SYNTHESE / *Volume 148 No. 3 February 2006*

PROOF-THEORETIC SEMANTICS

Editors:
Reinhard Kahle and Peter Schroeder-Heister

REINHARD KAHLE and PETER SCHROEDER-HEISTER

INTRODUCTION: PROOF-THEORETIC SEMANTICS

According to the model-theoretic view, which still prevails in logic, semantics is primarily denotational. Meanings are denotations of linguistic entities. The denotations of individual expressions are objects, those of predicate signs are sets, and those of sentences are truth values. The meaning of an atomic sentence is determined by the meanings of the individual and predicate expressions this sentence is composed of, and the meaning of a complex sentence is determined by the meanings of its constituents. A consequence is logically valid if it transmits truth from its premisses to its conclusion, with respect to all interpretations. Proof systems are shown to be correct by demonstrating that the consequences they generate are logically valid. This basic conception also underlies most alternative logics such as intensional or partial logics. In these logics, the notion of a model is more involved than in the classical case, but the view of proofs as entities which are semantically dependent on denotational meanings remains unchanged.

Proof-theoretic semantics proceeds the other way round, assigning proofs or deductions an autonomous semantic role from the very onset, rather than explaining this role in terms of truth transmission. In proof-theoretic semantics, proofs are not merely treated as syntactic objects as in Hilbert's formalist philosophy of mathematics, but as entities in terms of which meaning and logical consequence can be explained.

The programme of proof-theoretic semantics can be traced back to Gentzen (1934). Seminal papers by Tait, Martin-Löf, Girard and Prawitz were published in 1967 and 1971.[1] An explicit formulation of a semantic validity notion for generalized deductions with respect to arbitrary justifications was given by Prawitz (1973). Much of the philosophical groundwork for proof-theoretic semantics was laid by Dummett from the 1970s on, culminating in Dummett (1991). Martin-Löf's type theory, whose philosophical foundation is proof-theoretic semantics, became a full-fledged theory in the 1970s as well (see Martin-Löf 1975, 1982). The term "proof-theoretic semantics" was proposed by the second editor in a lecture in Stockholm in 1987.[2]

Synthese (2006) 148: 503–506
DOI 10.1007/s11229-004-6292-5

Since proof-theoretic semantics has reached some status of maturity, we considered it appropriate to organize a conference with that title at the University of Tübingen in January 1999.[3] The papers presented at this conference were the following:

- Dag Prawitz: Meaning explained in terms of proofs: A comparison of some approaches
- Lars Hallnäs: Defining the semantics
- Patrizio Contu: The justification of the logical laws revisited
- Gabriele Usberti: Towards a semantics based on the notion of justification
- Michael Dummett: Reply to Warren Goldfarb
- Göran Sundholm: Inference *versus* consequence
- Roy Dyckhoff: Permutation-free sequent calculi
- Jörg Hudelmaier: A semantical sequent calculus for intuitionistic logic
- Robert Stärk: Proof-theoretic semantics of logic programs
- Grigori Mints: Partial proofs and cut introduction
- Per Martin-Löf: The distinction between sense and reference in constructive semantics
- Kosta Došen: Models of proofs
- Peter Schroeder-Heister: Frege's sequent calculus
- Reinhard Kahle: A proof-theoretic view of intensionality
- Michael Rathjen: The role of ordinals in proof theory
- William Tait: Beyond the axioms: The question of objectivity in mathematics

The present collection grew out of this conference but is not intended as a volume of proceedings. Our idea was, by means of various basic papers, to shed some light on central topics of proof-theoretic semantics to enable researchers from other branches of logic to gain some insight into a subject which we think has a bright future.

The first topic of these papers are approaches giving proofs a semantic value without reference to denotations: Prawitz philosophically elucidates his meaning theory based on proofs, and Schroeder-Heister, Contu and Hallnäs deal affirmatively and critically with validity notions developed in the tradition created by Prawitz. Tait, in a type-theoretic framework, shows that a non-denotational approach does not necessarily lead to non-classical (intuitionistic) logic. Then there are contributions which reflect on the framework in which proofs should be dealt with: Sundholm compares different forms of natural deduction from a

meaning-theoretic point of view, and Došen puts forward categorical logic as a framework particularly appropriate for proof-theoretic semantics. Two papers develop applications: Kahle uses proof-theoretic semantics in order to clarify the notion of necessity, while Usberti carries over proof-theoretic semantics to the justification of empirical sentences. Finally we have two contributions dealing with the background to proof-theoretic semantics: Mints presents some basic ideas of Russian constructivism, and Rathjen gives an overview of theories of ordinals which have dominated proof theory for quite some time.

Due to various circumstances, editing this collection stretched over a period of several years. We received the first manuscripts in 1999 and the last update of a paper in 2004. We apologize for this delay to those authors who submitted their contributions early.

We should like to thank the reviewers for their efforts, Wilfried Sieg for valuable comments on a previous version, and Janah Putnam for her help with language editing. Special thanks are due to Thomas Piecha, who prepared the final manuscript, for his careful editorial work.

NOTES

[1] See Tait (1967), Girard (1971), Martin-Löf (1971), Prawitz (1971).

[2] First in press in Schroeder-Heister (1991). Whether this term had already occasionally been used in Stockholm at that time he cannot recall, although he does not want to rule this out. – As early as 1968 Kutschera used the term "Gentzen semantics" ["Gentzensemantik"] (see Kutschera 1968).

[3] Supported by DFG grant Schr 275/12-1.

REFERENCES

Dummett, M.: 1991, *The Logical Basis of Metaphysics*, Duckworth, London.

Gentzen, G.: 1934, 'Untersuchungen über das logische Schließen', *Mathematische Zeitschrift* **39** (1934/35), 176–210, 405–431, English translation ('Investigations into Logical Deduction') in M. E. Szabo (ed.), *The Collected Papers of Gerhard Gentzen*, North Holland, Amsterdam 1969, 68–131.

Girard, J.-Y.: 1971, 'Une extension de l'interprétation de Gödel à l'analyse, et son application à l'élimination des coupures dans l'analyse et la théorie des types', in J. E. Fenstad (ed.), *Proceedings of the 2nd Scandinavian Logic Symposium* (*Oslo 1970*), North Holland, Amsterdam, pp. 63–92.

Kutschera, F. von: 1968, 'Die Vollständigkeit des Operatorensystems $\{\neg, \wedge, \vee, \supset\}$ für die intuitionistische Aussagenlogik im Rahmen der Gentzensemantik', *Archiv für mathematische Logik und Grundlagenforschung* **11**, 3–16.

Martin-Löf, P.: 1971, 'Hauptsatz for the Intuitionistic Theory of Iterated Inductive Definitions', in J. E. Fenstad (ed.), *Proceedings of the 2nd Scandinavian Logic Symposium* (*Oslo 1970*), North Holland, Amsterdam, pp. 179–216.
Martin-Löf, P.: 1975, 'An Intuitionistic Theory of Types: Predicative Part', in H. E. Rose and J. Shepherdson (eds.), *Logic Colloquium '73*, North Holland, Amsterdam, pp. 73–118.
Martin-Löf, P.: 1982, 'Constructive Mathematics and Computer Programming', in L. J. Cohen et al. (eds.), *Logic, Methodology and Philosophy of Science VI [1979]*, North Holland, Amsterdam, pp. 153–175.
Prawitz, D.: 1971, 'Ideas and Results in Proof Theory', in J. E. Fenstad (ed.), *Proceedings of the 2nd Scandinavian Logic Symposium* (*Oslo 1970*), North Holland, Amsterdam, pp. 235–308.
Prawitz, D.: 1973, 'Towards a Foundation of a General Proof Theory', in P. Suppes et al. (eds.), *Logic, Methodology, and Philosophy of Science IV [1971]*, North Holland, Amsterdam, pp. 225–250.
Schroeder-Heister, P.: 1991, 'Uniform Proof-Theoretic Semantics for Logical Constants'. Abstract. *Journal of Symbolic Logic* **56**, 1142.
Tait, W. W.: 1967, 'Intensional Interpretations of Functionals of Finite Type I', *Journal of Symbolic Logic* **32**, 198–212.

Reinhard Kahle
Departamento de Matemática
Universidade de Coimbra
Apartado 3008
P-3001-454 Coimbra
Portugal
E-mail: kahle@mat.uc.pt

Peter Schroeder-Heister
Wilhelm-Schickard-Institut
Universität Tübingen
Sand 13
72076 Tübingen
Germany
E-mail: psh@informatik.uni-tuebingen.de

DAG PRAWITZ

MEANING APPROACHED VIA PROOFS

ABSTRACT. According to a main idea of Gentzen the meanings of the logical constants are reflected by the introduction rules in his system of natural deduction. This idea is here understood as saying roughly that a closed argument ending with an introduction is valid provided that its immediate subarguments are valid and that other closed arguments are justified to the extent that they can be brought to introduction form. One main part of the paper is devoted to the exact development of this notion. Another main part of the paper is concerned with a modification of this notion as it occurs in Michael Dummett's book *The Logical Basis of Metaphysics*. The two notions are compared and there is a discussion of how they fare as a foundation for a theory of meaning. It is noted that Dummett's notion has a simpler structure, but it is argued that it is less appropriate for the foundation of a theory of meaning, because the possession of a valid argument for a sentence in Dummett's sense is not enough to be warranted to assert the sentence.

1. INTRODUCTION

The term proof-theoretic semantics would have sounded like a *contradictio in adjecto* to most logicians and philosophers half a century ago, when proof theory was looked upon as a part of syntax, and model theory was seen as the adequate tool for semantics. Michael Dummett is one of the earliest and strongest critics of the idea that meaning could fruitfully be approached via model theory, the objection being that the concept of meaning arrived at by model theory is not easily connected with our speech behaviour so as to elucidate the phenomenon of language. Dummett pointed out at an early stage that Tarski's T-sentences, i.e., the various clauses in Tarski's definition of truth, cannot simultaneously serve to determine both the concept of truth and the meaning of the sentences involved. Either one must take the meaning as already given, which is what Tarski did, or one has to take truth as already understood, which is the classical approach from Frege onwards.

This latter alternative amounts to an account of meaning in terms of truth conditions depending on a tacit understanding of truth. In the

Synthese (2006) 148: 507–524
DOI 10.1007/s11229-004-6295-2

case of a construed formal language, the T-sentences become postulated semantic rules that are supposed to give the formulas a meaning (a representative presentation of this view is in *Introduction to Mathematical Logic* by Alonzo Church (1956)). If the T-sentences are to succeed in conferring meaning to sentences, this must be because of some properties of the notion of truth. A person not familiar with the notion of truth would obviously not learn the meaning of a sentence by being told what its truth condition is. It therefore remains to state what it is about truth that makes the semantic rules function as genuine meaning explanations – the semantics has to be embedded in a meaning theory as Dummett puts it.

In the case of an already given natural language, the T-sentences become instead hypotheses, which must somehow be connected with speech behaviour. Here one may follow Donald Davidson's suggestion which may roughly be put: if "A is true iff C" is a correct T-sentence for the sentence A in a language L, then a speaker of L who asserts A normally believes that the truth condition C is satisfied; cases when a speaker is noticed both to observe that C is satisfied and to assert A therefore constitute data supporting the T-sentence.

By making this connection between T-sentences and speech behaviour for at least observation sentences, one begins spelling out the concept of truth, which is needed to support the claim that the T-sentences give the meaning of the sentences of a language. However, as argued by Dummett (e.g., in Dummett 1983), it is only a beginning, because the assertion of sentences is only one aspect of their use. If the T-sentences are really to be credited with ascribing meaning to sentences, they must be connected with all aspects of the use of sentences that do depend on meaning. In other words, there are further ingredients in the concept of truth that must be made explicit, if the truth condition of a sentence is to become connected with all features of the use of the sentence that do depend on meaning. One such feature is the use of sentences as premisses of inferences. When asserting a sentence we are not only expected to have grounds for the assertion, we also become committed to certain conclusions that can be drawn from the assertion taken as a premiss.

I shall leave the prospects of rightly connecting the meanings of expressions with our use of them within a theory of meaning developed along these lines, and shall instead review some approaches to meaning that are based on how we use sentences in proofs. One advantage of such an approach is that from the beginning meaning is connected with aspects of linguistic use.

One very simple version of an approach of this kind is to take meaning to be determined by all the rules for a language. Restricting oneself to deductive uses of language and thinking of proofs as determined by a set of inference rules, meaning simply becomes determined by all the inference rules of the language. This way of literally following the slogan "meaning is use" – the inference rules that determine the use of sentences also determine their meaning – fell in some disrepute, when Prior (1960) introduced a sentential operator tonk governed by rules similar to the introduction rule for disjunction and the elimination rule for conjunction. Since the effect of adding tonk to a language is to make all sentences derivable, a person who adheres to the idea that an arbitrary set of inference rules determines meaning must be prepared to allow that even inconsistent languages are entirely meaningful.

An interesting defence of such a standpoint is given by Cozzo (1994). He develops a theory in which the meaning of a sentence is given by arbitrary argumentation rules concerning the terms that occur in the sentence. The theory is interesting because, in spite of the fact that it gives a meaning to tonk and thus to inconsistent languages, it makes meaning compositional, it rejects semantic holism but respects epistemological holism, and it allows criticism of a language; a meaningful language may not be a good language (or in Cozzo's terminology: a "correct" language). However, this is not a line of thought that I shall follow here.

The approaches that I shall discuss are inspired by the main idea behind Gentzen's systems of natural deduction, and take a quite different view with respect to the kind of inference rules that are considered as a possible basis for a theory of meaning. I shall mainly restrict myself to an approach that I first proposed in Prawitz (1973) and to a somewhat modified approach suggested by Dummett (1991). These two approaches will be compared and problems concerning the possibility of embedding them into a full theory of meaning will be discussed.

2. GENTZEN'S IDEA OF INFERENCE RULES DETERMINING MEANING

There is a remark by Gentzen (1934) which he made after having constructed his system of natural deduction and which I have quoted before both as a key to the normalization theorem for natural deduction (or the *Hauptsatz*) and as a basis for a proof-theoretic semantics. It reads:

The introductions constitute, as it were, the 'definitions' of the symbols concerned, and the eliminations are, in the final analysis, only consequences of this, which may be expressed something like this: At the elimination of a symbol, the formula with whose outermost symbol we are dealing may be used only 'in respect of what it means according to the introduction of that symbol'.

In contrast to meaning theories inspired by model theory, meaning is now given not by truth conditions but by certain ways in which truth is established, what Gentzen calls *introductions*. The truth of a sentence may however be established also in other ways, which is to say that a sentence may also occur in other connections than introductions. Gentzen is now careful to stress something which was noted to be absent but needed in the approach to meaning based on truth conditions, viz., that the other uses of a sentence are accounted for in terms of (or as Gentzen expresses it: are 'consequences of') the meaning ascribed.

To develop Gentzen's idea we have thus firstly to state more exactly how the introductions determine the meaning of the logical constants; the phrase saying that the introductions represent definitions is clearly not meant to be taken literally. The view that I am taking is that the introductions represent what we may call the canonical ways of inferring a sentence. Other ways of inferring a sentence have to be justified by reducing them to the canonical ways.

Gentzen considers besides introductions certain specific inferences that he calls *eliminations*. We cannot expect these eliminations to be derivable from the introductions in the ordinary sense of being derived inference rules in the system given by the introduction rules. Instead, we have to show that they can be justified in some semantic way, which is to say that they can be shown to be valid in view of the meaning of the sentences involved.

The task is thus to develop an appropriate notion of validity and to show that certain legitimate forms of reasoning are valid in the sense defined. We should then not restrict ourselves to the eliminations given by Gentzen but consider what it is for any non-canonical inference to be valid.

3. DEFINING THE VALIDITY OF ARGUMENTS

3.1. *Argument Skeletons*

Validity has thus to be defined not only for derivations in some given system but for more arbitrary ways of reasoning.[1] I shall

therefore define validity for what I call arguments. Furthermore, it seems strange to speak of the validity of proofs. A false sentence is still a sentence, but an invalid proof is not really a proof. In contrast, an argument may be valid or invalid – if it is valid, it represents a proof.

By an *argument skeleton* I shall understand a tree arrangement of formulas; if all the formulas are sentences (i.e., closed formulas), the arrangement is to be understood as claiming for each sentence in the tree except the ones at the top that it follows from the sentences (premisses) standing immediately above. For each top sentence of the tree there is to be indicated whether it is claimed outright as holding (to follow from zero premisses) or if it is entered as an assumption made for the sake of the argument, in which case there may also be an indication at which step in the argument the assumption is discharged or *bound* as I shall say. An example of a step that is allowed to bind assumptions is implication introduction, i.e., an inference of the form

$$\frac{\begin{array}{c}[A]\\ \mathcal{D}\\ B\end{array}}{A \supset B} \tag{1}$$

There may also be indications that a variable that is free in the formulas in which it occurs is *bound by some step in the argument*. An example of a step that binds variables is universal introduction, i.e., an inference of the form

$$\frac{\begin{array}{c}\mathcal{D}\\ A(x)\end{array}}{\forall x A(x)} \tag{2}$$

An inference binds only occurrences of an assumption or a variable that appear in the part of the tree that is above the conclusion of the inference. When a variable is bound by a step it must not occur in the conclusion of the step or in assumptions that are not bound by the step or by some step higher up in the tree (cf. the conditions on so called *Eigenvariablen*). An occurrence that is not bound is said to be *free*.

An argument skeleton is *closed*, if all occurrences of assumptions are bound and likewise all occurrences of variables that are free in the formulas are bound in the argument skeleton; it is *open* otherwise. An open argument skeleton is to be understood as a schema, from which closed argument skeletons can be generated by first substituting

closed terms for the free variables and then closed argument skeletons for the free assumptions (in the argument skeleton resulting from the first substitution); the result is said to be an *instance* of the open argument skeleton.

I have been speaking about argument skeletons because what I shall take to be arguments will contain something in addition to the trees of formulas (with indications of how assumptions and variables are bound) which we have been considering so far. The notion of validity will be defined for arguments, i.e., argument skeletons supplemented in a way that remains to be specified. The need for this supplementation does not arise in connection with the introduction inferences, which will now be considered in more detail.

3.2. *Canonical Forms*

An argument skeleton whose last step is an introduction will be said to be in *canonical form*. For each sentence there are given forms of arguments for the sentence which count as canonical. The idea is that these forms determine the meaning of the sentence. The sentence is to be understood as standing for something whose canonical proof, if there is a proof at all, is of the form specified. An argument step that has the form of an introduction is therefore valid by the very meaning of the sentence occurring as conclusion. We shall take care of this idea by saying that an argument whose skeleton is closed and is in canonical form is valid provided its immediate subarguments (i.e., the arguments for the premisses of the last inference step) are valid.

Closed arguments whose skeleton has the form exhibited in (1), (2) or

$$\frac{\begin{array}{cc}\mathcal{D}_1 & \mathcal{D}_2\\ A_1 & A_2\end{array}}{A_1 \& A_2} \qquad \frac{\begin{array}{c}\mathcal{D}\\ A_i\end{array}}{A_1 \vee A_2} \qquad \frac{\begin{array}{c}\mathcal{D}\\ A(t)\end{array}}{\exists x A(x)} \tag{3}$$

are thus valid provided the immediate subarguments resulting from leaving out the last step are valid.

When a skeleton has one of the forms shown in (3), one could impose a more stringent requirement on the canonical forms, namely that the skeletons of the immediate subarguments are themselves in canonical form. However, when the last inference step is an implication introduction or a universal introduction as in (1) or (2), then, as we have seen, it binds occurrences of an assumption or of a variable, respectively. Therefore, the argument skeleton obtained by leaving out the last step may not be closed, and it would be too

stringent to impose on the canonical forms that also such an open part of the skeleton is canonical.

3.3. *Open Argument Skeletons*

In line with the understanding of open argument skeletons as schemata, we shall adopt the principle that an open argument is valid provided those instances are valid that are obtained by substituting closed terms for the free variables (supposed to denote objects that belong to the range of the variable) and valid closed arguments for the free assumptions. Let us call such an instance an *appropriate instance*. We have thus the following

Principle of validity for open arguments: An open argument is valid if and only if all its appropriate instances are valid.

3.4. *Justifications of Non-Canonical Arguments*

How is then an inference step that is not an introduction to be justified with reference to the meaning of the sentences involved? Consider a closed argument whose skeleton ends with modus ponens:

$$\dfrac{\begin{array}{cc}\mathcal{D}_1 & \mathcal{D}_2 \\ A & A \supset B\end{array}}{B} \tag{4}$$

Suppose that the immediate subarguments are valid. Their skeletons $\mathcal{D}_1$ and $\mathcal{D}_2$ that end with A and $A \supset B$ are closed, and by the meaning of $A \supset B$, it should be possible to bring the valid argument $\mathcal{D}_2$ for $A \supset B$ into canonical form with a skeleton as exhibited in (1) above. It should remain valid, and its immediate subargument with skeleton $\mathcal{D}$ should then also be valid. $\mathcal{D}$ is open, but by substituting $\mathcal{D}_1$ for the open occurrences of the assumption A in $\mathcal{D}$ we obtain

$$\begin{array}{c}\mathcal{D}_1 \\ [A] \\ \mathcal{D} \\ B\end{array} \tag{5}$$

i.e., a closed argument for B, which should also be valid, being an appropriate instance of a valid closed argument schema.

This is a rough outline of how modus ponens is justified in terms of a notion of validity that is not yet defined. The main idea is that there is an operation that transforms an argument skeleton of the form (4) where the part $\mathcal{D}_2$ is in canonical form into another argument skeleton (5) still ending with B but from which the

exhibited application of modus ponens is eliminated. An operation of this kind I shall call a *justification* (strictly speaking one should say an alleged justification) in this case of modus ponens. A justification of modus ponens should show that a closed argument for B whose skeleton has the form exhibited in (4) and whose immediate subarguments are valid could be brought into a valid closed canonical argument for B.

My approach is now to let the arguments for which validity is defined consist of argument skeletons together with proposed justifications of all the inferences that are non-canonical. A bare argument skeleton is not regarded in itself as a valid argument. In other words, it is not enough that there exist effective means for finding another argument skeleton for A in canonical form for counting a given argument skeleton for a sentence A as a valid argument. It is the skeleton together with such effective means, operating on the given skeleton, that constitute an argument for A, as I see it.

An (alleged) justification is any operation that is defined for argument skeletons of some form and transforms them to other argument skeletons for the same formulas without introducing additional free variables or free assumptions. In addition we only need to impose a few formal requirements on the operations such as commuting with substitutions. A set of such operations with mutually disjoint domains of definitions will be said to be a *consistent* justification set. By an *argument* I shall understand an argument skeleton together with a consistent justification set.

An argument consisting of an argument skeleton $\mathcal{D}$ and a justification set $\mathcal{J}$ will be said to reduce to another argument consisting of the skeleton $\mathcal{D}'$ and the justification set $\mathcal{J}$, if $\mathcal{D}$ reduces to $\mathcal{D}'$ in the same way as natural deductions are said to reduce to each other in connection with normalizations, but now using the justification set $\mathcal{J}$ instead of the reductions defined for natural deductions. Notions introduced for argument skeletons may be carried over to arguments in the obvious way. In particular, an argument consisting of an argument skeleton $\mathcal{D}$ and a justification set $\mathcal{J}$, written $\langle \mathcal{D}, \mathcal{J} \rangle$, will be said to be open or closed, if $\mathcal{D}$ is open or closed, respectively. Similarly, an immediate subargument of $\langle \mathcal{D}, \mathcal{J} \rangle$ is an argument $\langle \mathcal{D}', \mathcal{J}' \rangle$ where $\mathcal{D}'$ is an initial part of $\mathcal{D}$ ending with a premiss of the last inference step in $\mathcal{D}$ and $\mathcal{J}'$ is the subset of $\mathcal{J}$ obtained by leaving out justifications of steps not occurring in $\mathcal{D}'$.

3.5. *Principles of Validity*

We may now sum up the ideas outlined above by stating three principles that the notion of validity for arguments should satisfy.

Principle 1. A closed argument in canonical form is valid iff its immediate subarguments are valid.

Principle 2. A closed argument not in canonical form is valid iff it reduces to a valid argument in canonical form, i.e., to an argument that is valid by principle 1.

Principle 3. An open argument $\langle \mathcal{D}, \mathcal{J} \rangle$ is valid iff all those instances $\langle \mathcal{D}', \mathcal{J}' \rangle$ of $\langle \mathcal{D}, \mathcal{J} \rangle$ are valid where $\mathcal{J}'$ is a consistent extension of $\mathcal{J}$ and $\mathcal{D}'$ is an appropriate instance of $\mathcal{D}$, i.e., the argument is to be appropriate in the sense that for any argument $\mathcal{E}$ that is substituted for a free assumption in $\mathcal{D}$ in order to form $\mathcal{D}'$ it should hold that $\langle \mathcal{E}, \mathcal{J}' \rangle$ is valid.

Principles 1 and 2 articulate the idea that the meaning of a sentence is given by what counts as a canonical proof of the sentence: the use of an introduction in an argument preserves validity by the very meaning of the inferred sentence, and non-canonical arguments are valid if and only if they reduce to valid canonical ones. Principle 3 articulates the idea that an open argument is seen as an argument schema.

Together the three principles also constitute an inductive definition of the notion of validity, provided that a set of valid canonical arguments for atomic sentences is given as an induction base. Principle 3 refers the validity of open arguments to the validity of closed arguments as determined by principles 1 and 2. Principle 2 refers to validity as determined by principle 1. Principle 1 finally refers the validity of an argument for a given sentence to the validity of other arguments as determined by all the principles but with respect to formulas of lower complexity than the given one. Clearly we can extend the present approach to other sentence forming operations, if we can formulate introduction rules for the operations in such a way that each inference that proceeds according to the rule satisfies the requirement that the premisses of the inference and the assumptions that are bound by the inference are of lower complexity than the conclusion of the inference.

When the induction base B is made explicit by a set of inference rules for atomic formulas (both conclusion and premisses are to be atomic), I shall speak about validity relative to an *atomic base* B. A logically valid argument must be valid relative to an arbitrary base B.

But we should require more. Otherwise the atomic formulas get a special status, not congruent with the idea that when speaking of logical validity the atomic formulas are thought to stand for arbitrary propositions. Of logical validity one should therefore require that the validity is invariant also for substitutions for atomic formulas.

3.6. *The validity of inference rules*

An inference rule may be defined as valid relative to a justification j if it preserves validity. More precisely it is to hold for each argument skeleton $\mathcal{D}$ whose last inference step is an application of the rule and for each consistent extension $\mathcal{J}$ of $\{j\}$ that the argument $\langle \mathcal{D}, \mathcal{J} \rangle$ is valid if its immediate subarguments are. A rule whose validity is invariant for variations of atomic base and substitutions for the atomic formulas may be said to be logically valid.

All arguments where the skeleton is formed according to the intuitionistic rules of natural deduction for predicate logic and where the justifications assigned to the elimination steps consist of the ordinary reductions defining normalizations for natural deductions are (logically) valid. This is most easily proved by first showing that each intuitionistic elimination rule is (logically) valid with respect to its reduction operation.

To exemplify we may again look at modus ponens to which we assign an operation j that transforms a skeleton of form

$$\dfrac{\mathcal{D}_1 \quad \begin{array}{c}[A]\\ \mathcal{D}\\ B\end{array}}{\dfrac{A \qquad A \supset B}{B}} \tag{6}$$

to a skeleton of the form exhibited in (5) above. We want to show that an argument is valid when its skeleton has the form exhibited in (4) and its set $\mathcal{J}$ of justifications is a consistent extension of $\{j\}$ given the assumption that its immediate subarguments $\langle \mathcal{D}_1, \mathcal{J} \rangle$ and $\langle \mathcal{D}_2, \mathcal{J} \rangle$ are valid. We first note that the validity of $\langle \mathcal{D}_2, \mathcal{J} \rangle$ implies that it reduces to a valid argument in canonical form. Remembering that the justifying operations commute with substitutions, it follows that the given argument reduces to an argument whose skeleton has the form (6), which in turn reduces to (5) by applying j. That the argument $\langle (5), \mathcal{J} \rangle$ is valid, then follows from principle 3 by the fact that it is an instance of a valid open argument $\mathcal{D}$ obtained by substituting the valid argument $\langle \mathcal{D}_1, \mathcal{J} \rangle$ for the free assumption A.

4. DUMMETT'S PROOF THEORETIC JUSTIFICATIONS

In his book *The Logical Basis of Metaphysics*, Dummett (1991) describes and discusses what he calls proof-theoretic justifications of logical laws, which in many respects follow the approach presented in the previous section. There are a couple of noteworthy differences however. The major difference is that Dummett defines validity for what I call argument skeletons. Another minor one is that the canonical forms are defined slightly differently. To facilitate a comparison of Dummett's treatment with mine I shall keep the terminology of the previous section when I state Dummett's definitions. (I shall ignore some other small differences such as one concerning the definition of what I call an instance of an argument skeleton. In Dummett's definitions of corresponding notions, he happens to pay no attention to atomic sentences that occur as free assumptions or to free variables that do not occur in the conclusion or in some free assumption. The difference may be due to the fact that Dummett does not operate explicitly with the notion of open and closed argument (skeleton).)

4.1. *Hereditary Canonical Form*

Dummett makes the more stringent requirement on the canonical forms that was noted above (in Section 3.2) to be possible to make in some cases. He then obtains something that we may call *hereditary canonical forms* defined inductively as follows: an argument skeleton is in hereditary canonical form iff (a) its last step is an introduction, and (b) in case the introduction does not bind any assumption or variable, its immediate subarguments are also in hereditary canonical form. (It is to be assumed that the atomic base specifies introduction rules for atomic formulas.)

It is easy to see (by induction over the definition of validity) that if we replace "canonical form" by "hereditary canonical form" in the definition of validity (i.e., in principles 1 and 2), the extension of the notion of validity stays the same.[2] We may therefore disregard this difference between Dummett's definition and mine.

4.2. *Leaving out the Justifications*

Returning to the major difference between the two definitions, we may try to phrase in my terminology Dummett's definition of what it

is for an argument skeleton to be valid by stating three principles similar to the ones in Section 3.5:

*Principle 1**. A closed argument skeleton in canonical form is valid if and only if its immediate parts are valid.

*Principle 2**. A closed argument skeleton for a sentence A that is not in canonical form is valid if and only if a closed valid argument skeleton for A in canonical form can be found effectively.

*Principle 3**. An open argument skeleton $\mathcal{D}$ is valid if and only if all those instances of $\mathcal{D}$ are valid that are obtained by substituting closed valid canonical argument skeletons for free assumptions.

These three principles constitute as before an inductive definition. Leaving out the justifications, as I called them, the notion of validity now defined is much simpler. It may be simpler than the notion intended by Dummett, however. If we apply principles 2* and 3* together to an open non-canonical argument skeleton $\mathcal{D}$ for the formula A, we find that $\mathcal{D}$ is now defined to be valid if and only if for any closed instance $\mathcal{D}^\sigma$ of $\mathcal{D}$ obtained by a substitution σ that substitutes closed valid canonical arguments for free assumptions in $\mathcal{D}$ and terms for free variables in $\mathcal{D}$, we can find effectively a closed argument skeleton $\mathcal{D}'$ for A^σ. Here it is not required that the argument skeleton $\mathcal{D}'$ to be found for A^σ is in any way related to $\mathcal{D}^\sigma$. It is only required that for any σ there is an effective method to find a valid closed canonical argument skeleton $\mathcal{D}'$ for A^σ, not that there is an effective uniform method which applied to any $\mathcal{D}^\sigma$ finds such a $\mathcal{D}'$.

This may not be intended, and perhaps Dummett's notion of validity is instead rendered by principle 1* and a modified combination of Principles 2* and 3* as follows:

*Principles (2–3)**. An argument skeleton $\mathcal{D}$ for a formula A that is not in canonical form is valid if and only if there is an effective method M such that for any closed instance $\mathcal{D}^\sigma$ of $\mathcal{D}$ obtained by a substitution σ that substitutes terms for free variables in $\mathcal{D}$ and closed valid canonical arguments for free assumptions in $\mathcal{D}$, M applied to $\mathcal{D}^\sigma$ yields a valid canonical argument skeleton for A^σ.

How are the two notions of validity related to each other? Could it be that an argument skeleton $\mathcal{D}$ is valid as defined by principles 1* and (2–3)* if and only if there are justifying operations $\mathcal{J}$ to assign to the non-introductory steps of $\mathcal{D}$ so that the argument $\langle \mathcal{D}, \mathcal{J} \rangle$ is valid as defined by principles 1, 2 and 3? If a closed argument $\langle \mathcal{D}, \mathcal{J} \rangle$ for a sentence A reduces to a valid canonical argument $\langle \mathcal{D}', \mathcal{J} \rangle$ (as required for $\langle \mathcal{D}, \mathcal{J} \rangle$ to be valid by principle 2), then $\mathcal{D}'$ can of course be found effectively and $\mathcal{D}'$ is an argument skeleton for A in

canonical form (as required for $\mathcal{D}$ to be valid by principle (2–3)*). But conversely, it is not obvious that given an effective method for finding a canonical argument skeleton $\mathcal{D}'$ for a sentence A, the existence of which makes any closed argument skeleton $\mathcal{D}$ for A valid provided that $\mathcal{D}'$ is valid, we can find justifying operations $\mathcal{J}$ to assign to the non-introductory steps of such a $\mathcal{D}$ so that $\langle \mathcal{D}, \mathcal{J} \rangle$ reduces to $\langle \mathcal{D}', \mathcal{J} \rangle$. The two notions are therefore not easily compared to each other.

For Dummett's notion of validity it holds, as he himself remarks, that an argument skeleton $\mathcal{D}$ for a sentence A from open premisses $A_1, A_2, \ldots, A_n$ is valid if and only if the one step argument

$$\frac{A_1, A_2, \ldots, A_n}{A}$$

is valid. Thus, regardless of how irrelevant the steps of $\mathcal{D}$ are for inferring A from $A_1, A_2, \ldots, A_n$, $\mathcal{D}$ is valid if the corresponding one step argument is valid. In other words, it is the existence of effective means for finding a closed valid canonical argument skeleton for A, given closed valid canonical argument skeletons for $A_1, A_2, \ldots, A_n$, that makes $\mathcal{D}$ valid, not what goes on in the skeleton $\mathcal{D}$. It is these means and not the skeleton alone that carries epistemic force, and this was my motivation for including them, i.e., what I have called justifying operations, in the arguments.

However, it remains to discuss whether the notions developed so far form a reasonable basis for a theory of meaning which is supposed to represent the idea that the meaning of a sentence is determined by how it is established as true.

5. RELATIONS TO VERIFICATIONISM

The verificationism of the logical positivists was an early attempt to relate the meaning of a sentence to how we establish its truth, i.e., how we verify it. The slogan "the meaning of a sentence is its method of verification" is not very apt however. It seemed both from this slogan and from what some of the early verificationists said that knowing the meaning of a sentence involved knowing how to decide the truth of the sentence in principle.

That a viable verificationism cannot require that a meaningful sentence is decidable but should relate the understanding of a sentence with the ability to recognize a verification of the sentence when presented by one was pointed out long ago by Michael Dummett.[3]

What determines the meaning of a sentence is thus not its method of verification but rather what it is to verify it, or what counts as a verification of it, as Per Martin-Löf (1985) formulates it. Furthermore, there is now a need to single out a class of direct or canonical verifications – in the first place, they are what is related to meaning.[4]

The basic idea of verificationism as construed here is thus that the meaning of a sentence is given by what counts as a direct verification of it. Gentzen's suggestion that the meanings of the logical constants are determined by their introduction rules can be seen as a special case of this verificationist idea. So as to conform better to this way of expressing the general verificationist idea, the suggestion may be slightly reformulated as saying, firstly, that the meaning of a compound sentence in the language of first order predicate logic is given by what counts as a direct verification of it, and, secondly, that the forms of these direct verifications are given by the introduction rules, i.e., a direct verification has the form of an argument whose last step is an introduction.

Now, an argument cannot count as a direct verification just because its last step is an introduction, something more must be required. What must be added is something about the validity of the argument. The validity of the last step is of course not called in question – that is part of the essence of Gentzen's suggestion. What must be added is thus only that the rest of the argument is valid, i.e., that the immediate subarguments are valid. This is precisely how the validity of an argument in canonical form is defined both by me and by Dummett, except that an argument for Dummett is what I call an argument skeleton.

We arrive in this way at the following formulation of Gentzen's suggestion: A direct verification of a compound sentence A is the same as a valid argument in canonical form, i.e., an argument ending with an introduction whose immediate subarguments are valid, and this is what determines the meaning of A. In other words, it is the inductive definition of what it is to be a valid argument for A, which follows the inductive built up of A, that is proposed to be constitutive for the meaning of A.

I said in the introduction to this paper that the approach to meaning that I was going to review had the advantage that the meaning of a sentence is directly connected with aspects of its use. There is an obvious connection between assertions and verifications or valid arguments. Roughly speaking the assertion of a sentence is warranted iff a verification of the sentence is known. A fundamental

requirement on the definition of validity of arguments is that it respects this equivalence: a person should be warranted in asserting a sentence iff she is in the possession of a valid argument for A and knows it to be a valid argument for A.

