DAG PRAWITZ

REMARKS ON SOME APPROACHES TO THE CONCEPT OF LOGICAL CONSEQUENCE

A foundation of mathematics, I understand as something different from just a systematic exposition of mathematical theories and also as something more than just a philosophical discussion of the nature of that subject. Rather, it has to combine these two things by showing in accordance with some philosophical principles how mathematics can be built up systematically.

The concept of logical consequence must be of importance in any such foundation. The attention it has received in foundational discussions has varied, however. While it was of course a fundamental concept in a foundational scheme such as logicism, its relevance was sometimes denied by intuitionists. But then the notion of proof was instead fundamental, and logical consequence comes out at least as an important derivative notion. It is clear that any fairly complete foundations of mathematics has to account for its partly deductive character and hence, somehow, for the concept of logical consequence.

A discussion of different approaches to this concept may therefore be appropriate in a workshop "The Present State of the Problem of Foundations of Mathematics". I shall discuss three main approaches, the classical model theoretical approach due to Bolzano–Tarski, a conception of logical consequence as a relation between rules, which goes back to Lorenzen, and a proof-theoretical/intuitionistic approach, which goes back to Gentzen.*

I. THE EXPLICATION BY BOLZANO AND TARSKI

The well-known explication of the concept of logical consequence that was proposed by Bolzano and taken up again by Tarski [14] is based upon a distinction between logical and nonlogical or descriptive constants, although, curiously enough, both Bolzano and Tarski were unsure about how to draw the line between them and doubted that it was an essential matter how it was drawn.

Let $A(c_1, \ldots, c_n)$ and $B(c_1, \ldots, c_n)$ be sentences where $c_1, \ldots, c_n$ stand for the nonlogical constants in A and B, and let $A(v_1, \ldots, v_n)$ and

Synthese **62** (1985) 153–171. 0039–7857/85.10

$B(v_1, \ldots, v_n)$ be the corresponding open formulas obtained by replacing the constants c_i by variables v_i of appropriate kind. Then, as we all know, the explication is in Tarski's words that $B(c_1, \ldots, c_n)$ is a logical consequence of $A(c_1, \ldots, c_n)$ if and only if

(1) every assignment or model satisfying $A(v_1, \ldots, v_n)$ also satisfies $B(v_1, \ldots, v_n)$.

In the paper [14] where Tarski presented his analysis, he first criticized, essentially on basis of Gödel's incompleteness theorem, the then perhaps rather widespread view that logical consequence could be identified with derivability in an appropriate logical calculus. I think that there is no doubt that both this criticism and the material equivalence asserted between logical consequence and (1) are correct, and that, almost regardless of foundational view-point, the Tarskian analysis must be accepted as far as it goes.

However, it is plain that the analysis does not go very far. If we make the domain of the quantifiers explicit as Bolzano did, (1) can be written

(2) $(\forall v_1 \in D_1) \ldots (\forall v_n \in D_n)(A(v_1, \ldots, v_n) \rightarrow B(v_1, \ldots, v_n))$ is true regardless of how independent domains are chosen

(where an example may clarify the notion 'independent domain': in a second order formula, the individual domain D is independent, while the domain of the 1-place predicate variables is not independent, since it is thought of as fixed by D). The content of the explication is thus that a sentence S is logically true if the logical sentence got by taking the closure of the formula arising from S by replacing nonlogical constants by variables is true (regardless of choice of independent domains). But the interesting question in this context must instead be what it means that a logical sentence is true, i.e., what is meant by (2). If A and B are sentences (e.g., in predicative second order logic) that are already without descriptive constants, then the explication says only that B is a logical consequence of A (or that B is logically true) if and only if $A \rightarrow B$ is true (respectively, B is true). A variation of descriptive constants, which is the whole point in the idea of Bolzano and Tarski, cannot be a main step in the analysis of logical consequence (and logical truth) – all the more so when not even any importance is attached to the distinction between logical and nonlogical constants.

Tarski also has an analysis of truth, to be sure, and hence, of what is meant by (2). But that analysis makes no distinction between logical sentences (containing only logical constants) and factual sentences (containing also descriptive constants). The effect is that a logical sentence is understood as logically true just in case it is true in the same sense as factual sentences are true. In other words, no analysis is made of the necessity involved in logical truth – not to mention that no answers are attempted to questions like what is the *ground* for a universal truth like (2), or how can we come to *know*, even with certainty, that a logical sentence is true in all domains.

It is plain that if we want to cast some light upon such questions about the necessity of logical truths, the grounds for them, and how we can come to know them, and in this way cast some light upon the demonstrative element of mathematics, then we must somehow bring in the notions of inference rule and proof – notions that were always in the centre of logicians interests up to the time of Tarski and model theory. At the same time, we must remember Tarski's justified criticism of the identification of logical truth with provability in some given logical calculus.

