if A.

Does that mean that our definition of truth for L_{simple} is deficient, too, since it entails all instances of the truth scheme for sentences in L_{simple}?

Not really, fortunately. In our definition from before, we defined the truth predicate 'true' for all the sentences of L_{simple}. But the truth predicate itself was not a member of the vocabulary of L_{simple}. Hence, P2 is not an instance of the truth scheme for L_{simple}.

If we had considered a language other than L_{simple}, if we had stated what might have looked like a definition of truth for that language, where the definition would have been materially adequate, where the language would have contained the truth predicate that we were aiming to define, and where the language would also have included a Liar sentence, then, by the Liar paradox, a contradiction would have followed from this purported definition of truth. Any such definition would not be formally correct then, for consistency is part of what one means by formal correctness. But none of this applies to our definition of truth for L_{simple}: that language did not include the truth predicate that we defined for it. The truth predicate, and its definition, do not belong to L_{simple} but to a different language, the language in which we stated our definition of truth for L_{simple}.

We find that anyone who is in the business of defining truth, is actually dealing with two languages at the same time:

(Slide 108/1)

- The object language: the language *for* which one defines truth.

That is: for the sentences of which one defines truth. One defines under what conditions a sentence in the object language is true.

(Slide 108/2)

- The metalanguage: the language *in* which one defines truth for the object language.

For instance: in our toy example from before, L_{simple} was the object language, and a fragment of English – augmented by a couple of mathematical symbols – was the metalanguage. All the basic expressions and sentences of the object language were also contained in the metalanguage, but not the other way around. In particular, the truth predicate that we defined for L_{simple} was not itself part of the vocabulary of L_{simple}, nor were all of the terms by which we defined truth for L_{simple}: for instance, the expression 'for all sets Y' played a major role in our definition of truth for L_{simple}, but there are no sentences in L_{simple} that contain that expression. More generally, whenever one succeeds in defining truth satisfactorily for an object language in a metalanguage in the Tarskian sense of the word, then the metalanguage will have to be more expressive than the object language: one can say something in the metalanguage – like ascribing truth to whatever sentence of the object language – that the object language itself cannot say.

This all goes back to Tarski's work on truth: he had analysed the Liar paradox in detail, and he intentionally set up his example definitions of truth so that any paradoxical conclusions were avoided from the start. For instance: if the truth predicate 'true' is defined for L_{simple}, P2 is simply not an instance of the truth scheme for the object language in question, that is, L_{simple};

(Slide 109)

> (P2) 'The sentence in Lecture 2 that is introduced by a star symbol is not true' is true if and only if
>
> the sentence in Lecture 2 that is introduced by a star symbol is not true.

in fact P2 is false then: even granted the existence of the starred sentence on my slides, the left-hand side of P2,

'The sentence in Lecture 2 that is introduced by a star symbol is not true' is true

is not the case, as no sentence involving the truth predicate 'true' is contained in the object language, while the right-hand side of P2,

the sentence in Lecture 2 that is introduced by a star symbol is not true.

is the case, because the sentence in Lecture 2 that is introduced by a star symbol is not true, as it is not even part of the object language.

In this way, any disastrous effects of the paradox are avoided. One of the premises of the Liar paradox is false, even though it might have looked right initially.

One question remains: say, we accept the existence of Liar sentences; and we manage to define the predicate 'true' for an object language in a formally correct and material adequate Tarskian manner. It follows that the object language does not itself include the predicate 'true', and the truth equivalence for the Liar sentence is not derivable from the definition, which is good. However, in fact, no sentence whatsoever that includes the defined predicate 'true' is then a member of the object language. For instance:

(Slide 110)

'Plato is a teacher of Aristotle' is a sentence of L_{simple}.

' 'Plato is a teacher of Aristotle' is true' is not a sentence of L_{simple}.

We *can* derive from our definition of truth for L_{simple}:

- 'Plato is a teacher of Aristotle' is true if and only if Plato is a teacher of Aristotle.

We *cannot* derive from our definition of truth for L_{simple}:

- ' 'Plato is a teacher of Aristotle' is true' is true if and only if 'Plato is a teacher of Aristotle' is true.

It seems that by the same move by which we avoid the truth equivalence for the Liar sentence we also throw away many plausible and unproblematic truth equivalences.

Tarski does have a reply to this.

(Slide 111)

L_{simple}: object language.

O.k., so we have defined 'true' for L_{simple} in a fragment of English plus a couple of formal symbols. Now determine precisely those symbols and grammatical rules that were needed to state that definition of 'true' – including of course the term 'true' itself – and build up a new formal language L_1 precisely from these symbols and grammatical rules.