Does the definition of validity satisfy this fundamental requirement? Consider the case of a simple argument for a closed sentence $A \supset B$, whose skeleton is

$$\dfrac{\dfrac{A}{B}}{A \supset B}$$

Dummett counts this skeleton as a valid argument for $A \supset B$ if there is an effective method M for finding a closed valid argument for B given a closed valid argument for A. As already remarked in Section 4, to be in possession of such a skeleton does not amount to very much, certainly not to be entitled in asserting $A \supset B$. It is true that if we know that it is a valid argument in Dummett's sense, then we know that there exists such a method M. But what Dummett calls an argument, i.e., the skeleton shown above, plays virtually no role here.

This supports my more involved notion of argument, according to which a valid argument for $A \supset B$ whose skeleton is as shown above also contains as a second ingredient a method M that applied to any valid argument for A yields a valid argument for B. To be in possession of an argument is now to be in possession of such a method M. But again it can be said that it is not sufficient to be just in possession of M, we must also know that M is a method which applied to any valid argument for A yields a valid argument for B.

These considerations may be taken to speak in favour of counting a demonstration of the fact that M is such a method as an additional ingredient of a real argument for the truth of $A \supset B$, which was the approach of G. Kreisel (1962). A different response to these concerns is given by Per Martin-Löf (not in the paper by him quoted above but in later papers such as Martin-Löf 1995 and 1998). He separates what he calls proofs or proof objects from demonstrations. A proof (object) is an object in the type theory developed by Martin-Löf, while a demonstration is something which shows that an object is of a specific type. For instance, a canonical proof of $A \supset B$ is an object of the form $\lambda x b(x)$ such that $b(a)$ is a proof of B given that a is a proof of A. What in this way counts as a canonical proof of $A \supset B$ determines the meaning of $A \supset B$. But it is the act of demonstrating that something is a proof

of $A \supset B$ that warrants the assertion of the truth of $A \supset B$. This approach differs from the verificationist idea that meaning is determined by how we establish truths. A more detailed comparison with the approach that I have outlined would take us outside the scope of this essay. We have therefore to leave it at that, although it must be admitted that the problematic feature of my approach noted above has not been resolved here.

The discussion so far has concerned the question whether knowledge of a valid argument for a sentence A is sufficient for the warranted assertion of A. But what about the necessity of such knowledge for being entitled to asserting A? Knowing a valid argument for A implies knowing how to find a valid argument for A in canonical form. But is it right that when we are entitled to assert a complex sentence A, we could in principle have arrived at that position by constructing a canonical argument for A? That the answer is yes is what Dummett (1991) calls the *fundamental assumption* of this approach to meaning. The answer is required to be yes, if the definition of validity is to respect the equivalence stated above between an assertion being warranted and a corresponding valid argument being known.

Dummett (1991) devotes a chapter to a discussion of this fundamental assumption, pointing out reasonable doubts that one can have about it. The doubts have the form of examples of sentences A with predicates that relate to ordinary empirical discourse and where it seems reasonable to say that the assertion of A may be warranted although the speaker knows no valid argument for A (or argument skeleton for A, the examples function equally well regardless which definition we choose).

Some of the examples are related to the fact that when we are concerned with tensed empirical sentences, the possibility of a having direct verification may be lost or may not yet be at hand. It is obvious that the notion of valid argument for empirical sentences has to be more lax than for mathematical ones. We cannot require that the argument is to give us a method for finding a valid canonical argument, but have to be satisfied if it demonstrates for sentences in the past time that a valid canonical argument could have been had at the time in question, and for sentences in future tense that a valid canonical argument will be possible to have at the future time in question.

There are counterexamples that cannot be dealt with in this way however. In my opinion, the most serious ones concern universal

sentences in empirical discourse.[5] As may be expected, the discussions of these examples do not result in a suggestion that the canonical forms of arguments for the various kinds of sentences can be specified in some different way, which would be to replace Gentzen's introduction rules by some other introduction rules. The examples must rather be understood as casting doubts on the whole idea that it is possible to specify canonical forms of arguments (or verifications) such that the truth of a sentence can be identified with the existence of a valid canonical one. In other words, it is the whole verificationist project that is in danger when the fundamental assumption cannot be upheld. An essential prerequisite for this project is the distinction between direct and indirect verification as I have argued elsewhere (e.g., Prawitz 1995).

The discussion in this section indicates that the development of Gentzen's idea into a full theory of meaning along the lines considered here is not unproblematic. However, it should be recalled that here I have essentially confined myself to a review of two closely related lines of thought, and have only in passing considered alternative ways of developing Gentzen's idea or the general idea of approaching meaning via proofs.

NOTES

[1] When in Prawitz (1971) I started to use the term validity in this connection it was defined for derivations in given formal systems. To define it for arguments in general was one of the main ideas of Prawitz (1973).

[2] It is assumed in Dummett (1991) that the stronger notion of hereditary canonical form is needed when one is not confined to justify only given elimination inferences but is considering arbitrary inferences. As follows from the claim made above (easily proved by showing that a closed valid argument in canonical form reduces to one in hereditary canonical form), there is actually no such need.

[3] Most explicitely in for instance Dummett (1976).

[4] As pointed out in connection with proofs already by, e.g., Dummett (1973) and Prawitz (1974).

[5] I have briefly discussed them in for instance Prawitz (1987).

ACKNOWLEDGEMENTS

I thank Professor Per Martin-Löf for comments to an earlier version of the paper.

REFERENCES

Church, A.: 1956, *Introduction to Mathematical Logic*, Princeton University Press, Princeton.

Cozzo, C.: 1994, *Meaning and Argument*, Almqvist & Wiksell, Stockholm.

Dummett, M.: 1973, *The Justification of Deduction*, The British Academy, London (republished in M. Dummett, *Truth and Other Enigmas*, Duckworth, London 1978).

Dummett, M.: 1976, 'What is a Theory of Meaning II', in G. Evans et al. (eds.), *Truth and Meaning*, Oxford, pp. 67–137 (republished in M. Dummett, *The Seas of Language*, Clarendon Press, Oxford 1993).

Dummett, M.: 1983, 'Language and Truth', in R. Harris (ed.), *Approaches to Language*, Oxford (republished in M. Dummett, *The Seas of Language*, Clarendon Press, Oxford 1993).

Dummett, M.: 1991, *The Logical Basis of Metaphysics*, Duckworth, London.

Gentzen, G.: 1934, 'Untersuchungen Über das Logische Schließen', *Mathematische Zeitschrift* **39**, 176–210 and 405–431.

Kreisel, G.: 1962, 'Foundations of Intuitionistic Logic', in E. Nagel et al. (eds.), *Logic, Methodology and Philosophy of Science*, Stanford University Press, Stanford, pp. 198–210.

Martin-Löf, P.: 1985, 'On the Meanings of the Logical Constants and the Justification of the Logical Laws', in *Atti degli Incontri di Logica Matematica*, Siena, **Vol. 2**, pp. 203–281 (reprinted in *Nordic Journal of Philosophical Logic* **1**, 11–60).

Martin-Löf, P.: 1995, 'Verificationism Then and Now', in E. Köhler et al. (eds.), *The Foundational Debate: Complexity and Constructivity in Mathematics and Physics*, Kluwer, Dordrecht, pp. 187–196.

Martin-Löf, P.: 1998, 'Truth and Knowability: on the Principles *C* and *K* of Michael Dummett', in H. G. Dales and G. Olivieri (eds.), *Truth in Mathematics*, Clarendon Press, Oxford, pp. 105–114.

Prawitz, D.: 1971, 'Ideas and Results in Proof Theory', in J. E. Fenstad (ed.), *Proceedings of the Second Scandinavian Logic Symposium*, North-Holland, Amsterdam, pp. 235–307.

Prawitz, D.: 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.

Prawitz, D.: 1974, 'On the Idea of a General Proof Theory', *Synthese* **27**, 63–77.

Prawitz, D.: 1987, 'Dummett on a Theory of Meaning', in B. Taylor (ed.), *Michael Dummett, Contributions to Philosophy*, Kluwer, Dordrecht, pp. 117–165.

Prawitz, D.: 1995, 'Quine and Verificationism', *Inquiry* **37**, 487–494.

Prior, A. N.: 1960, 'The Runabout Inference-Ticket', *Analysis* **24**.

Department of Philosophy, Stockholm University
106 91 Stockholm
Sweden
E-mail: dag.prawitz@philosophy.su.se

PETER SCHROEDER-HEISTER

VALIDITY CONCEPTS IN PROOF-THEORETIC SEMANTICS

ABSTRACT. The standard approach to what I call "proof-theoretic semantics", which is mainly due to Dummett and Prawitz, attempts to give a semantics of proofs by defining what counts as a valid proof. After a discussion of the general aims of proof-theoretic semantics, this paper investigates in detail various notions of proof-theoretic validity and offers certain improvements of the definitions given by Prawitz. Particular emphasis is placed on the relationship between semantic validity concepts and validity concepts used in normalization theory. It is argued that these two sorts of concepts must be kept strictly apart.

1. INTRODUCTION: PROOF-THEORETIC SEMANTICS

Proof-theoretic semantics is an alternative to truth-condition semantics. It is based on the fundamental assumption that the central notion in terms of which meanings can be assigned to expressions of our language, in particular to logical constants, is that of *proof* rather than *truth*. In this sense proof-theoretic semantics is inherently inferential in spirit, as it is the inferential activity of human beings which manifests itself in proofs.

Proof-theoretic semantics has several roots, the most specific one being Gentzen's (1934) remarks that the introduction rules in his calculus of natural deduction define the meanings of logical constants, while the elimination rules can be obtained as a consequence of this definition. More broadly, it is part of the tradition according to which the meaning of a term should be explained by reference to the way it is *used* in our language.

Although the "*meaning as use*" approach has been quite prominent for half a century now and has provided one of the cornerstones of the philosophy of language, in particular of ordinary language philosophy, it has never prevailed in the *formal* semantics of artificial and natural languages. In formal semantics, the *denotational* approach, which starts with interpretations of singular terms and predicates, then fixes the meaning of sentences in terms of truth conditions, and finally defines logical consequence as truth preservation under all interpretations, has

Synthese (2006) 148: 525–571
DOI 10.1007/s11229-004-6296-1

always been dominant. The main reason for this, as I see it, is the fact that from the very beginning, denotational semantics received an authoritative rendering in Tarski's (1933) theory of truth, which combined philosophical claims with a sophisticated technical exposition and, at the same time, laid the ground for model theory as a mathematical discipline. Compared to this development, the "meaning as use" idea was a slogan supported by strong philosophical arguments, but without much formal underpinning.

There has been a lot of criticism of classical model-theoretic semantics from the denotational side itself. Examples are partial logics such as situation semantics, and dynamic approaches such as discourse representation theory and dynamic semantics. These logics reject the idea that *total* information about the world is always available and evaluate formulas with respect to certain information states.[1] Another example is Etchemendy's (1990) critique of classical consequence, which attracted much attention. However, in mainstream semantics, there has never been a fundamental reorientation, which could have turned the "meaning as use" idea into something that resembles a formalized theory.

Proof-theoretic semantics, as a sidestream development, attempts to achieve exactly this. As one would expect, it uses ideas from *proof theory* as a mathematical discipline, similar to the way truth-condition semantics relies on model theory. However, just this is the basis of a fundamental misunderstanding of proof-theoretic semantics. To a great extent, the development of mathematical proof theory has been dominated by the formalist reading of Hilbert's program as dealing with *formal* proofs exclusively, in contradistinction to model theory as concerned with the (denotational) *meaning* of expressions. This dichotomy has entered many textbooks of logic in which "semantics" means model-theoretic semantics and "proof theory" denotes the proof theory of formal systems. The result is that "proof-theoretic semantics" sounds like a contradiction in terms even today.

When I first used this term in the 1980s,[2] it was not very common, although the idea behind it was there in the Swedish school of proof theory established by Prawitz and Martin-Löf (see Kahle and Schroeder-Heister 2006). In the meantime, it has gained some ground and there have been some occasional references to it. Perhaps it will become more popular within general philosophy in the backwater of inferentialist approaches such as Brandom's[3], which more explicitly than ordinary language philosophy attempt to derive

denotational meaning from inferential meaning, i.e., use the idea that meaning is rooted in proofs as their starting point.

Strictly speaking, the formalist reading of proof theory is not any more foreign to the understanding of 'real' argumentation than model theory is to the interpretation of natural language. In order to apply proof-theoretic results, one has to consider formal proofs to be *representations* of proper arguments, just as, in order to apply model-theoretic methods, one has to consider formulas to be *representations* of proper sentences of a natural language like English. English is not *per se* a formal language, and arguments are not *per se* formal derivations. In this sense, the term "proof-theoretic semantics" is not any more provocative than Montague's (1970) conception of "English as a formal language". Both proof-theoretic semantics and model-theoretic semantics are *indirect* in that they can only be applied via a *formal reading* of aspects of natural language. The basic difference lies in what these aspects are: proof-theoretic semantics starts with arguments and represents them by derivations, whereas model-theoretic semantics starts with names and sentences and represents them by individual terms and formulas.

As indicated above, it was the Swedish school of proof theory, which paved the way for a non-formalist philosophical understanding of proofs. Although originally dealing with problems of the proof-theory of formal systems, Prawitz and Martin-Löf soon realized that many of the concepts and methods developed there had a non-technical counterpart when looking at formal proofs as formal representations of "genuine" proofs. In taking Gentzen's remarks on the definitional significance of introduction and elimination rules seriously, they developed the cornerstones of proof-theoretic semantics.

An immediate predecessor of proof-theoretic semantics was Tait (1967), who, in his work on the *convertibility* of terms, developed concepts which are closely related to those later employed in proof-theoretic semantics. Another predecessor was Lorenzen (1955), who, in his *operative logic*, used arbitrary production rules as definitional rules from which, by means of an inversion principle[4], corresponding elimination rules can be obtained.

In this paper I shall deal with *proof-theoretic validity* as one of the basic technical tools developed within proof-theoretic semantics. As this notion was essentially developed by Prawitz, my exposition is to a great extent a re-interpretation and, I hope, an improvement

to his approach. I shall not deal with the broader philosophical background of "anti-realism" and "verificationism" into which the concept of validity may be embedded, but mainly with the technical constructs and their (narrower) philosophical motivation. The reason for this is, besides lack of space, the fact indicated above that the desideratum of proof-theoretic semantics is not so much a *general* philosophical understanding of its position, but the formal development and philosophical clarification of its fundamental concepts. One result of this restriction is that I cannot give Dummett's work the attention it deserves, since his technical notions do not differ considerably from Prawitz's. I am well aware that he has made enormous contributions to the philosophical understanding of proof-theoretic semantics in general. To a considerable extent it is due to his work that the general climate is now more in favour of proof-theoretic semantics than it used to be.[5]

Validity is a property of derivations, or more general "derivation structures", which are considered to be representations of arguments. The format of these derivations is Gentzen-style natural deduction. In defining validity, attempts are made to justify arguments by turning certain proof-theoretic methods and results into semantic conditions, most prominently the following two: (1) Derivations can be simplified (or made more "direct") by certain reduction methods (terminating in *normal* derivations). (2) Assumption-free derivations in normal form are canonical (or "direct") in the sense that they apply an introduction rule in the last step. Valid arguments are then defined as derivation structures which exhibit properties like (1) and (2). However, I shall strictly distinguish between genuine semantic features and technical properties used in normalization proofs. This is extremely important, as Prawitz originally developed his semantic notion of validity along with adapting certain proof-theoretic concepts proposed by Tait and Martin-Löf to proofs of strong normalization. My main criticism of Prawitz will be that in his earlier writings on validity (Prawitz 1971, 1973, 1974) he does not sufficiently distinguish between semantic concepts and concepts used in proofs of (strong) normalization. I shall argue that they differ in fundamental respects.[6]

In spite of much philosophical discussion about meaning and theories of meaning, no thorough investigation of Prawitz's validity concept has been undertaken so far, although this concept is based on very elementary principles which are very close to Gentzen's original programme of justifying natural deduction. This is

why I chose this notion as my topic here. I want to leave open the question of whether validity should be taken as the *ultimate basis* of proof-theoretic semantics. I myself tend to favour a different approach which chooses rules as the unity of semantic investigation. Whereas in proof-theoretic validity in Prawitz's sense, derivations or arguments come first, and rules or consequences are regarded as steps which preserve the validity of arguments, a rule-based approach would first distinguish certain individual proof steps and then compose derivations or arguments from them. Whereas the first approach is *global*, dealing with proofs as a whole and imposing requirements on them, the second approach is *local*, as it interprets individual proof steps without demanding from the very onset that a proof composed of such single steps have special features. The rule-based approach has the advantage that the dependency of global features of arguments on local features of rules can be investigated separately, which makes this approach more flexible and capable of dealing with phenomena such as circular reasoning. Ideas in this direction have been developed in the context of logic programming jointly with Hallnäs[7], and will be dealt with in subsequent work.

This paper starts with recalling Gentzen's characterization of natural deduction and the way this characterization is turned into an inversion principle by Prawitz. The semantic validity concepts proposed are contrasted with concepts used in proofs of (strong) normalization, which were originally introduced by Tait and Martin-Löf. Special emphasis is placed on the difference between these concepts and semantic concepts, by calling those used for normalization "computability" and only the semantic ones "validity". Various forms of validity are defined and compared, among them notions of strict and strong validity which go beyond Prawitz's definitions. These notions are then extended to general derivation structures with arbitrary reductions serving as justifications, where the definition of a justification differs slightly from that of Prawitz. Finally, it is argued that *proof-theoretic* validity and the resulting notion of consequence is different from, and in a sense more specific than, *constructive* validity and consequence based on the notion of a constructive function.

For lack of space, Martin-Löf's meaning theory, which may be correctly viewed as carrying out a whole programme of proof-theoretic semantics, cannot be dealt with here (see e.g. Martin-Löf 1995, 1998). For the particular purpose of elucidating proof-theoretic

validity, this seems to me to be justified, since Martin-Löf's semantics is not explicitly concerned with formal notions of proof-theoretic validity. I cannot discuss Lorenzen's "operative logic" (1955) either, although it is very close to Gentzen's programme (at least "in spirit"). Furthermore, I do not consider categorical approaches to proof-theoretic semantics. The discussion about classical vs. intuitionistic logic is left out as well.[8] Even a rudimentary account of these items would turn this paper into a substantial monograph.

As a general framework, I use the implicational fragment of intuitionistic propositional logic, i.e. positive implicational logic, which suffices to demonstrate and exemplify all basic ideas. An adequate account of implication provides strong guidelines for the handling of other logical operators. Implication is the most complicated propositional operator, sharing crucial properties with universal quantification. The distinction between open and closed derivations, which will turn out to be semantically fundamental, is to a great extent due to its presence. It is intertwined with the notion of "assumption", which Gentzen gave a prominent role in logical calculi, and whose proper treatment is the cornerstone of proof-theoretic semantics.

Logical Preliminaries and Notational Conventions

In this paper, I stick to Prawitz's tree-based proof notation and do not use a term calculus via the Curry–Howard correspondence (although the typed λ-calculus would be a natural candidate). The tree-based proof notation is philosophically more natural, as proof terms obtain their philosophical significance through their reading as codes for "real" proofs.

Following Prawitz, I shall use the following conventions: If a derivation $\mathcal{D}$ ends with A, I shall also write $\begin{matrix}\mathcal{D}\\A\end{matrix}$. If it depends on an assumption B, I shall write $\begin{matrix}B\\\mathcal{D}\end{matrix}$ or $\begin{matrix}B\\\mathcal{D}\\A\end{matrix}$. This means that the notations $\mathcal{D}$, $\begin{matrix}\mathcal{D}\\A\end{matrix}$, $\begin{matrix}B\\\mathcal{D}\end{matrix}$ and $\begin{matrix}B\\\mathcal{D}\\A\end{matrix}$ do not denote different derivations, but just differ in what they make explicit. The *open assumptions* of a derivation are the assumptions on which the end-formula depends. A derivation is called *closed* if it contains no open assumptions, otherwise it is called *open*.

The system of natural deduction I shall use is that described by Gentzen (1934) and Prawitz (1965). Its positive implicational fragment contains only the schemata $\rightarrow$-introduction and $\rightarrow$-elimination (*modus ponens*):

$$\frac{\begin{array}{c}[A]\\ B\end{array}}{A \rightarrow B} \rightarrow \mathrm{I} \qquad \frac{A \rightarrow B \quad A}{B} \rightarrow \mathrm{E}$$

The reduction of a *maximum formula*, which is a conclusion of an application of an introduction inference and at the same time the major premiss of an elimination inference, is in our restricted framework represented as the schema of $\rightarrow$-reduction:

$$\begin{array}{cc} A & \\ \mathcal{D} & \\ B & \mathcal{D}' \\ \hline A \rightarrow B & A \\ \hline B & \end{array} \quad \text{reduces to} \quad \begin{array}{c} \mathcal{D}' \\ A \\ \mathcal{D} \\ B \end{array}$$

Occasionally I shall also refer to the reductions for other connectives as described in Prawitz (1965). These reductions will be called the "standard reductions" (in contradistinction to arbitrary reductions for generalized derivation structures).

A derivation is in normal form if it cannot be further reduced, which means that it contains no maximum formula. Prawitz (1965) showed that by iterated application of reduction steps, every derivation in intuitionistic logic can be normalized, i.e., can be rewritten as a derivation in normal form.[9] One corollary of this result is that every closed derivation in intuitionistic logic can be reduced to one using an introduction rule in the last step, as a closed normal derivation is of exactly that form. I call this the *fundamental corollary* of normalization theory. As seen below, the fundamental corollary is philosophically interpreted by *requiring* that a valid closed derivation be reducible to one using an introduction inference in the last step. In this sense, introduction rules describe the basic meaning-giving inferences.

The normalization result mentioned is also called *weak normalization*. The *strong* normalization result says that any reduction sequence terminates in a normal derivation, regardless of the order in which reductions are performed. Methods used to prove strong normalization have provided the basis for semantic validity concepts.

2. GENTZEN'S PROGRAMME AND PRAWITZ'S INVERSION PRINCIPLE

Proof-theoretic semantics in the sense discussed in this paper goes back to certain programmatic remarks in Gentzen's *Investigations into Natural Deduction*, where he gives a semantic interpretation of his inference rules.

Gentzen's remarks deal with the relationship between introduction and elimination inferences in natural deduction.

> The introductions represent, as it were, the 'definitions' of the symbols concerned, and the eliminations are no more, in the final analysis, than the consequences of these definitions. This fact may be expressed as follows: In eliminating a symbol, we may use the formula with whose terminal symbol we are dealing only 'in the sense afforded it by the introduction of that symbol'. (Gentzen 1934, p. 80)

This cannot mean, of course, that the elimination rules are *deducible* from the introduction rules in the literal sense of the word; in fact, they are not. It can only mean that they can be *justified* by them in some way.

> By making these ideas more precise it should be possible to display the E-inferences as unique functions of their corresponding I-inferences, on the basis of certain requirements. (Gentzen 1934, p. 81)

So the idea underlying Gentzen's programme is that we have "definitions" in the form of introduction rules and some sort of semantic reasoning which, by using "certain requirements", validate the elimination rules.

As indicated in the introduction, I shall not discuss in detail the philosophical reasons which might support Gentzen's programme. For that I would have to refer to Dummett's work and in particular to his claim that there are two different aspects of language use: one connected with 'directly' or 'canonically' asserting a sentence, and another one with drawing consequences from such an assertion.[10] The first is the primary or 'self-justifying' way corresponding to reasoning by introduction rules, whereas the second one, which corresponds to reasoning by elimination rules, is in need of justification. This justification relies on the *harmony* which is required to hold between both aspects: The possible *consequences* to be drawn from an assertion are determined by the *premisses* from which the assertion can possibly be inferred by direct means.

Prawitz, in an "inversion principle"[11] formulated in his classic monograph on *Natural Deduction* of 1965, tried to make Gentzen's remarks more precise in the following way.

> Let α be an application of an elimination rule that has B as consequence. Then, deductions that satisfy the sufficient condition [...] for deriving the major premiss of α, when combined with deductions of the minor premisses of α (if any), already 'contain' a deduction of B; the deduction of B is thus obtainable directly from the given deductions without the addition of α. (Prawitz 1965, p. 33)

Here the sufficient conditions are given by the premisses of the corresponding introduction rules. Thus the inversion principle says that a derivation of the conclusion of an elimination rule can be obtained without an application of the elimination rule if its major premiss has been derived using an introduction rule in the last step, which means that a combination

$$\dfrac{\dfrac{\mathcal{D}}{A}\text{I-inference} \qquad \{\mathcal{D}_i\}}{B}\text{E-inference}$$

of steps, where $\{\mathcal{D}_i\}$ stands for a (possibly empty) list of deductions of minor premisses, can be avoided.

At first glance, this simply states the fact that maximum formulas, i.e. formulas being conclusions of an I-inference and at the same time major premiss of an E-inference (in the example: A), can be removed by means of certain reductions, which leads to the idea of a normal derivation. However, it also represents a *semantical* interpretation of elimination inferences by saying that nothing is gained by an application of an elimination rule if its major premiss has been derived *according to its meaning* (i.e., by means of an introduction rule). So the reductions proposed by Prawitz for the purpose of normalization are at the same time semantic justifications of elimination rules with respect to introduction rules. His inversion principle elaborates Gentzen's idea of "special requirements" needed for this justification, by demanding that elimination rules invert introduction rules in a precise sense.

That it corresponds indeed to what Gentzen had in mind can be seen by taking a closer look at the example Gentzen gives:

> We were able to introduce the formula $A \to B$ when there existed a derivation of B from the assumption formula A. If we then wished to use that formula by eliminating the $\to$-symbol (we could, of course, also use it to form longer formulae, e.g., $(A \to B) \vee C$, $\vee$-I), we could do this precisely by inferring B directly, once A has been proved, for

what $A \to B$ attests is just the existence of a derivation of B from A. (Gentzen 1934, pp. 80–81)

This may be read as follows: Given the situation

$$\dfrac{\dfrac{\begin{array}{c}A\\ \mathcal{D}\\ B\end{array}}{A \to B} \quad \begin{array}{c}\mathcal{D}'\\ A\end{array}}{B}$$

where $\mathcal{D}$ is "a derivation of B from the assumption formula A", and $\mathcal{D}'$ is the derivation showing that "A has been proved", so that we can use $A \to B$ to obtain B "by eliminating the $\to$-symbol". Then by means of

$$\begin{array}{c}\mathcal{D}'\\ A\\ \mathcal{D}\\ B\end{array}$$

we can infer "B directly, once A has been proved [by means of $\mathcal{D}'$]", as "$A \to B$ attests [...] the existence of a derivation [viz. $\mathcal{D}$] of B from A". According to this reading, Gentzen describes the standard reduction for implication later made explicit by Prawitz (1965) and used in his normalization proof.

However, although Gentzen's remarks are correctly read as outlining a semantic programme, he himself takes a more formalistic stance, which is clear from his writings in general and from the continuation of the passage quoted above:

> Note that in saying this we need not go into the 'informal sense' ['inhaltlicher Sinn'][12] of the $\to$-symbol. (Gentzen 1934, p. 81)

Prawitz (1965) deserves credit to have drawn our attention to the genuine semantic content of Gentzen's remarks, though this is not spelled out in detail in his monograph. Only later in Prawitz (1971) and in particular in Prawitz (1973, 1974) is it turned into a full-fledged semantic theory.

3. NORMALIZATION, COMPUTABILITY AND VALIDITY

3.1. *Normalization and Computability*

Normalization plays a prominent role in the formal background of proof-theoretic semantics, in particular the result that normal closed

proofs are in introduction form, i.e., use an introduction inference in the last step.

Of equal importance is a technical method within normalization theory, which is especially used in proofs of *strong* normalization. By means of this method, a certain predicate P of proofs is defined which has the property that it entails (strong) normalizability. The predicate P has some flavour of a semantic predicate, and in a kind of correctness proof it can be shown that every derivation satisfies P, yielding as a corollary that every derivation is (strongly) normalizable. Such a predicate was first defined by Tait (1967) under the name "convertibility" and used to demonstrate (weak) normalizability of terms. Martin-Löf (1971) carried Tait's idea over from terms to derivations and defined a corresponding predicate which he called "computability", proving (weak) normalization for an extension of first-order logic, called the theory of iterated inductive definitions. At the same time, Girard (1971) used this method to prove (weak) normalization for second-order logic. Again at the same time, it was Prawitz (1971) who emphasized its particular usefulness for proving *strong* normalization, calling it "strong validity". Since then, it has served as the basis of proofs of strong normalization for a variety of systems.[13]

In the following I shall speak of *computability* predicates or *the computability* predicate when dealing with this notion as it is used in normalization proofs, thus adopting Martin-Löf's terminology. The term "valid" will be reserved for genuinely semantic notions. I consider the terminology of Prawitz, who speaks of "validity based on the introduction rules" (1971, p. 284) in contradistinction to "validity used in proofs of normalizability" (1971, p. 290), somewhat unfortunate. It is one of the basic claims of this paper that there are fundamental differences between these two concepts.

I restrict Prawitz's notion of computability ("validity used in proofs of normalizability") to positive implicational logic $\mathcal{L}$, i.e., to the system with only introduction and elimination rules for implications as primitive rules of inference. Under this restriction, Prawitz's computability notion is basically the same as Martin-Löf's.

A derivation is in *I-form* if it uses an introduction rule in the last step, i.e., if it is of the form

$$\dfrac{\begin{array}{c}[A]\\ \mathcal{D}\\ B\end{array}}{A \to B}.$$

Using a term employed by Prawitz (1974) and Dummett (1975) and now common in proof-theoretic semantics, such a derivation is also called *canonical*. Let $\mathcal{D} \succ_1 \mathcal{D}'$ mean that the derivation $\mathcal{D}$ reduces to the derivation $\mathcal{D}'$ by applying a single reduction step to a subderivation of $\mathcal{D}$.

DEFINITION OF COMPUTABILITY

(i) A derivation of the form $\dfrac{\begin{matrix}[A]\\ \mathcal{D}\\ B\end{matrix}}{A \to B}$ is *computable*, if for every *computable* $\begin{matrix}\mathcal{D}'\\ A\end{matrix}$, $\begin{matrix}\mathcal{D}'\\ A\\ \mathcal{D}\\ B\end{matrix}$ is *computable*.

(ii) If a derivation $\mathcal{D}$ is not in I-form and is normal, then it is *computable*.

(iii) If a derivation $\mathcal{D}$ is not in I-form and is not normal, then $\mathcal{D}$ is *computable*, if every $\mathcal{D}'$, such that $\mathcal{D} \succ_1 \mathcal{D}'$, is *computable*.

This is a generalized inductive definition. It uses induction on the degree of the end formula of the derivation (clause i), and, within each degree, induction on the reducibility relation[14] (clauses ii and iii).

The proof of strong normalization then proceeds by establishing the following two propositions:

PROPOSITION 1. Every computable derivation is strongly normalizable.

PROPOSITION 2. Every derivation is computable.

Proposition 1 is a (nearly) immediate consequence of the definition of computability. Proposition 2 is based on a kind of correctness proof, verifying step by step that computability is carried over from the premisses to the conclusion of an inference step. Other formulations of "computability" differ slightly from the one given here. However, the basic features remain the same. The resulting normalization proofs all proceed via Propositions 1 and 2.

Computable derivations are closed under substitution with computable derivations, i.e., the following lemma holds:

SUBSTITUTION LEMMA FOR COMPUTABILITY

If $\begin{matrix} A_1 \dots A_n \\ \mathcal{D} \\ B \end{matrix}$ is computable, where all open assumptions of $\mathcal{D}$ are among $A_1, \dots, A_n$, then for any list of computable derivations $\begin{matrix} \mathcal{D}_i \\ A_i \end{matrix}$ $(1 \leq i \leq n)$, $\begin{matrix} \mathcal{D}_1 & & \mathcal{D}_n \\ A_1, \dots, A_n \\ \mathcal{D} \\ B \end{matrix}$ is computable.

Note that the converse direction of the lemma is trivial, as every assumption A_i is itself a normal, and therefore computable, derivation of A_i from A_i.

If closure under substitution with computable derivations is called *computability under substitution*, the lemma says that computability implies computability under substitution.

Weaker versions of computability entail (weak) normalization. Instead of requiring in clause (iii) that *every* $\mathcal{D}'$, such that $\mathcal{D} \succ_1 \mathcal{D}'$, be computable, we might demand that a certain $\mathcal{D}'$, which is obtained from $\mathcal{D}$ *in a particular way* (i.e., by performing a particular reduction step) be computable. This yields the notion defined by Martin-Löf (1971). We might even weaken this by not referring to a particular procedure and just postulate in (iii) that $\mathcal{D}$ reduces to a computable $\mathcal{D}'$, without specifying the procedure in the definition (it must then be specified in the normalization proof, of course).

3.2. *From Computability to Validity*

Validity is a core notion of proof-theoretic semantics. Prawitz introduced it as a semantic predicate for derivations, in analogy to truth as a semantic predicate of propositions in model-theoretic semantics. He developed it in connection with computability predicates, to which it bears a strong resemblance. As his terminology ("strong validity" for "computability" in our sense) suggests, Prawitz actually considers computability and validity to be concepts on one scale, computability being the stronger one. There are several remarks in his 1971, 1973, 1974 papers, where he deals with both notions, which indicate that computability is obtained by augmenting validity, some of them even stating that these extensions make the concept of validity more plausible or convenient.[15] However, Prawitz never explains the exact relationship between these two concepts. In

particular, he never attempts to formally prove that computability (strong validity) implies validity – a result one should expect to hold if the relationship is as simple as the terminology suggests. In his publications after 1974, Prawitz never returns to computability and its relation to validity.

In the following I shall argue that, in spite of many similarities, and contrary to Prawitz's opinion, semantically useful validity notions must differ considerably from computability. Crucial modifications are necessary to turn computability into validity. I shall make the following points:

(1) The notion of computability is not suitable as a foundational semantic notion, because it stipulates normal derivations as computable without further justification.
(2) In order to adjust the notion of computability to serve foundational purposes, *closed* derivations must be given a distinguished role in the justification of irreducible (= normal) derivations.
(3) This distinguished role of closed derivations includes, as a semantic condition, their reducibility to canonical form.

Ad (1): Computability is not a semantic notion

According to clause (ii) in the definition of computability, every normal derivation which is not in I-form, is computable.[16] This could be counted as a semantic clause only if in proof-theoretic semantics we are prepared to consider non-canonical normal derivations as valid *by definition*. However, as we have seen, it is one of the ideas of proof-theoretic semantics in the sense of Gentzen's programme to consider introduction inferences as basic and to *justify* all other inferences by them. In other words, only derivations based on introduction rules should be taken for granted. In any other case the definition of validity should rely on some *justification* procedure rather than on the syntactic form of derivations. This is obviously violated by clause (ii), which simply stipulates irreducible non-canonical derivations as valid. There is no semantic reason whatsoever to consider non-canonical irreducibility as a definition case of validity. According to such a definition, the derivation

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

would be valid *by definition* and not *by justification*, which is not what is intended. *Modus ponens*, as an elimination rule, definitely needs semantic justification. Of course, for the purpose of proving normalizability, clause (ii) is absolutely natural, as normal derivations are trivially (strongly) normalizable. For semantic purposes, however, we would have to argue that non-canonical irreducible derivations have some special status, which exempts them from justification. Since there is no argument at hand to support this, using normal derivations as a starting point in defining validity is an ill-guided approach.

In contradistinction to clause (ii), clauses (i) and (iii) make good semantic sense. In terms of validity, clause (i) says that a canonical derivation of $A \to B$ is valid if its immediate predecessor, a derivation of B from A, provides a way of transferring every valid derivation of A into a valid derivation of B, which corresponds to the meaning one wants to associate with $A \to B$. Furthermore, clause (iii) says that a non-canonical derivation may be considered as valid if it reduces to a valid derivation. This reflects the idea that non-canonical derivations are valid if they reduce to derivations which are already justified as valid (such as canonical ones).

Therefore the *basic flaw* in computability, understood as a semantic notion, is the following implicit assumption:

If $\mathcal{D}$ is non-canonical and irreducible ($=$normal), then $\mathcal{D}$ is valid.

Ad (2): Semantically modified computability: open assumptions and closed derivations

One could try to modify the definition of computability to make it suitable for a definition of semantic validity. This would mean that clause (ii) of the definition is dropped and replaced with something which justifies non-canonical irreducible derivations as valid. An obvious possibility would be to consider such a derivation as valid if the replacement of open assumptions with valid derivations yields a valid derivation of the end formula. This idea would follow the substitution lemma for computability, according to which computability is the same as computability under substitution. More formally, clause (ii) would then read as follows (where we now use the term "valid", as we are dealing with turning the computability notion into a semantic concept):

(ii)* A non-canonical irreducible derivation

$$\begin{array}{c} A_1 \ldots A_n \\ \mathcal{D} \\ B \end{array}$$

where all open assumptions of $\mathcal{D}$ are among $A_1, \ldots, A_n$, is *valid*, if for every list of *valid* derivations $\begin{array}{c}\mathcal{D}_i \\ A_i\end{array}$ $(1 \leq i \leq n)$,

$$\begin{array}{c} \mathcal{D}_1 \quad \mathcal{D}_n \\ A_1 \ldots A_n \\ \mathcal{D} \\ B \end{array}$$

is *valid*.