II. LOGICAL CONSEQUENCE AS A RELATION BETWEEN RULES

The view that logical consequence is a relation between rules has been advocated by Lorenzen and several followers. Lorenzen illustrates the nature of this relation by the example that in chess, the possibility of moving the horse four squares in any direction by two consecutive moves is a consequence of the rules of the game in the same way as the implication $A \to C$, which can be understood as a rule, is a consequence of $A \to B$ and $B \to C$.

This idea can of course be developed in different ways. I shall consider one such way, which seems to have some initial plausibility, and which was stimulated by reading a recent dissertation by Schroeder-Heister [13], but which is not identical with the way followed by Lorenzen nor with the different way chosen by Schroeder-Heister.

Let us consider a language of iterated rules built up from atomic formulas with the help of $\to$ and universal quantification. We write for any $m, n = 0, 1, \ldots$ and distinct variables $x_1, x_2, \ldots, x_n$

(*a*) $\quad (A_1, A_2, \ldots, A_m \xrightarrow[x_1 \ldots x_n]{} B)$

for the rule to pass to B^σ from $A_1^\sigma, A_2^\sigma, \ldots, A_m^\sigma$, where σ is any substitution for the variables $x_1, x_2, \ldots, x_n$ (i.e., A_i^σ and B^σ are the results of replacing all free occurrences of the variables $x_1, x_2, \ldots, x_n$ in A_i and B by appropriate expressions assigned to the variables by σ with usual precautions to avoid conflicts between free and bound variables – the variables below the arrow being understood as indication of a universal quantification binding the variables in question). We allow iterations of this way of writing so that A_i and B need not be atomic formulas but may already be of the form (a) themselves. In this way, we get rules of higher order allowing us to pass from one rule to another.

We want to say what it means that a rule A follows logically from a finite set of rules Γ, in symbols

$$\Gamma \Rightarrow A,$$

and we do this inductively by saying what it means that A follows logically from Γ in k steps, in symbols

$$\Gamma \overset{k}{\Rightarrow} A$$

(I) For atomic formulas A:

(1) $\Gamma \overset{k}{\Rightarrow} A$ if and only if either $A \in \Gamma$, or

(2) there is a rule of the form $(A_1, A_2, \ldots, A_m \underset{x_1 \ldots x_n}{\longrightarrow} B)$ in Γ and a substitution σ such that for each $i \leq m$, $\Gamma \overset{k'}{\Rightarrow} A_i^\sigma$ for some $k' < k$, and $\Gamma, B^\sigma \overset{k'}{\Rightarrow} A$ for some $k' < k$.

(II) For nonatomic formulas A of the form $(A_1, A_2, \ldots, A_m \underset{x_1 \ldots x_n}{\longrightarrow} B)$:

(3) $\Gamma \overset{k}{\Rightarrow} A$ if and only if $\Gamma, A_1', A_2', \ldots, A_n' \overset{k}{\Rightarrow} B'$, where A_1' and B' are like A_i and B except for possibly containing distinct free variables $y_1, y_2, \ldots, y_n$ not occurring free in Γ in the place of $x_1, x_2, \ldots, x_n$.

The last clause (II) is a deviation from Lorenzen and follows instead Schroeder-Heister. It expresses the idea that a rule R follows from a set of rules Γ if and only if the conclusion of R follows from the rules occurring as premisses of R together with the rules of Γ. When R is just an atomic formula A, clauses (1) and (2) agree with the usual definition of what it means that a formula is derivable by means of a set of inference rules except that one may wonder why in clause (2), it is not simply required that $B^\sigma = A$ instead of $\Gamma, B^\sigma \overset{k'}{\Rightarrow} A$; indeed, if B^σ is atomic, this comes to the same thing, but we have to remember that B may now be nonatomic. (We could instead have defined $\Gamma \overset{k}{\Rightarrow} A$ by induction as follows:

(1′) $\Gamma \overset{k}{\Rightarrow} A$, if $A \in \Gamma$;

(2′) $\Gamma \overset{k+1}{\Longrightarrow} A$, if there is a rule R of the form $(A_1, A_2, \ldots, A_m \underset{x_1 \ldots x_n}{\longrightarrow} B)$ such that $\Gamma \overset{k'}{\Rightarrow} R$ for some $k' < k$, and a substitution σ such that for each i it holds that $\Gamma \overset{k'}{\Rightarrow} A_i$ and such that $B^\sigma = A$.

(3′) like II except that "if and only if" is replaced by "if".)