(Slide 112)

L_{simple}: object language.

L_1: metalanguage of L_{simple} (includes 'true' for L_{simple}).

L_1 is a metalanguage of L_{simple} in which we can define the truth predicate 'true' for L_{simple}. Now apply the Tarskian method of constructing a definition of truth again, but this time for the language L_1: the truth predicate that we define for this purpose cannot be 'true' again, since that predicate already occurs in L_1; so let's use a new predicate, say 'true$_1$', instead which does not occur in L_1.

The language in which we carry out this definition is a metalanguage of L_1 and hence a metametalanguage of the original object language L_{simple}. Call that language L_2.

(Slide 113)

L_{simple}: object language.

L_1: metalanguage of L_{simple} (includes 'true' for L_{simple}).

L_2: metametalanguage of L_{simple} (includes 'true$_1$' for L_1).

And so on: iterate the procedure through L_3, L_4, ...

(Slide 114)

L_{simple}: object language.

L_1: metalanguage of L_{simple} (includes 'true' for L_{simple}).

L_2: metametalanguage of L_{simple} (includes 'true$_1$' for L_1).

$\vdots$

In this way,

(Slide 115)

- 'Plato is a teacher of Aristotle' is true if and only if Plato is a teacher of Aristotle

is derivable from the definition of truth for L_{simple} in the metalanguage L_1.

- ' 'Plato is a teacher of Aristotle' is true' is true$_1$ if and only if 'Plato is a teacher of Aristotle' is true

is derivable from the definition of truth for L_1 in the metametalanguage L_2.

- ' ' 'Plato is a teacher of Aristotle' is true' is true$_1$' is true$_2$ if and only if ' 'Plato is a teacher of Aristotle' is true' is true$_1$

is derivable from the definition of truth for L_2 in the metametametalanguage L_3.

And so on.

If we start from the language L_{simple}, this is all doable: we can construct such a hierarchy of metameta...languages and Tarskian truth definitions on top of it. And the same can be done for much more complex languages than L_{simple}.

In this way, the Tarskian method of defining truth does allow us to derive, as it were, reflective truths: sentences that say something about the truth of truth of truth of...; however, one needs to employ a new truth predicate at each step of reflection, and the more metas are accumulated, the more expressive the corresponding languages become, the more can be said in these languages: while none of these languages is expressive enough to speak about the truth or falsity of all of its own sentences, they can speak about the truth or falsity of all the sentences at the previous stages in the Tarskian hierarchy.

Quiz 18:
Does the sentence 'it is not true that Tarski is a philosopher' belong to L_{simple} or to a metalanguage of L_{simple}?
Solution

2.12 Conclusions (10:44)

What is truth? We have given an answer to this question, Alfred Tarski's answer, in the case of a simple toy language for the sentences of which we defined truth; and we managed to define it in familiar terms. In order to do so it was necessary to determine the vocabulary and grammar of that language in precise terms: in other words, the language needed to be a formal language. At first we stated recursive definitions for sentencehood and truth for that language, but then we turned these recursive definitions into more standard explicit definitions; for that purpose we employed talk about sets. With these definitions in place, we were able to prove general claims about truth, claims about infinitely many sentences, by a similar method as the one that mathematicians use when they prove statements about natural numbers. We presented the Liar paradox and explained why the definition of truth for our toy language does not fall prey to it. And we saw that by building a whole hierarchy of languages on top of our toy language, we would also be able to define reflective truth and to derive truths about truths about... about truths about sentences of the toy language. Mathematical methods have been crucial to all of that.

Let me conclude by addressing a couple of potential worries.

First of all: Our toy language is but – a toy language. What one can say in that language is enormously restricted. Why should we be interested in it at all? The answer is: ultimately we are not interested in that silly toy language. But it is a good starting point. Methodologically, the situation is not different from the one in science: say, we want to study some phenomenon; in our case: truth. At first one studies the target phenomenon in a highly restrictive setting, in which it is easy to keep certain parameters fixed and under control:

in our case, that's the toy language setting. Once one understands the phenomenon there, one starts to relax parameters: that's when we turn to more complex languages. And that is exactly what Tarski and other philosophers did: for instance, in principle, it is possible to state a Tarskian definition of truth for the language of modern physics. That's not so trivial anymore! Or for some big chunks of natural language, maybe after some cleaning up, that is, some formalization of that part of natural language. And sometimes linguists have already formalized the relevant parts of natural language in their own work, and we can simply build upon their formalizations when we want to define truth.