For example, the one step non-canonical irreducible derivation

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

would be considered as valid, if for each pair of valid derivations $\begin{array}{c}\mathcal{D}_1 \\ A \to B\end{array}$ and $\begin{array}{c}\mathcal{D}_2 \\ A\end{array}$, the derivation $\dfrac{\begin{array}{cc}\mathcal{D}_1 & \mathcal{D}_2 \\ A \to B & A\end{array}}{B}$ is valid. However, a clause like (ii)* would then no longer proceed by induction on the complexity of the end formula but on the complexities of the open assumptions plus that of the end formula, in the example: on the complexities of $A \to B$, A and B. But then the quantification over all valid derivations of the open assumptions is no longer feasible, since these derivations may depend on assumptions of arbitrary complexity. Therefore this is no viable solution.[17]

The way out of this problem used in semantic definitions of validity is to use *closed* valid proofs rather than arbitrary valid proofs as a basis. Instead of (ii)*, one would then propose the following clause.

(ii)** A non-canonical non-reducible derivation

$$\begin{array}{c} A_1 \ldots A_n \\ \mathcal{D} \\ B \end{array}$$

where all open assumptions of $\mathcal{D}$ are among $A_1, \ldots, A_n$, is *valid*, if for every list of *closed valid* derivations $\begin{matrix}\mathcal{D}_i \\ A_i\end{matrix}$ $(1 \leq i \leq n)$,

$$\begin{matrix}\mathcal{D}_1 & \mathcal{D}_n \\ A_1 \ldots & A_n \\ & \mathcal{D} \\ & B\end{matrix}$$

is *valid*.

However, even now we are proceeding by induction on the joint complexity of $A_1, \ldots, A_n, B$ rather than only the complexity of B, even if we only quantify over *closed* valid derivations. This is not compatible with clause (i), where we proceed by induction on the end formula only. In order to cope with that, we would also have to change clause (i) to

(i)** A derivation of the form $\dfrac{\begin{matrix}[A] \\ \mathcal{D} \\ B\end{matrix}}{A \to B}$ is *valid*, if for every *closed valid* $\begin{matrix}\mathcal{D}' \\ A\end{matrix}$, $\begin{matrix}\mathcal{D}' \\ A \\ \mathcal{D} \\ B\end{matrix}$ is *valid*

where this is understood as proceeding by induction on the joint complexity of open assumptions plus end-formula of a derivation.

The definition based on (i)**, (ii)** and (iii) may be called *validity**. In passing from computability to validity* we have interpreted *open assumptions as placeholders for closed derivations*.

Ad (3): The reducibility of closed derivations

Unfortunately, validity* does not yet eliminate the possibility that irreducible (= normal) derivations are considered valid without any further justification. In the case of open derivations, this possibility has been removed, but not so in the case of closed derivations. Suppose $\mathcal{D}$ is a closed non-canonical derivation which is irreducible. Then clause (ii)** applies, and, as there are no open assumptions, $\mathcal{D}$ is (vacuously) valid*.

One might argue that there are no closed non-canonical irreducible derivations. However, this is an accidental property of first-order logic with the standard reductions. Since the notion of validity should in principle be applicable to more general notions of derivations and reductions, the formal possibility of closed non-canonical irreducible derivations must be taken into account. Such a derivation should simply turn out to be invalid by definition. This is accomplished by transforming a corollary of the normalization of proofs into a semantic condition:

A closed non-canonical derivation is valid, if it is reducible to a valid closed canonical derivation.

It was Dummett in particular who repeatedly stressed as a fundamental epistemological principle[18] that, if something is known in an indirect (non-canonical) way, it must be possible to turn this indirect knowledge into direct (canonical) knowledge. This is part of the reason why this sort of semantics is also called verificationist, and it is part of the interpretation of Gentzen's programme of the primacy of introduction rules: In the closed case an I-rule derivation can always be found. With this motivation we arrive at Prawitz's definition of the validity of derivations.

3.3. *Validity of Derivations*

We follow Prawitz (1971) in defining validity with respect to atomic systems S, which are given by production rules for atomic formulas. Let then $\mathcal{L}(S)$ be implicational logic over S, i.e., the system given by introduction and elimination rules for implication plus the production rules of S. We may identify $\mathcal{L}(\mathcal{S})$ with the set of all derivations in this system. A system S' is an *extension* of S ($S' \geq S$) if S' is S itself or results from S by adding further production rules. As a limiting case, we consider the empty atomic system S_0 without any inference rules and with propositional variables as formulas, and correspondingly $\mathcal{L}(S_0)$ as standard implicational logic over propositional variables. Obviously, as a formal system, $\mathcal{L}(\mathcal{S}_0)$ is the same as $\mathcal{L}$. It will turn out that validity with respect to S_0 is the same as *universal* validity when defined in an appropriate way. We say that $\mathcal{D}$ *reduces to* $\mathcal{D}'$ ($\mathcal{D} \geq \mathcal{D}'$), if $\mathcal{D}'$ can be obtained from $\mathcal{D}$ by applying a (finite) number of reduction steps. As a limiting case, $\mathcal{D}$ reduces to itself. In the context of atomic systems, we also extend the notion of a canonical derivation. A canonical derivation of an

atom of S is a derivation in S, whereas, as before, a canonical derivation of a complex formula is a derivation in I-form, i.e., a derivation using an introduction rule in the last step.

Then our first definition of validity corresponding to the one given in Prawitz (1971) runs as follows:

DEFINITION OF S-VALIDITY (1)

(i) For atomic A, a closed derivation of A is *S-valid*, if it reduces to a derivation in S.

(ii) A closed derivation $\begin{matrix}\mathcal{D}\\ A\rightarrow B\end{matrix}$ is *S-valid*, if $\mathcal{D}$ reduces to a derivation of the form $\dfrac{\begin{matrix}[A]\\ \mathcal{D}'\\ B\end{matrix}}{A\rightarrow B}$ such that for every $S'\geq S$ and every closed S'-*valid* $\begin{matrix}\mathcal{D}''\\ A\end{matrix}$, $\begin{matrix}A\\ \mathcal{D}'\\ B\end{matrix}$ is S'-*valid*.

(iii) An open derivation $\begin{matrix}A_1\ldots A_n\\ \mathcal{D}\\ B\end{matrix}$, where all open assumptions of $\mathcal{D}$ are among $A_1,\ldots,A_n$, is *S-valid*, if for every $S'\geq S$ and every list of closed S'-*valid* $\begin{matrix}\mathcal{D}_i\\ A_i\end{matrix}$ $(1\leq i\leq n)$, $\begin{matrix}\mathcal{D}_1\quad\mathcal{D}_n\\ A_1\ldots A_n\\ \mathcal{D}\\ B\end{matrix}$ is S'-*valid*.

This inductive definition proceeds on the joint complexities of the open assumptions and the end formula of the given derivation.

In view of clause (iii), clause (ii) can be changed to

(ii) A closed derivation of $A\rightarrow B$ is *S-valid* if it reduces to a canonical derivation of $A\rightarrow B$ whose immediate subderivation is *S-valid*.

By putting reduction into a clause of its own, the whole definition can then be equivalently stated as follows:

DEFINITION OF S-VALIDITY (2)

(I) Every closed derivation in S is *S-valid*.

(II) A closed canonical derivation of $A\rightarrow B$ is *S-valid*, if its immediate subderivation is *S-valid*.

(III) A closed non-canonical derivation is *S-valid*, if it reduces to an *S-valid* canonical derivation.

(IV) An open derivation $\begin{matrix} A_1 \dots A_n \\ \mathcal{D} \\ B \end{matrix}$, where all open assumptions of $\mathcal{D}$ are among $A_1, \dots, A_n$, is *S-valid*, if for every $S' \geq S$ and for every list of closed S'-*valid* $\begin{matrix} \mathcal{D}_i \\ A_i \end{matrix}$ $(1 \leq i \leq n)$, $\begin{matrix} \mathcal{D}_1 \quad \mathcal{D}_n \\ A_1 \dots A_n \\ \mathcal{D} \\ B \end{matrix}$ is S'-*valid*.

The equivalence of these two definitions of S-validity is easy to prove. Obviously, every (not necessarily closed) derivation in S is S-valid, since every closed S-valid derivation of an atom reduces to a derivation in S. The second definition corresponds to the one proposed by Prawitz (1974, 2006). As explained in the last subsection, the philosophical motivation behind this definition is that, in the closed case, derivations in S as well as introduction steps are self-justifying (clauses I and II), whereas all other steps are justified on the basis that they *reduce* to something which is already justified (clause III), or, in the open case, produce justified closed derivations when combined with such derivations (clause IV).

The reason for considering arbitrary extensions S' of S, is to block arguments for S-validity based on the underivability of certain formulas in S. Otherwise, for example, every derivation in $\mathcal{L}$ starting with a propositional variable as an open assumption, should be counted as S_0-valid, because there is no closed derivation of a propositional variable in S_0. In this sense, the consideration of extensions $S' \geq S$ is a *monotonicity* condition for S-validity. S-valid derivations should remain *S-valid* if one's knowledge incorporated in the atomic system S is increased.[19] In fact, it is easy to show that we have a

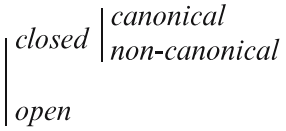
MONOTONICITY THEOREM FOR S-VALIDITY
A derivation $\mathcal{D}$ in $\mathcal{L}(S)$ is S-valid iff for every $S' \geq S$, $\mathcal{D}$ is S'-valid.

Investigating the consequences of permitting non-monotonicity of S-validity is beyond the scope of this paper.

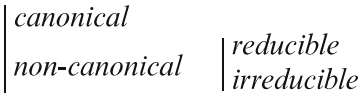
As compared to computability, this definition relies on two crucial insights:

(1) The distinction between closed and open derivations is primary as compared to that between canonical and non-canonical derivations. The latter plays the role of a subdistinction within

closed derivations. In the definition of S-validity, we proceed according to the concept tree



whereas the definition of computability rests on



In S-validity, closed canonical derivations are self-justifying, carrying the burden of semantic justification. In computability, this holds of non-canonical irreducible (= normal) derivations.[20]

(2) The reduction clause for closed derivations (clause III) uses an existence condition corresponding to *weak* normalization, which is again due to the self-justifying character of closed canonical derivations. Whereas in computability, self-justifying derivations are by definition tied to the reducibility concept, viz. as derivations which are *irreducible*, in S-validity self-justifying derivations are defined independently of reducibility and are not trivially available when a derivation is not reducible, which means that we have to postulate their existence as a result of reduction.

For our case of implicational logic we can easily show the following:

SOUNDNESS THEOREM FOR S-VALIDITY
For any S, every derivation in $\mathcal{L}(S)$ is S-valid.

3.4. *Validity and Universal Validity*

Universal validity will be defined for derivations in $\mathcal{L}$. Intuitively, a derivation in $\mathcal{L}$ should be universally valid if it is S-valid for every S. For that, we must interpret derivations of $\mathcal{L}$ in $\mathcal{L}(S)$. Let an S-assignment v be a mapping of propositional variables to S-formulas. Then for an $\mathcal{L}$-derivation $\mathcal{D}$, $\mathcal{D}^v$ is the $\mathcal{L}(S)$-derivation resulting from $\mathcal{D}$ by replacing every propositional variable with the corresponding S-formula assigned to it via v. We can then say that $\mathcal{D}$ is *valid in S under* v, if $\mathcal{D}^v$ is S-valid in the sense defined in the previous section. $\mathcal{D}$ is then called *valid in S* if it is valid in S under every

v, and it is called *universally valid*, if it is valid in S for every S. Now the following can be shown to hold:

PROPOSITION Let $\mathcal{D}$ be a derivation in $\mathcal{L}$. Then $\mathcal{D}$ is universally valid iff $\mathcal{D}$ is S_0-valid.

Proof. We use the fact that when $\mathcal{L}$ is interpreted in $\mathcal{L}(S)$, every extension $S' \geq S$ can be viewed as an interpretation of an extension of S_0 via an assignment. □

Therefore, from now on we shall use the term "valid" terminologically as meaning universal or S_0-validity.

Then as a corollary of the soundness theorem for S-validity we have the following:

SOUNDNESS THEOREM FOR VALIDITY
Every derivation in $\mathcal{L}$ is valid.

As we have a corresponding theorem for computability (Proposition 2), and as we are so far only considering derivations in implicational logic, computability and validity coincide in the sense that any computable derivation (i.e., any derivation in implicational logic) is a valid derivation (i.e., a derivation in implicational logic) and vice versa. So extensionally, computability and validity coincide. We can differentiate between them when we consider more general notions of derivation structures. Then we can give actual counterexamples which show that computability and validity differ not only with respect to their contents, but are in fact *extensionally* different concepts (see Section 6). This further substantiates our claim that, contrary to Prawitz, computability is at best a forerunner to validity but not a semantic concept in itself.

3.5. *Validity Concepts which Imply Normalizability: Strict and Strong Validity*

Our basic semantic argument against computability and for validity was that irreducible derivations should never be counted as valid without further justification, i.e., the implication

irreducible implies valid

should *not* hold by definition. One might, however, expect that the implication

valid implies normalizable

holds.[21] According to the present definition of validity, normalizability is not implied by validity. If we consider intuitionistic logic with no introduction rule for absurdity $\bot$, then according to our definition of validity, $\dfrac{\bot}{\mathcal{D}}$ is vacuously valid for any $\mathcal{D}$ with $\bot$ as the only open assumption, even if $\mathcal{D}$ is not normalizable. Now one might argue that a semantic justification of open derivations in terms of substitution with closed valid derivations should only be applied if the derivation is reduced as far as possible, and not already in a situation, where $\mathcal{D}$ can still be reduced. This means that the substitution justification in clause (IV) of the definition of *S*-validity should be put into action only if all possibilities of obtaining a justification by means of reduction are exhausted, i.e., when the derivation in question is irreducible. Calling this notion "strict *S*-validity" (or "strict validity" [simpliciter] for the universal concept), we reach the following definition:

DEFINITION OF STRICT *S*-VALIDITY

(I) Every closed derivation in S is *strictly S-valid*.
(II) A closed canonical derivation of $A \to B$ is *strictly S-valid*, if its immediate subderivation is *strictly S-valid*.
(III) A closed non-canonical derivation is *strictly S-valid*, if it reduces to a *strictly S-valid* canonical derivation.
(IV) An open reducible derivation is *strictly S-valid*, if it reduces to a *strictly S-valid* derivation.
(V) An open irreducible derivation $\dfrac{\begin{matrix}A_1 \dots A_n\\ \mathcal{D}\\ B\end{matrix}}{}$, where all open assumptions of $\mathcal{D}$ are among $A_1, \dots, A_n$, is *strictly S-valid*, if for every $S' \geq S$ and for every list of closed and *strictly S'-valid* $\dfrac{\mathcal{D}_i}{A_i}$ $(1 \leq i \leq n)$, $\dfrac{\begin{matrix}\mathcal{D}_1 \quad \mathcal{D}_n\\ A_1 \dots A_n\\ \mathcal{D}\\ B\end{matrix}}{}$ is *strictly S'-valid*.

The difference to the definition of *S*-validity is that clause (IV) is split up into clauses (IV) and (V), where the new clause (IV) demands the reduction of reducible open derivations, while the new clause (V) is the old clause (IV), but applied only to the irreducible case. So the conceptual tree of this definition is the following one

$$\begin{array}{|l|l|} \hline \multirow{2}{*}{\textit{closed}} & \textit{canonical} \\ & \textit{non-canonical} \\ \hline \multirow{2}{*}{\textit{open}} & \textit{reducible} \\ & \textit{irreducible} \\ \hline \end{array}$$

which contrasts sharply with computability, where the reducible/irreducible distinction is a subdistinction of non-canonical derivations.

I speak of "strict" rather than "strong" *S*-validity to distinguish it from Prawitz's notion of strong validity, which corresponds to computability, and from associations with strong normalization. Furthermore, I should like to reserve "strong *S*-validity" for a notion defined below for which this association is justified. Strict *S*-validity as considered here is indeed a notion on the same scale as *S*-validity. It is obvious that strict *S*-validity implies *S*-validity, but not necessarily vice versa.[22] The corresponding universal notion of strict validity (simpliciter) is defined as in Subsection 3.4.

Let us define (weak) normalizability inductively as follows:

DEFINITION OF NORMALIZABILITY

(i) Every canonical derivation is *normalizable* if its immediate subderivation is *normalizable*.
(ii) Every non-canonical normal derivation is *normalizable*.
(iii) Every non-canonical reducible derivation is *normalizable*, if it reduces to a *normalizable* derivation.

We can then formulate as a theorem that strict validity implies (weak) normalizability.

THEOREM. Every strictly valid derivation is normalizable.

By *strong S-validity* we denote a further strengthened concept, which implies strong normalization.

DEFINITION OF STRONG *S*-VALIDITY

(I) Every closed derivation in *S* is *strongly S-valid*.
(II) A closed canonical derivation of $A \to B$ is *strongly S-valid*, if its immediate subderivation is *strongly S-valid*.

(III) A closed non-canonical derivation $\mathcal{D}$ is *strongly S-valid*, if $\mathcal{D}$ is reducible, and if every $\mathcal{D}'$, such that $\mathcal{D} \succ_1 \mathcal{D}'$, is *strongly S-valid*.

(IV) An open reducible derivation $\mathcal{D}$ is *strongly S-valid*, if every $\mathcal{D}'$, such that $\mathcal{D} \succ_1 \mathcal{D}'$, is *strongly S-valid*.

(V) An open irreducible derivation $\begin{matrix} A_1 \dots A_n \\ \mathcal{D} \\ B \end{matrix}$, where all open assumptions of $\mathcal{D}$ are among $A_1, \dots, A_n$, is *strongly S-valid*, if for every $S' \geq S$ and for every list of closed and *strongly S′-valid* $\begin{matrix} \mathcal{D}_i \\ A_i \end{matrix}$ $(1 \leq i \leq n)$, $\begin{matrix} \mathcal{D}_1 \quad \mathcal{D}_n \\ A_1 \dots A_n \\ \mathcal{D} \\ B \end{matrix}$ is *strongly S′-valid*.

Obviously, strong S-validity implies strict S-validity.

A corresponding universal notion of strong validity (simpliciter) is defined as in Subsection 3.4. We extend the definition of normalizability to a definition of strong normalizability by replacing "if it reduces to" with "if every derivation it reduces to in a single step is" in clause (iii) of this definition. In analogy with the case of strict validity, we can show that strong validity implies strong normalizability.

THEOREM. Every strongly valid derivation is strongly normalizable.

There are also soundness theorems for strict and strong [S-]validity.

SOUNDNESS THEOREMS FOR STRICT AND STRONG S-VALIDITY
All [S-]derivations are both strictly and strongly [S-]valid.

With strict and strong validity we have obtained concepts which are semantically satisfying and at the same time imply weak and strong normalization, respectively.

4. VALIDITY AND COMPUTABILITY BASED ON ELIMINATION RULES

A central idea of proof-theoretic semantics is to consider one set of rules as basic and justify derivations based on other rules with respect to this first set of rules as valid. The standard approach is to consider the introduction rules as primitive or "self-justifying" (Dummett).

However, as envisaged by Prawitz[23], one might try an approach from the opposite direction, starting with elimination inferences. Prawitz's presentation is very sketchy. I reconstruct it as follows:

According to the I-rule conception, if in $\begin{matrix}\mathcal{D}\\A\end{matrix}$ the formula A is the conclusion of an introduction rule whose premiss derivation is S-valid, then $\mathcal{D}$ is S-valid by definition. If A is not derived by an introduction rule, $\mathcal{D}$ is S-valid if it can be *reduced* to an S-valid derivation. Analogously, one might postulate within an E-rule conception that, if all applications of elimination rules to the end-formula A of $\mathcal{D}$ yield S-valid derivations, then $\mathcal{D}$ is itself S-valid by definition. If no elimination rule can be applied to A, then $\mathcal{D}$ is S-valid if it can be *reduced* to an S-valid derivation. (Obviously, the latter case only arises when A is atomic.)

This suggests the following definition.

DEFINITION OF S-VALIDITY BASED ON ELIMINATION RULES

(I) Every closed derivation in S is S-*valid*$_E$.

(II) A closed derivation $\begin{matrix}\mathcal{D}\\A\to B\end{matrix}$ of $A\to B$ is S-*valid*$_E$, if for every $S'\geq S$ and every closed S'-*valid*$_E$ $\begin{matrix}\mathcal{D}'\\A\end{matrix}$, the (closed) derivation $\dfrac{\begin{matrix}\mathcal{D}\\A\to B\end{matrix}\quad\begin{matrix}\mathcal{D}'\\A\end{matrix}}{B}$ is S'-*valid*$_E$.

(III) A closed derivation $\begin{matrix}\mathcal{D}\\A\end{matrix}$ of an atomic formula A, which is not a derivation in S, is S-*valid*$_E$, if it reduces to a derivation in S.

(IV) An open derivation $\begin{matrix}A_1\ldots A_n\\\mathcal{D}\\B\end{matrix}$, where all open assumptions of $\mathcal{D}$ are among $A_1,\ldots,A_n$, is S-*valid*$_E$, if for every $S'\geq S$ and for every list of closed S'-*valid*$_E$ $\begin{matrix}\mathcal{D}_i\\A_i\end{matrix}$ $(1\leq i\leq n)$, $\begin{matrix}\mathcal{D}_1&&\mathcal{D}_n\\A_1\ldots A_n\\\mathcal{D}\\B\end{matrix}$ is S'-*valid*$_E$.

Clause (IV) is identical with clause (IV) in the definitions of S-validity in Section 3.3, i.e., open assumptions in derivations are

interpreted in the same way as they were previously. Clauses (I) and (III) can be conjoined to form the single clause

(I/III) A closed derivation $\begin{matrix}\mathcal{D}\\ A\end{matrix}$ of an atomic formula A is S-$valid_E$, if it reduces to a derivation in S.

Using the main reductions, it can again be shown that all derivations in $\mathcal{L}(S)$ are S-valid$_E$.

As Prawitz remarks, this approach only works for logical constants with "direct" elimination rules such as $\rightarrow$, $\wedge$ and $\forall$. There is no way to extend this to constants like $\vee$ and $\exists$ with "indirect" elimination rules.

Corresponding to the procedure in Section 3.5, notions of strict S-validity$_E$ and strong S-validity$_E$ can be defined such that strict S-validity$_E$ implies weak normalizability and strong S-validity$_E$ implies strong normalizability.[24]

There is also a corresponding notion of computability based on elimination rules for the purpose of strong normalization proofs. Actually, this notion is more common in today's presentations than computability based on introduction rules, as long as one does not have to deal with $\exists$ or $\vee$. For example, Troelstra and Schwichtenberg (1996) define computability as follows:

DEFINITION OF COMPUTABILITY BASED ON ELIMINATION RULES

(1) For atomic A, $\begin{matrix}\mathcal{D}\\ A\end{matrix}$ is $computable_E$, if $\begin{matrix}\mathcal{D}\\ A\end{matrix}$ is strongly normalizable.

(2) $\begin{matrix}\mathcal{D}\\ A\rightarrow B\end{matrix}$ is $computable_E$, if for every $computable_E$ $\begin{matrix}\mathcal{D}'\\ A\end{matrix}$, $\dfrac{\begin{matrix}\mathcal{D}\\ A\rightarrow B\end{matrix}\quad\begin{matrix}\mathcal{D}'\\ A\end{matrix}}{B}$ is $computable_E$.

Similar to computability based on introduction rules, this notion again has the feature that normal derivations – here even normalizable ones – are considered computable$_E$ without further justification, which is natural for proving normalization, but cannot be used for a semantic foundation.

As a characteristic feature of the defintions of validity$_E$ and computability$_E$, it might be noted that the notion of reduction does not come in until the atomic stage is reached (in the definition of computability$_E$ in the form of a derivation being strongly normalizable).

In the terminology of terms, one might say that everything is played down to the atomic level by means of term application, whereas the I-rule conceptions were based on what corresponds to term substitution.

The approach sketched here is not the only possible and perhaps not even the most genuine way of putting elimination rules first. If one really tried to dualize the I-rule approach by putting "deriving from" rather than "deriving of" in front, one should develop ideas such as the following: A *closed derivation from* $\begin{matrix}A\\ \mathcal{D}\\ B\end{matrix}$ should be a derivation of absurdity from A, and a derivation $\begin{matrix}A\\ \mathcal{D}\\ B\end{matrix}$ should be justified, if, for every closed valid derivation $\begin{matrix}B\\ \mathcal{D}'\end{matrix}$ from B, $\begin{matrix}A\\ \mathcal{D}\\ B\\ \mathcal{D}'\end{matrix}$ is a closed valid derivation from A, etc. This, however, would be in conflict with the asymmetry of derivations, which usually have exactly one end formula, but possibly more than one open assumption. So full dualization would perhaps lead to some variant of a single-premiss/multiple-conclusion logic. A genuine E-rule approach might be desirable if one wanted to logically elaborate ideas like Popper's falsificationism by establishing refutation as the basis of reasoning.[25]

5. DERIVATION STRUCTURES, JUSTIFICATIONS AND ARGUMENTS

The soundness theorems for derivations in $\mathcal{L}$ are interesting metalogical facts. However, of a semantic notion of validity we expect more than that. Validity should be a *distinguishing* feature, telling that some derivations are valid while others are not. This is quite analogous to the notion of truth which states that some propositions are true, whereas others are not true. A result showing that every proposition is true, making truth a general feature of propositions, would be considered inadequate. Similarly, there should be a more general notion of derivation within which the notion of validity determines a *subclass*. It is easy to construct such derivations by simply combining arbitrary rules, not only the rules which belong to $\mathcal{L}$. For example, a single-step derivation in $\mathcal{L}$ of the form

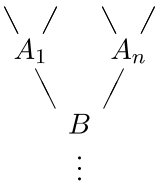
$$\frac{A \to B}{B}$$

should turn out not to be valid, because for certain $S \geq S_0$, not every closed S-valid derivation of $A \to B$ becomes a closed S-valid derivation of B when B is appended at the end of this derivation.[26] This means that we must be able to talk about arbitrary derivations which are not built according to a previously given set of rules. This is important particularly if one would like to pose the question of *completeness*, i.e., the question of whether every valid derivation can be represented in $\mathcal{L}$. As long as [S-]validity is only defined for $\mathcal{L}$ or $\mathcal{L}(S)$, completeness is absolutely trivial. It simply says that every [S-]valid derivation is a derivation, as there are no candidates for derivations which are not in $\mathcal{L}$ or $\mathcal{L}(S)$. This problem does not arise when we are dealing with computability and normalizability only. Computability, as an auxiliary concept to prove normalization, is not necessarily a concept which aims at classifying derivations as computable and non-computable, at least not primarily. In the context of computability, we would simply like to show that *all* derivations exhibit the property of being normalizable.

Since by "derivations" one normally understands derivations in a given system, one should choose a different term for candidates of derivations. I propose talking of *derivation structures*.[27] Hence, the purpose of this section is to define a notion of a derivation structure and of the [S-]validity of derivation structures in such a way that derivations in $\mathcal{L}$ or $\mathcal{L}(S)$ become special derivation structures generated by particular rules of inference. Such a definition will also require generalizing the notion of *reducing* a derivation, which in the standard case is only defined for elimination inferences (in the implicational fragment only *modus ponens*), provided its major premiss results from applying an introduction rule. In principle, reductions should be definable for derivation structures ending with any non-introduction inference.

In order to develop a notion of derivation in a generalized sense, we make use of concepts from the theory of natural deduction and extend them to arbitrary formula trees. A derivation structure over the language of implicational logic (and possibly over atomic systems S as well) can be defined as follows: A *derivation structure* is a *formula tree* together with a *discharge function*. A discharge function for a formula tree is a function which associates with every top formula[28] a formula occurring below (on the same branch in the

tree).[29] The intended reading is the following. Suppose $A_1, \ldots, A_n$ and B occur in the tree as follows:



where each A_i is the value of the discharge function f for top-formulas $C_{i1}, \ldots, C_{im_i}$, i.e. $f(C_{i1}) = \ldots = f(C_{im_i}) = A_i$. Then B is inferred as a conclusion from the premisses $A_1, \ldots, A_n$, where at this application, for each i $(1 \leq i \leq n)$, the assumptions $C_{i1}, \ldots, C_{im_i}$ in the derivation of each A_i are discharged. This means that the step depicted can be viewed as governed by the following inference rule:

$$\frac{\begin{array}{cccc}[C_{11}, \ldots, C_{1m_1}] & & & [C_{n1}, \ldots, C_{nm_n}] \\ A_1 & \ldots & & A_n\end{array}}{B} R.$$

Conversely, an inference rule of the form R can be used to create a step in a derivation structure, where the C_{ij} above the A_i describe the appropriate discharge function $f(C_{i1}) = \ldots = f(C_{im_i}) = A_i$, where if all C_{ij} are missing there is no discharge function with value A_i, and where in the absence of all A_i we are left with B as an axiom. In this way, inference rules can be *extracted* from a derivation structure, and we can check if a given set of inference rules allows us to *generate* this derivation structure. This means that every occurrence of a formula in a derivation structure uniquely *determines* a rule leading to it; in particular, it uniquely determines the rule applied in the last step. This rule is the most specific rule which just describes the derivation step in question. Borrowing a term from the philosophy of science, it may be called the *minimal covering rule* of this derivation step. We may then define a generality order on rules, according to which rules which are more general than the minimal covering rule allow one to generate this derivation step as well. For example, the rule $\dfrac{\begin{array}{c}[C] \\ A\end{array}}{B}$ may be considered more general than the rule $\dfrac{A}{B}$, as it not only allows

one to pass from A to B, but also to discharge the assumption C at the same time. So if $\frac{A}{B}$ is the minimal covering rule of a step in a derivation structure, this step may also be viewed as resulting from the application of $\frac{[C]}{\underline{A}}$. The possible generality orders depend on various parameters. In the given example, it is essential that "vacuous discharging" of assumptions be permitted. This implies in particular that issues of substructural logics may come into play. I cannot discuss these points here. In what follows I shall apply the usual structural conventions common in natural deduction, such as viewing sequences of assumptions $[C_{i1}, \ldots, C_{im_i}]$ as sets, permitting vacuous discharging of assumptions etc.[30]

It should be emphasized that rules are understood as "concrete rules" rather than rule schemata. Others (such as Prawitz 1973, p. 231) would speak of *instances* of rules instead. Thus, when talking of *modus ponens*

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

in a general fashion, I would refer to this as a *rule schema*, whereas, if *modus ponens* for particular formulas A and B is meant, I speak of a *rule*. There are various options for capturing the notion of a rule in relation to that of a derivation structure. In the terminology used here, a rule of the form R might be applied in the last step of different derivation structures. Such a derivation structure may be written as

$$\frac{\begin{array}{cc}[C_{11}, \ldots, C_{1m_1}] & \qquad\qquad [C_{n1}, \ldots, C_{nm_n}] \\ \mathcal{D}_1 & \qquad\qquad \mathcal{D}_n \\ A_1 & \qquad \ldots \qquad A_n\end{array}}{B}$$

(for concrete $\mathcal{D}_1, \ldots, \mathcal{D}_n$) and viewed as an *application* of R. This means that even a (concrete, non-schematic) rule is uniform in the sense that all applications follow the same pattern. A different view would be to define a rule simply as a set of such patterns, meaning that the applications of a rule are in no way structurally related.[31]

According to our definition,

$$\frac{A \to B}{B}$$

is a (very simple) derivation structure. Therefore, once we have defined S-validity and validity for derivation structures, we are in the position to state that this derivation structure is not valid. Now why should

$$\frac{A \to B}{B}$$

be invalid, whereas a corresponding instance

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

of *modus ponens* is obviously valid? Both rules share the feature that they are not (self-justifying) introduction rules. However, for *modus ponens* a reduction procedure is available which helps generate a valid derivation when valid derivations of the premisses are given, whereas for

$$\frac{A \to B}{B}$$

no such procedure is at hand. So non-validity is due to the lack of appropriate reductions. On the other hand, the one-step derivation structure

$$\frac{A \to (B \to C)}{B \to (A \to C)} \qquad (\star)$$

should be counted as valid, even if it is not a derivation in $\mathcal{L}$ and no standard reduction applies. Here, to ensure validity, we must add a new reduction, which is different from the standard reduction. For example, if we use the reduction step

$$\frac{\begin{array}{c}\mathcal{D}\\ A \to (B \to C)\end{array}}{B \to (A \to C)} \quad \text{reduces to} \quad \dfrac{\dfrac{\dfrac{\begin{array}{c}\mathcal{D}\\ A \to (B \to C)\end{array} \quad [A]\,(1)}{B \to C} \quad [B]\,(2)}{C}\,(1)}{A \to C}\,(2)\,, \qquad (\star\star)$$

then the derivation ($\star$) would indeed turn out as valid according to our definition of validity.[32]

So what needs to be changed is not so much the notion of validity itself, but the notion of reduction the definition of validity refers to. This fits in very well with the general idea of reduction. Reductions serve as justifying procedures for non-canonical steps, i.e., for steps, which are not self-justifying. When we consider validity for arbitrary derivation structures, we should not only consider the topological structure of derivations, but also generalize their reductions. This means that a more appropriate concept would be that of a derivation structure *combined* with a set of permitted reductions, which need not coincide with the set of standard reductions used in the normalization of derivations in $\mathcal{L}$.

This is exactly the step taken by Prawitz (1973) and in his later publications (see Prawitz 2006). I present it in modified form, where the modifications do not only affect terminology. By an *argument* I understand a pair $\langle \mathcal{D}, \mathcal{J} \rangle$, where $\mathcal{D}$ is a derivation structure and $\mathcal{J}$ is a *justification* consisting of a set of reductions. This conforms with our previous talking of derivations in a particular system like $\mathcal{L}$, as in such a system certain standard reductions are available.

A *reduction* is a pair

$$\mathcal{D}_1 \rhd \mathcal{D}_2,$$

also read as

$$\mathcal{D}_1 \text{ reduces to } \mathcal{D}_2,$$

which associates a derivation structure $\mathcal{D}_2$ with the derivation structure $\mathcal{D}_1$, such that $\mathcal{D}_2$ has the same end formula as $\mathcal{D}_1$ and no open assumptions beyond those in $\mathcal{D}_1$ (but possibly less assumptions). We say that this reduction is *assigned* to $\mathcal{D}_1$, or that it is a reduction *for* $\mathcal{D}_1$. If it belongs to a justification $\mathcal{J}$, we say that it is *assigned* to $\mathcal{D}_1$ *via* $\mathcal{J}$. If $\mathcal{D}_1$ is an application of a rule R, we call $\mathcal{D}_1 \rhd \mathcal{D}_2$ a reduction *for* R (remember that an application of a rule R is a derivation applying this rule in its last step). By $\mathcal{J}(R)$ or $\mathcal{J}_R$ we denote the subset of $\mathcal{J}$ containing all reductions for R via $\mathcal{J}$. Conversely, if j_R is a set of reductions for R, then we may compose some $\mathcal{J}$ as the union of all j_R for a given set of rules R. It is not expected that j_R comprise reductions for *all* potential applications of R; as a limiting case, j_R may even be empty. Furthermore, it is not excluded that there might be more than one reduction for the same derivation structure. This corresponds to the idea that there might be "alternative justifications" for the same derivation structure. Reductions for

introduction rules are also not excluded in principle, although they are of no real use as introduction rules are self-justifying without any need for reduction. The only constraint we have to impose on a justification $\mathcal{J}$ is that it be closed under substitution in the following sense.

Closure under substitution:

If the reduction

$$\begin{array}{c} A_1 \ldots A_n \\ \mathcal{D} \end{array} \quad \rhd \quad \begin{array}{c} A_1 \ldots A_n \\ \mathcal{D}' \end{array}$$

is in $\mathcal{J}$, then for any $\begin{array}{c}\mathcal{D}_1\\ A_1\end{array}, \ldots, \begin{array}{c}\mathcal{D}_n\\ A_n\end{array}$,

$$\begin{array}{c} \mathcal{D}_1 \quad \mathcal{D}_n \\ A_1 \ldots A_n \\ \mathcal{D} \end{array} \quad \rhd \quad \begin{array}{c} \mathcal{D}_1 \quad \mathcal{D}_n \\ A_1 \ldots A_n \\ \mathcal{D}' \end{array}$$

is in $\mathcal{J}$ as well.

So a justification $\mathcal{J}$ in our sense is nothing but a proof reduction system, for which closure under substitution holds.[33] If $\mathcal{J}$ is a justification, then $\mathcal{J}'$ is called an extension of $\mathcal{J}$ ($\mathcal{J}' \geq \mathcal{J}$) if $\mathcal{J}'$ results from $\mathcal{J}$ by adding reductions such that closure under substitution continues to hold. In other words, an extension of $\mathcal{J}$ is any superset of $\mathcal{J}$ which is itself a justification.