To make this language as expressible as possible, we may admit quantification in expressions of the form (a) not only for individual variables for which individual terms may be substituted but also for n-ary predicate variables for which formulas with n free variables may be substituted. We may then express e.g., the principle of induction as the rule

$$X0, (Nx, Xx \underset{x}{\rightarrow} Xx') \underset{X}{\rightarrow} (Nx \underset{x}{\rightarrow} Xx).$$

In this language, we may now define the usual logical constants. For instance, $A \vee B$ may be understood as the rule

$$(A \rightarrow X), (B \rightarrow X) \underset{X}{\rightarrow} X$$

and $\exists x A(x)$ as the rule

$$(A \underset{x}{\rightarrow} X) \underset{X}{\rightarrow} X$$

It could now be claimed that all mathematical statements can be understood as rules in this way, and it could be thought that therefore we have an analysis of logical consequence in terms of the relation

$\Gamma \Rightarrow A$ by which is to be meant that $\Gamma \overset{k}{\Rightarrow} A$ for some k as defined above. In other words, a sentence B would be said to be a logical consequence of a sentence A if and only if $A^* \Rightarrow B^*$ holds where A^* and B^* are the translations of A and B into rules. One could say that the meaning of A and B is made fully explicit by this translation into A^* and B^*, and B is then understood as following from A, when B follows from A on the basis of this meaning, i.e., when B^* follows from A^*. The latter is in turn not analyzed by presupposing the validity of any rules. On the contrary, one just explicates what it means that a rule follows from other rules in a seemingly straightforward manner on the basis of what is meant by a rule.

Since a rule R may occur in two contexts in connection with the relation $\Rightarrow$ – on the one hand, R may be asserted to hold (or to follow from a set Γ of rules), and on the other hand, applications of R (possibly together with a set Γ of other rules) may be asserted to yield a certain result – one must require in order that the analysis is to be at all plausible that there is a *harmony*, to use a term by Dummett, between the uses of R in these two contexts, i.e., one must require that if R holds (or follows from Γ) and applications of R (together with Γ) yield C, then also C holds (or follows from Γ, respectively). This is to require that a Gentzen Hauptsatz holds for $\Rightarrow$, i.e., that

$$\Gamma \Rightarrow R \quad \text{and} \quad R, \Gamma \Rightarrow C \text{ imply } \Gamma \Rightarrow C.$$

(It comes to the same thing to require that the two inductive definitions of $\overset{k}{\Rightarrow}$ accomplished by (I) and (II) and by (1′)–(3′) are equivalent.) This requirement is in fact satisfied as we can see by adapting results already established in the literature.

The analysis of logical consequence proposed here may seem quite attractive at first sight: there is no problem about how we get to know a logical consequence understood in this way, and the necessity involved in logical consequence comes out clearly – the logical consequences of given premisses are forced on us as what is just obtained by applying rules expressed in the assumed premisses.

Nevertheless there is a decisive objection to this analysis. Let Γ be the second order Peano axioms expressed as rules and let G be the Gödel sentence corresponding to Γ and the definition of $\Rightarrow$. Then, $\Gamma \Rightarrow G$ does not hold, but by simply extending the language considered above, now allowing also predicates that takes predicates as arguments and quantifications over such predicates, but otherwise leaving the definition as it were, we get a new relation $\acute{\Rightarrow}$ such that $\Gamma \acute{\Rightarrow} G$. Since $\acute{\Rightarrow}$ is defined just in the same way as $\Rightarrow$ except that a richer language is used, it must be said that G is really a logical consequence of Γ according to the idea underlying the proposed explication – only, it did not come out as such in the definition of $\Rightarrow$ because of having restricted the language. As we know, no matter how we extend the language, there will always be sentences G such that G does not come out as a logical consequence when this concept is defined as above, and such that, nevertheless, G must be admitted to follow logically from Γ in accordance with the general idea of the definition, since by a still further extension of the language, G does turn out to be a logical consequence of Γ in the sense defined. Of course, this is essentially Tarski's point in his criticism of certain approaches to the concept of logical consequence. I am not implying that the general idea of Lorenzen and Schroeder-Heister should not be viable, only that this particular way of approaching the concept of logical consequence (based on their ideas but not identical to anything proposed by them) is not possible.

III. A PROOF-THEORETICAL APPROACH

The definition of the consequence relation $\Rightarrow$ in the section above amounts formally, as you may have noted, to a variant of Gentzen's calculus of sequents (or his system of natural deduction, if the clauses (1′)–(3′) are used instead of (I) and (II)). But Gentzen cannot be said to have proposed his system in this way as an analysis of the concept of logical consequence. Instead, in his early work [3] where he presents a system of natural deduction and a calculus of sequents for the first time, there is implicit a quite different approach to logical consequence, or, more general, to a theory about the meaning of logically compound sentences and to a justification of deduction.

1. *Introductory Explanations and Examples*

The main idea behind Gentzen's system of natural deduction, as I have understood it, is an analysis of inferences as in the end consisting in applications of introduction and elimination rules, where the elimination rules are in a sense justified by the introduction rules, which are understood as conferring a meaning on the logical constants by stating what forms proofs of different sentences are to have (and which, in their turn, are therefore self-justifying). The way in which the elimination rules are justified by the introduction rules can be seen to be what makes possible normalizations of proofs in natural deduction or, equivalently, elimination of cuts in the calculus of sequents as established in Gentzen's Hauptsatz.