> Remark on natural language and formal languages: If you want to know more about how some linguistic theories treat natural languages as formal languages, check out, e.g., http://plato.stanford.edu/entries/montague-semantics/.

Secondly: You might have the nagging feeling that not much has happened really: we wanted an answer to the question 'What is truth?' and all we got was trivial truth equivalences and a definition of truth for sentences with certain expressions that showed up again on the right-hand side of that very definition. If that is on your mind, then you should go back to the beginning of this lecture and ask yourself: what kind of answer did I expect to our initial question? Reconsider: What is grandfatherhood? Well, define it in familiar terms. What is truth? Well, define it in familiar terms. That's what we did. If that is not good enough, why?

Thirdly: I should add that Tarski's work from the 1930s and 1940s is of course not the end of the story. For instance, some philosophers after Tarski thought that one should not aim to define truth explicitly at all: obviously, all definitions must come to an end somewhere, some terms at least must be taken to be primitive, undefined, and since the term 'true' seems so fundamental, why not take it to be such a primitive term? Of course, that in itself is not yet good enough: we still need to determine the meaning of 'true' somehow. But to some extent that can be done without stating a definition: for example, the membership predicate in set theory – something is a member of something – is not defined in set theory in terms of anything else, but rather one states axioms, postulates, in which the membership predicate figures. These axioms at least constrain the meaning of the membership predicate to some extent. Accordingly, one might constrain the meaning of 'true' by some axioms in which that term figures: for example, why not take some of the truth equivalences as the axioms of truth? And philosophers have worked this out in detail.

> Remark on truth equivalences as axioms of truth: This is pretty much what the deflationary theory of truth suggests that I had already mentioned before: http://plato.stanford.edu/entries/truth-deflationary/.

A different criticism of Tarski's theory concerns the branching of the concept of truth into many such concepts, which we observed at the end of this lecture when we investigated how we would be able to define truth of truths of... Each metameta... language did have its own level or type of truth predicate. Wouldn't it be possible to have something like a Tarskian theory of truth of truths of... but where there is just one truth predicate that does the job on all levels. Famously, the US-American philosopher Saul Kripke suggested a way of doing better than Tarski along these lines in a paper from 1975, and since that time various other theories of truth have been developed and defended along similar lines: but they all build on Tarski, and they all use mathematical methods. Some of these theories also suggest a more radical way out of the Liar paradox than the one that Tarski had proposed: the idea is to accept both P1 and P2 but to change the rules of logic so that the conclusion might well follow from P1 and P2 but it is not contradictory anymore in a bad sense: in a logic different from the standard so-called classical one, maybe a sentence of the form 'A if and only if not-A' is not actually problematic. That's what's so cool about philosophy: you are even allowed to question the underlying logic of your reasoning! But in many of these approaches one also softens the blow somehow: e.g., things might be set up so that one may still reason classically with normal "everyday" sentences and sentences in mathematics and science, and the logic gets changed only when one reasons with paradoxical sentences such as the Liar sentence.

Remark on theories of truth that do not invoke the object language, metalanguage, metameta... distinction: The classic reference is
Kripke, Saul, "Outline of a Theory of Truth", *Journal of Philosophy* 72 (1975), 690-716.
In that paper, Kripke develops a formal theory of truth for a language that is at the same time its own metalanguage. According to Kripke's theory, the Liar sentence turns out to be neither true nor false.
And here is an example of a formal theory of truth according to which the conclusion of the Liar paradox is not contradictory in any bad sense, since in the underlying non-classical logic of that theory, sentences of the form 'A if and only if A' are perfectly acceptable:
Field, H., *Saving Truth from Paradox*, Oxford: Oxford University Press, 2008.

A fourth point: What if your understanding of the term 'true' differs crucially from what I had presupposed in this lecture? Then clearly having Tarski's definition of truth available won't be of big help for you. Why think that Tarski has any right to say that his notion of truth is the right one? Why do I have the right to say so? In principle, obviously, I am perfectly happy with discussing other accounts of truth: make them sufficiently clear to me, so that I can understand them, and then let's talk about it. But I wonder how different your understanding of truth will be: for instance, do you accept the truth equivalences for everyday sentences, like "Snow is white' is true if and only if snow is white'? And do you understand 'if and only if' as it is understood in standard classical logic, and thus, e.g., in present-day mathematics? If so, then your view is at least in the same ballpark as Tarski's. The difference can't be that big. Now, say, you are fine with classical logic, but you do not buy the truth equivalences: your understanding of 'true' does not support the truth equivalences. Then I would still suggest that you should develop your view in all proper detail, I would only wonder if you should call it an account of truth: Why not use a different term than 'true' for the concept that you are interested in, if only to avoid possible misunderstandings with people who use the term 'true' in a way that does support the truth equivalences? And there is great potential of misunderstanding here, for many people do accept the truth equivalences: they may find them trivially right, but even more firmly they must accept them.