This definition differs considerably from Prawitz's, as he uses a so-called "consistency" requirement for justifications which restricts the formation of extensions, and perhaps even disallows alternative reductions for the same derivation structure. As I see it, this consistency requirement plays a role only if strong normalization is the aim, which is not in the centre of interest when the semantic concept of validity is defined. Adding reductions for the same derivation structure may only be detrimental to *strong* validity, as this introduces new reduction sequences not previously considered.

Normally, the set j_R of reductions for R will be given schematically, which means that it does not depend on the particular application of R. When R is an instance of a uniform rule schema, the reductions for R are often given schematically in the more general sense that they are independent of the particular formula which

occurs in R, as in the case of the standard reduction for implication

$$mp: \quad \dfrac{\begin{array}{c} A \\ \mathcal{D} \\ B \end{array}}{A \to B} \quad \dfrac{\mathcal{D}'}{A} \quad \rhd \quad \begin{array}{c} \mathcal{D}' \\ A \\ \mathcal{D} \\ B \end{array}.$$

(If we wanted to specialize this to particular formulas A and B, we would write $mp_{A \to B}$.) However, this is not mandatory, and it is not excluded that different applications of the same rule or applications of different rules which are instances of the same rule schema receive entirely different reductions. Only with respect to the substitution of derivation structures for open assumptions are reductions schematic, as required by closure under substitution.

A derivation structure $\mathcal{D}$ *reduces in one step* to a derivation structure $\mathcal{D}'$ ($\mathcal{D} \succ_1 \mathcal{D}'$), if $\mathcal{D}'$ results from $\mathcal{D}$ by finitely often applying a reduction to a substructure of $\mathcal{D}$. Here, a substructure of $\mathcal{D}$ is a subtree of $\mathcal{D}$ with the discharge function f restricted to assumptions whose values under f occur above the end formula of $\mathcal{D}$. A derivation structure $\mathcal{D}$ *reduces to* a derivation structure $\mathcal{D}'$ ($\mathcal{D} \succeq \mathcal{D}'$), if $\mathcal{D}'$ is identical with $\mathcal{D}$ or results from $\mathcal{D}$ by a finite number of one-step reductions. It is important to note that reductions apply to derivation structures, given a justification $\mathcal{J}$. So we could more explicitly write $\mathcal{D} \succeq_{\mathcal{J}} \mathcal{D}'$ rather than $\mathcal{D} \succeq \mathcal{D}'$. Reductions cannot change or generate justifications, which means that a notion such as $\langle \mathcal{D}, \mathcal{J} \rangle \succeq \langle \mathcal{D}', \mathcal{J}' \rangle$ is not defined.

Let $\mathcal{L}^*(S)$ be the *logic of arguments* over S, which may be identified with the set of arguments $\langle \mathcal{D}, \mathcal{J} \rangle$, where the derivation structure $\mathcal{D}$ is built up from implicational formulas over formulas of S as atoms, and $\mathcal{J}$ is a justification whose reductions are defined for such derivation structures. As a limiting case, we again have $\mathcal{L}^*(S_0)$, in short $\mathcal{L}^*$, which uses only propositional variables as atoms. Standard implicational logic $\mathcal{L}$ would then be obtained by considering the set of all $\langle \mathcal{D}, \mathcal{J} \rangle$ such that $\mathcal{D}$ is a derivation in standard implicational logic, whereas $\mathcal{J}$ is fixed for all derivations and comprises exactly the standard reductions.

Then the S-validity of arguments $\langle \mathcal{D}, \mathcal{J} \rangle$, which is the same as the S-validity of derivation structures $\mathcal{D}$ with respect to justifications $\mathcal{J}$, is defined as follows:

DEFINITION OF *S*-VALIDITY FOR ARGUMENTS

(I) Every closed derivation in S is *S-valid* with respect to $\mathcal{J}$ (for every $\mathcal{J}$).

(II) A closed derivation structure $\dfrac{\begin{array}{c}A\\ \mathcal{D}\\ B\end{array}}{A \to B}$ is *S-valid* with respect to $\mathcal{J}$, if its immediate substructure $\begin{array}{c}A\\ \mathcal{D}\\ B\end{array}$ is *S-valid* with respect to $\mathcal{J}$.

(III) A closed non-canonical derivation structure is *S-valid* with respect to $\mathcal{J}$, if it reduces, with respect to $\mathcal{J}$, to a canonical derivation structure, which is *S-valid* with respect to $\mathcal{J}$.

(IV) An open derivation structure $\begin{array}{c}A_1 \ldots A_n\\ \mathcal{D}\\ B\end{array}$, where all open assumptions of $\mathcal{D}$ are among $A_1, \ldots, A_n$, is *S-valid* with respect to $\mathcal{J}$, if for every $S' \geq S$ and $\mathcal{J}' \geq \mathcal{J}$, and for every list of closed derivation structures $\begin{array}{c}\mathcal{D}_i\\ A_i\end{array}$ $(1 \leq i \leq n)$, which are *S′-valid* with respect to $\mathcal{J}'$, $\begin{array}{c}\mathcal{D}_1 \quad \mathcal{D}_n\\ A_1 \ldots A_n\\ \mathcal{D}\\ B\end{array}$ is *S′-valid* with respect to $\mathcal{J}'$.[34]

In clause (IV), the reason for considering extensions $\mathcal{J}' \geq \mathcal{J}$ of justifications, in addition to extensions $S' \geq S$ of atomic systems, is again, in my view, a monotonicity constraint. It is obvious that the following holds:

MONOTONICITY OF *S*-VALIDITY (FOR ARGUMENTS)

An argument $\langle \mathcal{D}, \mathcal{J} \rangle$ in $\mathcal{L}^*(S)$ is S-valid iff for every $S' \geq S$ and $\mathcal{J}' \geq \mathcal{J}$, $\langle \mathcal{D}, \mathcal{J}' \rangle$ is S'-valid.

The corresponding universal concept is then defined as follows: If v is an assignment of S-formulas to propositional variables, then for a $\mathcal{J}$ comprising reductions for arguments in $\mathcal{L}^*$, $\mathcal{J}^v$ is defined as the set of reductions which acts on derivations $\mathcal{D}^v$ in the same way as $\mathcal{J}$ acts on $\mathcal{D}$ (i.e., $\mathcal{J}^v$ is the homomorphic image of $\mathcal{J}$ under v). Then an argument $\langle \mathcal{D}, \mathcal{J} \rangle$ in $\mathcal{L}^*$ is defined *universally valid* iff for every S and every v, $\langle \mathcal{D}^v, \mathcal{J}^v \rangle$ is S-valid. Again we can prove:

PROPOSITION Let $\langle \mathcal{D}, \mathcal{J} \rangle$ be in $\mathcal{L}^*$. Then $\langle \mathcal{D}, \mathcal{J} \rangle$ is universally valid iff $\langle \mathcal{D}, \mathcal{J} \rangle$ is S_0-valid.

This means that we can continue to use the term "valid" (now with respect to some $\mathcal{J}$) interchangeably for both universal and S_0-validity.

It is obvious how notions of strict S-validity for $\mathcal{L}^*(\mathcal{S})$ and of strict validity for $\mathcal{L}^*$ can be defined. We can also prove (weak) normalization from strict validity. However, as mentioned above, for strong validity, problems arise with our unrestricted notions of justifications $\mathcal{J}$ and extensions $\mathcal{J}' \geq \mathcal{J}$.

Suppose $\mathcal{J}$ is the set of standard reductions. Then it is obvious that all derivations in $\mathcal{L}(S)$ are valid arguments with respect to $\mathcal{J}$, and all derivations in $\mathcal{L}$ are universally valid with respect to $\mathcal{J}$. So the "new" concept of validity is a generalization of the "old" concept, which yields the same results for derivations in standard implicational logic.

The basic difference between derivations in the old sense and arguments is, of course, that soundness no longer holds in every case; it simply depends on the justifications provided, as was intended by the introduction of the general notion of an argument.

Returning to our previous example, we can now specify what is meant by the validity of the one-step derivation

$$\frac{A \to (B \to C)}{B \to (A \to C)}. \qquad (\star)$$

This derivation is obviously valid with respect to the standard reductions of implicational logic extended with the reduction given by $(\star\star)$.[35] We may ask whether completeness of intuitionistic logic, or at least of minimal or positive implicational logic holds in the sense that for any derivation structure, which *can be justified* as valid (i.e., which is valid with respect to *some* justification), a derivation in $\mathcal{L}$ of its end formula from its open assumption formulas can be found. That this is indeed the case was posed by Prawitz as a conjecture (1973, p. 246), without his being able to indicate so far how it might be proved.

In addition to *validity* in the sense sketched here, Prawitz also defines a notion of *computability* for arguments, which he (unfortunately) calls *strong validity*. It is not surprising that he is able to establish strong normalization of minimal logic with respect to the general context of arbitrary justifications, given a notion of "consistent" extensions of justifications. I cannot present this here. From

my point of view, his general computability concept suffers from the same defect as did the less general computability concept dealt with in Sections 3 and 4. Again, Prawitz has to consider irreducible non-canonical arguments as strongly valid, the only difference being that irreducibility is now taken with respect to a justification $\mathcal{J}$, which is not confined to the standard reductions (1973, p. 239; 1974, p. 74).

6. THE RELATIONSHIP BETWEEN COMPUTABILITY, VALIDITY AND NORMALIZABILITY: COUNTEREXAMPLES

In Section 3, we claimed that computability and validity are crucially different, particularly by arguing that normal derivations need to be justified semantically. However, at that stage we were not able to give counterexamples establishing this difference, as extensionally the concepts were identical, comprising all derivations in $\mathcal{L}$. Now with respect to the generalized concept of an argument, we can provide counterexamples.

We understand the computability of an argument $\langle \mathcal{D}, \mathcal{J} \rangle$, i.e. the computability of $\mathcal{D}$ with respect to a justification $\mathcal{J}$, in the following sense, which leads to weak normalization, and compare it with the validity of $\mathcal{D}$ with respect to $\mathcal{J}$.

DEFINITION OF (WEAK) COMPUTABILITY OF ARGUMENTS

(i) A derivation structure of the form $\dfrac{\begin{matrix}[A]\\ \mathcal{D}\\ B\end{matrix}}{A \to B}$ is *computable* with respect to $\mathcal{J}$, if for every $\mathcal{J}' \geq \mathcal{J}$ and every $\begin{matrix}\mathcal{D}'\\ A\end{matrix}$ *computable* with respect to $\mathcal{J}'$, $\begin{matrix}\mathcal{D}'\\ A\\ \mathcal{D}\\ B\end{matrix}$ is *computable* with respect to $\mathcal{J}'$.

(ii) If a derivation structure $\mathcal{D}$ is not in I-form and is normal (= irreducible) with respect to $\mathcal{J}$, then it is *computable* with respect to $\mathcal{J}$.

(iii) If a derivation is not in I-form and is not normal with respect to $\mathcal{J}$, then $\mathcal{D}$ is *computable* with respect to $\mathcal{J}$, if $\mathcal{D}$ reduces with respect to $\mathcal{J}$ to a $\mathcal{D}'$ which is *computable*.

Counterexample 1: Computability of $\langle \mathcal{D}, \mathcal{J} \rangle$ does not imply validity of $\langle \mathcal{D}, \mathcal{J} \rangle$

We construct an argument $\langle \mathcal{D}, \mathcal{J} \rangle$ in such a way that $\langle \mathcal{D}, \mathcal{J} \rangle$ is closed, non-canonical and normal, and therefore computable, but not valid. Choose a closed non-canonical derivation structure $\mathcal{D}$ ending with a non-atomic formula, e.g.,

$$\dfrac{\dfrac{\overset{(1)}{[B \to C]}}{(B \to C) \to (B \to C)}\,(\to \mathrm{I})^{(1)}}{B \to C}.$$

Choose $\mathcal{J}$ in such a way that with $\mathcal{D}$ no reduction is associated (e.g., take $\mathcal{J}$ to be empty). Then $\langle \mathcal{D}, \mathcal{J} \rangle$ is computable, because it is irreducible. However, $\langle \mathcal{D}, \mathcal{J} \rangle$ is not valid, because it cannot, as required for validity, be reduced to a canonical derivation structure, since no reduction for $\mathcal{D}$ is available in $\mathcal{J}$.

Counterexample 2: Validity of $\langle \mathcal{D}, \mathcal{J} \rangle$ does not imply computability of $\langle \mathcal{D}, \mathcal{J} \rangle$

We consider $\langle \bot, \to \rangle$-logic, i.e., a system with a logical constant $\bot$, for which there is no introduction rule. In such a system, the derivation $\dfrac{\bot}{A}$, and therefore $\dfrac{\dfrac{\overset{(1)}{[\bot]}}{A}}{\bot \to A}{}^{(1)}$ is valid with respect to any $\mathcal{J}$. Now for some B, let $\mathcal{J}$ be chosen in such a way that $\dfrac{B}{\bot}$ is irreducible. Let $\mathcal{J}$ furthermore be chosen such that $\dfrac{\dfrac{B}{\bot}}{A}$ reduces to itself with respect to $\mathcal{J}$, i.e., the reduction for $\dfrac{\dfrac{B}{\bot}}{A}$ is non-terminating. Then $\dfrac{\dfrac{\overset{(1)}{[\bot]}}{A}}{\bot \to A}{}^{(1)}$ is not computable with respect to $\mathcal{J}$, because for computable $\dfrac{B}{\bot}$, $\dfrac{\dfrac{B}{\bot}}{A}$ is not computable (with respect to $\mathcal{J}$).[36]

It can easily be seen that these counterexamples also hold for strict validity instead of validity. Furthermore, they hold for strong validity, when computability is defined in the strong sense (demanding in clause (iii) that all one-step reductions lead to computable derivations).

It might be added that Counterexample 1 is at the same time a counterexample showing that normalizability does not imply validity.[37] Similarly, Counterexample 2 shows that normalizability does not imply computability. The latter is not surprising, as computability is a stronger concept than normalizability, using infinite branching when quantifying over substitution instances of open derivation structures.

7. LOGICAL CONSEQUENCE AND THE VALIDITY OF INFERENCE RULES

It is natural that the S-validity of an inference rule

$$\frac{A_1 \quad \ldots \quad A_n}{A}$$

with respect to a justification $\mathcal{J}$, should mean that the one-step derivation structure of the same form is S-valid with respect to $\mathcal{J}$. We can even define the S-validity of an inference rule which allows the discharging of assumptions, such as the generalized rule

$$\frac{\begin{array}{cccc}[C_{11}, \ldots, C_{1m_1}] & & & [C_{n1}, \ldots, C_{nm_n}] \\ A_1 & & \ldots & A_n\end{array}}{B}.$$

This rule is called *S-valid* with respect to $\mathcal{J}$, if for all $S \geq S_0$, all $\mathcal{J}' \geq \mathcal{J}$, and every list of derivation structures $\begin{array}{c}C_{i1}, \ldots, C_{im_i} \\ \mathcal{D}_i \\ A_i\end{array}$ $(1 \leq i \leq n)$, which are *S-valid* with respect to $\mathcal{J}'$, the derivation structure $\dfrac{\mathcal{D}_1 \quad \ldots \quad \mathcal{D}_n}{B}$ is *S-valid* with respect to $\mathcal{J}'$.

This gives rise to a corresponding notion of consequence.[38] Instead of saying that the rule

$$\frac{A_1 \quad \ldots \quad A_n}{A}$$

is S-valid with respect to $\mathcal{J}$, we may say that A is a *consequence of $A_1, \ldots, A_n$ with respect to S and $\mathcal{J}$* $(A_1, \ldots, A_n \models_{S,\mathcal{J}} A)$; if we

consider universal validity with respect to $\mathcal{J}$, we say speak of *consequence with respect to* $\mathcal{J}$ $(A_1, \ldots, A_n \models_{\mathcal{J}} A)$; and finally, if there is some $\mathcal{J}$ such that universal validity holds for $\mathcal{J}$, then we may speak of *logical consequence* $(A_1, \ldots, A_n \models A)$. Corresponding to the case of rules discharging assumptions, we obtain a notion of consequence

$$\Gamma_1 \Rightarrow A_1, \ldots, \Gamma_n \Rightarrow A_n \models_{S,\mathcal{J}} A$$

for sets of formulas Γ_i. This is to express that the rule

$$\frac{\begin{array}{ccc}[\Gamma_1] & & [\Gamma_n] \\ A_1 & \ldots & A_n\end{array}}{A}$$

is S-valid with respect to $\mathcal{J}$, i.e., we have some notion of implication in the antecedent of $\models$, which is independent of whether the logical constant of implication is available in our language.

This goes crucially beyond any classical notion of consequence. In proof-theoretic semantics, we use (mostly implicitly) some *structural* notion of implication throughout, which is due to the fact that rules can discharge assumptions. As a structural concept it is comparable to the comma as a structural conjunction. This structural notion of implication ("$\Rightarrow$" in my terminology) has been used in generalized concepts of inference rules. It is also important for the formulation of a basic sequent calculus in theories of definitional reflection (see Hallnäs 1991, 2006, Schroeder-Heister 1991b, 1993).

It should be emphasized that it is extremely misleading to write a valid rule or consequence as

$$\frac{A_1 \quad \ldots \quad A_n}{A}\, j$$

with j being understood as the justification of the step from $A_1, \ldots, A_n$ to A. In simple (or "direct") cases like *modus ponens*

$$\frac{A \to B \quad A}{B}\, mp$$

the reduction mp (which is actually a reduction schema) is indeed a justification of this single step. However, in a case like

$$\frac{A \to (B \to C)}{B \to (A \to C)}\, j$$

with j being the reduction given by ($\star\star$), j alone does not suffice to justify this step, as the result of using j, given a valid derivation of the premiss $A \to (B \to C)$, uses *modus ponens*. So the result of applying j is valid only if the *modus ponens* reduction mp is available. This again means that the step

$$\frac{A \to (B \to C)}{B \to (A \to C)}$$

is justified only with respect to some $\mathcal{J}$, where $\mathcal{J}$ comprises both j and the *modus ponens* reduction mp. Thus

$$\frac{A \to (B \to C)}{B \to (A \to C)}\{j, mp\}$$

or

$$A \to (B \to C) \models_{\{j,mp\}} B \to (A \to C)$$

would be an appropriate notation. What is involved in the justification of single inference steps is often a whole reduction system, not a single justifying reduction.

This makes *proof-theoretic* consequence differ from *constructive* consequence according to which

$$\frac{A_1 \quad \ldots \quad A_n}{A}$$

might be defined as valid with respect to a *constructive function* f, if f transforms valid arguments of the premisses $A_1, \ldots, A_n$ into a valid argument of the conclusion A. Actually, it is not always possible to extract such a constructive function from our proof reduction system, as a reduction system $\mathcal{J}$ serving as a justification need not be deterministic, which means that it merely generates a constructive relation on arguments. In any case, the notion of a justification as a proof reduction system presents an intensional analysis of the transformation of arguments which is more fine-grained and more specific than approaches based on the abstract notion of a constructive function.

ACKNOWLEDGEMENTS

I should like to thank Dag Prawitz for many discussions on the topics of this paper, and for detailed comments on a previous version.

NOTES

[1] See Muskens et al. (1997), Blamey (2002) and the references therein.
[2] First in print in Schroeder-Heister (1991c).
[3] See especially Brandom (2000), where the relationship to Dummett's and Gentzen's approaches is expressed very clearly.
[4] Actually, the term "inversion principle" was coined by Lorenzen.
[5] For similar reasons I do not deal with projects like that of Tennant, who combines an anti-realist meaning theory with an alternative approach to relevant logic (see Tennant 1987, 1997).
[6] In his later writings, in which he focuses on semantic aspects, Prawitz does not explicitly return to the relationship with normalization.
[7] See Hallnäs (1991, 2006), Hallnäs and Schroeder-Heister (1990, 1991), Schroeder-Heister (1991a, 1992, 1993, 1994b).
[8] Tait (2006) presents some ideas on how to deal with classical logic in proof-theoretic semantics.
[9] In addition to Prawitz (1965), the monographs by Tennant (1978), Troelstra and Schwichtenberg (1996) and Negri and von Plato (2001b) can be recommended as introductory references.
[10] See especially Dummett (1991).
[11] Named following Lorenzen (1955).
[12] Quotes by Gentzen.
[13] See the recent paper by Joachimski and Matthes (2003), which contains many references to the literature.
[14] More precisely: induction given by the operator associating with a set of derivations X of a formula the set of those derivations which reduce in one step to a derivation in X.
[15] See Prawitz (1971, p. 289); Prawitz (1973, p. 238).
[16] In other renderings of computability, all normal derivations are computable by definition, not only those which are not in I-form. For the definition of computability chosen here, this follows as a lemma.
[17] It is bound to fail due to the impredicative character of the substitution lemma, when it is turned into a definition. "Impredicative" here means that computability is defined by quantifying over all substitution instances obtained by substituting arbitrary computable derivations.
[18] In the context of natural deduction derivations it is called the "fundamental assumption", see Dummett (1991, p. 254).
[19] I suppose that Prawitz had something similar in mind (see Prawitz 1971, p. 276). In later papers he ceases to consider extensions $S' \geq S$, considering only extensions of justifying procedures (see Section 5).
[20] This does not mean that the S-validity of closed and of open derivations is defined separately. These two cases occur intertwined in the same derivation. This is due to the fact that the immediate subderivation of a closed canonical derivation of $A \rightarrow B$ is a derivation of B *from* the assumption A.
[21] This is not exactly the converse "*valid implies normal*", which is, of course, wrong.
[22] Again some emphasis has to be placed on "necessarily", as in the case of intuitionistic logic, all derivations are strictly *S-valid*, i.e., strict S-validity and S-validity coincide in this case.

[23] Prawitz (1971, p. 289f. [= appendix A.2]).
[24] However, in the case of strict S-validity$_E$ (but not in the case of strong S-validity$_E$), we would have to distinguish between reducible and irreducible derivations not only in the open case, but also in the closed case, i.e., clause (II) should only be applicable if $\begin{array}{c}\mathcal{D}\\ A\to B\end{array}$ has been reduced as far as possible, meaning that it is irreducible. Otherwise, we cannot prove that $\begin{array}{c}\mathcal{D}\\ A\to B\end{array}$ is weakly normalizable given that $\dfrac{\begin{array}{cc}\mathcal{D} & \mathcal{D}'\\ A\to B & A\end{array}}{B}$ is weakly normalizable. (In the case of strong normalizability this is trivial.)
[25] Dummett (1991, Ch. 13, pp. 283–286) attempts to develop some kind of a "genuine" E-rule approach (within the standard setting of derivations with multiple premisses and single conclusions).
[26] For example, if A and B are propositional variables, we may choose S as having no axiom and $A \Rightarrow B$ as the only inference rule. Then there is an S-valid derivation of $A\to B$, but no S-valid derivation of B.
[27] Prawitz (1973) speaks of "argument schemata" (with arguments being closed argument schemata), Prawitz (2006) of "argument skeletons" (with arguments being argument skeletons together with justifications).
[28] More precisely, we should talk of top formula *occurrences*. I do not always terminologically distinguish between formulas and their occurrences. It will always be clear from the context what is meant.
[29] The use of discharge functions was introduced by Prawitz (1965, pp. 20–31). Here it is used in the generalized form as proposed in Schroeder-Heister (1984a).
[30] A corresponding notion of *rule* and *derivation structure* is spelled out in detail in Schroeder-Heister (1984a,b).
[31] Prawitz (1973, p. 31) follows such a general approach, calling $(\langle \mathcal{D}_1, \ldots, \mathcal{D}_n\rangle, A)$ an inference (= rule instance), whenever $\dfrac{\mathcal{D}_1 \quad \ldots \quad \mathcal{D}_n}{A}$ is a derivation structure.
[32] Step ($\star$) is reduced not directly, but indirectly by invoking *modus ponens* in the reduction result. See the final remarks in Section 7.
[33] Technically, this proof reduction system can be viewed as a higher-order term rewriting system (it is of higher order due to the assumption structure corresponding to λ-abstraction).
[34] See Prawitz (1973, p. 236; 1974, p. 73; 2006). Prawitz does not consider extensions of atomic systems S.
[35] It is not valid with respect to ($\star\star$) alone – see the final remarks in Section 7.
[36] The intuitive reason for this behaviour is the following: $\dfrac{\bot}{A}$ is always valid as there is no *closed* valid derivation of $\bot$. However, for open normal derivations $\dfrac{B}{\bot}$, the reduction of $\dfrac{\dfrac{B}{\bot}}{A}$ can be made non-terminating by means of an appropriate $\mathcal{J}$. (Note that we do not choose $\dfrac{B}{\bot}$ to be simply $\bot$, because then the example would not work for *strict* validity, as the reduction for $\dfrac{\bot}{A}$ would not terminate.)

[37] Actually, normalizability is implied by computability, but this fact is not used in the counterexample.

[38] See also Prawitz (1985).

REFERENCES

Blamey, S.: 2002, 'Partial Logic', in D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, *2nd edn.*, Volume 5, Kluwer, Dordrecht, pp. 261–353.

Brandom, R.B.: 2000, *Articulating Reasons: An Introduction to Inferentialism*, Harvard University Press, Cambridge Mass.

Dummett, M.: 1975, 'The Philosophical Basis of Intuitionistic Logic', in H. E. Rose and J. C. Shepherdson (eds.), *Logic Colloquium '73*, North Holland, Amsterdam, pp. 5–40, repr. in M. Dummett, *Truth and Other Enigmas*, Duckworth, London 1978, pp. 215–247.

Dummett, M.: 1991, *The Logical Basis of Metaphysics*, Duckworth, London.

Etchemendy, J.: 1990, *The Concept of Logical Consequence*, Harvard University Press, Cambridge Mass.

Gentzen, G.: 1934, 'Untersuchungen über das logische Schließen', *Mathematische Zeitschrift* **39** (1934/35), 176–210, 405–431, English translation ('Investigations into Logical Deduction') in M. E. Szabo (ed), *The Collected Papers of Gerhard Gentzen*, North Holland, Amsterdam 1969, pp. 68–131. Quotations are according to Szabo's translation.

Girard, J.-Y.: 1971, 'Une extension de l'interprétation de Gödel à l'analyse, et son application à l'élimination des coupures dans l'analyse et la théorie des types', in J. E. Fenstad (ed.), *Proceedings of the 2nd Scandinavian Logic Symposium (Oslo 1970)*, North Holland, Amsterdam, pp. 63–92.

Hallnäs, L.: 1991, 'Partial Inductive Definitions', *Theoretical Computer Science* **87**, 115–142.

Hallnäs, L.: 2006, 'On the Proof-Theoretic Foundation of General Definition Theory', *Synthese* (this issue).

Hallnäs, L. and P. Schroeder-Heister: 1990, 'A Proof-Theoretic Approach to Logic Programming. I. Clauses as Rules', *Journal of Logic and Computation* **1** (1990/91), 261–283.

Hallnäs, L. and P. Schroeder-Heister: 1991, 'A Proof-Theoretic Approach to Logic Programming. II. Programs as Definitions', *Journal of Logic and Computation* **1** (1990/91), 635–660.

Joachimski, F. and R. Matthes: 2003, 'Short Proofs of Normalization for the Simply-Typed λ-calculus, Permutative Conversions and Gödels T', *Archive for Mathematical Logic* **42**, 59–87.

Kahle, R. and P. Schroeder-Heister: 2006, 'Introduction: Proof-Theoretic Semantics', *Synthese* (this issue).

Lorenzen, P.: 1955, *Einführung in die operative Logik und Mathematik*, Springer, Berlin, 2nd edn., 1969.

Martin-Löf, P.: 1971, 'Hauptsatz for the Intuitionistic Theory of Iterated Inductive Definitions', in J. E. Fenstad (ed.), *Proceedings of the 2nd Scandinavian Logic Symposium (Oslo 1970)*, North Holland, Amsterdam, pp. 179–216.

Martin-Löf, P.: 1995, 'Verificationism Then and Now', in W. DePauli-Schimanovich et al. (eds.), *The Foundational Debate: Complexity and Constructivity in Mathematics and Physics*, Kluwer, Dordrecht, pp. 187–196.

Martin-Löf, P.: 1998, 'Truth and Knowability: On the Principles C and K of Michael Dummett', in H. G. Dales and G. Oliveri (eds.), *Truth in Mathematics*, Clarendon Press, Oxford, pp. 105–114.

Montague, R.: 1970, 'English as a Formal Language', in B. Visentini et al. (eds.), *Linguaggi nella società e nella tecnica*, Milano. Repr. in R. H. Thomason (ed.), *Formal Philosophy: Selected Papers of Richard Montague*, Yale University Press, New Haven 1974, pp. 188–221.

Muskens, R. A., J. van Benthem and A. Visser: 1997, 'Dynamics', in J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Language*, Elsevier, Amsterdam, pp. 587–648.

Negri, S. and J. von Plato: 2001, *Structural Proof Theory*, Cambridge University Press.

Prawitz, D.: 1965, *Natural Deduction: A Proof-Theoretical Study*, Almqvist & Wiksell, Stockholm.

Prawitz, D.: 1971, 'Ideas and Results in Proof Theory', in J. E. Fenstad (ed.), *Proceedings of the 2nd Scandinavian Logic Symposium (Oslo 1970)*, North Holland, Amsterdam, pp. 235–308.

Prawitz, D.: 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.

Prawitz, D.: 1974, 'On the Idea of a General Proof Theory', *Synthese* **27**, 63–77.

Prawitz, D.: 1985, 'Remarks on Some Approaches to the Concept of Logical Consequence', *Synthese* **62**, 152–171.

Prawitz, D.: 2006, 'Meaning Approached Via Proofs', *Synthese* (this issue).

Schroeder-Heister, P.: 1984a, 'A Natural Extension of Natural Deduction', *Journal of Symbolic Logic* **49**, 1284–1300.

Schroeder-Heister, P.: 1984b, 'Generalized Rules for Quantifiers and the Completeness of the Intuitionistic Operators &, $\vee$, $\supset$, $\curlywedge$, $\forall$, $\exists$', in M.M. Richter et al., *Computation and Proof Theory. Proceedings of the Logic Colloquium held in Aachen, July 1983, Part II.* Springer, Berlin, LNM, Vol. 1104, pp. 399–426.

Schroeder-Heister, P.: 1991a, 'Hypothetical Reasoning and Definitional Reflection in Logic Programming', in P. Schroeder-Heister (ed.), *Extensions of Logic Programming. International Workshop, Tübingen, December 1989, Proceedings.* Springer, Berlin, LNCS, Vol. 475, pp. 327–340.

Schroeder-Heister, P.: 1991b, 'Structural Frameworks, Substructural Logics, and the Role of Elimination Inferences', in G. Huet and G. Plotkin (eds.), *Logical Frameworks*. Cambridge University Press, pp. 385–403.

Schroeder-Heister, P.: 1991c, 'Uniform Proof-Theoretic Semantics for Logical Constants. Abstract.' *Journal of Symbolic Logic* **56**, 1142.

Schroeder-Heister, P.: 1992, 'Cut-Elimination in Logics with Definitional Reflection', in D. Pearce and H. Wansing (eds.), *Nonclassical Logics and Information Processing. International Workshop, Berlin 1990, Proceedings.* Springer, Berlin, LNCS, Vol. 619, pp. 146–171.

Schroeder-Heister, P.: 1993, 'Rules of Definitional Reflection', in *8th Annual IEEE Symposium on Logic in Computer Science (Montreal 1993)*. IEEE Computer Society Press, Los Alamitos, pp. 222–232.

Schroeder-Heister, P.: 1994b, 'Definitional Reflection and the Completion', in R. Dyckhoff (ed.), *Extensions of Logic Programming. Proceedings of the 4th International Workshop, ELP '93, St. Andrews, March/April 1993*, Springer, Berlin, LNCS, Vol. 798, pp. 333–347.

Tait, W. W.: 1967, 'Intensional Interpretations of Functionals of Finite Type I', *Journal of Symbolic Logic* **32**, 198–212.

Tait, W. W.: 2006, 'Proof-Theoretic Semantics for Classical Mathematics', *Synthese* (this issue).

Tarski, A.: 1933, 'Der Wahrheitsbegriff in den formalisierten Sprachen', *Studia Philosophica* **1** (1935), 261–405 (translated from the Polish original of 1933, with a postscript). Reprinted in K. Berka and L. Kreiser (eds.), *Logik-Texte*, Berlin 1971. English translation of the German version in A. Tarski, *Logic, Semantics, Metamathematics*, Clarendon Press, Oxford, 1956.

Tennant, N. W.: 1978, *Natural Logic,* Edinburgh University Press.

Tennant, N. W.: 1987, *Anti-Realism and Logic,* Clarendon Press, Oxford.

Tennant, N. W.: 1997, *The Taming of the True,* Clarendon Press, Oxford.

Troelstra, A. S. and H. Schwichtenberg: 1996, *Basic Proof Theory.* Cambridge University Press, 2nd edn. 2000.

Wilhelm-Schickard-Institut, Universität Tübingen
Sand 13
72076 Tübingen
Germany
E-mail: psh@informatik.uni-tuebingen.de

PATRIZIO CONTU

THE JUSTIFICATION OF THE LOGICAL LAWS REVISITED

ABSTRACT. The proof-theoretic analysis of logical semantics undermines the received view of proof theory as being concerned with symbols devoid of meaning, and of model theory as the sole branch of logical theory entitled to access the realm of semantics. The basic tenet of proof-theoretic semantics is that meaning is given by some rules of proofs, in terms of which all logical laws can be justified and the notion of logical consequence explained. In this paper an attempt will be made to unravel some aspects of the issue and to show that this justification as it stands is untenable, for it relies on a formalistic conception of meaning and fails to recognise the fundamental distinction between semantic definitions and rules of inference. It is also briefly suggested that the profound connection between meaning and proofs should be approached by first reconsidering our very notion of proof.

1. INTRODUCTION

Logical semantics can be thought of as a system of principles which purport to justify a certain set of logical laws or rules of inference. The justification takes place by showing the laws in question to be valid by virtue of the meanings of the logical constants. Thus, a semantic theory contains the specification of the meanings of the logical constants and the justification of a certain set of rules of inference. By its very nature, logical semantics answers at least the following questions:

The extensional one: which rules are valid?
The intensional one: why are they valid?

In the following, I will sketch how the proof-theoretic approach to semantics justifies the logical laws, and eventually I will point to what I take to be some of its main problems. By the proof-theoretic approach to semantics I mean roughly the attempt at providing an explanation of the meanings of the logical constants and of the concept of logical consequence in terms of proofs rather than truth.

This approach originates on the one hand from the foundations of constructive mathematics, and on the other hand from the verificationistic theory of meaning first put forward by Michael Dummett.

Synthese (2006) 148: 573–588
DOI 10.1007/s11229-004-6270-y

Thus the emphasis will be mainly on the constructive meaning of the logical constants, and I will concentrate on the intensional question above.

2. SEMANTIC DEFINITIONS AND RULES OF INFERENCE

The most obvious feature of a justification is that two things are involved: something which is justified, and something in terms of which it is justified. For the time being, let me say without further discussion that what is justified by a semantic theory is a certain set of rules of inference or logical laws, and that the justifying principles are *semantic definitions*.

A semantic definition is the specification of the meaning of a logical constant in terms of some distinguished semantic notion.

In model-theoretic semantics, for example, the meaning of a logical constant is defined to be its behaviour in compound statements having the constant under consideration as the principal operator, with respect to the notion of truth in a model.

When we explain the intended meanings of constructive logical constants, we take the notion of proof to be the basic semantic notion instead. Such explanations usually go under the name of *BHK* interpretation. A clause of the *BHK* interpretation has in general the form

$$\mathrm{Pr}\_\varphi\_(x) \Leftrightarrow \alpha(x)$$

(to be read as follows: x is a proof of the statement form φ if and only if x satisfies the condition α).

For example, if φ is of the form $A \rightarrow B$, the condition α states that x must be a method or a function transforming any proof of A into a proof of B.

Now the question arises of which rules of inference are justified by these semantic definitions. Of course, the definition yields immediately the following rules:

$$I\frac{\alpha(x)}{\mathrm{Pr}\_\varphi\_(x)} \qquad E\frac{\mathrm{Pr}\_\varphi\_(x)}{\alpha(x)}$$

Call these rules *semantic* (*or definitional*) *rules*. Given the close resemblance of Gentzen's rules of natural deduction for intuitionistic logic to the clauses of the *BHK* interpretation, one can think of such rules as a formal translation of the semantic principles, provided one keeps in mind the following facts.

First, in general the translation is not always faithful to the *content* of the semantic definition, that is, the intended meaning may fail to be preserved by the rules that build up a formal system. One example is given by the standard *I*-rule for implication. As has been pointed out on several occasions by Dag Prawitz, while for the assertability of $A \to B$ it is intuitionistically required that *any* function be available which transforms each proof of A into a proof of B, the formal translation confines the constructibility of such a function to the conceptual resources of a given formal system. Thus, one main result of the translation process is the switch from a genuinely semantic viewpoint to a formalistic one.