As an illustration, let us recall Gentzen's introduction and elimination rules for existential quantification, which may be stated as follows:

$$(\exists\mathrm{I}) \quad \frac{A(t)}{\exists x A(x)} \qquad (\exists\mathrm{E}) \quad \frac{\exists x A(x) \qquad \begin{matrix}(A(a))\\ B\end{matrix}}{B}\,(a).$$

In applications of the rule for ∃-elimination (the ∃E-rule), assumptions $A(a)$ that the second premiss B depends on may be discharged (so that the B occurring as conclusion is independent of these assumptions) subject to usual restrictions on the so-called proper parameter a, which must not occur in B or in assumptions other than $A(a)$ that the second premiss depends on.

It is now possible to justify the ∃E-rule by showing how, given derivations of the premisses, the conclusion B in any application of ∃E can be derived without this application of ∃E provided that the major premiss $\exists x A(x)$ has been derived in accordance with the rule for ∃-introduction (the ∃I-rule). Indeed, given an application of the ∃E-rule in which the first premiss $\exists x A(x)$ has been inferred in the way stipulated by the ∃I-rule, i.e. from an instance $A(t)$, and which may therefore be written

$$\frac{\dfrac{\begin{matrix}\mathscr{D}_1\\ A(t)\end{matrix}}{\exists x A(x)} \qquad \begin{matrix}[A(a)]\\ \mathscr{D}_2(a)\\ B\end{matrix}}{B},$$

we proceed as follows to obtain the desired derivation of B:

(i) take the derivation $\mathscr{D}_2(a)$ of the second premiss B;

(ii) substitute the term t for the parameter a in $\mathscr{D}_2(a)$, which, because of the restriction on the proper parameter a, does not affect B or assumptions that B depends on other than $A(a)$; and then

(iii) substitute the derivation $\mathscr{D}_1$ of $A(t)$ for each assumption $A(t)$ in

$\mathscr{D}_2(t)$ that (before the substitution) was discharged by the application of the $\exists$E-rule.
The result obtained by this operation may be written

$$\begin{array}{c} \mathscr{D}_1 \\ [A(t)] \\ \mathscr{D}_2(t) \\ B \end{array}$$

The operation described above is one of the operations by which a derivation is reduced to normal form (as defined in [7]) (and which occurs also in a more disguised form as a step in the elimination of cuts in the proof of Gentzen's Hauptsatz), but what interests me here is how this operation can be seen as justifying the $\exists$E-rule.

2. *Validity of Arguments*

To give some substance and precision to the ideas that certain inference rules, the introduction rules, are self-justifying – namely, in virtue of laying down what is to be reckoned as proofs of the conclusion, and thereby determining the meaning of the conclusion – and that the other rules are justified in terms of this meaning, one has at least to be able to explain in terms of convincing semantical notions derived in the way suggested what a proof is and what is meant by saying that an inference or inference rule is valid.

The way in which I have tried to develop these general ideas is by defining a notion of *validity* for derivations (in [8], pp. 284–289) and later, more generally, for arguments (in [9]). By a *derivation*, I understand something that proceeds according to inference rules in a given formal system, while here I reserve the term *proof* for the intuitive notion of what establishes a proposition – the question of the adequacy of a formal system is then, among other things, the question whether the derivations in the system really represent proofs. From a proof we may abstract its *argument structure* by which I mean the way the sentences involved in the proof are linked to each other in the sense that one sentence is inferred from some other sentences, i.e., asserted on the grounds of these other sentences. By an *argument* I just understand such a structure (regardless of whether it has been abstracted from a proof or not).

An argument is thus an arbitrary collection of linked inferences preferably arranged in tree form to make the links explicit: it is to be clear stepwise what is asserted, whether it is asserted as something holding or as an assumption; in the former case, it is to be clear whether it is asserted on the grounds of something else, and if so, what the grounds are; finally, it is assumed that only one sentence does not serve as the ground for another sentence, and the argument is said to be an argument *for* that final sentence (for a more exact definition, see [9], p. 229 – what is here called an argument is there called an argument-schema).

Some of the inferences in an argument may *bind* (discharge dependency of) assumptions, and some may be associated with proper parameters (subject to the usual restrictions that they are not to occur in the conclusion of the inference or in assumptions that the conclusion depends on), which are also said to be *bound* by the inference – both phenomena are exemplified by the $\exists$E-rule above. If all assumptions and parameter occurrences in an argument are bound by some inference, the argument is said to be *closed*; otherwise, it is said to be *open*. An open argument is to be understood as a schema for obtaining closed arguments in the sense that by replacing free (not bound) occurrences of parameters by terms and free assumptions by closed arguments, one obtains a closed argument.

An argument may be valid or invalid (while it is usually taken as a part of the meaning of the word proof that a proof is correct: if what was thought to be a proof turns out to be incorrect, it was not a proof). Obviously, to know only a closed argument for a sentence is not enough for having a proof of the sentence. What more do we then need to know in order to be in possession of a proof? In other words, the question is what do we need to know that an argument is valid.