Which leads me to the final worry: we ended up with a definition of truth that is not much harder to get than the definition of, say, multiplication; well, not very much harder anyway. Where is the deepness of the original question gone? 'What is truth?' It's like when the magic of the infinite seemed to have disappeared at the end of the first lecture. But neither is actually the case. Remember Cantor's theory of infinities: there are sets of ever greater sizes of infinity. I mentioned at the end of the first lecture that all sets whatsoever taken together are too many to form a set; in other words: there is no universal set. Accordingly, consider a universal language: a language in which one can express everything that is meaningful at all. We found today that it is not possible to state a formally correct and materially adequate definition of truth for such a universal language: for such a definition would have to be carried out in some language again, hence that definition of truth could be formulated in some way in the universal language itself, which would lead to the contradictory conclusion of the Liar paradox. So there is no universal language for which one could state a satisfactory definition of truth in the sense of Tarski. If a natural language such as English is universal, then we cannot state a satisfactory definition of truth for it as a whole, only for fragments of it. And yet we are comfortable speaking that language. Isn't that magical enough?

Here are some monographs on truth and the Liar paradox (all of which, I should emphasize, presuppose a very thorough formal-logical training):

Barwise, J. and Etchemendy, J., *The Liar*, Oxford: Oxford University Press, 1987.

Field, H., *Saving Truth from Paradox*, Oxford: Oxford University Press, 2008.

Gupta, A. and Belnap, N., *The Revision Theory of Truth*, Cambridge: The MIT Press, 1993.

Maudlin, T., *Truth and paradox. Solving the riddles*, Oxford: Clarendon Press, 2004.

McGee, V., *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*, Indianapolis and Cambridge: Hackett Publishing, 1991.

Appendix A

Quiz Solutions Week 2: Truth

Quiz 09:

What is the truth equivalence for the sentence 'The Munich Center for Mathematical Philosophy is part of LMU Munich'?

SOLUTION Quiz 09:

The truth equivalence for the sentence 'The Munich Center for Mathematical Philosophy is part of LMU Munich' is the following sentence:

'The Munich Center for Mathematical Philosophy is part of LMU Munich' is true if and only if the Munich Center for Mathematical Philosophy is part of LMU Munich.

(Never mind about using capital letters or not using them, as in 'The' vs. 'the'. Things like that we will simply ignore. And since we are on it: We will also simplify things throughout our lectures by suppressing e.g. temporal parameters: e.g., we should actually say 'The Munich Center for Mathematical Philosophy is part of LMU Munich at time t', where t is a particular point or period of time, and similarly for many other example sentences in our leture; but for simplicity we will leave this out.)

Back to quiz

Quiz 10:

Is it possible to replace the sentence ‘For every sentence x, either x is true or the negation of x is true’ by a sentence which does *not* include the truth predicate, but which says the same as the given sentence?

SOLUTION Quiz 10:

No (or at least not easily).

You might think that ‘For every sentence x, either x or negation of x’ would do the job, but in natural language this would correspond to saying ‘For every sentence, either it or the negation of it’, which is not quite a proper English sentence (where is the predicate?). This said, certain formal logical systems of higher-order logic do allow one to say, and to derive, ‘For every proposition p, p or not p’. In such formal languages, some of the usual applications of the truth predicate can be expressed without invoking the truth predicate, which is achieved by applying universal quantification (‘for all’) over propositions (‘p’).

Another idea would be to express ‘For every sentence x, either x is true or the negation of x is true’ in terms of the scheme ‘A or not A’. However, a scheme is not itself a sentence; it is only turned into a sentence once the schematic placeholder ‘A’ is itself replaced by a concrete sentence. For that reason, the scheme ‘A or not A’ does not mean the same as the given sentence ‘For every sentence x, either x is true or the negation of x is true’. By the way: more on schemata can be found at

http://plato.stanford.edu/entries/schema/.