Secondly, the rules of inference may not be faithful to the *form* of the semantic definition. In other words, some rules – specifically, elimination rules for disjunction and the existential quantifier – do not correspond directly to the rules that can be obtained in an obvious way from the semantic definition. The reason is that while semantic definitions are meant to fix the meaning of the logical operators, rules of inference are devised to serve the purposes of deduction, i.e., they are formulated so as to be useful in deduction. Consider for example the definition of the existential quantifier:

$$\mathrm{BHK}(\exists)\; p \text{ proves } \exists x A(x) \text{ iff } p = \langle a \in D, \pi \rangle \text{ and } \pi \text{ proves } A(a).$$

The natural semantic elimination rule from this definition is the following:

$$\frac{p \text{ proves } \exists x A(x)}{\pi \text{ proves } A(a)}$$

where π is obtained simply by inspection of p. Now, such a rule would not be very useful in inference, because we would be able to apply it only in those cases in which we already possess the proof p having the required properties. This shows a general feature which applies to all elimination rules: their usefulness lies in the fact that they are applicable to those propositions for which no proof is available, or whose proof is not in *canonical form*, that is, in the form specified by the definition.

We might then drop the requirement that p be in canonical form, and thus formulate a more acceptable rule of inference. Call this rule a *generalised semantic* (*or definitional*) rule. But by so doing, we would completely lose the connection between p and π, that is, we would have no clue of how to obtain π once given p. This illustrates the

other important feature of elimination rules (and of rules of inference in general), i.e., the fact that they must provide a simple method for obtaining the proof in the conclusion given the proof(s) in the premisses. For some logical constants, e.g., implication, this can be achieved in quite a direct way from the definitional elimination rule. Given assumptions $A \rightarrow B$ and A, simply take the first assumption to be a function that when applied to the second assumption yields the conclusion B, and you have obtained a formal derivation of B from A even though you did not already have it in the premisses. Thus in the case of implication elimination the formal rule can be seen as encoding an abbreviation, or rather a simulation, of the process which is described by the generalised semantic rule. In the case of disjunction and the existential quantifier, on the other hand, the need to specify the proof in the conclusion calls for some structural modification of the semantic rule. Assuming that there is a proof of $\exists x A(x)$, we reason as follows. By the definitional rule, there is a proof of $A(a)$ for some a. Then, if any B not containing a is provable from $A(a)$, it is provable from the sole premiss $\exists x A(x)$ (or more precisely, from the premiss that there is a proof of $\exists x A(x)$). Since we exploit the definition only as a means of justifying the assumption of the existence of a proof of $A(a)$, we do not need to know how to construct that proof, for the piece of reasoning that we are carrying out will end up by discharging such an assumption.

Thus we have provided an informal justification of the standard rules of inference on the basis of the semantic definitions. For the purposes of formal deduction, the syntactic viewpoint is trivially the one to be preferred. Rules of inference build up a formal system, so that the kind of theoretical framework in which we are reasoning is precisely determined; and such rules are formulated in such a way as to be useful in inference. The main consequence of this, however, is that at the formal level, elimination rules allow for the construction of non-canonical proofs, thus leading to a remarkable asymmetry between (formal) introduction and elimination rules.

3. GENTZEN–PRAWITZ JUSTIFICATIONS

Thus elimination rules do not in general act on, and do not give rise to, canonical proofs, i.e., proofs in the form specified by the definition. So one might conclude that they do not have a definitional character after all, and this conclusion may have been the reason that

has led to the fairly widespread conviction that introduction rules are somehow more basic than elimination rules.

This view dates back to Gentzen's remark that *I*-rules are the definitions (in a loose, non-literal sense of the word) of the logical operators, whereas *E*-rules are consequences thereof, in that they exploit the meaning fixed by *I*-rules. It is to be noted, however, that Gentzen understood this asymmetry in a purely formalistic way, for he thought that the systematic relation between rules only depends on their formal structure. As he explicitly pointed out, we do not need to refer to the meaning of the symbols involved: "Es braucht hierbei nicht auf einen 'inhaltlichen Sinn' des Zeichens $\supset$ Bezug genommen zu werden" (Gentzen 1935, p. 189). This implies that the meaning that can be read off from a rule is only what the rule says within the formal system it belongs to. Call this the *formal meaning* of a sign. Later writers have extended this kind of argument to the general theory of meaning, in which what is involved is not the structural features of a formal system but the 'real' (or intended) meaning of logical signs (see e.g., Prawitz 1985; Martin-Löf 1985).

The view that *I*-rules are more basic gives rise to a different kind of justification, not a justification of rules from semantic definitions, but a justification of rules from a distinguished class of other rules, which is identified with the class of *I*-rules. The justification runs as follows.

(1) First we assume that every inference represented by an application of an introduction rule is automatically valid.
(2) Secondly, we assume that every complex statement, if assertible at all, must be capable of being asserted by means of an argument whose last step is given by an application of one of the introduction rules for the principal operator of the statement in question. This second clause, which has been called the "fundamental assumption" by Dummett (cf. Dummett 1991), ensures that introduction rules exhaust the meaning.

Under these assumptions we are able to justify elimination rules. By way of example, suppose that we have a proof π of $A \to B$. Then by (2) π is a method that yields a new proof π' in canonical form, that is

$$\dfrac{\begin{array}{c}[A]\\ \pi'\\ B\end{array}}{A \to B}$$

Now any application of an *E*-rule turns out to be eliminable, i.e.,

$$\dfrac{\dfrac{\begin{array}{c}[A]\\ \pi'\\ B\end{array}}{A \to B} \quad \begin{array}{c}\pi''\\ A\end{array}}{B} \quad \pi'' \rightsquigarrow \quad \begin{array}{c}\pi''\\ A\\ \pi'\\ B\end{array}$$

And this serves as a justification of the elimination rule, for the consequences that it entitles to draw are already contained in the proof in canonical form. The claim is therefore that *logical validity originates from introduction rules alone* (see e.g., Prawitz (1965), p. 165). The main question to be tackled here is whether this claim is warranted by the justification above.

4. THE STATUS OF THE FUNDAMENTAL ASSUMPTION

We have seen that the fundamental assumption plays an essential role in justification procedures. First of all, observe that it consists of two steps.

(1) First, postulate that the relevant definitional construction can always be obtained, that is, the construction which defines the meaning of the logical constant in question.
(2) Then apply the *I*-rule to such construction.

Since the *I*-rule is already assumed to hold, the only relevant step is the postulational one. As already mentioned, what is at stake for most authors is not the formal meaning of these rules, but their relevance in determining the 'real' meaning of the logical signs. As such, our problem is a genuinely semantic one. So in order to see what step (1) amounts to, it seems useful to consider some semantic construals of inference rules.

The general requirement that an inference rule should meet is, of course, that it be truth-preserving. As Per Martin-Löf has often pointed out, the explicit form of an inference rule is thus the following:

$$\frac{A_1 \text{ true}, \ldots, A_n \text{ true}}{A \text{ true}}$$

(where $A_1, \ldots, A_n, A$ are propositions, possibly depending on other assumptions).

There is at present no general agreement on how the notion of truth should be understood from a constructive viewpoint.

4.1. *Existence-based Interpretation*

Some people hold the view that truth should be identified with *provability in principle*, in order to be able to espouse constructivism while escaping relativism. According to this view, we have the equation

$$\begin{aligned}\text{Truth of } A &= \text{Provability in principle of } A\\ &= \text{There exists a (canonical) proof of } A\end{aligned}$$

Here the notion of existence is not codified by means of an existential quantifier, because the realm of proofs is not a well-defined domain of quantification.

On this construal of the notion of truth, we can say that an inference rule must be *existence-preserving*, whereby existence means existence of a proof. Thus the general form of a rule is

$$\frac{\text{There exists a proof of } A_1, \ldots, \text{ There exists a proof of } A_n}{\text{There exists a proof of } A}$$

Interpreted in this way, an inference rule is *postulational* in character, for it postulates the existence of a proof of the conclusion under the assumption of the existence of proofs of the premisses.

The reason why we state the rule for arbitrary proofs and not only for canonical proofs, is that, as already pointed out, in general, rules are applied when no canonical proofs are available.

4.2. *Possession-based Interpretation*

An alternative view of the constructive notion of truth identifies the truth of a proposition with the actual possession of a proof of it. That is, we have the equation

$$\text{Truth of } A = \text{Possession of a proof of } A$$

If we subscribe to such a view, then rules of inference are rather required to be *possession-preserving*, where again by possession we mean possession of a proof. Thus the general form of a rule is

$$\frac{\text{I possess a proof of } A_1, \ldots, \text{ I possess a proof of } A_n}{\text{I possess a proof of } A}$$

Assuming that I possess the proofs as stated in the premisses, I can inspect such proofs and see (given some definitions) that I actually possess a proof as claimed in the conclusion. Therefore we could also say that this is an *inspection-based* interpretation of rules of inference.

4.3. *Now which Construal should be Preferred?*

As far as *I*-rules are concerned, both interpretations seem to be acceptable. Consider for example implication again. Assuming that there is a hypothetical proof of B from A, it is of course correct to conclude that there is a proof of $A \to B$; and assuming that we possess a hypothetical proof of B from A, we can also safely conclude that we possess a proof of $A \to B$.

Things are different when it comes to *E*-rules. While *E*-rules can correctly be interpreted as existence-preserving, they are not, in general, possession-preserving. We can interpret modus ponens as follows:

$$(a) \qquad \frac{\text{There exists a proof of } A \to B}{\text{There exists a hypothetical proof of } B \text{ from } A}$$

But we cannot in general formulate modus ponens as a possession-preserving rule, because from the fact that we possess a hypothetical proof of $A \to B$ it does not follow that we also possess a hypothetical proof of B from A. This is only correct when the proof of $A \to B$ that we possess is in canonical form, but then the application of the *E*-rule gives rise to a redundant step. The usefulness of *E*-rules lies precisely in the fact that they are applicable in those cases in which no canonical proof, or any proof at all, is available. This is reflected by the fact that they are not possession-preserving. This, at least, holds good when we concern ourselves with the genuine semantic interpretation of rules of inference. As already pointed out, the formal translation of the semantic rule (i.e., usual modus ponens) does provide us with a hypothetical proof on the assumption that we possess a proof of $A \to B$. But this is only an abbreviation (from the point of view of the constructive semantics based on the *BHK* interpretation) of the real semantic rule for the purposes of formal deduction.

It is then apparent that (a) is none other than one of the usual formulations of the fundamental assumption (for the reason explained above, I am identifying the fundamental assumption with the sole step (1)), that is

$$(a^*) \qquad \frac{\text{There is a proof of } A \to B}{\text{There is a canonical proof of } A \to B}$$

Therefore on the first construal of rules of inference, the *E*-rule and the fundamental assumption seem to amount to the same thing.

There is, however, a more careful formulation of the fundamental assumption, since it is not only postulated that the canonical proof

exists, but it is also required that I know how to construct it if I possess an arbitrary proof. Therefore we have

$$(b) \qquad \frac{\text{I possess a proof of } A \to B}{\text{I know how to construct a canonical proof of } A \to B}$$

Or alternatively,

(c) (There is a proof of $A \to B$)/(There is a method, i.e., the proof in the premiss of providing a canonical proof of $A \to B$)

What is added in (b) is a better explanation of the relationship between premiss and conclusion form an epistemic viewpoint, whereas in (c) a better explanation is added from a nonepistemic viewpoint. In both cases the mere postulational character of the rule (a^*) is changed into a more precise explanation of how to obtain the postulated proof. But besides this, nothing is added to the fact that at the semantic level modus ponens is a postulational rule. So (a^*) on the one hand, and (b)–(c) on the other hand, are just two different ways of making explicit the semantic content of the formal rules, given some assumptions; the difference lies essentially in the amount of information which is provided.

Therefore it seems that the conclusion that the semantic content of the *E*-rule is tantamount to the fundamental assumption holds good for the more accurate formulation as well.

If the previous conclusion is correct, the very idea of justifying rules from other rules should be abandoned, for in order to carry out this justification we need to resort to full definitions anyway. In other words, the justifications of *E*-rules presupposes both rules obtained by a semantic definition, and the justification is in this sense circular. One might perhaps object that some sort of circularity is inherent in any kind of justification of the logical laws, so that the previous argument is beside the point. But the main purpose of the justification procedure was to obtain the content of elimination rules from the sole content of the introduction rules, as pointed out before, and it seems to me that this cannot be achieved. So the justifying principles are semantic definitions after all, as I suggested in Section 2, and not some selected class of rules.

Before proceeding any further it seems useful to take stock of this argument and examine the relationship between semantic definitions and the rules they give to in its full generality. As we have seen, from a *BHK* clause we obtain the obvious semantic rules

$$I\frac{\alpha(x)}{\mathrm{Pr}\_\varphi\_(x)} \qquad E\frac{\mathrm{Pr}\_\varphi\_(x)}{\alpha(x)}$$

It is of course understood that the variable x refers to the same object in the definiens and in the definiendum. But when we formulate the definition in terms of rules of inference, this can no longer be maintained as far as the E-rule is concerned. This is due to the fact that an E-rule is only useful if it is applicable in those cases in which no canonical proof is available. So the elimination rule should rather be formulated as

$$E\frac{\mathrm{Pr}\_\varphi\_(x)}{\alpha(y)}$$

where x and y in general denote different objects. This is what we called the generalised semantic rule in Section 2. We also saw that the main stumbling block for the usefulness of such rules is the fact that they do not provide any method for constructing the proof in the conclusion. Thus the two more careful formulations of the fundamental assumption are two ways of making sense of the E-rule, by saying what the relation is between x and y.

So we can conclude that the E-rule is equivalent to the fundamental assumption, when its semantic content is made fully explicit.

One can perhaps distinguish between what an E-rule *does* and what it *claims*.

What it does is provide a non-canonical means of proving propositions. This is the usual syntactic interpretation.

What an E-rule claims, is that a canonical proof can always be constructed, provided there is a proof at all. To justify this claim, we need to resort to the same claim. So in this sense nothing is gained.

5. A POSSIBLE OBJECTION ANSWERED

An objection might be raised against this conclusion by arguing that the principle that we have called the fundamental assumption is not a real assumption, but arises instead from our *reflection* in the metalanguage on the fact that the introduction rule exhausts the meaning. According to this account, we just *see* that the introduction rule exhausts the canonical assertability conditions, and we express this in the form of a principle. Therefore, although the elimination rule may be recognised to be equivalent, in its semantical formulation, to the fundamental principle, the justification is not circular because it arises by reflection on meaning and not by making any assumptions.

However, the claim that the fundamental principle arises by *pure reflection* is rather dubious. Here, by pure reflection I mean reflection in which no further assumptions are involved. For this is tantamount to saying that reflection is infallible, that there are no other interpretations of the rule but the chosen one. But it seems that there are many other interpretations, depending on our intentions and purposes. Here, of course, I am rejecting the view that there is just one correct logic that can be discovered by pure philosophical reflection. If we refuse to resort to this conception of pure reflection, we should recognise that there are as many correct logics as there are sound semantic principles.

6. GENTZEN'S ORIGINAL CLAIM

So far, by examining the semantic content of rules, I have argued against the view that logical validity originates solely from the meaning fixed by the introduction rules. But as observed, Gentzen meant his remark to apply to the formal meaning of logical signs. Should we then say that his claim holds good at least as far as the formal meaning of rules is concerned? First of all, note that if we confine ourselves to e.g., the standard formal systems for intuitionistic logic, no fundamental assumption is involved anymore, for by means of normalisation procedures it can be *proved* that a proof in normal form can always be obtained. This, however, does not imply that elimination rules do not contribute to determining the meaning of the logical signs. Another feature of the formal system is that elimination rules cannot be dispensed with in spite of normalisation. This is the formal counterpart to the fact that their usefulness lies in being applicable to assumptions for which no proof is provided. If we want to analyse assumptions, we have to resort to elimination rules. Thus, what a formal elimination rule does is to say what the case would be if the assumptions were true, i.e., if they had a proof, or, as in the case of disjunction and the existential quantifier, to exploit the consequences of this. So we can take elimination rules as mimicking at the formal level the content of (postulational) semantic rules. Given some assumptions, instead of constructing the relevant canonical proof(s), we suppose we already have such constructions, and in our subsequent reasoning we build upon such hypothetical possession. This makes deduction possible. Without this method we could not prove relevant theorems in our formal systems. Insofar as

we are concerned with the formal meaning of logical signs, i.e., the content that can be read off from the rules of the formal system they belong to, it is clear that if any rule is not eliminable, then it is essential to determining the formal meaning. I take this to be a tautology. The significance of the non-eliminability of *E*-rules has been generally overlooked, probably due to the fact that their role in deduction has not been adequately emphasised. Thus, even according to the formalistic approach, it is not true that logical validity originates solely from the meaning fixed by the introduction rules.

Gentzen's "formalistic" view may be related to the observation that elimination rules are the formal *inverse* of introduction rules: given this feature, we need only to take into account their structural relationship, and therefore can disregard their meanings. On the other hand, the claim about the definitional character of introduction rules seems to be really intensional and not easily reconciled with this formalistic stance. We face here what is arguably the original sin of the proof-theoretic approach to semantics as we know it: the attempt at drawing meaning-theoretic consequences from the peculiar features of certain formal systems. Since these features are contingent – there are alternative systems – a deeper conceptual analysis is needed.

Let us consider e.g., the notion of inversion. Its general structure can be represented by means of the abstract algebraic properties of the logical operations, which are made explicit in the language of category theory by the concept of *adjointness* (see e.g., MacLane 1998). An adjoint situation determines a pair of symmetric rules, such as (in the case of exponentiation):

$$I\frac{f : C \times A \to B}{f^* : C \to (A \Rightarrow B)} \qquad E\frac{f^* : C \to (A \Rightarrow B)}{f : C \times A \to B}$$

where, roughly speaking, the *E*-rule corresponds to modus ponens (for modus ponens may be interpreted as saying that if there is a proof of $A \Rightarrow B$ then there is a proof of B from A). Now, while the language of category theory does not yield a satisfactory account of the complex semantics of intuitionistic proofs that was outlined above (e.g., no distinction is made between semantic and syntactic rules of inference), it does provide a different framework within which some ideas about rules and inversion no longer appear plausible. Within such a framework it would be very unnatural to claim that some rules are more basic than others, and this suggests that the bias towards some subclass of the rules originated from the peculiar

syntactic characteristics of the formalism adopted, rather than from a truly intensional analysis of meaning.

7. THE PRAGMATISTIC APPROACH

Let us now briefly examine the view that assumes the elimination rules as basic and attempts to justify the introduction rules. Just as the previous approach is related to a verificationistic theory of meaning, this alternative view is inspired by the pragmatistic principle according to which the consequences of our assertions determine the meaning of the asserted sentences. More precisely, we have here the following assumptions.

(1) Every inference represented by an application of an elimination rule is valid.
(2) Pragmatistic fundamental assumption (PFA): Any consequence of a statement must be capable of being derived by means of an argument whose first step is given by an application of one of the elimination rules for the principal operator of the statement in question (where the statement is the main premiss of the inference).

The justification of introduction rules is now fairly simple: whenever we draw any consequences from a statement A, by PFA we can draw them by first applying an elimination rule for that statement, so that if we now append the application of a standard introduction rule for A we obtain a redundant step which can be dispensed with. This is, of course, simply a different interpretation of the situation described in Section 3.

In spite of the duality between the PFA and the verificationistic fundamental assumption (VFA), it should be noted that there is a striking difference with respect to the previous justification argument. Indeed, in analogy with VFA, we can analyse PFA as consisting of two steps:

(A) Apply one of the elimination rules to a certain statement A.
(B) Postulate the existence of a proof of any consequence of A starting with the result of the application of an E-rule.

Unlike the downward procedure, however, here there is no relation whatsoever between the postulated proof and the definitional construction, so that in the case at hand the FA does not coincide (in any

sense) with the rule being justified. Therefore the upward justification procedure does not result in the form of circularity that we detected in the former justification structure.

Elimination rules are concerned with definitional constructions only insofar as they simulate them in their conclusions, which means that we have to turn to (A) itself rather than (B). In other words, the downward justification is related to the semantic definitions only in the derived way, which is the main characteristic of *E*-rules.

What our assumption says is that the *E*-rules for a certain statement A – besides being valid – are the only rules by means of which we are entitled to draw consequences from it. Given A, we can infer in one step a certain set χ of simulated constructions, which by virtue of PFA-(A) is the maximal set of constructions inferred in one step from A. Not only do we know that given A, a corresponding set χ of simulated constructions must hold; we also know that no other constructions can be so inferred in one step. We are then actually claiming that χ makes explicit the whole content of A, or in other words, that χ is definitional and thus equivalent to A. Therefore the inverse rule must hold as well, and the upward justification just confirms this.

Unlike the downward justification, here we are not confronted with a real form of circularity: it is only a matter of choosing axioms. But we cannot claim that the *E*-rules alone bear all the content of the logical operators, for we always need a further closure principle.

The remarkable fact about *E*-rules is that by means of them we have switched from genuine constructions to simulated ones. Whenever the premisses are not facts but hypotheses, the consequences are simulations.

8. CONCLUDING REMARKS

To my mind, the preceding discussion results in two basic remarks.

(1) Introduction and elimination rules are symmetric when derived straightforwardly from definitions, but they become asymmetric when they are meant as rules of *inference*. The application of an introduction rule is an act of *synthesis*: given a certain structure, we recognise that it amounts to a previously defined notion. Elimination rules, on the other hand, allow us to carry out the *analysis* of a defined notion[1], by dissecting the content of that notion according to the definition. If we now suppose that some object x

falls under the defined concept α, we have no information about it other than our assumption of its property α, so that any piece of reasoning deriving from our assumption can only deal with a fictitious x. Consequently, the elimination rule(s) derived from our definition must be formulated in such a way as to allow for fictitious objects and simulated constructions in inferences. Thus we see that the asymmetry applies not only to the rules derived from the clauses of the *BHK* interpretation, but also to the inference rules derived from definitions in general, as long as they are understood in a constructive way.

(2) As a consequence, the programme initiated by Gentzen cannot be maintained in its original form. The asymmetry of inference rules ultimately depends on the gap between semantic definitions and their translation for the purposes of deduction, and it does not imply that some rules are more basic than others. More generally, Gentzen's approach to proof theory requires a deep revision. Under the influence of Hilbert's formalism, Gentzen analysed proofs as syntactic objects in a formal system whose properties can be laid bare by means of the techniques of cut elimination. This was a great achievement, with far-reaching consequences both for our understanding of the formal structure of proofs and for the study of mathematical theories. Nevertheless, this approach causes a fundamental distortion of *real* proofs, which are based on large unitary steps called *lemmas* and conceptual constructions called *definitions*. Standard proof theory does not have much to say about them, for it systematically gets rid of lemmas by cut elimination and ignores the role of definitions in mathematical reasoning because of the eliminability of explicit definitions. Therefore I claim that although proofs can be dealt with in a purely syntactic way, there is strong evidence to the effect that that is by no means the most appropriate way to deal with them. Instead, proofs are best thought of as a true semantic phenomenon. The idea of *proof-theoretic semantics*, as a theory of truth and logical consequence that is based on the notion of proof, is but a starting point towards the construction of a real *semantic* (*i.e., conceptual, non-formalistic*) *theory of proofs*.

NOTE

[1] This is of course different from the analysis of a concept aimed at carving out its definition.

REFERENCES

Dummett, M.: 1991, *The Logical Basis of Metaphysics*, Duckworth, London.
Gentzen, G.: 1935, 'Untersuchungen über das logische Schließen', *Mathematische Zeitschrift* **39**, 176–210, 405–431.
MacLane, S.: 1998, *Categories for the Working Mathematician*, 2nd edn, Springer, Berlin.
Martin-Löf, P.: 1984, *Intuitionistic Type Theory*, Bibliopolis, Napoli.
Martin-Löf, P.: 1985, 'On the Meanings of the Logical Constants and the Justifications of the Logical Laws', repr. in *Nordic Journal of Philosophical Logic* **1** (1996), 11–60.
Prawitz, D.: 1965, *Natural Deduction: A Proof-Theoretical Study*, Almqvist & Wiksell, Stockholm.
Prawitz, D.: 1985, 'Remarks on Some Approaches to the Concept of Logical Consequence', *Synthese* **62**, 153–171.

Wilhelm-Schickard-Institut
Universität Tübingen
Sand 13
72076 Tübingen
Germany
E-mail: patrizio_contu@hotmail.com

LARS HALLNÄS

ON THE PROOF-THEORETIC FOUNDATION OF GENERAL DEFINITION THEORY

ABSTRACT. A general definition theory should serve as a foundation for the mathematical study of definitional structures. The central notion of such a theory is a precise explication of the intuitively given notion of a definitional structure. The purpose of this paper is to discuss the proof theory of partial inductive definitions as a foundation for this kind of a more general definition theory. Among the examples discussed is a suggestion for a more abstract definition of lambda-terms (derivations in natural deduction) that could provide a basis for a more systematic definitional approach to general proof theory.

1. INTRODUCTION

A general definition theory should serve as a foundation for the precise mathematical study of definitional structures (cf. General Pattern Theory (Grenander 1994)). The central notion in such a theory is a precise explication of the intuitively given notion of a definitional structure. Definitional structures occur in all theory building and in all formalization. Induction principles show that global definitional structures have an intrinsic methodological value. Definitional structures can also be interesting in their own right as well as being a tool for the intensional classification of mathematical objects.

The purpose of this paper is to discuss a more general perspective of earlier work on a proof-theoretic foundation for partial inductive definitions (Hallnäs 1991; Schroeder-Heister 1993). A first formulation was given in terms of a sequent calculus for a sort of infinitary propositional language (Hallnäs 1991). Originally, I considered this calculus to be rather provisional in search of a more general and elementary semantical formulation. It now seems to me that this search for a general *semantical* formulation just led astray. I think that the initial proof-theoretic formulation is already essentially an elementary one that serves its purpose as a foundation for a general definition theory well.

Synthese (2006) 148: 589–602
DOI 10.1007/s11229-004-6291-6

By a definitional structure we mean the local logic of a definition, i.e., a notion of consequence, or, more generally, of connections, that depends only on local information. Information is local if it only refers to the meaning of defining conditions and to the way atomic components are defined in a given definition D, i.e., without any considerations concerning assumptions about global closure-properties of the definition. This also means that defining conditions will be given an interpretation that is local to a definition.

The intended reading of defining conditions implicitly rests on certain general closure properties, i.e., a global logic invariant across different definitions. Consider the following definition which gives a simple canonical example of the puzzles and paradoxes hidden here:

$$D\{ \quad P = P \to Q.$$

Assume P, then by definition $P \to Q$ and therefore Q (with a reasonable interpretation of $\to$, which includes *modus ponens*). Thus Q follows from P. So $P \to Q$ (again by a reasonable interpretation of $\to$, which includes *conditionalization*). By definition, this gives P. So we have derived both P and $P \to Q$ using only local information about P and about $P \to Q$ with respect to D. If we now continued to use the intended reading of $\to$ as "if ... then", we would obtain Q. But this is paradoxical. Q is not defined in D, which means that it is not possible to infer Q on the basis of local information alone. The closure property needed to infer Q can be written as

$$\vdash_D P \quad \text{and} \quad \vdash_D P \to Q \quad \text{imply} \quad \vdash_D Q,$$

where $\vdash_D$ means derivability with respect to D. This cannot hold, since it would contradict the basic general closure property of a definitional structure, i.e., the basic definitional structure axiom:

An atom a *holds with respect to a definition* D *iff some condition* A *defining* a *in* D *holds with respect to* D.

So the interpretation of $\to$ with respect to D differs from the intended global reading. What the example shows is that we can not assume that defining conditions in a definition always can be given a global extensional interpretation. The local logic of a definition will conform to the intended – global and extensional – reading of defining conditions only if certain non-elementary closure properties of the definition are satisfied – the kind of properties that are closely connected with basic assumptions or foundational axioms.

A definition is an intensional object. It is the result of an act where we intend to define something. Even if the resulting definition does not satisfy the intended properties, it is still there generating a definitional structure of some sort. This is the basic reason why an explication of the structure has to be local in nature. Implication is not an absolute notion in the present context, it might have a meaning locally with respect to a definition which is different from the intended global reading. Now the main question for the foundation of a general definition theory is how all this should be made mathematically precise.

A definition is something where we have a definiendum–definiens relation

$$a =_{\text{def}} A$$

– from the context it is clear that this is a definition, so we can skip the "def" notation – aesthetics is important – and write

$$a = A.$$

It is difficult to say something in general about the fine structure of A. The notion of a definitional structure we have in mind is based on two basic general structural properties of A:

(i) A may directly depend on the definition of atoms $a_1 \dots a_n \dots$ which gives structural components of A.
(ii) The act of defining introduces a duality between right – definiens – and left – definiendum – which gives a possible duality structure of A, expressed by some sort of implication.

In the definition

$$\mathit{True}(A \to B) = \text{if } \mathit{True}\, A, \text{ then } \mathit{True}\, B$$

there are not only the components $\mathit{True}\, A$ and $\mathit{True}\, B$ but also a duality between them: left($\mathit{True}\, A$) and right($\mathit{True}\, B$), i.e., a duality between *assuming* and *proving*. To establish $\mathit{True}(A \to B)$ we have by definition to prove $\mathit{True}\, B$ under the assumption $\mathit{True}\, A$.

So definitions introduce a definiens $D(a)$ for an atom a, components $D(A)$ for a defining condition A and a duality between left and right. The first issue – $D(A)$ – is a simple combinatorial property, so foundational matters are mostly connected with an interpretation of the duality between left and right that a definition introduces. It is this aspect of a definitional structure that we try to study here. It is only natural to give a logical interpretation of this structure, as duality is a logical notion after all.

The duality between left and right is intuitively a duality between two dual acts, the act of replacing a definiens with a corresponding definiendum–a local and deterministic act – and the act of replacing a definiendum with its definiens – a global and possible indeterministic act. The duality left–right that a definition introduces is given by two basic aspects of a definition D. Given an atom a; we may open D with a (left) or we may close D with a (right). In the first case, we approach D with a and in the second case we leave D with a. This duality can be interpreted in many different ways; as a duality between assumptions and conclusions, between generating and evaluating, between reflection and closure, between initiate and finish etc. It is a basic structure of an intensional topology; false–true, closed–open, sets–individuals, global–local etc. If we open D with a we start with a and ask for its definition in D, if we close D with a we end up with a using a single definitional clause in D.

So the basic properties of a definitional structure considered here are the following:

(i) connections generated by a given definition – a definition D as a combinatorial object,
(ii) a specific duality generated by a given definition – a D-closed duality notion $\vdash$.

A main point here is that if we want to study definitional structures with respect to this duality – B follows from A – then a simple sequent calculus for an infinitary language seems to give an adequate foundation. The proof-theoretic foundation is furthermore elementary and natural with respect to these structural properties. The notion of local information establishing that B follows from A with respect to some definition D is closely related to the sub-formula property of a cut-free sequent calculus. The finiteness property given by the fact that the basic proof-theoretic concepts are introduced by inductive definitions gives furthermore a natural foundation of a definition theory. There is no reason to think that a more extensional model-theoretic interpretation giving a more elementary foundation is lurking in the background waiting for discovery. The notion of a local logic is in a certain sense an elementary one, so it is perhaps not too wrong to compare the situation we have here with recursion theory where the notion of a partial recursive function gives a good elementary foundation.

So we will think of a definitional structure on U as given by

(i) a definition D over U,
(ii) a D-closed duality notion $\vdash$.

2. DEFINITIONS

By a *definition* D as a mathematical object we mean a set of equations

$$a = A$$

where $a \in U$ for some given universe of discourse and where A is a condition built up from objects in U, $\top$ and $\bot$ using (possibly infinitary) conjunctions $\bigwedge_I$ and implications $\rightarrow$. We let $D(a)$ be the set of conditions defining a in D if there are any and $\{\bot\}$ otherwise. We interpret a definition D in terms of a *local* notion of consequence $\vdash_D$ given as follows

$$\Gamma, a \vdash_D a$$

$$\frac{}{\Gamma \vdash_D \top} \qquad \frac{}{\Gamma, \bot \vdash_D C}$$

$$\frac{\Gamma \vdash_D A_i \quad (i \in I)}{\Gamma \vdash_D \bigwedge_I A_i} \qquad \frac{\Gamma, A_i \vdash_D C}{\Gamma, \bigwedge_I A_i \vdash_D C}(i \in I)$$

$$\frac{\Gamma, A \vdash_D B}{\Gamma \vdash_D A \rightarrow B} \qquad \frac{\Gamma \vdash_D A \quad \Gamma, B \vdash C}{\Gamma, A \rightarrow B \vdash_D C}$$

$$\frac{\Gamma \vdash_D A}{\Gamma \vdash_D a}(A \in D(a)) \qquad \frac{\Gamma, A \vdash_D C \quad (A \in D(a))}{\Gamma, a \vdash_D C}$$

We let $Def(D) = \{a \mid \vdash_D a\}$. Let $\mathrm{Cov}(D) = \{A \mid \vdash_D A\}$ – intuitively $\mathrm{Cov}(D)$ is the set of conditions that *covers* the set of *true* objects and *true* – i.e., intended – connections.

A definition D is said to be *total* if the *cut rule* is derivable, i.e., if

$$\frac{\Gamma \vdash A \quad \Gamma, A \vdash B}{\Gamma \vdash B}$$

holds for all Γ, A, B. Let $\leq$ be the reflexive and transitive closure of $<^*$, where $<^*$ is given by

$$A <^* a \quad \text{if} \quad A \in D(a)$$
$$A_i <^* \bigwedge_I A_i.$$

D gives the elementary combinatorial structure of a definition and $\vdash$ interprets this structure as a definition. Such a notion of a definitional structure has a clear topological flavor.

Now it is not immediately clear in every case that the cover of D, $\mathrm{Cov}(D)$, directly gives the *intended* duality interpretation of D. Rather, the basic assumption is that it *covers* it. As a typical example take the following recursive definition of plus.

$$(n + 0) = n$$
$$(n + s(m)) = s(n + m)$$
$$s(n + m) = \bigwedge_N(((n + m) \to k) \to s(k)).$$

Due to an elementary logical reading of $\vdash$, which includes the contraction rule, it turns out that $\mathrm{Cov}(D)$ is not the correct evaluation relation. Using contraction we can show $(2 + 2) \vdash m$ for all $m \geq 4$.[1] The correct definition is given by: $(n + m)$ is the *smallest* k such that $(n + m) \vdash k$.

To look for a definitional structure on a set U is to look for a definitional structure $(D, \vdash)$ where some $M \subset U$ is included in $Cov(D)$ and *elementarily* definable in $\vdash$. What exactly *elementarily* should mean will depend on the context.

The basic idea here is that we introduce a certain duality on U with respect to a definition, i.e., a logic that gives an intensional presentation of some $M \subset U$. It is a structural classification of M in intensional terms – local proof theory with a flavor of concrete topology, i.e., a theory about the *form* of definitional presentations.

The interpretation we give of the notion of *duality* in the present context is then the following:

A *duality structure* on a set X is a triple (Δ, R, L) where

$$\Delta\colon X \to P(X)$$

$$R, L\colon P(X) \to P(X)$$

$$(\exists A \in \Delta(a) : A \in \Gamma R) \Rightarrow a \in \Gamma R$$

$$(\forall A \in \Delta(a) : A \in L\Gamma) \Rightarrow a \in L\Gamma.$$

Given a definition D, this is then interpreted as follows

$$\Gamma R \ - \ \text{the open cover of } \Gamma \ - \ \{A \mid \Gamma \vdash_D A\}$$

$$L\Gamma \ - \ \text{the closed cover of } \Gamma \ - \ \{A \mid A \vdash_D \Gamma\}.$$

Intuitively, we think of a definition D as generating connections on a set and that our reading of these connections as *definitional* ones will introduce a duality on this connection structure. The duality between *open* and *closed* covers is simply the duality between $\exists$ and $\forall$; $\bot \vdash a$, $a \vdash \top$ etc.

3. SOME CHALLENGES

3.1. *Induction Principles – Closure Properties: The Functional Closure*

Given a set $X \subset U$ and functions $f_1 \ldots f_n$ with arities $k_1 \ldots k_n$ over U, the *function closure* with respect to $(X, f_1 \ldots f_n)$ is the set obtained by starting with X and closing it under the given functions, i.e., the set given by the definition $D(X, f_1 \ldots f_n)$

$$a = \top \quad (a \in X)$$
$$f_i(x_1 \ldots x_{k_i}) = (x_1 \ldots x_{k_i}) \quad (i \leq n).$$

It is then easy to see that $Def(D(X, f_1 \ldots f_n))$ is the smallest set containing X and being closed under the functions $f_1 \ldots f_n$.

Similarly, given a set $X \subset U$, functions $f_1 \ldots f_n$ with arities $k_1 \ldots k_n$ over U, a functional $F : [U \to U] \to U$ and a set $\Phi \subset [U \to U]$, the *functional closure* with respect to $(f_1 \ldots f_n, F, \Phi, X)$ is given by the following definition $D(X, f_1 \ldots f_n, F, \Phi)$ (DF for short in what follows)

$$a = \top \quad (a \in X)$$
$$f_i(x_1 \ldots x_{k_i}) = (x_1 \ldots x_{k_i}) \quad (i \leq n)$$
$$F(f) = \bigwedge_U (x \to f(x)) \quad (f \in \Phi).$$

The functional closure is a generic structure for studying higher order induction principles and foundational axioms. Induction principles are the traditional way in which we use definitional structures and make them explicit. These principles represent various foundational axioms. The challenge here is to use the functional closure as a guiding structure in the search for new axioms and new formulations of old ones. Take the notion of a closed term in the λ-calculus (which corresponds to the notion of a closed derivation in natural deduction) as an example. If we view this notion as given by a functional closure – the rule for λ-abstraction being the functional – we get a more abstract presentation of syntactical notions. What does this mean in terms of induction principles, axioms for syntactic structures etc.?