Viewing an open argument as an argument-schema, the question of its validity should be reduced to that of a closed argument: an open argument is to be valid just in case all its closed instances obtained by substituting valid closed argument for its free assumptions are valid.

To answer the question what makes a closed argument valid, I make two assumptions: firstly, for each form of sentence occurring in the argument, I suppose that there is specified an *introduction rule* for that form of sentence, and secondly, for each inference in the argument that is not an application of an introduction rule, I suppose that there is associated a *justifying operation* of the kind exemplified in the case of the ∃E-rule above, which transforms certain arguments to another argument for the same formula depending on not more assumptions than the transformed one (for a more precise definition, see [9], p. 234, (a)–(c)). The general idea is as before that an introduction rule for a certain form of sentences determines the meaning of sentences of that form, or, more precisely, determines how the meaning of the sentences depends on the meaning of their constituents. The significance of first singling out certain inferences in an argument as applications of introduction rules and then assigning justifying operations to the other inferences, henceforth called *elimination inferences*, is furthermore to specify how these inferences are to be taken, i.e., on what grounds they are understood to be correct. These general remarks require some further elucidations, which I shall try to give before defining validity of arguments in 2.3 below.

2.1. *Canonical forms of arguments*. Strictly speaking, it is not correct to say that an introduction rule states the condition for how to prove sentences of the form in question: a proof may of course also be obtained by application of elimination rules. Rather, the idea is that by stating introduction rules, one is laying down the *canonical forms*[1] of arguments for the sentences in question, and that by doing so one determines the meaning of the sentences in terms of the meaning of their constituents.

To say that certain arguments or argument forms are canonical implies two things: firstly, that these forms *are* correct forms of arguments for the sentences in question, and secondly, that any correct closed argument for the sentences in question *could* always be given in these forms. For instance, if somebody asks why the rule for ∧-introduction

$$\frac{A \quad B}{A \wedge B}$$

is a correct inference rule, one can answer only that this is just part of the meaning of conjunction: the meaning is determined partly by laying down that a conjunction is proved by proving both conjuncts, and partly by the understanding that a proof of a conjunction could always be given in that way.

Another way of expressing the same thing is to say that

(i) an argument in canonical form is valid if (and only if) its immediate subarguments (i.e., the arguments for the premisses of the last inference of the argument) are valid, and that

(ii) if a sentence is provable at all, its proof must be capable of being written in canonical form; in other words, an argument not in canonical form can be valid only if it could be written in canonical form. (Compare the canonical forms of natural numbers: it is understood without further ado that any expression in that form denotes a natural number; and secondly, every natural number can always be given in that form.)

2.2. *The justification of elimination rules.* What has been said so far leaves open when an argument not in canonical form is to be reckoned as valid. Clause (ii) above gives only a necessary condition since it can clearly not be sufficient that there merely exists an argument satisfying clause (i) – to be valid, the given argument must establish just this existence. The operations assigned to elimination inferences or more generally to elimination rules are supposed to show just this, i.e., how arguments not in canonical form can be written canonically.

Given the canonical arguments, I thus propose to answer the question of what we have to know in addition to a closed argument in order to be in possession of a proof by saying that we must know some procedures attached to the elimination inferences – what I have called justifying procedures – and must know that by carrying out these procedures the given closed argument is transformed to a valid canonical argument. Given the forms of the canonical arguments, we can thus say what it means that an argument *together* with justifying operations is valid. In sections 4 and 5, I shall return to a discussion of this way of analyzing validity of arguments.

2.3. *Definitions.* We may now just sum up what has been said above by the following definition of what it means that an argument $\mathscr{D}$ together with an assignment of justifying operations $\mathscr{J}$ is valid; I shall express this by saying either that $(\mathscr{D}, \mathscr{J})$ is valid or that $\mathscr{D}$ is valid with respect to $\mathscr{J}$.

When $\mathscr{D}$ is a *closed argument,* $(\mathscr{D}, \mathscr{J})$ is *valid* if and only if either

(i) $\mathscr{D}$ is in canonical form and each immediate subargument $\mathscr{D}'$ of $\mathscr{D}$ is valid with respect to $\mathscr{J}$, or

(ii) $\mathscr{D}$ is not in canonical form, but by successively applying the operations in $\mathscr{J}$, $\mathscr{D}$ is transformed to an argument for which (i) holds.

When $\mathscr{D}$ is an *open argument*, $(\mathscr{D}, \mathscr{J})$ is *valid* if and only if all closed instances $\mathscr{D}'$ of $\mathscr{D}$ that are obtained by substituting for free parameters closed terms and for free assumptions closed arguments for the assumptions valid with respect to an extension $\mathscr{J}'$ of $\mathscr{J}$, are valid with respect to $\mathscr{J}'$.

Note that the immediate subarguments of a closed argument may be open, namely if the last inference binds assumptions or parameters. Provided that the premisses of an introduction and the assumptions bound by it (if any) are always of lower complexity than that of its conclusion, the definitions above can be understood as proceeding by a simultaneous induction.