Back to quiz

Quiz 11:

By applying the grammatical rules of L_{simple}, show that 'Socrates is a teacher of Tarski or it is not the case that Socrates is a teacher of Tarski' is a sentence of L_{simple}.

SOLUTION Quiz 11:

By applying the rule 'If we put a name (in L_{simple}) before 'is a teacher of' and another one after it, we end up with a sentence of L_{simple}', we get: (1) 'Socrates is a teacher of Tarski' is a sentence of L_{simple}.

By applying the rule 'If we put 'it is not the case that' in front of a sentence of L_{simple}, we get a sentence of L_{simple} again', it follows from (1) that also the following holds: (2) 'it is not the case that Socrates is a teacher of Tarski' is a sentence of L_{simple}.

Finally, by applying the rule 'For every two sentences of L_{simple}, if we put an 'and' or an 'or' between them, then we get sentences of L_{simple}', we can derive from (1) and (2) that also this is the case: (3) 'Socrates is a teacher of Tarski or it is not the case that Socrates is a teacher of Tarski' is a sentence of L_{simple}.

Back to quiz

Quiz 12:

Derive the following from our definition of truth: ‘Socrates is a teacher of Tarski or it is not the case that Socrates is a teacher of Tarski’ is true if and only if Socrates is a teacher of Tarski or it is not the case that Socrates is a teacher of Tarski.

SOLUTION Quiz 12:

‘Socrates is a teacher of Tarski or it is not the case that Socrates is a teacher of Tarski’ is an or-sentence that results from putting together ‘Socrates is a teacher of Tarski’ with the logical symbol ‘or’ and with ‘it is not the case that Socrates is a teacher of Tarski’. By the part of our definition of truth that deals with ‘or’-sentences, it follows: (1) ‘Socrates is a teacher of Tarski or it is not the case that Socrates is a teacher of Tarski’ is true if and only if ‘Socrates is a teacher of Tarski’ is true or ‘it is not the case that Socrates is a teacher of Tarski’ is true.

‘Socrates is a teacher of Tarski’ is the result of putting together the name ‘Socrates’ with the predicate ‘is a teacher of’ and with the name ‘Tarski’, so by the respective part of our definition of truth it follows: (2) ‘Socrates is a teacher of Tarski’ is true if and only if Socrates is a teacher of Tarski.

Accordingly, ‘it is not the case that Socrates is a teacher of Tarski’ is the result of putting together the logical symbol ‘it is not the case that’ with ‘Socrates is a teacher of Tarski’, which is why the part of our definition of truth that deals with ‘not’-sentences entails: (3) ‘it is not the case that Socrates is a teacher of Tarski’ is true if and only if ‘Socrates is a teacher of Tarski’ is not true. By applying what we know from (2) (and negating both sides of (2)), we can reformulate the right-hand side of (3) equivalently, so that (3) becomes: (4) ‘it is not the case that Socrates is a teacher of Tarski’ is true if and only if it is not the case that Socrates is a teacher of Tarski.

Finally, using (2) and (4), we can substitute equivalents for equivalents on the right-hand side of (1), so that we get: ‘Socrates is a teacher of Tarski or it is not the case that Socrates is a teacher of Tarski’ is true if and only if Socrates is a teacher of Tarski or it is not the case that Socrates is a teacher of Tarski.

(Again, for simplicity, we never actually care about capital letters when we consider sentences of L_{simple}; we simply treat ‘It is not the case that...’ and ‘it is not the case that...’ in the same way.)

Back to quiz

Quiz 13:

Find a recursive definition of the function f on numbers $1, 2, 3, \ldots$ so that $f(n) = 1 \cdot 2 \cdot \ldots \cdot n$.

(So $f(n)$ is such that it is the product of the numbers $1, 2, 3, \ldots, n$.)

SOLUTION Quiz 13:

$f(1) = 1$.

$f(n+1) = f(n) \cdot (n+1)$.

Back to quiz

Quiz 14:

Let a function f be defined recursively on the numbers $0, 1, 2, \ldots$ as follows:

$f(0) = 1$. $f(n+1) = f(n) + 2$.

Can you find a way of defining the same function f explicitly, that is, a way of assigning the same value to '$f(n)$' but without mentioning in the definition the function f itself again?

SOLUTION Quiz 14:

Define: $f(n) = 2n + 1$.

(We will return to this example later.)

Back to quiz

Quiz 15:

In a previous quiz, a function f was defined recursively on the numbers $0, 1, 2, ...$ as follows:

$f(0) = 1$. $f(n + 1) = f(n) + 2$.