3.1.1. *Abstract Derivations and Terms*
Using the notion of a functional closure, we may model derivations in natural deduction in a more abstract manner. Thus, rules like $\wedge$-introduction and $\rightarrow$-elimination are *function* rules, while typically rules like $\rightarrow$-introduction are *functional* rules. This means that the *premise* of the $\rightarrow$-*I* rule is a function, i.e., an abstract object. So the notion of a bound variable will disappear. The problem of modelling rules involving discharging of assumptions in a concrete manner will here correspond to finding the right principles for characterizing the functional in question and the set of *concrete* functions. Such principles will have to ensure that there is a sufficient number of such functions, but also that there are not too many. The functionals must also be given the correct type of restrictions. Following the Curry–Howard *isomorphism* between derivations and λ-terms we can concentrate on terms, which makes notation easier, without losing basic structural information.

Let U be a set having enough closure properties for building our terms. Let $X \subset U$ be a set of *variables*, Ap a binary function on U, λ a functional in $[U \rightarrow U] \rightarrow U$ and Φ a subset of $[U \rightarrow U]$. The λ-terms are now given by the functional closure with respect to (X, Ap, λ, Φ) – $D\lambda X$ for short. Then a term t is *closed* if $f \in Def(D\lambda\emptyset)$ and open if $t \in Def(D\lambda X)$ implies that X is non empty. We write $D\lambda$ for short if $X = \emptyset$.

We may then consider the following principles:

(i) Assume that ρ is a finite partial function from X to $Def(D\lambda X)$. Then one natural axiom here is to state that the following *definition* makes sense

$$x\rho = \rho[x]$$
$$Ap(t,r)\rho = Ap(t\rho, r\rho)$$
$$\lambda(g)\rho = \lambda(g\rho)$$

where $\rho[x] = \rho(x)$ if ρ is defined for x otherwise $\rho[x] = x$ and where $g\rho(t) = g(x)\rho + \{x = t\}$ for some given *new* $x \in X$, i.e., some x for which ρ is not defined.

(ii) We also want to have some principle of combinatory completeness at our disposal. Given a term $t \in Def(D\lambda X)$ we have a function $\Lambda x.t(r) = t\{x = r\}$. Now we want $\Lambda x.t$ to be in Φ for all terms t.

Principle (i) gives restrictions on the definition $D\lambda X$ by stating that a certain principle of recursion on the given definition is valid, i.e., the

definition is deterministic, there are not too many functions in Φ etc. It also states that a certain completeness property holds:

$$\text{If } \lambda(g) \in \mathit{Def}(D\lambda X), \text{ then } g\rho \in \Phi.$$

Principle (ii) states that there are enough functions in Φ.

3.1.2. *Induction on a Term (Derivation)*
Let us assume that a property is given by a definition D'. We have the following principle of induction on the functional closure DF in general – what amounts to *induction plus co-induction*.

If for all $a \in U$

(i) $A \in DF(a)$ implies $A \vdash_{D'} a$
(ii) and there is a $A \in DF(a)$ such that $a \vdash_{D'} A$

then $\mathit{Def}(DF) \subset \mathit{Def}(D')$.

For the particular functional closure $D\lambda$ it is natural to think of a much stronger principle of induction:

If E is a set such that

(i) E is closed under Ap
(ii) and $\lambda(f) \in E$ if E is closed under f for $f \in \Phi$

then $E \subset \mathit{Def}(D\lambda)$.

3.1.3. *Local Definability and the Notions of Redex and Cut*
Given a definition D we say that a function $f: U \to U$ is *locally definable* in D, if for all $x \in U$

$$x \vdash_D f(x).$$

The idea is simply that we always can reach $f(x)$ from x by reasoning in D. That is to say D contains sufficient information to compute $f(x)$ given x, i.e., the value of f at x is local to D. We may also think of D as introducing a certain structure on a given set. That a function is locally definable in D then means that $f(x)$ can be described in terms of the structural operators of D as a formal D expression. f is then pointwise an elementary operation with respect to D.

If t is a closed term ($t \in \mathit{Def}(D\lambda)$), then

$$\lambda(g) \leq t$$

for some g, hence either t is $\lambda(g)$ or $Ap(\lambda(g), r) \leq t$ for some r. So $\lambda(g)$ is a *normal* form in the sense that $\lambda(g)$ is a $\leq$-minimal form of a closed term.

We may compare a *concrete* version of the β-rule $c\beta$

$$Ap(\lambda x.t, r) \Rightarrow t(r/x)$$

and an *abstract* version $a\beta$

$$Ap(\lambda(g), r) \Rightarrow g(r).$$

The $c\beta$ rule is not locally definable with respect to the definition of terms. This means that not all rules of contraction are locally definable with respect to the definition of derivations. The reason for this is that when we substitute a term in the scope of a variable binding operation like λ-abstraction, we sometimes have to change the name of the bound variable by introducing a *fresh* variable. This makes substitution a non-local operation with respect to the definition of terms. It is a direct consequence of the, in a certain respect, too concrete presentation of the basic ideas involved in the rules of λ-abstraction and $\rightarrow$-introduction. On the other hand, the $a\beta$ rule is locally definable with respect to $D\lambda X$, which means that all rules of contraction for the abstract derivations are locally definable with respect to the functional closure definition of derivations. This is one way of expressing that the basic definition of derivations is given on the right level of abstraction, namely that the basic structural operations of contractions all are locally definable with respect to the definition of derivations.

3.1.4. *Normalization*

Assume that the following definition on $\mathrm{Def}(D\lambda\emptyset)$ makes sense;

$$\mathrm{Con}(Ap(\lambda(f), t)) = f(t)$$
$$Con(Ap(Ap(r, u), t)) = Ap(Con(Ap(r, u), t).$$

Tait's (1967) notion of convertibility for closed terms may then be given by the following modification of $D\lambda\emptyset$

$$Ap(r, t) = \mathrm{Con}(Ap(r, t))$$
$$\lambda(f) = \textstyle\bigwedge_U(t \rightarrow f(t)).$$

Let us denote this transformation of $D\lambda\emptyset$ by $T\lambda\emptyset$. Now, normalization follows if certain basic logical and structural properties of $D\lambda\emptyset$ are invariant under this transformation.

3.2. *Definitional Structures per se: Introducing a Proof Theoretic Perspective*

Using a proof-theoretic foundation of a general definition theory means that we introduce a proof-theoretic perspective concerning central concepts and basic methods. The basic challenge here is to develop this approach through extensive studies of examples, looking for definitional structures in several areas of mathematics. The distinction between definability theory and definition theory here is similar to the distinction between reductive proof theory and general proof theory (see Prawitz 1971) or between computability theory and a structure theory of computations.

Example. Solving an equation $E(x_1 \ldots x_n) = k$ can structurally be viewed as looking for a definition D such that

$$E(x_1 \ldots x_n) \vdash_D k.$$

That is, we introduce definitional structures not only in terms of assigning values to variables, but also with respect to methods for finding solutions.

3.3. *Intensional Classification of Mathematical Objects: Classification with respect to Presentational Structure*

A general definition theory is concerned with principles of intensionality, i.e., with presentational structures. So a very basic challenge is the classification of mathematical objects with respect to various presentational structures. Given the extensional notion of an object, we look for different concrete presentations of it. In definition theory we study the structure of the definitions of objects. An example of this is Fredholm's use of a definitional approach to study the intensional classification of primitive recursive functions (Fredholm 1995). The general recursive definition of the *min* function is structurally symmetrical, i.e., the definition of $\min(n, m)$ has the same form as the definition of $\min(m, n)$ for all pairs (n, m). In contrast, there are no symmetrical primitive recursive presentations of the *min* function. This is one way of reading the results obtained by Colson (1991) and Fredholm (1995).

At this point it is natural to try to develop definition theory as a concrete topology of presentational structures. Several basic notions of topology have a natural definitional interpretation. A definition introduces connections on a space – where $a = (b \to c)$ can be thought of as a second order connection etc. So, naturally, a space X is

connected with respect to D if $a \vdash b$ for all $a, b \in X$ – a distinction between *right connected* and *left connected* may of course be introduced.

3.4. *The Structure of Proofs: to Classify Propositions with Respect to Definitional Structures Inherent in their Proofs*

General proof theory (Prawitz 1971, 1973a, b) is concerned with the structure of proofs. Studying the structure of proofs of given propositions could be viewed as a sort of intensional inverse mathematics. A canonical example is the connection between consistency and the normalization of proofs in natural deduction systems. Ekman (1994) studied the possible structure of proofs of Cantor's theorem in natural deduction systems of set theory. His investigations led, among other things, to insights concerning the impredicative nature of proofs of certain simple propositions in intuitionistic propositional logic. Here is a twofold challenge in pursuing Ekman's investigations towards of more complex set theoretic propositions such as the axiom of choice and the continuum hypothesis, and to introduce more explicit definition theoretic notions. It is clearly possible to formulate several of Ekman's notions in terms of definitional structures of proofs.

Ekman shows that all proofs of $\neg(p \leftrightarrow \neg p)$ in intuitionistic propositional logic reduce to a proof containing the following structure: a derivation step

$$\frac{\neg p \quad \neg p \rightarrow p}{p}$$

is immediately followed by the derivation step

$$\frac{p \quad p \rightarrow \neg p}{\neg p}$$

Let us say that a definition D is assigned to the derivation step

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

if $\vdash_D B$ and $A \vdash_D C$. In the present case, the definition

$$D = \{\ p = p \rightarrow \bot$$

can be assigned to both derivation steps in Ekman's example. It can be seen that for any such assignment (provided there is at least some atom which is not defined), D cannot be a total definition. So we have a kind of intensional characteristics of the proof theoretic structure needed to prove $\neg(p \leftrightarrow \neg p)$. The idea here is simply that the pairs

$(D, \neg p \vdash_D p)$, $(D, p \vdash_D \neg p)$ describe a definitional structure that is intrinsic to the original proof. This structure is then, of course, strongly connected with the structure of Russell's paradox etc. and the proof structure we need to prove Cantor's theorem.

4. CONCLUDING REMARKS

A definitional structure is a presentation of some notion of a definition over a given structure. We use the local logic $\vdash$ with respect to a definition D to study and characterize logical properties of this presentation. The type of properties we think of are then typically invariant with respect to the structure of presentations. It is not completely obvious how to make this concept of invariance precise. One suggestion is to use a mapping

$$i: U_1 \to U_2$$

preserving the basic definitional and logical structure, i.e.,

$$i(\top) = \top, i(\bot) = \bot$$

$$i(A \to B) = i(A) \to i(B) \text{ etc.}$$

$$i[D_1(a)] = D_2(i(a))$$

to define definitional isomorphism. But it is not clear whether this "structural" solution is too simple minded. Perhaps a more "logical" equivalence is a better candidate for the given intuitive notion of definitional invariance. In any case, it is a very appealing idea to capture definitional properties in terms of invariance with respect to some "natural" notion of equivalence.

ACKNOWLEDGEMENTS

This work has been supported by grants from the Swedish Natural Science Research Council (NFR), the Swedish Research Council for Engineering Sciences (TFR) and the National Board for Industrial and Technical Development (NUTEK).

NOTE

[1] This is spelled out in detail in Hallnäs (1991), p. 132.

REFERENCES

Colson, L.: 1991, 'About Primitive Recursive Algorithms', *Theoretical Computer Science* **83(1)**.

Ekman, J.: 1994, *Normal Proofs in Set Theory*, Ph.D. thesis, Department of Computing Science, Chalmers University of Technology.

Fredholm, D.: 1995, 'Intensional Aspects of Function Definitions', *Theoretical Computer Science* **152(1)**.

Grenander, U.: 1994, *General Pattern Theory, A Mathematical Study of Regular Structures*, Oxford University Press.

Hallnäs, L.: 1991, 'Partial Inductive Definitions', *Theoretical Computer Science* **87**.

Lorenzen, P.: 1955, *Einführung in die operative Logik und Mathematik*, Springer, Berlin.

Prawitz, D.: 1971, 'Ideas and Results in Proof Theory', in J. E. Fenstad (ed.) *Proceedings of the Second Scandinavian Logic Symposium*, North Holland, Amsterdam.

Prawitz, D.: 1973, 'Towards a General Proof Theory', in P. Suppes (ed.) *Logic, Methodology and the Philosophy of Science IV*, North Holland, Amsterdam.

Prawitz, D.: 1973, 'On the Idea of a General Proof Theory', *Synthese* **27**.

Schroeder-Heister, P.: 1993, 'Rules of Definitional Reflection', in *Proceedings of the 8th Annual IEEE Symposium on Logic in Computer Science, Montreal 1993*, Los Alamitos.

Tait, W.: 1967, 'Intensional interpretations of functionals of finite type I', *Journal of Symbolic Logic* **32**.

Department of Computing Science
Chalmers University of Technology
412 96 Göteborg
Sweden
E-mail: lars@cs.chalmers.se

WILLIAM W. TAIT

PROOF-THEORETIC SEMANTICS FOR CLASSICAL MATHEMATICS

ABSTRACT. We discuss the semantical categories of *base* and *object* implicit in the Curry-Howard theory of types and we derive derive logic and, in particular, the comprehension principle in the classical version of the theory. Two results that apply to both the classical and the constructive theory are discussed. First, compositional semantics for the theory does not demand 'incomplete objects' in the sense of Frege: bound variables are in principle eliminable. Secondly, the relation of extensional equality for each type is definable in the Curry-Howard theory.

The picture of mathematics as being about constructing objects of various sorts and proving the constructed objects equal or unequal is an attractive one, going back at least to Euclid. On this picture, what counts as a mathematical object is specified once and for all by effective rules of construction.

In the last century, this picture arose in a richer form with Brouwer's intuitionism. In his hands (for example, in his proof of the Bar Theorem), proofs themselves became constructed mathematical objects, the objects of mathematical study, and with Heyting's (1959) development of intuitionistic logic, this conception of proof became quite explicit. Today it finds its most elegant expression in the Curry–Howard theory of types, in which a proposition may be regarded, at least in principle, as simply a type of object, namely the type of its proofs. When we speak of 'proof-theoretic semantics' for mathematics, it is of course this point of view that we have in mind.

On this view, objects are given or constructed as objects of a given type. *That* an object is of this or that type is thus not a matter for discovery or proof. One consequence of this view is that equality of types must be a decidable relation. For, if an object is constructed as an object of type A and A and B are equal, then the object is of type B, too, and this must be determinable.

One pleasant feature of the type theoretic point of view is that the laws of logic are no longer 'empty': The laws governing the type

$$\forall x{:}A.F(x) = \Pi_{x:A}F(x)$$

Synthese (2006) 148: 603–622
DOI 10.1007/s11229-004-6271-x

simply express our general notion of a function, and the laws governing

$$\exists x{:}A.F(x) = \Sigma_{x:A}F(x)$$

express our notion of an ordered pair.

Much of my discussion applies equally to constructive mathematics. But the type-theoretic point of view remains, for many people, restricted to the domain of constructive mathematics. The term 'classical' is included in the title to indicate that, on the contrary, classical mathematics can also be understood in this way and does not need to be founded on an inchoate picture of truth-functional semantics in the big-model-in-the-sky, a picture that can in any case never be coherently realized.

Of course, no particular system of types is going to capture all of classical – or for that matter, constructive – mathematics. In the classical case, the open-endedness can be expressed by the possibility of introducing ever larger systems of transfinite numbers as types. But here I will discuss only the elementary theory of types, omitting even the introduction of the finite numbers.

One thing to be noticed, and this is independent of whether or not one admits classical reasoning, is that Frege's simple ontology of function and object must be abandoned. In particular, his notion of a concept, as a truth-valued function, won't do: it must be replaced by the notion of a propositional or type-valued function. This is obvious in the case of constructive mathematics; but it applies equally to the classical case. What we prove are propositions, not truth-values: propositions may be said to *have* truth-values; but that is in virtue of being provable or refutable. Moreover, our concepts, i.e., proposition- or type-valued functionals, range not over the 'universe of all objects', as for Frege, but over specific types.

1.

Another element of Frege's picture that I want to at least avoid is his notion of an 'incomplete object': the notion of a function as what is denoted by an open term and the notion of a propositional function as what is expressed by an open sentence. This very ugly idea has raised its head again in recent times because of the apparent need for bound variables in formulas such as

$$Qx{:}A.F(x)$$

where Q is a quantifier, and in terms

$$\lambda x{:}A.t(x)$$

expressing universal quantifier introduction. ($x{:}A$ expresses the restriction of x to objects of type A.)

However useful in practice, variables are dispensible in the compositional semantics of type theory. Providing that we can always bring the formula $F(v)$ into the form F', where F' contains only variables in F other than v and denotes a propositional function defined on A, and similarly we can always bring $t(v)$ into the form $t'v$, where t' contains only variables in t other than v and denotes a function defined on A, then we may eliminate bound variables as primitive notation and write

$$Qx : A.F(x) \;:=\; QF' \qquad\qquad \lambda x : A.t(x) := t'.$$

That we can eliminate bound variables in this way I proved in Tait (1998b), building on the work of Schönfinkel in (1924). Let me describe the ontology upon which the elimination is founded.

More generally, consider a sentence of the form

$$Q_1x_1 : A_1\ Q_2x_2 : A_2(x_1)\cdots$$
$$Q_nx_n : A_n(x_1,\ldots,x_{n-1}.F(x_1,\ldots,x_n)$$

where the Q_k are quantifiers. Iterating the above procedure, we obtain

$$Q_1x_1 : A'_1\ Q_2x_2 : A'_2x_1\cdots Q_nx_n : A'_nx_1\cdots x_{n-1}.A'x_1\cdots x_n$$
$$= Q_1\cdots Q_nA'$$

where

$$A'_kx_1\cdots x_{k-1} = A_k(x_1,\ldots,x_{k-1})$$
$$F'x_1\cdots x_n = F(x_1,\ldots,x_n).$$

The sequence

$$A'_1,\ldots,A'_n,F'$$

is a type-base in the sense of the following.

DEFINITION. The notion of a *type-base* or simply a *base* and of an *argument* for a base is defined by induction.

- The null sequence is a base and its only argument is the null sequence of objects.

- For $n \geq 0$, the sequence

  $$F_0, \ldots, F_n$$

  is a base iff $F_0, \ldots, F_{n-1}$ is a base and F_n is a type-valued function defined on the arguments of $F_0, \ldots, F_{n-1}$. An argument for this base is a sequence $b_0, \ldots, b_n$ such that $\mathbf{b} = b_0, \ldots, b_{n-1}$ is an argument for $F_0, \ldots, F_{n-1}$ and b_n is of type $F_n\mathbf{b}$. In particular, the unit sequence consisting of a type A is a base. Its arguments are the unit sequences consisting of objects of type A.

Thus, every initial segment $F_0, \ldots, F_k$ $(k < n)$ of a base $\mathbf{B} = F_0, \ldots, F_n$ is a base. If $b_0, \ldots, b_n$ is an argument for $\mathbf{B}$, then $F_k b_0 \cdots b_{k-1}$ is a type. We will call the terms in a base *functionals*. When $\mathbf{B}, F$ is a base, we call $\mathbf{B}$ the *base* of the functional F. When $\mathbf{B}$ is a base and $\mathbf{b}$ is an argument for it, we write

$$\mathbf{b} : \mathbf{B}.$$

As a special case, when B is a type

$$b : B$$

means that b is an object of type B.

2.

We assume given an initial stock of functionals, closed under bases, and, for each included type, all the objects of that type. We now extend this stock by means of base-forming operations. In §3, we extend the class of objects by means of object-forming operations

INSTANTIATION. We may, à la Schönfinkel, regard a functional of $n+1$ variables as a function of the first variable, whose values are functions of the remaining variables:

$$Fb_0b_1 \cdots b_n = (Fb_0)b_1 \cdots b_n.$$

Thus, when $\mathbf{B} = A, F_0, \ldots, F_n$ is a base and $b : A$, then

$$\mathbf{B}b = F_0b, \ldots, F_nb$$

defines a base whose arguments are those sequences $b_0, \ldots, b_n$ such that $b, b_0, \ldots, b_n$ is an argument for $A, F_0, \ldots, F_n$. $\mathbf{B}b$ is called an *instantiation* of $\mathbf{B}$.

We assume, too, that when $\mathbf{B}$ is in the initial stock of bases, then so is its instantiations.

QUANTIFICATION. When F has base A, $\forall F$ and $\exists F$ are types.

We may extend quantification QF, defined initially for a functional whose base is a type, to arbitrary functionals with non-null bases as follows: when F has base $\mathbf{B}, G$, then QF has base $\mathbf{B}$. If $\mathbf{b}$ is an argument for $\mathbf{B}$, then $(QF)\mathbf{b}$ is defined point-wise by

$$(QF)\mathbf{b} = Q(F\mathbf{b}).$$

If the base of the functional F is of length n, then we define its *universal closure* to be

$$F^* := \forall \cdots \forall F$$

with n occurrences of $\forall$. So F^* is a type.

If $b_0, \ldots, b_n$ is an argument for the base $F_0, \ldots, F_n$, then the type $F_k b_0 \cdots b_{k-1}$ of b_k depends upon $b_0, \ldots, b_{k-1}$. But sometimes we may wish to consider propositional functions F of n arguments of independent types $D_1, \ldots, D_n$, respectively. For example, in first or higher order predicate logic, we have variables ranging over the type of individuals, the type of sets of individuals, the type of sets of these, and so on. To deal with this, we need the following operation:

DUMMY ARGUMENTS. If A is a type and $F_0, \ldots, F_n$ is a base, the base $A, F_0[A], \ldots, F_n[A]$ is defined by

$$F_k[A]a = F_k.$$

We extend this operation point-wise: If $\mathbf{B}, G$ and $\mathbf{B}, H_0, \ldots, H_n$ are bases, then so is

$$\mathbf{B}, G, H_0[G], \ldots, H_n[G]$$

where, for each argument $\mathbf{b}$ for $\mathbf{B}$

$$H_k[G]\mathbf{b} := H_k\mathbf{b}[G\mathbf{b}].$$

Now, given the list $D_1, \ldots, D_n$ of types, we may form the base $D^1, \ldots, D^n$ where $D^1 = D_1$ and

$$D^{k+1} = D_{k+1}[D^1][D^2] \cdots [D^k].$$

Then, if $\mathbf{b} = b_1, \ldots, b_n$ is an argument for this base, then

$$D^k b_1 \cdots b_{k-1} = D_k.$$

Hence $\mathbf{b}$ is an argument for the base iff $b_k : D_k$ for each $k = 1, \ldots, n$.[1]

In terms of the quantifiers and dummy arguments, we can define implication and conjunction: Let F and G have base $\mathbf{B}$. then

$$F \longrightarrow G \;:=\; \forall G[F] \qquad F \wedge G \;:=\; \exists G[F].$$

Then $F \longrightarrow G$ and $F \wedge G$ have base $\mathbf{B}$. Note that, if $\mathbf{b}:\mathbf{B}$, then

$$(F \longrightarrow G)\mathbf{b} = (F\mathbf{b} \longrightarrow G\mathbf{b}) \qquad (F \wedge G)\mathbf{b} = (F\mathbf{b} \wedge G\mathbf{b}).$$

Coimplication is defined between functionals with the same base by

$$F \longleftrightarrow G \;:=\; (F \longrightarrow G) \wedge (G \longrightarrow F).$$

There is one more operation on bases that we need, besides quantification, instantiation and introducing dummy arguments:

TRANSPOSITION. Let

$$F, G, H_0, \ldots, H_n$$

be a base. $\forall G$ is a type and $F[\forall G]$ has base $\forall G$; so $\forall G, F[\forall G]$ is a base. If c, d is an argument for this base, then $c:\forall(G)$ and $d:F[\forall G]c = F$. So cd is defined and is of type Gd. Thus d, cd is an argument for F, G. It follows that we can form a new base

$$\forall G, F[\forall G], H_0\{G\}, \ldots, H_n\{G\}$$

where the functionals $H_k\{G\}$ are defined by

$$H_k\{G\}cd := H_k d(cd).$$

Again, we may extend the operation point-wise: Let $\mathbf{B}, F, G, H_0, \ldots, H_n$ be a base. Then the base $\mathbf{B}, \forall G, F[\forall G], H_0\{G\}, \ldots, H_n\{G\}$ is defined by

$$H_k\{G\}\mathbf{b} = H_k\mathbf{b}\{G\mathbf{b}\}.$$

3.

We turn now to object-forming operations.

The LAW OF $\forall$-ELIMINATION is

$$f : \forall F \;\;\&\;\; b : A \;\;\Rightarrow\;\; fb : Fb.$$

The LAW OF $\exists$-INTRODUCTION is

$$b : A \;\;\&\;\; c : Fb \;\;\Rightarrow\;\; (b, c) : \exists F$$

and the LAW OF $\exists$-ELIMINATION takes the form of projections:

$$p : \exists F \;\;\Rightarrow\;\; p1 : A \;\;\&\;\; p2 : F(p1).$$

Note that 1 and 2 do not count here as objects.[2]

In order to obtain the law of $\forall$-Introduction without using lambda-abstraction, we need to introduce a generalization of Schönfinkel's combinators.

COMBINATOR K. If A and B are types, then

$$K(A, B) : A \longrightarrow (B \longrightarrow A)$$

where

$$K(A, B)ab = a.$$

This is the typed version of one of Schönfinkel's combinators.

We extend K to functionals F and G with a common base $\mathbf{B}$: if $\mathbf{b} : \mathbf{B}$, then

$$K(F, G) : (F \longrightarrow (G \longrightarrow F))^*$$

is defined point-wise by setting

$$K(F, G)\mathbf{b} = K(F\mathbf{b}, G\mathbf{b}).$$

COMBINATORS $S_\forall$ and $S_\exists$. Let H have base F, G and let Q be a quantifier $\forall$ or $\exists$. Then $\forall QH$ and $Q\forall(H\{G\})$ both are types. We will define the combinator

$$S_Q(H) : (\forall QH \longrightarrow Q\forall(H\{G\}).$$

Let $c : \forall Q(H)$. Then

$$S_Q(H)c : Q\forall(H\{G\}).$$

$H\{G\}$ has base $\forall G, F[\forall G]$.

First, let $Q = \forall$.

$$S_\forall(H)c : \forall\forall(H\{G\}).$$

Let $d : \forall G$ and $e : F = F[\forall G]c$. $S_\forall(H)cde$ must be defined to be of type $H\{G\}de$, i.e., of type $He(de)$, which is the type of $ce(de)$. Hence, we define

$$S_\forall(H)cde = ce(de).$$

Thus, $S_\forall$ is the typed version of Schönfinkel's other combinator.

Now let $Q = \exists$.

$$S_\exists(H)c : \exists\forall(H\{G\}).$$

Thus we must have

$$S_\exists(H)c1 : \forall G$$

$$S_\exists(H)c2 : \forall H\{G\}(S_\exists(H)c1).$$

Let $d : F$. Then we must have

$$S_\exists(H)c1d : Gd$$

$$S_\exists(H)c2d : Hd(S_\exists(H)c1d).$$

But $cd : \exists Hd$ and so $cd1 : Gd$ and $cd2 : Hd(cd1)$. So we may define $S_\exists(H)$ by

$$S_\exists(H)c1d := cd1 \qquad S_\exists(H)c2d := cd2.$$

We again extend the combinators $S_Q(H)$ to the case in which H has a base $\mathbf{B}, F, G$ by point-wise definition:

$$S_Q(H) : (\forall QH \longrightarrow Q\forall H\{G\})^*.$$

Let $\mathbf{b} : \mathbf{B}$. Then

$$S_Q(H)\mathbf{b} := S_Q(H\mathbf{b}).$$

Notice that, if H has base $A, B[A]$, the type

$$\forall\exists H \longrightarrow \exists\forall H\{B[A]\}$$

of $S_\exists(H)$, which may be expressed by

$$\forall x : A\exists y : B.Hxy \longrightarrow \exists z : (A \longrightarrow B)\forall x : A.Hx(zx)$$

is an expression of the Axiom of Choice. Our definition of the combinator $S_\exists(H)$ amounts to a constructive proof of this axiom.

4.

We need now to discuss the notion of definitional equality. We are discussing a system Σ of bases and objects built up by means of certain operations: instantiation, quantification, introducing dummy arguments, transposition, and $\exists$ and $\forall$ introduction and elimination from a given stock of bases and objects, which are distinct from the newly introduced ones. We assume that equality between given functionals or objects is given and we assume that it is closed under instantiation and $\forall$-elimination. Thus, besides the defining equations given above for the new functionals and objects, we have all of the true equations

$$Fa = G \qquad fa = b$$

concerning given functionals F and G and given objects a and b (when f is of some type $\forall H$ and A, F is a given base). We can extend the equality relation to the new objects and types by taking it to be the equivalence relation generated by the defining equations and the true equations concerning the given objects and functionals. However, for functionals in general, we need a weaker (i.e., more inclusive) notion of equality, which we can define by induction on the length of their bases: two functionals F and G are equal if their bases have the same

length n and the members are pairwise equal and if $Fx_1 \cdots x_n = Gx_1 \cdots x_n$ follows purely formally from the given equations, where $x_1, \ldots, x_n$ are distinct symbols. We need to specify that, when an object c is of some type A and $A = B$, then c is of type B, too.

It can be shown that this definition of equality, called *definitional equality*, is decidable relative to the true equations concerning the given functionals and objects. It turns out, too, that when $b = c$ and $b:A$, then $c:A$. So each object has a unique type. The inelegant definition of equality between functionals with non-null bases is necessitated by the fact that we need below some special cases of the equations

$$H[G]\{G\} = H[\forall G]$$
$$(QH)[G] = Q(H[G])$$

whenever Q is a quantifier and the left-hand sides are meaningful. These equations are valid for the notion of equality we have just introduced; but it would be more satisfactory to be able to extend this list of equations to a 'complete' one, i.e., so that equality of objects or functionals in general could be taken to be the equivalence relation generated by all the equations. For example, besides the above equations, we would need

$$(QH)\{G\} = Q(HG)$$
$$H[F][G[F]] = H[G][F]$$
$$H\{F\}\{G\{F\}\} = H\{G\}\{F\}$$
$$H[F]\{G[F]\} = H\{G\}[F]$$

when, again, the left-hand sides make sense.

QUESTION. Is this system of equations complete? I suspect so; but if not, how is it to be extended to a complete system?

5.

Let G and H have base $\mathbf{B}, F$, so that $H[G]\{G\} = H[\forall G]$. Let $S = S_\forall(H[G])$. Then

$$S:(\forall\forall H[G] \longrightarrow \forall\forall(H[G]\{G\}))^*.$$

Hence

$$S:(\forall(G \longrightarrow H) \longrightarrow \forall\forall(H[\forall G]))^*.$$

But $\forall(H[\forall G]) = (\forall H)[\forall G]$ and so, finally

$$S\colon(\forall(G \longrightarrow H) \longrightarrow (\forall G \longrightarrow \forall H)). \tag{1}$$

Note that the types of $K(F, G)$ and S, i.e.

$$(F \longrightarrow \forall F[G])^* \qquad \text{and} \qquad (\forall(G \longrightarrow H) \longrightarrow (\forall G \longrightarrow \forall H))^*$$

are precisely Quine's (1951) axioms for the universal quantifier in first-order predicate logic, which have the property of avoiding λ-abstraction, i.e. hypothetical proof. The difference is that Quine retains bound variables in formulas. Both Quine (1960a) and Bernays (1959) discussed the question of eliminating bound variables in formulas in first-order logic; but neither presented an entirely adequate account from the point of view of proof theory, even for predicate logic.

If in the equation (1) we replace G and H by $B[A]$ and $C[A]$, respectively, where A, B, C are types, then we obtain Schönfinkel's combinator S of type

$$[A \longrightarrow (B \longrightarrow C)] \longrightarrow [(A \longrightarrow B) \longrightarrow (A \longrightarrow C)].$$

This type, together with the type of K(A,B):

$$A \longrightarrow (B \longrightarrow A)$$

are the axioms of positive implicational logic. This correspondence between the positive implicational logic and the typed theory of combinators seems to have been first noticed in Curry and Feys (1958) and is cited in Howard (1980) as one of the sources of the propositions-as-types point of view.

6.

So far, we've said nothing about eliminating bound variables, or what amounts to the same thing, eliminating the need for free variables. In order to show how we can eliminate variables, we first have to introduce them. Let A be a type in Σ and let v be a symbol that we take as an indeterminate of type A. We can construct the formal 'polynomial extension' $\Sigma[v]$ of Σ in the obvious way, formally closing it under the above operations and where equality is again the relation of definitional equality. For every $b\colon A$ in Σ, there is a homomorphism $b^* : \Sigma[v] \longrightarrow \Sigma$ obtained by 'substituting b for v in the formulas and terms of $\Sigma[v]$'. Restricted to Σ, b^* is the identity function. The following is proved in Tait (1998b).

EXPLICIT DEFINITION THEOREM. Let $F_1, \ldots, F_n$ be a base and t a term of type B in $\Sigma[v]$.

– There is a base $A, F'_1, \ldots, F'_n$ in Σ such that

$$F'_k v = F_k \qquad (k = 1, \ldots, n)$$

– There is an object $t' : \forall B'$ in Σ such that

$$t' v = t.$$

We can write

$$\lambda x : A.F(x) := F' \qquad\qquad \lambda x : A.t(x) := t'.$$

This process of taking formal extensions can be iterated: let $\Sigma[v_0, \ldots, v_n]$ be given and let B_{n+1} be a type in this system. Choose a new symbol v_{n+1} as an indeterminate of type B and form $\Sigma[v_0, \ldots, v_{n+1}] = \Sigma[v_0, \ldots, v_n][v_{n+1}]$.

Given a functional $F(v_1, \ldots, v_n)$ in the system $\Sigma[v_1, \ldots, v_n]$, we can construct the functional

$$F' = \lambda x_1 : B_1 \cdots \lambda x_n : B_b.F(x_1, \ldots, x_n)$$

in Σ such that

$$F' v_1 \cdots v_n = F(v_1, \ldots, v_n).$$

7.

Notice that we have not introduced disjunction in type theory: it is indeed an awkward operation. Were we to introduce the type **N** of the natural numbers with its corresponding introduction and elimination rules, the functional $= 0$ with base **N** can be defined and so disjunction can be defined by

$$A \vee B := \exists x : \mathbf{N}[(x = 0 \longrightarrow A) \wedge (x \neq 0 \longrightarrow B)].$$

But the most elementary way to deal with disjunction is to introduce the base

$$\mathbf{2}, \mathbf{T}$$

2 is the two-object type. **2**-Introduction specifies that the *truth-values* $\top$ and $\bot$ are of type **2**. **2**-Elimination asserts the existence, for any functional F with base **2**, of

$$\mathbf{N}_2(F) : [F\top \longrightarrow (F\bot \longrightarrow \forall F)]$$

where

$$\mathbf{N}_2(F)bc\top := b \qquad \mathbf{N}_2(F)bc\bot := c.$$

We take $\mathbf{T}\top$ to be the terminal type and $\mathbf{T}\bot$ to be the initial type, i.e.,

$$\mathbf{1} := \mathbf{T}\top \qquad \mathbf{0} := \mathbf{T}\bot$$

1-Introduction specifies that $\imath$ is an object of type **1** and **1**-Elimination asserts the existence, for any functional F of base **1**, of

$$\mathbf{N}_1(F) : [F\imath \longrightarrow \forall F]$$

where

$$\mathbf{N}_1(F)b\imath := b.$$

There is no **0**-Introduction, of course. **0**-Elimination asserts the existence, for any functional F of base **0**, of

$$\mathbf{N}_0(F) : \forall F.$$

Our types $\mathbf{k}$ are of course Martin-Löf's (1998) types N_k.

In order to preserve the Explicit Definition Theorem, the elimination rules for $\mathbf{k} = \mathbf{2}, \mathbf{1}$ and **0** need to be extended in the usual way by point-wise definition to functionals F with bases of the form

$$\mathbf{B}, \mathbf{k}[\mathbf{B}]$$

where, if $\mathbf{B}$ is $G_0, \ldots, G_n$, then $\mathbf{k}[\mathbf{B}] := \mathbf{k}[G_0][G_1]\cdots[G_n]$. We may do this as follows: Let F^+ with base $\mathbf{k}, \mathbf{B}[\mathbf{k}]$ be defined by

$$F^+ := \lambda x : \mathbf{k} \lambda y_1 : B_1 \cdots \lambda y_n : B_b . F y_1 \cdots y_n x$$

Then we define

$$\mathbf{N}_2(F) : [F^+\top \longrightarrow (F^+\bot \longrightarrow \forall F)]^*$$
$$\mathbf{N}_1(F) : [F^+\imath \longrightarrow \forall F]^*$$
$$\mathbf{N}_0(F) : \forall F$$

by

$$\mathbf{N}_\mathbf{k}(F)\mathbf{b} := \mathbf{N}_\mathbf{k}(F\mathbf{b})$$

for $\mathbf{b} : \mathbf{B}$.

