The importance of indirect arguments

Herman Ruge Jervell
University of Oslo

21st October 2002

1

In 1965 as a young student I read Georg Henrik von Wright : "Logik, filosofi och språk" for the first time. It was then one of my main motivation to study logic. Before these proceedings I found my old Aldus-paperback from 1965 and read it again. It is still as good as ever. I hope with the present paper to write something that could have been a chapter in the book.

2

An argument usually proceeds from assumptions to a conclusion. The argument is direct if the intermediate steps are either included in the assumptions or in the conclusion. Else it is indirect. Traditionally the indirect arguments have been connected with proofs by contradiction. This is so both by Theophrastus and Bolzano. We do not think this is important. The important thing about the indirect arguments is the liberal use of auxiliary notions and auxiliary theorems.

If we look at any mathematical text and its arguments it is striking how many auxiliary notions we need to prove what we are supposed to prove. We have lemmas as intermediate steps quite often proving things not needed in the end.

On the other hand in the explanation of what logic is about one most often talks about direct arguments. So there is a puzzle here — we use only direct arguments in our explanations but for the practical applications we resort to the indirect arguments.

3

The first impressive use of indirect arguments is in a small paper by Archimedes. In "The sand reckoner" he shows in 4 pages how to count the number of sand grains in the universe. To do that Archimedes must use some astronomy. But for our purposes here the important thing is how to count to large numbers. For Archimedes the largest concrete number was a myriad ($= 10^4$) — the number of ants we can see in an ant heap. Archimedes comes then to 10^8 — a myriad of myriads. The numbers

$$1, 2, \ldots, 10^8$$

are called the numbers of first order. The numbers

$$10^8, 2 \cdot 10^8, \ldots, 10^{16}$$

are called the numbers of second order. Going on to numbers of myriad-myriad-th order we obtain a number

$$A = 10^{8 \cdot 10^8}$$

This number will suffice. But Archimedes continues and calls the numbers up to A for the first period. A now becomes the unit of the second period. In this way he gets a myriad-myriad units of the myriad-myriad-th order of the myriad-myriad-th period which is

$$A^{10^8} = 10^{8 \cdot 10^{16}}$$

Now Archimedes ties this up with his astronomical ideas and counts the number of sand grains in the universe.

Let us now look at Archimedes argument from a logical point of view. Archimedes had to show that a gigantic number could be reached from below using concrete steps. We can think of Archimedes idea as the idea of building ladders. We start with the ordinary number series – starting with 1 and for each step we add 1. This ladder should be concrete so we should give an indication of how long the ladder is. Archimedes uses a ladder with a myriad-myriad steps. This is a ladder of order 1. The ladder of order 2 has equally many steps, but each step is a ladder of order 1. So we go on. This picture of a ladder has been reinvented a number of times, and seems to be well entrenched in our number system. (Karl Menninger: "Number words and number symbols", MIT Press 1969). The important point is that we use these auxiliary notions in our argument. The argument is indirect.

4

We have made a modern version of Archimedes argument. (For details see H.R.Jervell, W.Zhang: "Cut formulas for the Kalmar elementary formulas." Mittag-Leffler reports, 2000-01, no 33). We start with ordinary predicate logic with equality. To get the natural numbers we have one unary predicate $\mathcal{N}$, constant 0 and the unary successor function s. The natural numbers are given by the axioms

$$0 : \mathcal{N}$$

$$\forall x : \mathcal{N} . sx : \mathcal{N}$$

As a start that is all we know about them. Then there are notational axioms. Here are the ones for addition, multiplication and exponentiation:

$$+0y = y$$

$$+sxy = s + xy$$

$$\times 0y = 0$$

$$\times sxy = +y \times xy$$

$$e0y = sy$$

$$esxy = exexy$$

Note that the notational axioms does not involve $\mathcal{N}$. They are also true in the standard interpretation. We are free to introduce more notational axioms — as long as they are true and does not involve $\mathcal{N}$. Now we have the following important observation:

Observation 1 *For a closed term tmn a direct proof of $tmn : \mathcal{N}$ contains more than tmn steps.*

The only way we can give a direct proof of $tmn : \mathcal{N}$ is to use the axioms for $\mathcal{N}$ until we get $k : \mathcal{N}$ where k is the value of tmn.

The next observation is harder and we shall not go into the argument here

Observation 2 *For a closed term tmn there are indirect proofs of $tmn : \mathcal{N}$ which are linear in the parameters m, n.*

The precise notion used in the paper is to look at cut-free proofs, but any reasonable notion of direct proof will work. So what does Archimedes do? He uses the auxiliary notions connected with ladders. We do the same thing with proofs allowing cuts. This is perhaps easiest seen with the exponentiation:

$$exy = 2^x + y$$

The original ladder would be the axioms

$$0 : \mathcal{N}$$

$$\forall x : \mathcal{N}.sx : \mathcal{N}$$

A ladder using exy as steps where y is fixed for the ladder and x is the step number can be expressed by

$$\forall y : \mathcal{N}.e0y : \mathcal{N}$$

$$\forall x : \mathcal{N}.(\forall y : \mathcal{N}.exy : \mathcal{N} \rightarrow \forall y : \mathcal{N}.esxy : \mathcal{N})$$

This can be formulated closer to the original axioms by introducing a new unary predicate $\mathcal{E}$ by:

$$x : \mathcal{E} = \forall y : \mathcal{N}.exy : \mathcal{N}$$

and then observe that the formulas above can be written as

$$0 : \mathcal{E}$$

$$\forall x : \mathcal{E}.sx : \mathcal{E}$$

We call these for the inductive formulas for exponentiation. They express that we have an exponentiation ladder.