To get a notion of logical validity, we may relativize the notion of validity to a system of canonical arguments for atomic sentences, and say for a closed argument $\mathscr{D}$ that $(\mathscr{D}, \mathscr{J})$ is *logically valid* when it is valid relative to each system of canonical arguments for atomic formulas. The definition for an open argument $\mathscr{D}$ is similar: using the notations above, $(\mathscr{D}, \mathscr{J})$ is logically valid when for each system S of canonical arguments for atomic sentences, all $(\mathscr{D}', \mathscr{J}')$ are valid relative to S where the substituted arguments for free assumptions are now to be valid with respect to $\mathscr{J}'$ and relative to S. When $(\mathscr{D}, \mathscr{J})$ is logically valid, its validity thus depends only on the content of the logical constants determined by the introduction rules for compound sentences.

We may also say that an argument is (logically) valid when it is (logically) valid with respect to some justifying procedures.

An *inference rule* may be said to be *valid* when each application of it preserves validity of arguments. An introduction rule is then trivially valid (in view of clause (i) above), which is as it should be, if they are thought of as producing canonical forms of arguments. An elimination rule R is valid depending on whether there exists a justifying operation ϕ such that if $\mathscr{D}$ is any argument whose last inference is R and whose immediate subarguments are valid with respect to the justifying procedures $\mathscr{J}$, then $\mathscr{D}$ is also valid with respect to $\mathscr{J} \cup \{\phi\}$. If ϕ is independent of the system of canonical arguments for atomic formulas, R may be said to be logically valid.

3. *Logical Consequence*

Let us now consider the proposal (made in [10]) that a sentence A is said to be a *logical consequence* of a finite set Γ of sentences when there exists a logically valid argument for A from Γ (i.e., all its free assumptions belong to Γ), which is easily seen to be equivalent to the existence of an operation ϕ such that, independently of the system S of canonical arguments for atomic formulas, ϕ applied to arguments for the sentences in Γ valid relative to S yields an argument for A valid relative to S. This is again equivalent to saying that A is a logical consequence of $A_1, A_2, \ldots, A_n$ if and only if the one step argument

$$\frac{A_1 A_2 \cdots A_n}{A}$$

is logically valid with respect to some justifying procedure.

In contrast to the notion considered in section II, such a notion of logical consequence does not amount to derivability in some given formal system. A valid argument for a sentence B from a sentence A is not required to proceed in any given language or according to some given inference rules. All that is required is that the form of sentences used in the argument have been given a meaning by the specification of canonical arguments for them, and that the inferences in the argument that are not canonical have been assigned justifying procedures allowing us to transform the whole argument to canonical form when a valid argument has been substituted for A. This means e.g., that if G is a Gödel sentence in a formalization of Peano arithmetic with the axioms Γ for which we can see intuitively that G follows from Γ, then, provided that it can be seen that G follows from Γ with the help of a language that can be analyzed in the way proposed here, there is a logically valid argument for G from Γ, i.e., G is a logical consequence of Γ in the sense proposed here. This also means that when we extend a language L to L', we cannot in general expect L' to be a conservative extension of L, i.e., there may be valid arguments for sentences in L that are formulated in L' but that cannot be given in L.

4. *The Meaning of a Sentence*

How reasonable is it to think that the form of the canonical arguments for a sentence determines the meaning of the sentence? A satisfactory analysis of the concept of logical consequence must be related to other semantical notions such as meaning and truth.

In several works (e.g., [1] and [2]), Dummett has discussed the possibility that there are some central features of the use of a sentence that determine all other features of the use of that sentence in a *uniform manner* – to know these central features is then the only particular information about the sentence that is needed to master the sentence, and it can then be argued that such a feature or such features represent the meaning of the sentence. In particular, Dummett has considered the possibility that the condition under which a sentence is *correctly asserted* is such a central feature. In mathematics, it is quite uncontroversial that the condition for correctly asserting a sentence is to be in possession of a proof of it. (Of course, one also asserts sentences on the authority of others – what we are here concerned with is thus the conditions under which we, collectively, are entitled to assert a sentence.) If, on the other hand, what constitutes a proof depends on the meaning of the sentences involved in the proof, and, as is often the case, they are of greater complexity than that of the sentence proved, we get into a vicious circle that seems to endanger the project of taking the correctness of assertions as a central feature in a theory of meaning.

The notion of canonical argument offers hope of getting out of this circle, given that the canonical arguments for a sentence A can be specified in terms of the constituents of A. The possession of a closed valid argument in canonical form is of course not a necessary condition for asserting a sentence – it is quite sufficient that we know a method for finding one (cf. Prawitz [11], pp. 21–22 and 26–27), or, what comes to the same thing, that we know a closed valid argument (not necessarily in canonical form) for the sentence as defined above.

What we need is thus an explanation of what a proof of a sentence A is that does not depend on knowing what a proof is for all the sentences involved in the proof of A but only on knowing what a proof is for the constituents of A.

