The Logic of Provability, lecture 1

Joost J. Joosten

March 12, 2002

One of the most influential logicians of the previous century has been Kurt Gödel (1906-1987). In 1931 at the age of 25 he published a now famous article in which he presented his renowned incompleteness theorems. We can take his incompleteness theorems as a starting point of our course.

The title of this course is "Logical Techniques" which is rather misleading in the sense that it does not at all cover the material treated in this course. The course presented in these six weeks will have very little to do with the course as it is described in the studiegids. A better title of this course would have been something like "The logic of provability" or "Doing metamathematics in a formal setting" or as Boolos calls it "Proving the metatheory in the object theory". These latter titles all sound rather fancy but what do they actually say? Let us commit ourselves for example to explaining the last title "Proving the metatheory in the object theory". The prefix 'meta' is almost a guarantee for an exiting title and has become rather popular in the previous century. The use of the word 'meta' comes from the use of the word metaphisics. Originally this occurrence of 'meta' only meant 'after'. Alexander of Aphrodisias once made a catalogue of the works of Aristotle. A standard work on philosophy, prote filosofia, came just after a work on physics, Physics. The prote filosofia became better known as the book that comes after physics, that is, metafisica. In the previous century the prefix gained enormous popularity and has slightly been overused. The prefix meta in metatheory hints at a higher level theory, at a theory that speaks about the original theory. In a certain way we could still read it in the sense of "after" as the metatheory comes after one has gained some familiarity with the object theory itself.

So, we wanted to explain the title "Proving the metatheory in the object theory". The metatheory is here supposed to deal about the object theory. So, if T is the object theory, a metatheory U typically deals with matters like, what is the expressive power of T, what can T say and see about reality, what is it that T can not prove, is T consistent, is T efficient in proving a certain class of facts, and more like that. In some sense the metatheory U is a collection of facts about our object theory T. We talk of the object theory T as this theory is supposed to talk about the basic objects. Depending on our language and intended meaning the objects can be numbers, or polygons, or strings over a finite alphabet and so on. As the metatheory U is supposed to talk about the object theory T, in a certain sense T itself should be an object of which U can speak. Somewhat later on we will provide a proper definition of a (formal) theory. In the sense of this proper definition the U will often not be a formal theory. What we shall do is to single out a part of what is referred to as the (informal) metatheory and make that part formal. This very part is then to be proved in the object theory itself.

Here we see another problem arising. How can the object theory reason about itself? Of course this depends on the language of the theory. In our case T will deal with natural numbers. But T itself is not a natural number, nor is a proof in T. We could extend the language so that it could also talk about proofs and theories. This however does not feel very natural to do. We are interested in a theory that deals for example with natural numbers and not with some academic modification of it that also deals with proofs and theories. Moreover there is a feel that the metatheory always needs a richer language in a sense. If our object theory was to deal with numbers and proofs and theories that deal with numbers then the metatheory should deal with theories that deal with numbers and theories that deal with numbers. So, in a way, we would be in need for a universal language, a language that can unambiguously treat all possible different topics. Leibniz already dreamt of such a universal language. In a formal setting such a universal language would be a rather difficult if not impossible aim. It thus seems that we have a level shift in our languages.

The way out that was chosen by Kurt Gödel was found in coding techniques. As formulas and proofs and other syntactical objects are finite strings of symbols, they can be coded as natural numbers by means of a technique developed by Göbel. This coding can be done in such a way that the very basic facts about syntax can be proved in the object theory T. So, T itself just thinks it is proving facts about natural numbers but actually these facts *represent* facts about syntax. In this way we can at least "interpret" a part of the metatheory that is concerned with finite strings of syntax in the object theory.

We can see Gödels paper from 1931 as the starting point of our project. In this paper Gödel coded up syntax in the language of the natural numbers (actually he took a slightly different language but in this course we will present this work in a modified version) and then studied what a formal system can prove about its own metatheory. He restricted himself to the metatheory concerning statements about provability. Two famous results are stated in his incompleteness theorems the first of which states that whatever reasonable formal theory of the natural numbers you take, it is always possible to find a true sentence that is provable nor disprovable in that very theory. The second incompleteness theorem says that every reasonable arithmetical (that is, of the natural numbers) theory does not prove its own consistency.

In a very intuitive way we have thus explained the title "Proving the metatheory in the object theory". A natural question to ask is, why is this an interesting enterprise and how did interest in it arise? We will sketch five points here that could provide with an answer to these questions.