Observation 3 *We can give a direct proof of the inductive formulas for exponentiation.*

Assume $x : \mathcal{E}$. That is $\forall y : \mathcal{N}.exy : \mathcal{N}$. Now let $y : \mathcal{N}$. Then

$$exy : \mathcal{N}$$

and substituting this for y we get

$$exexy : \mathcal{N}$$

But then $esxy : \mathcal{N}$. This is for any $y : \mathcal{N}$. We get $\forall y : \mathcal{N}.esxy : \mathcal{N}$. And

$$\forall x : \mathcal{E}.sx : \mathcal{E}$$

This is the crucial argument. We can use this to show that for given numerals m, n we have proofs of length linear in m, n that

$$emn : \mathcal{N}$$

These proofs cannot be direct — a direct proof would be exponential in m. The point where it is not direct is by using the auxiliary lemma

$$\forall x : \mathcal{E}.sx : \mathcal{E}$$

and the auxiliary notion $\mathcal{E}$ — which expresses that we can climb up the numbers along an exponential ladder.

We have sketched the argument here for the exponential function, but the argument can be refined to do this for any Kalmar elementary function.

5

Let us make some simple observations about the argument in the previous section. The argument is robust. We have carried it out for classical sequential predicate calculus. It would have worked for other systems. The crucial part is to get the auxiliary notions. We have formulated this as a problem of how to get the right lemmas.

A more systematic formulation could have been to seek for the right cut formulas. We know a little about this. We have given an underlying alphabet. In our case with exponentiation it would consist of predicate calculus with $0, s, e, \mathcal{N}, =$. We can with some extra work do without equality $=$, but the rest is needed to express the problem. For the cut formulas we do not need to expand the alphabet. But we know that the cut formulas must be of a certain complexity — here measured in terms of nested quantifiers. The match up here between complexity needed and complexity used in the cut formulas are good and can be carried over to the treatment of the Kalmar elementary functions.

It is pleasing that our auxiliary notions are connected with a way of thinking about large numbers which have been used by a number of people through history. But so far we have not discussed whether there is a method of discovering the auxiliary notions. Let us say at once that we doubt that such a general method could be found. But the quest for it seems to have been going on for a long time.

6

In J.Hintikka, U.Remes:"The method of analysis" (D.Reidel 1974) the authors discuss the method of analysis. This was a method supposed to be known by the old geometry masters from antiquity — and Descartes meant that he had rediscovered. There were supposed to be two methods — the method of analysis and the method of synthesis. The latter followed the usual axiomatic method — deriving propositions from the axioms and postulates until one gets the wanted result. In the method of analysis we start with the problem, shows how to break the problem down in a systematic way and thereby have a way of getting desired auxiliary constructions.

For Descartes this was achieved by writing down the geometrical situation using what we now would call analytic geometry. It is important to remind ourselves that for Descartes the geometric situation was the primary. He worked hard to show that addition and multiplication could be defined for line segments. For the multiplication he noted that if a unit segment was given then the multiplication of two line segments resulted in a new line segment — previously one had ended up with an area.

The important point about the method of analysis is that it was supposed to be a discovery method. Using it one gets hints as to appropriate auxiliary notions for solving concretely given geometrical problems. For Descartes the method was

1. Write down the geometrical problem using equations
2. The auxiliary notions are given by subformulas of the equations
3. Do the calculations to solve the equations

Of course Descartes was too optimistic — but that is an afterthought. For Descartes and his contemporaries the method of analysis made a deep impression. Descartes himself was able to solve some outstanding problems.

Hintikka and Remes go a little further. Solving equations amounts to give solutions to existential quantifiers — and then Herbrand theory may give some extra information towards the solution or towards the appropriate auxiliary notions.

7

A better framework for describing the method of analysis is using sequential calculus with — or without — cut rule. The language for formulating the problems is predicate logic. Let us call this language for the system language. The system may or may not use extra axioms. After the work done by Negri and von Plato we know more about how to treat the cut rule in such extended systems. (Sara Negri, Jan von Plato: Structural proof theory. Cambridge University Press 2001).

Observation 4 *We can assume that the cut formulas are formulated in the system language. We do not need extra names for predicates etc. But the cut formulas may be logically complicated — in particular we can use complicated quantifier constructions.*

From our work with number systems we know much of how complicated quantifier constructions are needed for perspicuous treatment of various systems. This correspondence is to a great extent worked out. But the thing we know next to nothing about — is from a problem guess at what are appropriate cut formulas. This is *the problem of cut introduction* — and we know very little about it. For automation of proofs in computer science this is a key problem.