Such an explanation is now forthcoming, provided that knowledge of a valid closed argument – i.e., to have a closed valid argument $\mathscr{D}$ and a set of justifying procedures $\mathscr{J}$, and, in addition, to know that $(\mathscr{D}, \mathscr{J})$ is valid – is the right analysis of what it is to be in possession of a proof. The condition for asserting a sentence A is then knowledge of a closed valid argument for A. Since the only specific thing we need to know about a sentence A in order to know what is meant by a valid argument for A is what the canonical arguments for A are, the latter notion is a more central feature of a sentence than the condition under which it is correctly assertible.

To avoid misunderstandings[2] it may be remarked that it is not suggested that in order to be warranted in asserting a sentence A, one must know, in addition to a valid $(\mathscr{D}, \mathscr{J})$, also a proof that $(\mathscr{D}, \mathscr{J})$ is valid. On the pain of starting an infinite regress, it seems that also knowledge of proofs must in the end involve some implicit knowledge (compare Dummett's argument in e.g., [1] that knowledge of meaning must in the end be implicit).

Once we know the condition for correctly asserting a sentence, we also know when to accept an inference and when to accept that a sentence follows logically from a set of premisses. To know that B can be correctly inferred from the assumption A is to know that we are allowed to assert B in situations where we can correctly assert A. Therefore, in order to be right in accepting an inference as correct, it is necessary and sufficient to have a piece of knowledge which together with knowledge that allows us to assert A constitutes knowledge that allow us to assert B.

If the condition for asserting a sentence is to know a closed valid argument for it, such a piece of knowledge must therefore together with knowledge of a closed argument for A constitute knowledge of a closed valid argument for B. This piece of knowledge must therefore consist in knowing an argument $\mathscr{D}_2$ for B from A and justifying procedures $\mathscr{J}_2$ and knowing that if $\mathscr{D}_1$ is a closed argument for A valid with respect to justifying procedures $\mathscr{J}_1$, then the result of attaching $\mathscr{D}_2$ to $\mathscr{D}_1$ at the point of the assumption A is an argument valid with respect to $\mathscr{J}_1 \cup \mathscr{J}_2$. But this is just what it is to know that the argument $\mathscr{D}_2$ for B from A is valid with respect to J_2. When this knowledge is independent of the system of canonical arguments for atomic sentences, what we know is just that B is a logical consequence of A in the sense proposed here.

In this respect, the proposed analysis of the concept of logical consequence seems satisfactory – i.e., once one has accepted knowledge of valid arguments as the condition under which a sentence is correctly assertible. On this analysis, the grounds for a logical consequence consist essentially in the existence of procedures for transforming proofs of the premisses to proofs of the conclusion, (independent of the meaning of atomic sentences), and an element of necessity in logical consequence is also brought out in this way. But we have not said anything about how to get to know such procedures.

5. *Some Alternative Approaches*

If all the inferences of an argument are applications of valid inference rules (as defined at the end of section 2.3), then it is easily seen that also the argument must be valid, namely with respect to the justifying operations in virtue of which the rules are valid. But this is not the way we have defined validity of arguments. On the contrary, the validity of an inference rule is explained in terms of validity of arguments (although once explained in this way, an argument may be shown to be valid by showing that all the inference rules applied in the argument are valid). One may ask if this order of explanation can be reversed.

Given the principles that

(1) an argument is valid if and only if all inferences of the argument are application of valid inference rules, and that

(2) an inference rule is valid if and only if there is a procedure which applied to valid canonical arguments for the premiss of an application of the rule yields valid argument for the conclusion

together with the stipulation that all introduction rules are valid (the procedure demanded by principle (2) consisting in this case just in applying the rule) it follows easily (as pointed out in [12]) that all the usual elimination rules are indeed valid. If (1) is to serve as an explanation presupposing the explanation (2), one could hope that a valid canonical argument for a sentence A could be defined by an induction over the complexity of A with reference only to the stipulated validity of the introduction rules and without appeal to (1). But when the introduction rule binds an assumption (as is the case for the $\rightarrow$ I-rule), it is not possible to assume in general that the argument for the premisses are in canonical form, and, furthermore, an argument for the premiss may now contain elimination inferences of greater complexity than that of the conclusion. It is just this difficulty that made us bring in justifying procedures already in the definition of validity of arguments. By doing so, we were able to define validity of an argument (in contrast to canonical argument) for A by an induction over the complexity of A. The conclusion that I have drawn is thus that the validity of an argument is not to be explained by (1) above.

However, even if one accepts the idea of canonical arguments and thus clause (i) in the definition of validity for closed arguments (section 2.3), one may doubt that clause (ii) in the same definition is the right way to express the idea that to know a closed valid argument for A is to know how to find a closed valid argument for A in canonical form (recall principle (ii) of section 2.1 and the remarks at the beginning of section 2.2). In particular, one may consider different methods for finding the canonical arguments, leading to conceptions that replace what I have called justifying procedures. I shall roughly indicate how some of them may be described in relation to the approach developed above.