If F has base $\mathbf{B}$, then we define

$$\neg F := F \longrightarrow \mathbf{0}[\mathbf{B}].$$

Let F and G have the base $\mathbf{B}$. Set

$$\langle F, G\rangle := \exists x : \mathbf{2} . [(\mathbf{T}x \longrightarrow A) \wedge (\neg \mathbf{T}x \longrightarrow B)].$$

It is easy to deduce (i.e. find objects of type)

$$(F \longleftrightarrow \langle F, G\rangle \top)^* \qquad (G \longleftrightarrow \langle F, G\rangle \bot)^*. \tag{2}$$

So we may define

$$F \vee G := \exists x{:}\mathbf{2}\langle F, G\rangle x.$$

With this definition, using the deducibility of (2), the usual laws of $\vee$-Introduction and -Elimination are derivable.

8.

Up to now, we have been discussing the theory of types in general, with no particular reference to the classical theory. The latter is of course obtained by adding the law of

$\neg\neg$-ELIMINATION. Let F have base $\mathbf{B}$. Then

$$D(F){:}(\neg\neg F \longrightarrow F)^*.$$

Here again, when $\mathbf{B}$ is non-null, $D(F)$ is defined point-wise in terms of $D(A)$ for A a type:

$$D(F)\mathbf{b} := D(F\mathbf{b})$$

when $\mathbf{b}{:}\mathbf{B}$.

From the type-theoretic point of view, what characterizes classical mathematics is not truth-functional semantics, but the introduction of what, *somewhat* in the spirit of Hilbert, we may call the *ideal* objects $D(F)$.

It is interesting that the problem that writers have found with classical mathematics, that there exist undecidable propositions A, such as the continuum hypothesis, for which the law of excluded middle nevertheless is taken to be valid, takes a different form when one moves from the (anyway incoherent) point of view of truth-functional semantics to that of type theory. Even constructively we can produce a deduction p of

$$\neg\neg(A \vee \neg A).$$

So classically we have

$$q = D(A \vee \neg A)))p{:}A \vee \neg A.$$

i.e.,

$$q{:}\exists x{:}\mathbf{2}.\langle A, \neg A\rangle x.$$

Therefore, $q1 : \mathbf{2}$ and $q2 : \langle A, \neg A\rangle(q1)$. But we are unable to conclude that $q1$ is $\top$ and we are unable to conclude that it is $\bot$; and so we are unable to conclude that $q2$ is a proof of A or that it is a proof of $\neg A$. We will see below that, unlike the relation of definitional equality, we can define in type theory the notion of (extensional) equality $\equiv$. Hence, by **2**-Elimination (which states that anything true of both $\top$ and $\bot$ is true of all objects of type **2**)

$$q1 \equiv \top \ \vee \ q1 \equiv \bot$$

will be derivable, since

$$\top \equiv \top \ \vee \ \top \equiv \bot$$

and

$$\bot \equiv \top \ \vee \ \bot \equiv \bot$$

will both be derivable. Notice that this situation concerns not just undecidable propositions in classical mathematics, but *any* proposition A for which excluded middle cannot be proved constructively. Also, in the same way in which there are ideal objects of type **2** which cannot be said to be $\top$ and cannot be said to be $\bot$, but nevertheless are either one or the other, so there are ideal numbers for example, i.e., objects of type **N** (were we to introduce this type), which therefore are in the sequence $0, 1, 2 \ldots$ but which we cannot locate in this sequence. This is an interesting observation about how disjunction works in classical mathematics; but it seems paradoxical only when one assumes that classical mathematics is based on truth-functional semantics.

9.

In classical logic, the sets of elements of type A are not themselves types, but are precisely the **2**-valued functions on A, i.e., they are the objects of type

$$\mathbf{P}(A) := A \longrightarrow \mathbf{2}.$$

We may define the functional ϵ_A with base $A, \mathbf{P}(A)[A]$ by

$$a\epsilon_A f := \epsilon_A a f := \mathbf{T}(fa).$$

So when $fa = \top$, $a\epsilon f$ is the true proposition **1** and when $fa = \bot$, it is the false proposition **0**.

We have shown how, in classical logic, for any functional F with base A, to obtain an object

$$p : \forall x{:}A[F(x) \vee \neg F(x)].$$

Let $f = \lambda x{:}A(px1)$ and $g = \lambda x{:}A(px2)$. Then $f{:}A \longrightarrow \mathbf{2} = \mathbf{P}(A)$. Let $u{:}A$. Then gu is a deduction of $\langle F(u), \neg F(u)\rangle(fu) =$

$$[\mathbf{T}(fu) \longrightarrow F(u)] \wedge [\neg\mathbf{T}(fu) \longrightarrow \neg F(u)].$$

From the second conjunct, in classical logic, we obtain a deduction of $F(u) \longrightarrow \mathbf{T}(fu)$. Recalling that $u\epsilon_A f$ is $\mathbf{T}(fu)$, we thus have a deduction $r(u)$ of

$$u\epsilon_A f \longleftrightarrow F(u).$$

So $(f, \lambda x{:}A.r(x))$ is a deduction of the *COMPREHENSION PRINCIPLE*[3]

$$\exists z{:}\mathbf{P}(A)\forall x{:}A[x\epsilon z \longleftrightarrow F(x)].$$

10.

In this final section, I will show that extensional equality can be defined in the Curry–Howard theory.[4] Recall that equality between types is to be understood intensionally, as *definitional equality*. This is important for the type-theoretic point of view: what the type of an object is should be determined from the object.

If c is given as an object of type A and A is the same type as B, then we should be able to determine that c is of type B. Similarly, the identity of objects should be understood in terms of definitional equality: objects are given by terms and two terms denote the same object if they are definitionally equal. Of course, definitional equality as a relation among the objects of some type A is not expressible as a type; that is, there is no functional E with base $A, A[A]$ such that the type Ebc expresses the definitional equality of objects b and c of type A.

There is, however, the notion of *extensional equality* between objects of a given type, which we can express in type theory. Let me discuss this.

An immediate problem with defining extensional equality between objects of the same type is this: Let c and d be of type $\forall x{:}A.F$. Then clearly the extensional equality of c and d should imply that, for all extensionally equal a and b of type A, ca should be extensionally equal to db. But the types Fa and Fb of ca and db, respectively, are

not in general the same type, i.e., are not definitionally equal. So we must define extensional equality between object of certain different types.

In virtue of the Explicit Definition Theorem, every functional is of the form

$$Fb_1 \cdots b_n$$

where $n \geq 0$ and F is built up without using instantiation. If

$$Fc_1 \cdots c_n$$

is another functional, then we call these two functionals *congruent*. Obviously, congruence is an equivalence relation on the functionals. We need to define the functional $\equiv_{FG}$ of extensional equality for congruent functionals F and G. We will drop the subscript on $\equiv_{FG}$ when no confusion results. Let F have base $\mathbf{B}$ and G base $\mathbf{B}'$. $\equiv_{FG}$ will be defined as a functional with base $\mathbf{B}, \mathbf{B}'[\mathbf{B}], F[\mathbf{B}'[\mathbf{B}]], G[\mathbf{B}, F[\mathbf{B}'[\mathbf{B}]]]$, point-wise: for each $\mathbf{b}:\mathbf{B}$ and $\mathbf{b}':\mathbf{B}'$

$$\equiv_{FG} \mathbf{b}\mathbf{b}' \quad := \quad \equiv_{F\mathbf{b}G\mathbf{b}'} .$$

So it suffices to define $\equiv_{AB}$ for congruent types A and B. But for this we need only define what $a \equiv b$ means for objects $a:A$ and $b:B$ in some polynomial extension of $\mathbf{\Sigma}$. For then we obtain $\equiv$ as $\lambda x:A\lambda y:B.x \equiv y$.

We assume that the relation of extensional equality is defined for the basic types and that it is an equivalence relation. We define it now for the new types that we have introduced.

DEFINITION OF *EXTENSIONAL EQUALITY*. Let $a:A$ and $b:B$.

– $A = B = \mathbf{2}$. Recalling that $\mathbf{T}\top$ is the terminal type $\mathbf{1}$ and $\mathbf{T}\bot$ is the initial type $\mathbf{0}$, it clearly suffices to define $\equiv$ by

$$a \equiv b \quad := \quad \mathbf{T}a \longleftrightarrow \mathbf{T}b.$$

– $A = \mathbf{T}c, B = \mathbf{T}d$. Then c and d are of type $\mathbf{2}$.

$$a \equiv b \quad := \quad c \equiv d.$$

– $A = \forall F, B = \forall G$, where F and G have bases C and D, respectively. Then C and D are congruent and F and G are congruent.

$$a \equiv b \quad := \quad \forall x:C\exists y:D(x \equiv y) \wedge \forall y:D\exists x:C(x \equiv y)$$

$$\wedge\forall x:C\forall y:D(x \equiv y \longrightarrow ax \equiv by).$$

– $A = \exists F, B = \exists G$ where F and G have bases C and D, respectively. Again, C and F are respectively congruent to D and G.

$$a \equiv b \quad := \quad a1 \equiv b1 \wedge a2 \equiv b2.$$

It is easy to deduce that

$$\top \not\equiv \bot$$

$$\forall x{:}\mathbf{2}\ [x \equiv \top \ \vee \ x \equiv \bot]$$

$$\forall x{:}\mathbf{2} \forall y{:}\mathbf{T}x \forall z{:}\mathbf{T}xy \equiv z$$

and that extensional equality is transitive and symmetric. i.e.,

$$\forall x{:}A\ \forall y{:}A\ \forall z{:}A\ [x \equiv y \ \wedge \ y \equiv z \ \longrightarrow \ x \equiv z]$$

$$\forall x{:}A \forall y{:}A[x \equiv y \ \longrightarrow \ y \equiv x].$$

Moreover, for a given object a of type A *in* Σ, there is a deduction of $a \equiv a$. But, alas, we cannot deduce for an arbitrary type A that extensional equality on A is reflexive:

$$\forall x{:}A\ [x \equiv x]. \tag{3}$$

The problem case is, of course, when A is of the form $\forall F$, where F has some base B. It would be consistent to add non-extensional functions to Σ.

On the other hand, we could consider introducing the principle (3) as a 'regulative principle'. Thus, for each object b of a type A that is introduced, there must be a proof E'_b provided of $b \equiv b$. Then we would introduce the constant

$$E(A){:}\forall x{:}A.(x \equiv x).$$

by means of the definition

$$E(A)b := E_b.$$

However, we are not done, since according to our requirement for introducing objects, we have to prove the reflexivity of $E(A)$, itself: $E(A) \equiv E(A)$. I.e., we have to have a proof of

$$\forall x{:}A \forall y{:}A[x \equiv y \longrightarrow E(A)x \equiv E(A)y].$$

The proof of this turns out to be something of a *tour de force*.

PROPOSITION. In any polynomial extension on Σ, from $p{:}\ a \equiv a', q{:}\ b \equiv b'$ and $r{:}\ a \equiv b$, we can construct a proof of $p \equiv q$.

In particular, then, from proofs $r{:}u \equiv v$ we can construct a proof of $E(A)u \equiv E(A)v$. We prove the theorem by induction on the type of a. Note that from p, q and r we obtain a proof s of $a' \equiv b'$.

CASE 1. a is of type **2**. Then

$$p\colon \mathbf{T}a \longleftrightarrow \mathbf{T}a' \qquad q\colon \mathbf{T}b \longleftrightarrow \mathbf{T}b'.$$

Thus

$$p1\colon \mathbf{T}a \longrightarrow \mathbf{T}a' \qquad q1\colon \mathbf{T}b \longrightarrow \mathbf{T}b'.$$

We need $p1 \equiv q1$, i.e.

$$\forall x{:}\mathbf{T}a\ \exists y{:}\mathbf{T}b(x \equiv y) \qquad \forall y{:}Tb\ \exists x{:}\mathbf{T}a(x \equiv y)$$

and

$$\forall x\colon \mathbf{T}a\ \forall y\colon \mathbf{T}b[x \equiv y \longrightarrow p1x \equiv q1y]$$

$r1{:}\mathbf{T}a \longrightarrow \mathbf{T}b$ and $r2{:}\mathbf{T}b \longrightarrow \mathbf{T}a$. So, if $u{:}\mathbf{T}a$ and $v{:}\mathbf{T}b$, then $r1u{:}\mathbf{T}b$ and $r2v{:}\mathbf{T}a$. So the first two conditions are satisfied. The last condition is just $(\mathbf{T}a \longleftrightarrow \mathbf{T}b) \longrightarrow (\mathbf{T}a' \longleftrightarrow \mathbf{T}b')$, which is obtained from p and q. By symmetry, we also have $p2 \equiv q2$ and so $p \equiv q$.

CASE 2. $a\colon \mathbf{T}c$. Then $a'\colon \mathbf{T}c'$, $b\colon \mathbf{T}d$, $b'\colon \mathbf{T}d'$. Then $a \equiv a'$ is just $c \equiv c'$, etc.; and this case reduces to Case 1.

CASE 3. $a\colon \exists x\colon AF(x)$. Then $a1\colon A$ and $a2\colon F(a1)$, $p1\colon a1 \equiv a'1$, $q1\colon b1 \equiv b'1$, etc; and so, by the induction hypothesis, $p1 \equiv q1$ and $p2 \equiv q2$.

CASE 4. $a\colon (\forall x\colon A.F(x)), b\colon (\forall y\colon B.G(y)), a'\colon (\forall x'\colon A'.F'(x')), b'\colon (\forall y'\colon B'.G'(y'))$. In the following, I will drop the A in $x\colon A$, and similarly for x', y and y'. So p, as a proof of

$$\forall x \exists x'(x \equiv x') \wedge \forall x' \exists x(x \equiv x') \wedge \forall x x'[x \equiv x' \longrightarrow ax \equiv a'x']$$

has three components, p^0, p^1, p^2 which are proofs of the conjuncts, respectively, and similarly for q, r and s. We need to show for $i \equiv 0, 1, 2$ that $p^i \equiv q^i$. To prove $p^0 \equiv q^0$, we need

$$\forall x \exists y(x \equiv y) \qquad \forall y \exists x(x \equiv y)$$

which have the proofs r^0 and r^1, and

$$\forall xy[x \equiv y \longrightarrow p^0x \equiv q^0y].$$

Let $u\colon A, v\colon B$ and assume $u \equiv v$. $p^0u1\colon A'$, $q^0v1\colon B'$, $p^0u2\colon u \equiv p^0u1$ and $q^0v2\colon v \equiv q^0v1$. So $p^0u1 \equiv q^0v1$ and hence, by the induction hypothesis, $p^0u2 \equiv q^0v2$. Thus, $p^0 \equiv q^0$.

By symmetry, $p^1 \equiv q^1$.

As for p^2 and q^2, we have

$$p^2\colon \forall xx'[x \equiv x' \longrightarrow ax \equiv a'x'] \qquad q^2\colon \forall yy'[y \equiv y' \longrightarrow by \equiv b'y']$$

$p^2 \equiv q^2$ is clearly equivalent to the conjunction of

$$\forall x \exists y (x \equiv y) \qquad \forall y \exists x (x \equiv y)$$

$$\forall x' \exists y' (x' \equiv y') \qquad \forall y' \exists x' (x' \equiv y')$$

which have the proofs r^0 and s^0, and

$$\forall x x' y y' [x \equiv y \wedge x' \equiv y' \longrightarrow p^2 x x' \equiv q^2 y y'] \tag{4}$$

Let u, u', u^* be of types A, A' and $u \equiv u'$, respectively, and let v, v' and v^* be of types B, B' and $v \equiv v'$, respectively. Then $p^2 u u' u^* : au \equiv a'u'$ and $q^2 v v' v^* : bv \equiv bv'$. Let $w : u \equiv v$ and $w' : u' \equiv v'$. Then $r^2 uvw : au \equiv bv$. So by the induction hypotheses, $p^2 u u' u^* \equiv q^2 v v' v^*$ follows from the assumptions w and w'. This demonstrates (4).

NOTES

[1] We could avoid this admittedly artificial treatment of variables of independent types, but at the cost of a more complex structure of bases. Namely, we would define bases to be trees, where, besides the null base, trees of the form

$$\frac{\mathbf{B}_1 \cdots \mathbf{B}_n}{F}$$

are admitted as bases when $n \geq 0$, $\mathbf{B}_1, \ldots, \mathbf{B}_n$ are bases and, F is a type-valued function defined on $\mathbf{B}_1 \times \cdots \times \mathbf{B}_n$. An argument for this base is of the form $\mathbf{b}_1, \ldots, \mathbf{b}_n, c$, where $\mathbf{b}_k : \mathbf{B}_k$ for each $k = 1, \ldots, n$ and $c : F(\mathbf{b}_1, \ldots, \mathbf{b}_n)$. But, in the interests of simplicity, if not in the interests of efficient computation, we will continue to deal only with linear bases.

[2] This form of $\exists$-Elimination is different from that of Martin-Löf, e.g. in Martin-Löf (1998), although, as he notes, the two forms are equivalent. It would seem that projection more directly expresses what it means to have an object of type $\exists F$.

[3] Thus, the Comprehension Principle follows from the Law of Excluded Middle. The converse is also true: Let B be any type. Using the Comprehension Principle, there is a $b : \mathbf{P}(\mathbf{2})$ such that $\mathbf{T}(b\bot) \longleftrightarrow B$. Hence, $\neg\neg\mathbf{T}(b\bot) \longleftrightarrow \neg\neg B$. But even constructively, $\neg\neg\mathbf{T}(b\bot) \longrightarrow \mathbf{T}(b\bot)$ and so $\neg\neg B \longrightarrow B$.

[4] I discussed this in (1996); but the treatment, besides being unnecessarily complicated, contained an error which was discovered in conversations with Robert Harper and Christopher Stone at Carnegie-Mellon in Autumn 1999. Part of the complication in the earlier paper resulted from the fact that I thought that extensional equality could be treated only in the classical system. We shall see that this is false.

REFERENCES

Bernays, P.: 1959, 'Über eine Natürliche Erweiterung des Relationenkalküls', in A. Heyting (ed.), *Constructivity in Mathematics*, North-Holland, Amsterdam, pp. 1–14.

Curry, H. and Feys, R. 1958, *Combinatory Logic I, Studies in Logic and the Foundations of Mathematics*, North-Holland, Amsterdam; 2nd edition 1968.

Howard, W.: 1980, 'The formula-as-types notion of construction', in J. Hindley and J. Sheldon (eds.), *To H.B. Curry: Essays on Combinatorial Logic, Lambda Calculus and Formalism*, Academic Press, London, pp. 479–490.

Martin-Löf, P.: 1998, 'An intuitionistic theory of types', in G. Sambin and J. Smith (eds.), 1980, *Twenty-Five Years of Constructive Type Theory*, Oxford University Press, Oxford.

Quine, W.: 1951, *Mathematical Logic: Revised Edition*, Harvard University Press, Cambridge.

Quine, W.: 1960a, 'Variables explained away', *Proceedings of the American Philosophical Society*. Reprinted in Quine, (1966a), *Selected Logic Papers*, Random House, New York, pp. 227–235.

Schönfinkel, M.: 1924, 'Über die Bausteine der Mathematischen Logik', *Mathematische Annalen* **92**, 305–316.

Tait, W.: 1996, 'Extensional Equality in Classical Type Theory', in W. DePauli-Schimanovich, E. Köhler and F. Stadler (eds.), *The Foundational Debate: Complexity and Constructivity in Mathematics and Physics*, pp. 219–234.

Tait, W.: 1998b, 'Variable-free Formalization of the Curry–Howard Type Theory', in G. Sambin and J. Smith (eds.), 1980, *Twenty-Five Years of Constructive Type Theory*, Oxford University Press, Oxford, pp. 265–274.

Department of Philosophy
University of Chicago
5522 S. Everett Ave.
Chicago, IL 60637
U.S.A.
E-mail: wwtx@earthlink.net

GÖRAN SUNDHOLM

SEMANTIC VALUES FOR NATURAL DEDUCTION DERIVATIONS[1]

ABSTRACT. Drawing upon Martin-Löf's semantic framework for his constructive type theory, semantic values are assigned also to natural-deduction derivations, while observing the crucial distinction between (logical) consequence among propositions and inference among judgements. Derivations in Gentzen's (1934–5) format with derivable formulae dependent upon open assumptions, stand, it is suggested, for proof-objects (of propositions), whereas derivations in Gentzen's (1936) sequential format are (blue-prints for) proof-acts.

Contrasting Frege's logical systems, using (many) axioms and (few) rules of inference with those of Gentzen, using no axioms, but only rules of inference that may discharge open assumptions, Michael Dummett wrote:

> Frege's account of inference allows no place for a[n] ... act of supposition. Gentzen later had the highly successful idea of formalizing inference so as to leave a place for the introduction of hypotheses.

Indeed,

> it can be said of Gentzen that it was he who first showed how proof theory should be done.[2]

An evaluation of this claim concerning the merits of Gentzen depends on (i) in what sense *proof theory* is taken and on (ii) what, if any, was his contribution thereto. There are, at least, three relevant readings of the term *proof theory*:

(i) a "syntactic" manipulation-system that matches a prior ("semantic") consequence relation. Example: "Some consider proof theories using Polish notation less easy to use in practice; the formulae are so difficult to read".
(ii) investigations, or products of such investigations, carried out in furtherance of the Hilbert Programme. For example, "*Hilbertian* Proof Theory could not really prosper since Gödel's incompleteness theorem".
(iii) the branch of epistemology that deals with justifications of demonstrative truths.

Synthese (2006) 148: 623–638
DOI 10.1007/s11229-004-6298-z

Dummett's claim certainly seems just for the *first reading* (i). Gentzen's perspicuous systems that proceed according to his attractive method of *natürliches Schließen* work very well indeed under this heading. For example, standard propositional logic is complete, that is, every tautology is derivable. The proof of this completeness theorem that was offered by Paul Bernays (in his *Habilitationsschrift* from 1918, but only published in 1926) makes use of a method of transformation into **C**onjunctive **N**ormal **F**orm. A wff in CNF is a tautology if and only if each conjunct is a tautology. Such a conjunct, being a disjunction of (possibly negated) sentential letters will essentially have to be of the form

$$(A \vee \neg A),$$

possibly interspersed with side-formulae. But this formula is certainly derivable in classical logic, using indirect proof, and so, using repeated ($\vee$I) and (&I), is the CNF in question. Hence also the original wff is derivable. With the aid of Gentzen's Natural Deduction techniques, this proof is within the reach of philosophy undergraduates.[3]

Similarly, a Henkin-style proof of the *Completeness Theorem* for the predicate calculus is rendered very easy when one uses Natural Deduction rules for $\supset$, $\bot$ and the *universal* quantifier $\forall$ (instead of the customary existential one $\exists$).

The crucial *Saturation Lemma* that is central to Henkin's method then takes the form:

If Σ is a consistent set of sentences and the constant c does not occur in Σ, then also $\Sigma' =_{\text{def}} \Sigma \cup \{A[c/x] \supset \forall x A\}$ is consistent.

The proof runs very smoothly indeed:

Assume that Σ' is inconsistent. That is,

$$\Sigma, A[c/x] \supset \forall x A \vdash \bot$$

whence, by ($\supset$I),

$$\Sigma \vdash \neg(A[c/x] \supset \forall x A).$$

Thus, by propositional logic,

$$\Sigma \vdash A[c/x] \& \neg \forall x A$$

By (&E)

$$\Sigma \vdash \neg \forall x A$$
$$\Sigma \vdash A[c/x].$$

But the constant *c is new with respect to* Σ. Accordingly it behaves as an *eigen*-parameter and so an application of ($\forall$I) is admissible. Thus:

$$\Sigma \vdash \forall x A,$$

whence, by ($\neg$E), also Σ is inconsistent.
Thus, if Σ' is inconsistent, then so is Σ.
Therefore, if Σ is consistent, then so is Σ'. QED[4]

With the aid of this expository device, in my experience, also the completeness of predicate logic is within the reach even of *philosophy* undergraduates. Many more examples could be given, but surely the above two suffice to substantiate Dummett's claim under its first reading.

Also for the *second reading*, namely proof theory as a covering concept for contributions to the attempted execution of the Hilbert Programme, we can be brief. The claim makes equally good sense also here; Gentzen's Sequent Calculus, especially in a version that incorporates the infinitary ω-rule (Schütte), or its modern streamlined variant, using infinitary propositional logic after the fashion of Tait, remains the unequalled tool for such contributions.[5]

The *third reading*, though, poses a complex challenge. It must be remembered that Gentzen wrote *after* the "metalogical turn". His Natural Deduction systems, as well as their Sequent Calculus cousins, were designed primarily for contributions to the Hilbert programme. These formal systems are *metatheoretical* in character, whence, strictly speaking according to the letter of metamathematical legislation, in spite of their agreeable properties, they are objects of study, and not tools for use: Gentzen's logic is *docens* only, rather than *utens*. In particular, his formal languages are uninterpreted. His well-formed formulae, and other metamathematical "expressions", are (meta)mathematical objects. Owing to their lack of content, they do not serve the purpose of communication; they are objects *about which* one communicates. Gentzen's metamathematical "expressions", in fact, do not express, or have, content; on the contrary, they are expressed using *real* expressions.

This is in sharp contrast to how matters were prior to the metalogical turn around 1930. Frege, in particular, used a formal language for which he attempted to provide careful meaning-explanations so that the language would be adequate for the practice of mathematical analysis. His valiant efforts failed, owing to the emergence of Russell's paradox, whence the task still remains incumbent upon us to carry through a foundationalist project on a Fregean scale.

One prominent difference between Frege's way of doing things and our present (metalogical) mode of proceeding concerns the turnstile

"⊢". Today it is used as a theorem predicate among the well-formed formulae. The sign combination

$$\vdash \varphi$$

means that the well-formed formula φ is a theorem, that is, that there exists a certain (inductively defined) tree of well-formed formulae with φ as end formula. For Frege, though, the turnstile would be prefixed to a *meaningful* sentence (in use) only, where it has the task of making explicit the assertoric force of an assertion.

The first two readings of *Proof Theory* are squarely anchored to the metalogical perspective, and with respect to them it is clear that Dummett is right. When we consider the third, epistemological, reading of the term, Gentzen's own metamathematical manner of proceeding won't do, though, if we want to give Frege a fair hearing. That is, we cannot compare a Frege–Hilbert style *metamathematical* system with its corresponding *metamathematical* Gentzen system, because without content in the object language there is no knowledge there to be had in either calculus, be it a Frege–Hilbert style system or Gentzen's. Absent a contentual object-language, systems of both kinds are mere objects of study. Accordingly, for an evaluation of Dummett's claim under its third, epistemological reading, we need to consider an *interpreted* formal language. Such languages have to be supplied with meaning explanations. It is only against the background of such explanations that the comparison of the two approaches can worked out; if we stay at the level of metamathematical string-production machines the comparison won't be a fair one. Real inferences are what matters – not the formal productions effected by a syntactic theorem-engine.

Thus, the proper interpretation of the precise details of Gentzen's formalisms becomes a matter of paramount importance. Specifically, the following semiotic issues demand consideration:

(1) What is the proper syntax for the object language?
(2) What is the proper style for setting out Natural Deduction derivations? Specifically, are there significant differences between standard Natural Deduction and its sequential version?
(3) What is the proper semantic interpretation for the object-language?
(4) What notions are needed in order to do justice to the pragmatics of epistemic inference? Specifically, how does one cope with the pragmatics of "assumptions"?

The present paper seeks to answer this battery of questions.

With respect to syntax it is a remarkable fact – though one not always noted – that Gentzen uses not one, but *two* formulations of Natural Deduction. Our initial task, accordingly, is to settle which, if any, of these deserves preference above the other.

The syntax of the underlying first-order language of, say, arithmetic can be taken in the usual fashion, and so can its semantics. I prefer a constructive semantics, *in casu* that of Per Martin-Löf's constructive type theory, but most of what I say in this essay is neutral with respect to the constructivist issue and ought, *mutatis mutandis*, to be applicable without much trouble also within a Fregean framework of bivalent truth-value semantics. This, though, is not enough to settle the object-language syntax. The choice of what must be included in the object-language is dependent on the *semantics* of sequents and the *pragmatics* of assumptions. In particular, the object-theoretical status of the sequent arrow $\Rightarrow$ must be settled, as must the need for the inclusion of force-indicators for assertion and assumption.

The earliest version of Natural Deduction that Gentzen considered was the published (1934–1935) formulation of his Göttingen dissertation that uses assumptions and derivations D of the form:

$$\begin{array}{ll} & A_1, A_2, \ldots, A_k \\ & \vdots \\ D: & \\ & C \end{array}$$

Here the A_i's are undischarged assumption formulae and C is the conclusion "proved", or derived, by, or in, the derivation D. In the later (1936) version the derivable objects are *sequents*

$$A_1, A_2, \ldots, A_k \Rightarrow C. \qquad \text{(S)}$$

Standard-formulation assumptions $A_1, A_2, \ldots, A_k$ are turned into antecedent formulae of sequents. For both directions, mechanical procedures, transforming derivations in one style to the other, can be readily given, with quite low a recursion-theoretic complexity after Gödelization, say elementary in the sense of Kalmár. In order to carry out a fair comparison of Frege with Gentzen we have to indicate how, if at all, derivations set out according to the two styles can be interpreted as epistemic proofs.

On the basis of such considerations the Sequent Calculus formulation of Natural Deduction is seen as nothing but a stylistic variant of the standard format, for instance by Prawitz (1965, p. 102), (1971,

§1.6, p. 243, 'A trivial reformulation'), Dummett (1977, pp. 121–122: 'Another variation'), (1991, p. 248: 'We may therefore adopt the simpler notation'), as well as by myself (1983). This, as we shall see in the sequel, is not the full story.

The well-formed formulae of Gentzen's formalism do double duty. First, complex formulae are built up from other formulae; that is, propositional connectives join well-formed formulae. Accordingly, the wffs serve as formalistic simulacra for propositions. On the other hand, the end-formula of a (closed) derivation-tree without open assumptions is a *theorem* and thus, for Gentzen, wffs also serve as formalistic analogues to *assertions* (judgements made). The normal natural-language rendering of propositions is via that-clauses formed out of declaratives. Thus, in their first role, the wffs, when interpreted, should be read as that-clauses. Then, however, they cannot serve in their second role, since a single that-clause, on its own, cannot carry assertoric force. For their second role we need to use *statements* of the form

A is true

where A is a proposition. Thus for instance, an assertion to the effect that snow is white, is not effected by a single utterance of the propositional phrase (that-clause)

that snow is white.

For this we need to append the words

is true

in order to get the declarative statement

that snow is white is true,

an utterance of which will serve equally well as an utterance of the declarative

snow is white

in order to effect an assertion that snow is white.

Making use of this idea in order to disambiguate the double roles of wffs when interpreting the calculi transform the derivation D into

$$\begin{array}{l} A_1 \text{ is true}, A_2 \text{ is true}, \ldots, A_k \text{ is true} \\ \vdots \\ D': \\ \vdots \\ C \text{ is true} \end{array}$$

This step, though, is not enough to provide an interpretation for Gentzen's derivation-trees. We also need to account for phenomena pertaining to the pragmatics of assertion and assumption. In the degenerate case $k = 0$, the conclusion statement

C is true

is asserted outright, whence we need a force-indicator (for assertion) along the lines of the Fregean turnstile. However, this is still not enough, since the assumption statements

A_i is true $\quad i = 1, \ldots, k$

are not asserted but "assumed". Accordingly, we might attempt to use a reverse turnstile

$\dashv$

for "assumptory" force. The derivation tree D' is then transformed into a derivation tree D''

$$D'': \quad \frac{\dashv A_1 \text{ is true}, \dashv A_2 \text{ is true}, \ldots, \dashv A_k \text{ is true}}{\vdash C \text{ is true}}$$

This, *prima facie*, looks as if it might serve our purposes, but consideration of a more elaborate derivation shows that this is not so:

$$\dfrac{\dfrac{\dfrac{[A]}{B}}{A \supset B} \quad A}{B}$$

(Proof-theoretical *cognoscenti* will recognise this derivation as the one for which Dag Prawitz ([Prawitz (1965)], p. 37) defined his $\supset$-reduction.)

Dressing the derivation tree with truth- and force-indicators according to the above pattern yields

$$\dfrac{\dfrac{\dfrac{\dashv A \text{ is true } 1}{B \text{ is true } 2}}{\vdash A \supset B \text{ is true } 3} \quad \vdash A \text{ is true } 4}{\vdash B \text{ is true } 5}$$

Here the statements marked as 3, 4, and 5 are asserted, whereas 1 is only assumed. The statement 2 that serves as premise of the ($\supset$I) that has the statement 3 as conclusion, *is neither asserted nor assumed*. Thus, our force-indicators for assertion and assumption do not suffice to account for the phenomena of dependence that occur within standard Natural Deduction derivations. We must take note here that the statement

B is true

is not *asserted* outright, but only *conditionally*, under the assumption that A is true. What is the logical form of such conditionalization? One answer that suggests itself immediately is that of outright assertion of the categorical statement that the implication (al proposition) $A \supset B$ is true. This, however, will not do, because we would justify the categorical statement

$A \supset B$ is true

with an assertion of the conditional statement

B is true, on condition that A is true.

A conditional weakening of assertoric *force*, on the other hand, seems hardly possibly. The attempts in the literature—Belnap and others[6] – to treat of "conditional assertion", in my opinion, do not succeed in weakening the kind of assertoric force that is involved at the level of pragmatics, but alters the semantics of what is asserted (still categorically). I will opt for a treatment that keeps assertion categorical, whence there is no change in force, but *conditionalizes* the kind of *truth* that is ascribed to the propositional content in question.

Thus, the derivation tree D' above, where the assumptions

A_1 is true, A_2 is true, $\ldots$, A_k is true

are still open, does not allow for the ascription of outright truth to the proposition C, but only of truth on condition that A_1 is true,A_2 is true, $\ldots$, A_k is true.

The weakened, *conditional* truth in question will be

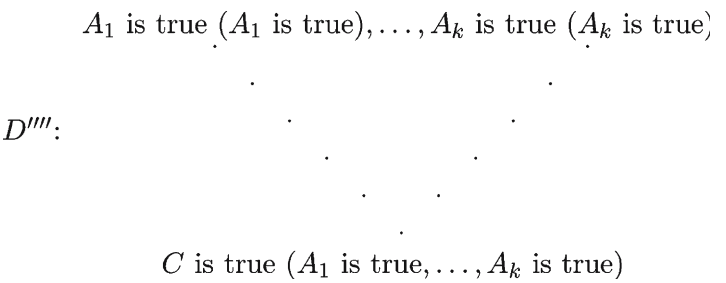
is true (A_1 is true, A_2 is true, $\ldots$, A_k is true).

This, however, means that nodes in derivation trees are not covered with statements of the form

A is true,

but with statements of the conditional form. Effecting this transformation, the derivation tree D ultimately takes the form D''''

D'''':



If we wish to include explicit force-indicators, the assertoric turnstile is enough, since at every step a categorical assertion of a conditional statement occurs. We must take care to distinguish between, on the one hand,

$A \supset B$ is true

and, on the other hand,

B is true (A is true).

The latter statement can be variously read as

B is true	under the assumption on condition given	that A is true

or even as

If A is true then B is true.

Our analysis thus reveals that when fully interpreted, taking into account also semantic and pragmatic features, Gentzen's standard Natural Deduction turns out to be nothing but the Sequent Calculus version in disguise: as a matter of semantical fact, at each node the assumptions are carried along. The transformation of

$$C \text{ is true } (A_1 \text{ is true}, A_2 \text{ is true}, \ldots, A_k \text{ is true}) \qquad (*)$$

into

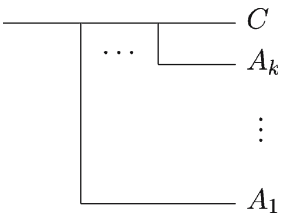
$$A_1 \text{ is true}, A_2 \text{ is true}, \ldots, A_k \text{ is true} \Rightarrow C \text{ is true} \qquad (**)$$

makes this explicit.

Thus, when the deductively relevant features are made explicit in order to account also for the pragmatic interaction of assumption with assertion, standard Natural Deduction is but a variant of the Sequent Calculus version, rather than the other way round. But the Sequent Calculus version is, in essence, a Frege–Hilbert system. Its derivable

objects, though, are not theorems that ascribe truth to (interpreted) well-formed formulae; they are *sequents* of interpreted well-formed formulae that, under the present reading, ascribe conditional, *dependent*, truth to a content, whence the difference with Frege is minute.