So we need indirect proofs — but we do not know of any reasonable method of how to introduce cuts. It may be that such a method would be unreasonable. For propositional calculus this question is connected with Cooks conjecture $P \neq NP$.

8

The big debate about the foundations of mathematics at the turn of the last century is well described in von Wrights book. Let us here go into David Hilberts proposal for a proof theory. The rough sketch would be as follows: We have two types of statements in mathematics

Real statements: These are the typical concrete statements from school mathematics.

Ideal statements: These are the more abstract statements used in higher mathematics.

To prove a real statement we often have an indirect argument involving ideal statements. We so to say go through a detour in the ideal universe in order to show properties in the real universe. Hilbert then observed that the ideal statements could be expressed in formal systems — like set theory. But the deductions in a formal system can be seen as real statements — as long as we only consider them as syntactical objects and not as interpreted objects.

Hilbert called this theory for proof theory as a parallel to number theory. In number theory we consider the data structure given by a start — zero — and a successor operation — $x \mapsto x + 1$. In proof theory there are more successors — the axioms — and the successor operations are given by the rules of the formal system. This makes for more book keeping in proof theory, but the development is perhaps not that different from number theory.

Hilbert introduced his ϵ-calculus as a way of explaining the insights behind proof theory. We start with the language of propositional calculus with free variables, predicates, equality and ϵ-terms

$$\epsilon x.Fx$$

which is read as a description — the x such that Fx. In the formal systems this is of course not interpreted, but taken as a new syntactical construction with new ϵ-axioms

$$Ft \rightarrow F(\epsilon x.Fx)$$

It was then shown that a derivation of an ϵ-free statement involving the extra axioms could be transformed into a derivation not involving ϵ-axioms.

Hilbert — or rather Hilbert students — did this for the ϵ-calculus variant of predicate calculus. They had hopes for doing it for much stronger systems. The ϵ-elimination theorem gave a formulation of the goal of proof theory. As we now know these hopes were dashed in 1930-31 with Kurt Gödels incompleteness theorem. But for our story we note that the indirect arguments pop up again — and also the hope that all arguments could be made in a direct way.

9

Let us look at the history of logic from the direct/indirect proof distinction. It is common to indicate two sources for logic

Algebra: The logical connectives are connected with addition and multiplication in algebra.

Proofs: Logic can be used to formalize actual mathematical proofs.

These sources give two different traditions in logic:

Algebra: George Boole, Ernst Schröder, Thoralf Skolem

Proofs: Gottlob Frege, Bertrand Russell, David Hilbert

For the two traditions we observe

Observation 5 *In the* algebra-*tradition the explanation of a logical expression is given by an explanation of its parts. This gives an emphasis to direct arguments. In the* proofs-*tradition we are forced to consider indirect arguments — a number of important mathematical arguments are indirect.*

The two traditions are brought together in one of the most important papers in logic — Gerhard Gentzen: Untersuchungen über das logische Schliessen. 1935.

Gentzen starts the paper by saying that the formalization of logical deduction as given by Frege, Russell and Hilbert is too far removed from the forms of deduction used in practice in mathematical proofs. So he gives a better way of formalizing logical deduction — by introducing his system of natural deduction. For our story here the important thing is the sequential calculus that he introduces. Gentzen shows that ordinary arguments can be formalized in the system as long as we use an extra rule — the cut rule. This rule is close to the traditional rule modus ponens

$$\vdash F \;,\; \vdash F \to G \;\Rightarrow \vdash G$$

The sequential calculus without cut rule is close to the *algebra*-tradition in logic. And Gentzen isolates the Hauptsatz which says that the cut rule can be eliminated and connects this with the elimination of indirect proofs. We get further results by estimating the growth of the length of the proofs when we eliminate cuts.

10

We have seen that David Hilbert tied his programme of proof theory with a particular kind of elimination of indirect arguments. Kurt Gödel showed with his incompleteness theorem that the argument for the elimination of indirect arguments in a system could not be carried out within the system — provided the system is rich enough to express a certain minimum of mathematics. Now we observe that for ordinary systems

Observation 6 *The system of direct proofs is consistent.*

There are no direct proof of a false statement. Let us now use provability predicates — $Prov(\ulcorner F \urcorner)$ "there is a proof of F", $DirProv(\ulcorner F \urcorner)$ "there is a direct proof of F

Observation 7 *We get for ordinary systems*

1. $Prov(\ulcorner F \urcorner) \wedge Prov(\ulcorner F \to G \urcorner) \to Prov(\ulcorner G \urcorner)$
2. $\neg DirProv(\ulcorner 0 = 1 \urcorner)$

and neither of those are provable for the other provability predicate.

Gerhard Gentzen went further in his investigation of direct proofs. He showed that a certain combinatorial statement $\mathcal{E}_0$ such that for elementary arithmetic (Peano arithmetic or Heyting arithmetic) we have that the following is equivalent

1. $\neg Prov(\lceil 0 = 1 \rceil)$
2. $\mathcal{E}_0 \wedge \neg DirProv(\lceil 0 = 1 \rceil)$

11

We leave the story after Gentzens work — but there is much more to be said. But to do that we need to more details and more work.