And then we determined the following equivalent explicit definition of f:

$f(n) = 2n + 1$.

But back then we did not actually *prove* the two definitions to be equivalent. Now please do prove by complete induction over natural numbers that for all n: if $f(0) = 1$ and for all m, $f(m + 1) = f(m) + 2$, then it holds that $f(n) = 2n + 1$. So the property P is in this case: n has P if and only if, if $f(0) = 1$ and for all m, $f(m + 1) = f(m) + 2$, then $f(n) = 2n + 1$.

SOLUTION Quiz 15:

First we show that 0 has the property P: Assume $f(0) = 1$ and for all m, $f(m+1) = f(m)+2$. Then it follows immediately from $f(0) = 1$ and $1 = 2 \cdot 0 + 1$ that $f(0) = 2 \cdot 0 + 1$.

Secondly, we show that if n has the property P, then also $n + 1$ has the property P. So assume n to have the property P, and assume also that $f(0) = 1$ and for all m, $f(m + 1) = f(m) + 2$.

It follows: $f(n+1) = f(n)+2 =$, since n has P by assumption, $= 2n+1+2 =$, by simple calculation, $= 2(n + 1) + 1$. Thus, $n + 1$ has property P as well.

Therefore, by the principle of complete induction over natural numbers, we have: for all n, if $f(0) = 1$ and for all m, $f(m + 1) = f(m) + 2$, then it holds that $f(n) = 2n + 1$.

Back to quiz

Quiz 16:

Fill in the missing part of the inductive proof of the theorem in the lecture as far as 'and'-sentences are concerned. That is, prove:

if (y is true or $\neg y$ is true) and (z is true or $\neg z$ is true), then $y\&z$ is true or $\neg(y\&z)$ is true.

(I write '$y\&z$' here as a shorthand for: the result of putting y together with 'and' and with z. By '$\neg$' I mean negation ('not').)

SOLUTION Quiz 16:

Assume that (y is true or $\neg y$ is true) and (z is true or $\neg z$ is true).

So there are four possible cases really:

(i) y is true, z is true: but then, since by our definition of truth it holds that $y\&z$ is true if and only if y is true and z is true, it follows: $y\&z$ is true. And by logic this entails: $y\&z$ is true or $\neg(y\&z)$ is true.

(ii) y is true, $\neg z$ is true: by our definition of truth, $\neg z$ is true if and only if z is not true; so we have: z is not true. But then, as by our definition of truth again it holds that $y\&z$ is true if and only if y is true and z is true, it follows: $y\&z$ is not true. From this we can conclude, since once again by our definition of truth it holds that $\neg(y\&z)$ is true if and only if $y\&z$ is not true: $\neg(y\&z)$ is true. And by logic this entails again: $y\&z$ is true or $\neg(y\&z)$ is true.

(iii) $\neg y$ is true, z is true: Analogously to (ii).

(iv) $\neg y$ is true, $\neg z$ is true: Analogously to (ii).

Since this covers all possible cases, and in each case we managed to show that $y\&z$ is true or $\neg(y\&z)$ is true, we can derive (independently of whatever case): $y\&z$ is true or $\neg(y\&z)$ is true. Which is what we were meant to prove.

Back to quiz

Quiz 17:

(1) Can you also derive a contradiction from the truth equivalences for these two sentences:

A: Sentence B is true.

B: Sentence A is not true.

(2) Can you also derive a contradiction from the truth equivalence for this sentence:

C: Sentence C is true.

SOLUTION Quiz 17:

Solution (1): Yes.

Sentence A is true if and only if 'Sentence B is true' is true if and only if Sentence B is true if and only if 'Sentence A is not true' is true if and only if Sentence A is not true.

Such a pair of sentences A, B is known as a Liar cycle (of length 2) in the literature. (There are also Liar Cycles of length greater than 2.)

Solution (2): No.

The only thing that one can derive is:

Sentence C is true if and only if 'Sentence C is true' is true if and only if Sentence C is true.

But that is not contradictory, it is merely trivial. Such a sentence C is called a Truth Teller in the relevant literature. The truth equivalence for a Truth Teller does not lead to a contradiction, but it does not tell us either whether the Truth Teller is true or whether it is not.

Back to quiz

Quiz 18:

Does the sentence ‘it is not true that Tarski is a philosopher’ belong to L_{simple} or to a metalanguage of L_{simple}?

SOLUTION Quiz 18:

It belongs to a metalanguage of L_{simple}, as the sentence includes the truth predicate for L_{simple}.

Back to quiz

List of Figures