1. Our enterprise is in a certain way concerned with a study to the relation between truth and provability, that is, truth and what a formal theory can see of the truth. This very study is carried out within that very formal theory. On itself this relation is a very fundamental and philosophically intersecting one. Although one could dispute on the philosophical content of these projects when they are carried out in a formal setting. Actually this course could give us better arguments for this critique as we will see the limitations of formal theories.

2. We mentioned before the universal language that Leibniz was dreaming of. This universal language was part of a more ambitious programme. The final aim of Leibniz was to develop a computational device (he actually started by making a sort of computer (abacus) that could deal with natural numbers) that could calculate the indisputable truth. These questions were to be asked in an unambiguous all pertaining universal language. Leibniz himself had a modest first question for his computer in mind: "does God exist?". Some philosophers have related the work of Gödel to this dream of Leibniz. There is something as the Church-Turing thesis that mathematically captures the computational power of such a vague concept as a computer. (Recall that in the time this notion was introduced no computers as we know today existed.

The word computer thus meant an an abstract computing device or a concrete one such as the human being. It has been rumored that the CT- thesis has to be revised in order to adept the latest developments in the field of quantum computing.) The CT-thesis can be captured by a formal theory so that one could invoke Gödels result. This whole discussion by the way is by no means transparent I think and we will not study it in further detail.

3. By the turn of the previous century mathematics went through a foundational crisis. (In retrospective we better speak of a flourishing period.) Important 'formal' systems that were freshly introduced in order to circumvent the known paradoxes and anomalities were seen to be inconsistent such as Frege's Begriffschrift, naive set theory, mathematical analysis and so on. A theory that is inconsistent literally proves everything and hence is not interesting. The notion of consistency became thus a topical one. Also it became of interest to demonstrate consistency within a formal setting.

4. Also during this 'foundational crisis' there was debate going on about which mathematical methods were justifiable. When dealing with finite objects and constructions (whatever vague this class of objects is) there is little problem in seeing the correctness of the constructions and reasonings. Things get more unclear when infinite objects and constructions are involved. Hilbert posed a programme that should justify the use of such infinite and intuitive non-trivial objects. Hilbert's programme was designed to in the end indeed justify the use of this advanced mathematics. Therefore Hilbert proposed to isolate a formal theory F which is unproblematic and only deals with finite and intuitively clear concepts. After that, one was to consider some extension R of real mathematics that is actually used by the working mathematician. Within the finite theory F it should then be shown that whatever is provable in R is actually true. Actually Hilbert only wanted this for a certain restricted class of statements that have unproblematic mathematical content. So, F should be able to state the phrase "whatever is provable in R". This again could be done by representing proofs and syntax in the object language by means of coding techniques. Let us denote the representation in T of "φ is provable in R" by $\Box_R\varphi$. Hilbert's programme thus was to show that $F \vdash \Box_R\varphi \to \varphi$. Gödel's second incompleteness theorem says that this is impossible, for take φ to be $\bot$. Gödel 2 then says that $F \nvdash \Box_F\bot \to \bot$. One could thus

receive the work of Gödel as a death-blow to Hilbert's programme.

5. The mathematical structure that comes as the result of the study of interpretability logic is of a compelling and overwhelming beauty. This in it self is a justification, if not the most important, to study the subject.

To come back to our theme, our project will be to study a part of the metatheory within the object theory. The tasks we thus see ahead of us are among the following.

- First we should specify our domain of discourse and our object theory. We will be talking about the natural numbers and our object theory will be PA in the form that we will define later on. In order to specify a theory we should determine three characteristics of it, the language, the axioms and the rules. So we shall do.
- We should isolate the part of the metatheory that we are interested in. This will be the part concerned with proofs and provability. So, first we should recall these notions precisely and choose a format suitable to our task.
- We should represent the isolated part of the metatheory in PA by means of coding. It should be proved that the coding has some desired properties and can be easily dealt with.
- Finally we should explore the maps of meta theoretic facts that are true and of facts that are provable.

The map of meta-theoretic facts that we mentioned should include Gödel's results. So, for example $\nvdash \bot \Rightarrow \nvdash \Box\bot$ should be part of it. The complete map will be more extensive and general and will also contain principles like $(\vdash \Box(A \to B) \wedge \Box A) \to \Box B$. In exploring this map and getting a better understanding of it we will make exhaustive use of a branch of logic that is called modal logic. A study of modal logic has already been initiated by Aristotle to deal with notions (modalities) such as "possibly" and "necessary".