The justifying procedures as conceived here operate on arguments, but it may seem a more natural alternative to let them operate not only on arguments but on arguments together with their justifying procedures, which are anyway parts of what we have to know to be in possession of a proof according to my analysis. This is the usual intuitionistic conception of a proof as outlined e.g., by Heyting [4] when he presents the intuitionistic meaning of the usual logical constants. A proof of B from A, e.g., is then a method which applied to a whole proof of A, so to say – not only to its argument structure – yields a proof of B. I followed this approach in [11] to define a notion of proof, thereby obtaining different assertibility conditions for sentences than the ones presented here.

In the intuitionistic type theory as developed by Martin-Löf in [6], canonical forms are defined for terms that represent proofs understood in this intuitionistic way. Parallel to clause (i) and (ii) in the definition of validity for closed arguments, a term t represents a proof if it either is in canonical form or reduces to such a form when evaluated according to the meaning of the operations described in t. In addition, rules are given for how to demonstrate that a term t is a proof of A in this sense.

A third possibility is to let the justifying procedures operate not only on arguments $\mathscr{D}$ together with justifying procedures $\mathscr{J}$ but also on the insight that $(\mathscr{D}, \mathscr{J})$ is valid, which insight is after all also a part of what we must have acquired to be in possession of a proof according to the main analysis considered here, and which insight is made explicit in Martin-Löf's type theory as just mentioned. A proof of B from A, e.g., is then a method M which applied to a proof of A yields a proof of B together with a proof that M has this property. This is the approach followed by Kreisel in [5].

The exact relationships between these approaches and the differences between their resulting assertibility conditions would be of great interest to get better clarified.

NOTES

* This paper differs from my talk at the workshop in essentially two respects: the discussion in section III of what I call a proof-theoretical approach has now been

expanded by the presentation of more details, and the discussion of alternative approaches has in return been much compressed.

[1] I use this term in [10] but the same idea is expressed in [9, p. 232]. The need of distinguishing canonical proofs is also recognized by Dummett [1]. In a similar context, Martin-Löf [6] speaks about canonical expressions (cf. section 4 below).

[2] And in response to a question by G. Kreisel at the workshop.

REFERENCES

[1] Dummett, Michael: 1975, 'The philosophical basis of intuitionistic logic', in H. E. Rose et al. (eds.), *Logic Colloquium '73*, North-Holland, Amsterdam, pp. 5–40.

[2] Dummett, Michael: 1974, 'What is a theory of meaning?' (II), in G. E. Evans and J. McDowell (eds.), *Truth & Meaning*, Oxford Univ. Press, Oxford, pp. 67–137.

[3] Gentzen, Gerhard: 1935, 'Untersuchungen über das logische Schliessen', *Mathematische Zeitschrift* **39**, 176–210, 405–431.

[4] Heyting, Arend: 1956, *Intuitionism*, North-Holland, Amsterdam.

[5] Kreisel, Georg: 1962, 'Foundations of intuitionistic logic', in E. Nagel et al. (eds.), *Logic, Methodology, and Philosophy of Science*, Stanford, pp. 192–210.

[6] Martin-Löf, Per: 1982, 'Constructive mathematics and computer programming, in L. J. Cohen et al. (eds.), *Logic, Methodology and Philosophy of Science VI*, North-Holland, Amsterdam, pp. 153–175.

[7] Prawitz, Dag: 1965, *Natural Deduction. A Proof-Theoretical Study*. Almqvist & Wiksell, Stockholm.

[8] Prawitz, Dag: 1975, 'Ideas and results in proof theory', in J. E. Fenstad (ed.), *Proceedings of the Second Scandinavian Logic Symposium*, North-Holland, Amsterdam, pp. 235–250.

[9] Prawitz, Dag: 1973, 'Towards a foundation of a general proof theory', in P. Suppes et al. (eds.), *Logic, Methodology, and Philosophy of Science IV*, North-Holland, Amsterdam, pp. 225–250.

[10] Prawitz, Dag: 1974, 'On the idea of a general proof theory', *Synthese* **27**, 63–77.

[11] Prawitz, Dag: 1977, 'Meaning and proofs: on the conflict between classical and intuitionistic logic', *Theoria* **43**, 1–40.

[12] Prawitz, Dag: 1978, 'Proofs and the meaning and completeness of the logical constants', in J. Hintikka et al. (eds.), *Essays on Mathematical and Philosophical Logic*, D. Reidel, Dordrecht, pp. 25–40.

[13] Schroeder-Heister, Peter: 1981, *Untersuchungen zur regellogischen Deutung von Aussagenverknüpfungen* (Inaugural-Dissertation), Bonn.

[14] Tarski, Alfred: 1936, 'Über den Begriff der logischen Folgerungen', in *Acte du Congrès International de Philosophie Scientifique VII*, Paris, pp. 1–11.

Filosofiska Inst.
Stockholm Univ.
106 91 Stockholm
Sweden