In fact, the link to Frege can be tied much closer than this. As the late Pavel Tichy (1988, pp. 248–252) observed, iterated Frege conditionals need not be seen as repeated implications among propositional contents.[7] Instead, they can equally well be taken as Gentzen(–Hertz) sequents. Customarily the repeated Frege conditional



is taken with the (same) meaning (as the well-formed formula)

$$(A_1 \supset (A_2 \supset (\ldots \supset (A_k \supset C)\ldots))).$$

However, rotating the Frege conditional 90° clockwise, while altering the notation only slightly, produces a familiar result, namely,

$$A_1, A_2, \ldots, A_k \Rightarrow C.$$

This correspondence between the calculi of Frege and Gentzen operates even down to the considerable fine-structure of rules, sometimes showing a surprising(?) resemblance of terminology.

The need to account for the pragmatics of assumptions forces us to give primacy to the Sequent Calculus version of natural deduction. What, then, do unadorned Sequent Calculus derivations, that begin with axioms of the form

$$B \text{ true} \Rightarrow B \text{ true},$$

and continue with applications of sequential introduction and elimination rules that operate to the right of the arrow, yielding sequents of the form $(**)$ above, express? What is their right interpretation? Kreisel (1971, 1971a, 1973), echoing Brouwer, holds that they stand for ('refer to') *abstract mental processes*. It is certainly right that proof, or better *demonstration*, ultimately pertains to mental acts (of getting to know). (According to the OED, a demonstration is that through which something is shown or made known.) Wittgenstein's reasons to reject private reference, however, apply here as well.

Accordingly, it does not seem correct to let derivation trees *refer* to individual mental acts. *Abstract species* of such acts (after the fashion of Husserl) constitute a better alternative. If that is what Kreisel means by the mental processes being *abstract*, it might be possible to go along with his proposal. Nevertheless, I prefer to let the Sequent Calculus derivation-trees express *blue-prints*, or manuals, for mental acts of knowledge, in the same way that a score of music is a blue-print for a performance or a chess score-sheet is a blueprint for a game that can be played. If one is in possession of such a manual, by following it through one can oneself carry out an act of knowledge that produces the theorem in question.

Our discussion has led to an account of the meaning of Gentzen's standard Natural Deduction formalism. The derivation trees are variants of derivations in the Sequent Calculus version and sequents express that a content is true, dependent on the truth of certain contents. This is one of the several original readings of sequents—why Gentzen, and before him Paul Hertz, dropped the *con-* part of the medieval *consequentia* I do not know—that Gentzen offered (1932, p. 330):

Ein Satz* hat die Form

$$u_1 u_2 \ldots u_\nu \rightarrow \nu \qquad (\nu \geq 1) \ldots$$

[M]an ... liest den "Satz" so: Wenn die Aussagen $u_1, \ldots, u_\nu$ richtig sind, so ist auch die Aussage ν richtig.

Richtig is how propositional truth was rendered by the Hilbert school, for instance in Hilbert–Ackermann's *Grundzüge*. One should note that the sequents express material rather than formal consequence, that is, validity under all substitution instances à la Bolzano is not demanded. This is repeated also in Gentzen ([Gentzen (1934–5)], pp. 89–90), where the sequent (S) is explained as

$$A_1 \ \& \ A_2 \ \& \ \ldots \ \& \ A_k \supset C.^8 \qquad (\#)$$

This, though, is not synonymous with the original explanation, even though the two explanations provide the sequents *with the same assertion conditions*. In Gentzen (1936, p. 512) this is put right. There (S) should be read:

Unter den Annahmen $A_1, A_2, \ldots, A_k$ gilt.B^9

This reading Gentzen retained also in (1938, p. 21).

My preferred contentual reading of the derivation trees gives pride of place to Sequential Natural Deduction and regards standard

Natural Deduction as a mere variant of this. Under this reading, derivation trees are blueprints for mental acts of knowledge; no *objectual* semantic values are assigned.

I now wish to consider another reading that *does* assign semantic values to standard Natural Deduction derivation trees. The previous discussion has been neutral with respect to the semantics used and the ensuing notion of proposition, presupposing only that the language is an interpreted one. Now, however, I shall avail myself of Heyting–Kolmogoroff's constructive notion of proposition, say, according to its formal rendering in Martin-Löf's constructive type theory.[10] A proposition is thus explained in terms of how its canonical proof-objects may be formed and when two such objects are equal canonical proofs. Note that what is at issue here are proofs of *propositions*. This is a notion that is novel with intuitionism. Previously, all proving throughout the history of logic has taken place at the level of judgement and not at that of their contents. Accordingly, within the constructivist framework, it is strictly necessary to keep apart *demonstrations*, that is, proof(-acts) of judgements (that propositions are true) and *proof*(-object)s of propositions. *Demonstration* is an epistemic notion, but the latter novel notion is *not*. Its closest analogue within (the framework of) classical semantics is that of a "truth maker" in a correspondence theory of truth for propositions. Thus it serves as a constructivist analogue of Tractarian *Sachverhalte*, the prototypical classical truth-makers, or perhaps better of (Husserlian) *moments*.

Heyting's meaning explanations draw upon canonical proofs that have to be cast in certain forms. In general, a *proof* of a proposition is a *method* for obtaining such a canonical proof by means of execution (evaluation). Thus, for instance, when A and B are propositions (whence we know how their canonical proofs may be put together), a canonical proof for the implication, that is, the proposition $A \supset B$, is of the introductory form:

$$\supset \mathrm{I}(A, B, (x)b),$$

where b is a proof-object for B, given that x is a proof-object for A, and $(x)b$ is a function, defined by (lambda-)abstraction, such that when a is a proof of A, then

$$(x)b(a) = b[a/x].$$

The second reading, then, that I want to offer for standard natural deduction derivation is that of their being (possibly dependent)

proof-objects for propositions. Thus, for instance, under this interpretation the derivation D above is a dependent proof object:

D is a proof of C,
given that x_1 is a proof of A_1, x_2 is a proof of $A_2, \ldots x_k$ is a proof of A_k.

Hence, when a_1 is a proof of A_1, a_2 is a proof of $A_2, \ldots a_k$ is a proof of A_k,

$$D[a_1/x_1, \ldots, a_k/x_k] \text{ is a proof of } C.$$

In terms of standard Natural Deduction, this substitution corresponds to putting the closed derivations D_1 of $A_1, \ldots, D_k$ of A_k on top of the open assumption formulae $A_1, A_2, \ldots, A_k$ in the derivation tree D, whence a closed derivation for the conclusion C results.

Above Hertz–Gentzen sequents were interpreted as statements ascribing conditional truth to propositions. The constructive semantical framework, nevertheless, allows also for another way of interpreting a sequent S. We then treat

sequent S holds

as a (novel) form of judgement that generalizes the common form of judgement

proposition A is true.

The semantical explanation of the statement that a sequent holds is–naturally enough–a generalization of the corresponding explanation for propositional truth:

$$A_1, A_2, \ldots, A_k \Rightarrow C \text{ holds}$$
$$= \text{ there exists a function from Proof}$$
$$(A_1), \ldots, \text{Proof}(A_k) \text{ to Proof } (C).$$

Identifying propositions with their proof-types we get a more compact expression:

the function type $(A_1, A_2, \ldots, A_k)C$ exists.

In order to distinguish these two readings which I call *open* and *closed*, respectively of the sequent S, also at the level of notation, the latter, closed reading that expresses the holding of a consequence relation between propositions, will be written:

$$(A_1, A_2, \ldots, A_k) \Rightarrow C \text{ holds.} \qquad (\text{S}')$$

It must be stressed that this holding of sequents is *material* only. The closed sequent S′ holds when the matching implication (#) is *true*. Logical truth of (#), on the other hand, is *not* required for the mere holding of (S′). Corresponding to the logical truth of (#), that is when (S′) holds under all variations, à la Bolzano, is the notion of it holding *logically* (come what may, independently of what is the case, under all variations, etc.).

Finally let me note that, while Gentzen and Hertz dealt only with *open* sequents, the theory of *closed* sequents has been explored primarily by Peter Schroeder-Heister in (1984), as well as in his two dissertations (1981) and (1987).

NOTES

[1] This paper develops in greater detail a line of thought that was adumbrated in my (1997), (1998a), (1998b), and (2000).

[2] Dummett (1973, p. 309, and p. 435, respectively).

[3] One only needs to derive the following laws that are needed for transforming a wff into a deductively equivalent CNF:

$$A \supset B \dashv\vdash \neg A \vee B \tag{1}$$
$$A \vee B \dashv\vdash B \vee A \tag{2}$$
$$A \,\&\, B \dashv\vdash B \,\&\, A \tag{3}$$
$$A \vee (B \vee C) \dashv\vdash (A \vee B) \vee C \tag{4}$$
$$A \,\&\, (B \,\&\, C) \dashv\vdash (A \,\&\, B) \,\&\, C \tag{5}$$
$$A \vee (B \,\&\, C) \dashv\vdash (A \vee B) \,\&\, (A \vee C) \tag{6}$$
$$A \,\&(B \vee C) \dashv\vdash (A \,\&\, B) \vee (A \,\&\, C) \tag{7}$$
$$\neg(A \vee B) \dashv\vdash \neg A \,\&\, \neg B \tag{8}$$
$$\neg(A \,\&\, B) \dashv\vdash \neg A \vee \neg B \tag{9}$$
$$\neg\neg A \dashv\vdash A \tag{10}$$

[4] Note that this is a constructive proof: inconsistency is a *positive* notion (Σ is inconsistent $= \Sigma$ does derive $\bot$), whereas consistency is a negative one (Σ is consistent $= \Sigma$ does *not* derive $\bot$). Thus the final contraposition goes in the constructively valid direction.

[5] Schütte (1950a), Tait (1968).

[6] Belnap(1973).

[7] Tichy's observation seems to have gone largely unnoticed. A decade later von Kutschera (1996) and Schroeder-Heister (1999) both discuss the matter in apparent unawareness of Tichy's earlier, very explicit treatment. Tichy's congenial Chapter 13 - **Inference** - definitely deserves to become more known, as does his paper 'On Inference' (1999).

[8] Other readings are possible, for instance,

$$\neg A_1 \vee \neg A_2 \vee \ldots \vee \neg A_k \vee C$$

or

$$(A_1 \supset (A_2 \supset (\ldots \supset (A_k \supset C)\ldots))).$$

Both were used by Schütte (1950) in his early reformulations of Gentzen's work. The first form has been streamlined by Tait (1968), who used finite sets of formulae, disjunctively read, with the use of negations rendered superfluous (de Morgan!) except in front of atomic formulae – Schwichtenberg (1977) gives a beautiful treatment of cut-elimination for the predicate calculus based on this approach. The second form is used for intuitionistic systems, where the de Morgan laws are not available, but has not proved more convenient to use than Gentzen's original formulation in terms of sequents.

[9] Where I have changed Gentzen's Fette Fraktur.

[10] Martin Löf (1984).

REFERENCES

Belnap, N.: 1973, 'Restricted Quantification and Conditional Assertion', in H. Leblanc (ed.), *Truth, Syntax and Modality*, North-Holland, Amsterdam, pp. 48–75.

Dummett, M.: 1973 (2nd edn. 1981), *Frege. Philosophy of Language*, Duckworth, London.

Dummett, M.: 1977, *Elements of Intuitionism*, Oxford University Press, Oxford.

Dummett, M.: 1991, *The Logical Basis of Metaphysics*, Duckworth, London.

Gentzen, G.: 1932, 'Über die Existenz Unabhängiger Axiomensysteme zu Unendlichen Satzsystemen', *Mathematische Annalen* **107**, 329–350.

Gentzen, G.: 1934–5, 'Untersuchungen über das logische Schließen', *Mathematische Zeitschrift* **39**, 176–210, 405–431.

Gentzen, G.: 1936, 'Die Widerspruchsfreiheit der reinen Zahlentheorie', *Mathematische Annalen* **112**, 493–565.

Gentzen, G.: 1938, 'Neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie', in *Forschungen zur Logik und zur Grundlegung der Exakten Wissenschaften* (N.F.) **4**, Felix Meiner, Leipzig, pp. 19–44.

Hertz, P.: 1929, 'Über Axiomensysteme für beliebige Satzsysteme', *Mathematische Annalen* **101**, 493–565.

Hilbert, D. and W. Ackermann: 1928, *Grundzüge der Theoretischen Logik*, Springer, Berlin.

Kreisel, G.: 1971, Review of *The Collected Papers of Gerhard Gentzen*, *Journal of Philosophy* **68**, 328–365.

Kreisel, G.: 1971a, 'A Survey of Proof Theory. II', in J. E. Fenstad (ed.), *Proceedings of the Second Scandinavian Logic Symposium*, Amsterdam, pp. 109–170.

Kreisel, G.: 1973, 'Perspectives in the Philosophy Pure Mathematics', in *Logic Methodology and Philosophy of Science* **IV**, North-Holland, Amsterdam, pp. 255–277.

von Kutschera, F.: 1996, 'Frege and Natural Deduction', in Matthias Schirn (ed.), *Frege: Importance and Legacy*, De Gruyter, Berlin, pp. 310–304.
Martin-Löf, P.: 1984, *Intuitionistic Type Theory*, Bibliopolis, Naples.
Prawitz, D.: 1965, *Natural Deduction*, Almquist & Wicksell, Uppsala.
Prawitz, D.: 1971, 'Ideas and Results in Proof Theory', in J. E. Fenstad (ed.), *Proceedings of the Second Scandinavian Logic Symposium*, North Holland, Amsterdam, pp. 235–307.
Schroeder-Heister, P.: 1981, *Untersuchungen zur regellogischen Deutung von Aussagenverknüpfungen* (diss.), Bonn, 1981.
Schroeder-Heister, P.: 1984, 'A Natural Extension of Natural Deduction', *Journal of Symbolic Logic* **49**, 1284–1300.
Schroeder-Heister, P.: 1987, *Structural Frameworks with Higher-Level Rules, Habilitationsschrift*, University of Konstanz, Department of Philosophy.
Schroeder-Heister, P.: 1999, 'Gentzen-style Features in Frege', in *Abstracts of the 11th International Congress of Logic, Methodology, and Philosophy of Science (Cracow, August 1999)*, Cracow, 1999, p. 499.
Schütte, K.: 1950, 'Schlußweisen-Kalküle der Prädikatenlogik', *Mathematische Annalen* **122**, 47–65.
Schütte, K.: 1950a, 'Beweistheoretische Erfassung der Unendlichen Induktion in der Zahlentheorie', *Mathematische Annalen* **122**, 369–389.
Schwichtenberg, H.: 1977, 'Proof Theory: Some Applications of Cut-Elimination', in J. Barwise (ed.), *Handbook of Mathematical Logic*, North-Holland, Amsterdam, pp. 867–895.
Sundholm, B. G.: 1983, 'Systems of Deduction', Chapter I:2 in D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, Vol. I. Reidel, Dordrecht, pp. 133–188.
Sundholm, B. G.: 1997, 'Implicit Epistemic Aspects of Constructive Logic', *Journal of Logic, Language and Information* **6**, 191–212.
Sundholm, B. G.: 1998a, 'Inference, Consequence, Implication: A Constructivist's Approach', *Philosophia Mathematica* **6**, 178–194.
Sundholm, B. G.: 1998b, 'Inference versus Consequence', in *The Logica Yearbook 1997*, Filosofia, Prague, pp. 26–35.
Sundholm, B. G.: 2000, 'Proofs as Acts versus Proofs as Objects: Some Questions for Dag Prawitz', *Theoria* **64** (for 1998, published in 2000): 2–3 (special issue devoted to the works of Dag Prawitz, with his replies), 187–216.
Tait, W. W.: 'Normal Derivability in Classical Logic', in J. Barwise (ed.), *The Syntax and Semantics of Infinitary Languages*, *Lecture Notes in Mathematics*, Springer, Berlin, pp. 204–236.
Tichy, P.: 1988, *The Foundations of Frege's Logic*, De Gruyter, Berlin.
Tichy, P. and J. Tichy: 1999, 'On Inference', *The Logica Yearbook 1998*, Filosofia, Prague, pp. 73–85.

Faculteit der Wijsbegeerte
Universiteit Leiden
Postbus 9515
2300 RA Leiden
E-mail: b.g.sundholm@let.leidenuniv.nl

KOSTA DOŠEN

MODELS OF DEDUCTION*

ABSTRACT. In standard model theory, deductions are not the things one models. But in general proof theory, in particular in categorial proof theory, one finds models of deductions, and the purpose here is to motivate a simple example of such models. This will be a model of deductions performed within an abstract context, where we do not have any particular logical constant, but something underlying all logical constants. In this context, deductions are represented by arrows in categories involved in a general *adjoint situation*. To motivate the notion of adjointness, one of the central notions of category theory, and of mathematics in general, it is first considered how some features of it occur in set-theoretical axioms and in the axioms of the lambda calculus. Next, it is explained how this notion arises in the context of deduction, where it characterizes logical constants. It is shown also how the categorial point of view suggests an analysis of propositional identity. The problem of propositional identity, i.e., the problem of identity of meaning for propositions, is no doubt a philosophical problem, but the spirit of the analysis proposed here will be rather mathematical. Finally, it is considered whether models of deductions can pretend to be a semantics. This question, which as so many questions having to do with meaning brings us to that wall that blocked linguists and philosophers during the whole of the twentieth century, is merely posed. At the very end, there is the example of a geometrical model of adjunction. Without pretending that it is a semantics, it is hoped that this model may prove illuminating and useful.

1. INTRODUCTION

According to the traditional vocation of logic to study deductive reasoning, deductions should indeed be of central concern to logicians. However, as an object of study, deductions have really a central place in a rather restricted area of logic called *general proof theory* – namely, proof theory done in the tradition of Gentzen. There, by studying normalization of logical deductions, one is led to consider criteria of identity of deductions. The goal of this brand of proof theory might be to find a mathematical answer to the philosophical question ''What is deduction?'', as recursion theory has found, with much success, a mathematical answer to the question ''What is computation?''.

Synthese (2006) 148: 639–657
DOI 10.1007/s11229-004-6290-7

In proof theory done in the tradition of Hilbert's program, where one is concerned with consistency proofs for fragments of mathematics, deducing is less central. The goal there is not to answer the question "What is deduction?", but to prove consistency by some particular means. Hilbertian proof theory, incorporating the fundamental lessons of Gödel, was for a long time dominant in proof theory, but nowadays it seems it may be yielding ground. It has become a rather secluded branch of mathematics, where one studies intricate problems about ordinals, not particularly appealing to other logicians, let alone other mathematicians.

However, following a general trend, seclusion has become the norm in logic, as well as elsewhere in mathematics. The trend is quite conspicuous in model theory, which was no doubt the dominant branch of logic during a long period in the second half of the twentieth century. As the century was drawing to its end, so model theory, which had drifted to some particular branches of algebra, appeared more and more esoteric. The remaining two great branches of logic, recursion theory and set theory, leave the same impression nowadays. Logicians of these various branches meet at congresses, and politely listen to each other's talks, but don't seem much moved by them.

Although we are speaking here of the dominant branches of *logic*, deductive reasoning hardly makes their subject matter. The study of deduction was for a long time confined to rather marginal fields of nonclassical logics. Perhaps the growth of general proof theory, and its connection with category theory and computer science, might bring deduction to the fore. (In that, the role of category theory and computer science would presumably not be the same, the former being otherworldly and the latter mundane, but – who knows – the two ways might end up by being in harmony.)

In standard model theory, deductions are not the things one models. But in general proof theory, in particular in categorial proof theory, one finds models of deductions, and my purpose in this talk is to motivate a simple example of such models. This will be a model of deductions performed within an abstract context, where we don't have any particular logical constant, but something underlying all logical constants. In this context, deductions are represented by arrows in categories involved in a general *adjoint situation*.

To motivate the notion of adjointness, one of the central notions of category theory, and of mathematics in general, we shall first consider how some features of it occur in set-theoretical axioms and in the axioms of the lambda calculus. Next, it will be explained how

this notion arises in the context of deduction, where it characterizes logical constants. We shall also see how the categorial point of view suggests an analysis of propositional identity. The problem of propositional identity, i.e., the problem of identity of meaning for propositions, is no doubt a philosophical problem, but the spirit of the analysis proposed here will be rather mathematical. Finally, we shall consider whether models of deductions can pretend to be a semantics. I merely ask this question, which, as so many questions having to do with meaning, brings us to that wall that blocked linguists and philosophers during the whole of the twentieth century. At the very end, we reach our example of a geometrical model of adjunction, for which I don't pretend that it is a semantics. Nevertheless, I hope that this model may prove illuminating and useful.

2. TERMS, PROPOSITIONS AND INVERSION

In logic, as well as in the philosophy of language, we are especially interested in two kinds of linguistic activity: referring and asserting. We engage in the first kind of activity with the help of *terms* (which abbreviates *individual terms*), while for the second we use *propositions* (or *formulae*). The two grammatical categories of terms and propositions are basic grammatical categories, with whose help other grammatical categories can be defined as functional categories: predicates map terms into propositions, functional expressions map terms into terms, and connectives and quantifiers map propositions into propositions.

The set-abstracting expression $\{x : \ldots\}$ maps a proposition A into the term $\{x : A\}$. This term is significant in particular when x is free in A, but it makes sense for any A, too. The expression $x \in \ldots$ is a unary predicate: it maps a term a into the proposition $x \in a$. The ideal set theory would just assume that $\{x : \ldots\}$ and $x \in \ldots$ are in some sense inverse to each other. Namely, we would have the following postulates:

$$\textit{Comprehension:} \quad x \in \{x : A\} \leftrightarrow A,$$
$$\textit{Extensionality:} \quad \{x : x \in a\} = a,$$

provided x is not free in a. In the presence of replacement of equivalents and of Comprehension, Extensionality is equivalent to the more usual extensionality principle

$$\forall x(x \in a_1 \leftrightarrow x \in a_2) \rightarrow a_1 = a_2,$$

provided x is not free in a_1 and a_2 (see Došen 2001, Section 2). We know that ideal set theory is inconsistent if in propositions we find negation, or at least implication. To get consistency, either $\{x : A\}$ will not always be defined, and we replace Comprehension by a number of restricted postulates, or we introduce types for terms.

Instead of $\{x : \ldots\}$ let us now write $(\lambda_x \ldots)$, and instead of $x \in \ldots$ let us write $(\ldots x)$. Then Comprehension and Extensionality become respectively

$$((\lambda_x A)x) \leftrightarrow A,$$
$$(\lambda_x(ax)) = a.$$

If we take that $(\lambda_x \ldots)$ maps a term a into the term $(\lambda_x a)$, while $(\ldots x)$ maps a term a into the term (ax), and if, furthermore, we replace equivalence by equality, and omit outermost parentheses, our two postulates become the following postulates of the lambda calculus:

$$\beta\textit{-equality:} \quad (\lambda_x a)x = a,$$
$$\eta\textit{-equality:} \quad \lambda_x(ax) = a,$$

provided x is not free in a in η-equality. The present form of β-equality yields the usual form in the presence of substitution for free variables. (The usual form of β-equality and η-equality imply α-equality.) The fact that the lambda calculus based on β-equality and η-equality is consistent is due to the fact that the language has been restricted, either by preventing anything like negation or implication to occur in terms, or by introducing types. Without restrictions, in type-free *illative* theories, we regain inconsistency.

So the general pattern of Comprehension and Extensionality, on the one hand, and of β and η-equality, on the other, is remarkably analogous. These postulates assert that a variable-binding expression Γ_x and application to a variable Φ_x are inverse to each other, in the sense that $\Phi_x\Gamma_x\alpha$ and $\Gamma_x\Phi_x\alpha$ are either equivalent or equal to α, depending on the grammatical category of α. It is even more remarkable that theories so rich and important as set theory and the lambda calculus are based on such a simple inversion principle.

3. DEDUCTIONS AND INVERSION

Besides referring and asserting, there is a third kind of activity of particular interest to logic: deducing, which is also linguistic, as far as it consists in passing from propositions to propositions.

To speak about deductions we may use *labelled sequents* of the form $f : \Gamma \vdash B$, where Γ is a collection of propositions making the premises, the proposition B is the conclusion, and the term f records the rules justifying the deduction. If the premises can be collected into a single proposition, and this is indeed the case when Γ is finite and we have a connective like conjunction, then we can restrict our attention to simple sequents of the form $f : A \vdash B$, where both A and B are propositions. We can take that $f : A \vdash B$ is an arrow in a category in which A and B are objects. When we don't need it, we omit mentioning the *type* $A \vdash B$ of $f : A \vdash B$, and write just the *arrow term* f.

Special arrows in a category are axioms, and operations on arrows are rules of inference. Equalities of arrows are equalities of deductions. For that, categorial equalities between arrows have to make proof-theoretical sense, as indeed they do, by following closely reductions in a normalization or cut-elimination procedure in intuitionistic and substructural logics.

In categorial proof theory we are not concerned with a consequence *relation*, but with a consequence *graph*, where more than one arrow, i.e., deduction, can join the same pair of objects, i.e., propositions. This should be the watershed between proof theory and the rest of logic. It is indeed a defect of traditional general proof theory, unaware of categories, that it is still very much under the spell of sequents understood in terms of consequence relations – as if all deductions with the same premises and conclusions were equal. The traditional theory has trouble in representing deductions and in coding them. It draws trees and has no clear criteria of identity of deductions. (Applying the typed lambda calculus in general proof theory usually brings awareness of categories.)

We shall now inquire whether there is something in the context of deductions, as they are understood in categories, which would be analogous to the inversion principle we encountered before in set theory and the lambda calculus.

Take a category $\mathcal{K}$ with a terminal object T (this object behaves like the constant true proposition), and take the *polynomial category* $\mathcal{K}[x]$ obtained by extending $\mathcal{K}$ with an *indeterminate arrow* $x{:}\mathrm{T} \vdash A$ (see Lambek and Scott 1986, Part I, Chapters 4–5.; Došen 2001). We obtain $\mathcal{K}[x]$ by adding to the graph of arrows of $\mathcal{K}$ a new arrow $x : \mathrm{T} \vdash A$, and then by imposing on the new graph equalities required by the particular sort of category to which $\mathcal{K}$ belongs. Note that $\mathcal{K}[x]$ is not simply the free category of the required sort generated by the new graph, because the operations on arrows of $\mathcal{K}[x]$ should coincide

with those of $\mathcal{K}$ on the arrows inherited from $\mathcal{K}$ (see Došen 2001, Section 5). We can conceive of $\mathcal{K}[x]$ as the extension of a deductive system $\mathcal{K}$ with a new axiom A.

Now consider the variable-binding expression Γ_x that assigns to every arrow term $f : C \vdash B$ of $\mathcal{K}[x]$ the arrow term $\Gamma_x f : C \vdash A \to B$ of $\mathcal{K}$, where $\to$, which corresponds to implication, is an operation on the objects of $\mathcal{K}$ (in categories, $A \to B$ is more often written B^A). Passing from f to $\Gamma_x f$ corresponds to the deduction theorem. Conversely, we have application to x, denoted by Φ_x, which assigns to an arrow term $g : C \vdash A \to B$ of $\mathcal{K}$ the arrow term $\Phi_x g : C \vdash B$ of $\mathcal{K}[x]$. Now, passing from g to $\Phi_x g$ corresponds to modus ponens. If we require that

$$(\beta) \quad \Phi_x \Gamma_x f = f,$$
$$(\eta) \quad \Gamma_x \Phi_x g = g,$$

we obtain a bijection between the hom-sets $\mathcal{K}(C, A \to B)$ and $\mathcal{K}[x](C, B)$. If, moreover, we require that this bijection be natural in the arguments B and C, we obtain an *adjunction*. The left-adjoint functor in this adjunction is the *heritage* functor from $\mathcal{K}$ to $\mathcal{K}[x]$, which assigns to objects and arrows of $\mathcal{K}$ their heirs in $\mathcal{K}[x]$, while the right-adjoint functor is a functor from $\mathcal{K}[x]$ to $\mathcal{K}$ that assigns to an object B the object $A \to B$. We find such an adjunction in cartesian closed categories, whose arrows correspond to deductions of the implication-conjunction fragment of intuitionistic logic, and also in bicartesian closed categories, whose arrows correspond to deductions of the whole of intuitionistic propositional logic.

In cartesian closed and bicartesian closed categories, as well as in cartesian categories tout court, we also have the adjunction given by the bijection between the hom-sets $\mathcal{K}(A \times C, B)$ and $\mathcal{K}[x](C, B)$. Here the heritage functor is right adjoint, and a functor from $\mathcal{K}[x]$ to $\mathcal{K}$ that assigns to an object C the object $A \times C$ is left adjoint. The binary product operation on objects $\times$ corresponds to conjunction, both intuitionistic and classical, as $\to$ corresponds to intuitionistic implication.

These adjunctions, which were first considered by Lambek, and which he called *functional completeness*, are a refinement of the deduction theorem (see Lambek 1974; Lambek and Scott 1986, Part I; Došen 1996, 2001). Through the categorial equivalence of the typed lambda calculus with cartesian closed categories, which was also discovered by Lambek, they are closely related to the so-called

Curry–Howard correspondence between typed lambda terms and natural-deduction proofs. They shed much light on this correspondence. The adjunctions of functional completeness may serve to characterize conjunction and intuitionistic implication.

4. LOGICAL CONSTANTS AND ADJUNCTION

Adjointness phenomena pervade logic, as well as much of mathematics. An essential ingredient of the spirit of logic is to investigate inductively defined notions, and inductive definitions engender free structures, which are tied to adjointness. We find also in logic the important model-theoretical adjointness between syntax and semantics, behind theorems of the "if and only if" type called semantical completeness theorems. However, adjunction is present in logic most specifically through its connection with logical constants.

Lawvere put forward the remarkable thesis that all logical constants are characterized by adjoint functors (see Lawvere 1969). Lawvere's thesis about logical constants is just one part of what he claimed for adjunction, but it is a significant part.

Actually, Lawvere didn't characterize conjunction and intuitionistic implication through the adjunctions of functional completeness we mentioned in the preceding section. Instead, there is for conjunction, i.e., binary product in cartesian categories, the adjunction between the diagonal functor $D : \mathcal{K} \to \mathcal{K} \times \mathcal{K}$ as left adjoint and the internal product bifunctor $\times : \mathcal{K} \times \mathcal{K} \to \mathcal{K}$ as right adjoint. Coproduct, i.e., disjunction, is analogously left adjoint to the diagonal functor. The terminal and initial objects, which correspond respectively to the constant true proposition and the constant absurd proposition, may be conceived as empty product and empty coproduct. They are characterized by functors right and left-adjoint, respectively, to the constant functor into the trivial category with a single object and a single identity arrow. Functors tied to the universal and existential quantifiers are, respectively, right adjoint and left adjoint to the substitution functor, which we find in hyperdoctrines, or fibered categories.

In all that, one of the adjoint functors carries the logical constant to be characterized, i.e., it involves the corresponding operation on objects and depends on the inner constitution of the category, while the other adjoint functor is a *structural* functor, which does not involve the inner operations of the category ("structural" is here used as in the "structural rules" of Gentzen's proof theory). The diagonal

functor and the constant functor are clearly structural: they make sense for any sort of category. The substitution functor may also be conceived as structural, and such is the heritage functor too. (We relied above on the presence of the terminal object to characterize implication through functional completeness, but this was done only to simplify the exposition, and is not essential.)

Lawvere's way to characterize intuitionistic implication through adjunction is by relying on the bijection between $\mathcal{K}(A \times C, B)$ and $\mathcal{K}(C, A \to B)$, which can be obtained by composing the two adjunctions with the heritage functor mentioned in the preceding section. The disadvantage of this characterization is that none of the adjoint functors $A\times$ and $A \to$ is structural (though the former resembles such a functor more than the latter).

In the late seventies (see Došen 1989, which summarizes the results of my doctoral thesis written ten years before), I was engaged in characterizing logical constants of classical, intuitionistic and substructural logics through equivalences between a sequent involving the logical constant in question at a particular place and a *structural*, purely schematic, sequent, not involving any logical constant. A typical such equivalence is

$$\Gamma \vdash A \to B \quad \textit{iff} \quad A, \Gamma \vdash B.$$

I called such equivalences *analyses*, and not *definitions*, because they may lack some essential traits of definitions, like conservativeness and replaceability by the defining expression in every context.

I realized more recently that my analyses were just superficial aspects of adjunctions. They pointed to the inversion principle, but didn't mention the naturalness condition of adjunctions. This last condition may perhaps be taken as implicit, but I lacked a clear idea of identity of deductions. However, this idea is also unclear in all of traditional general proof theory untouched by categorial proof theory. Gentzen's and Prawitz's inversion principle for natural deduction, which says that the elimination rules can be recovered from the introduction rules, amounts to analytical equivalence, and is in the same way a superficial aspect of adjointness (see Gentzen's *Untersuchungen über das logische Schließen*, II, § 5.13, and Prawitz 1965, Chapter II).

However, what I did brings something which I think should be added to Lawvere's thesis: namely, the functor carrying the logical constant should be adjoint to a *structural* functor recording some

features of deduction. With this amendment the thesis might serve to separate logical constants from other expressions.

I suppose that my notion of analysis corresponds to an adjoint situation that does not amount to an adjoint equivalence of categories, while ordinary definitions are based on an equivalence of categories.

In some poignant passages of his book on Frege, Dummett has argued very convincingly that the inversion principle of natural deduction, discovered by Gentzen and studied by Prawitz, operates in ordinary language too (see Dummett 1973, pp. 396–397, 454–455). With pejorative expressions this principle is broken so that sufficient conditions for an assertion are weaker than the conclusions we may draw from the assertion. "Long-winded" may be taken as a pejorative expression because conclusions one can infer from the assertion that somebody's performance is such, like the conclusion that the matter should be ignored, need not be warranted by a sufficient condition for the assertion, which can be merely that the thing is long. Actually, the point of using pejoratives is to licence some otherwise unwarranted inferences. (Unwarranted conclusions in the case of pejoratives are condemning, whereas the point of using flattery terms is to licence commending conclusions, which may also be unwarranted.)

To complete what Dummett is saying, one could add that with euphemisms, dually to what one has with pejoratives, the sufficient conditions for an assertion are stronger than the conclusions we are expected to draw from the assertion. A sufficient condition for asserting that some text is "not concise" might be that it is unbearably long, and the conclusion that it should be ignored, which could be drawn from this sufficient condition, is meant to be blocked by using the euphemism. The point of using euphemisms is to block unwanted inferences (or, at least, the speaker pretends he means to block them).

5. IDENTITY OF DEDUCTIONS AND PROPOSITIONAL IDENTITY

Many successful philosophical analyses are achieved by a shift in grammatical categories. Such is Frege's analysis of the predicate "exists" in terms of the existential quantifier, or Russell's analysis of definite descriptions. We find this shift in grammatical form in the analyses of logical constants mentioned in the preceding section. In

$$\Gamma \vdash A \to B \quad \textit{iff} \quad A, \Gamma \vdash B$$

the connective of implication is analyzed in terms of the turnstile, which stands for deducibility.

A simple example of a good analysis with shift in grammatical form, mentioned by Frege in *Die Grundlagen der Arithmetik* (§64–68), is the analysis of the notion of the direction of a line a as the equivalence class of lines parallel to a. This amounts to the analytical equivalence

The direction of a is equal to the direction of b iff a is parallel to b.

What is achieved in passing from the left-hand side of this equivalence to the right-hand side is that we have eliminated a spurious individual term "the direction of a" and used instead the uncontroversial binary predicate "is parallel to".

Leibniz's analysis of identity, given by the equivalence

a is identical to b iff "a" can always be replaced by "b" salva veritate,

achieves a fundamental grammatical shift. It assumes as given and uncontroversial propositional equivalence, i.e., identity of truth value, and analyzes in terms of it identity of individuals. It is because of this shift that Hide Ishiguro could find behind Leibniz's analysis a form of Frege's context principle, which says that we should explain the sense of a word in terms of the truth and falsity of propositions in which it may occur (see Ishiguro 1990, Chapter II). To put it in a nutshell, Frege's principle says that when it comes to explaining how language functions, asserting is more basic than referring (see Dummett 1973, pp. 3–7).

What about deducing? Is it less or more basic than asserting or referring? If we surmise that it is more basic than asserting, in the order of explaining how language functions, we have opened the way to analyze propositional identity in terms of an equivalence relation between deductions, much as Leibniz analyzed identity of individuals in terms of propositional equivalence. The most plausible candidate for an equivalence relation that would do the job is identity of deductions as codified in categories. We have said that this equivalence of deductions is motivated by normalization in natural deduction or by cut elimination.

Propositional equivalence, which in classical logic is defined by identity of truth value, is understood as follows in a proof-theoretical context:

A is equivalent to B iff there is a deduction $f : A \vdash B$ *and a deduction* $g : B \vdash A$.

This relation between the propositions A and B, which certainly doesn't amount to the stricter relation of propositional identity, does not rely on a criterion of identity of deduction.

By relying on such a criterion, we could analyze propositional identity as follows, quite in tune with how category theory understands identity of objects:

A is the same proposition as B iff A and B are isomorphic.

Isomorphism is here understood in the precise way how category theory understands isomorphism of objects: namely, there is a deduction, i.e., arrow, $f : A \vdash B$ and a deduction $g : B \vdash A$ such that g composed with f and f composed with g are equal respectively to the identity deductions from A to A and from B to B. That two objects are isomorphic means that they behave exactly in the same manner in deductions: by composing, we can always extend deductions involving one of them, either as premise or as conclusion, to deductions involving the other, so that nothing is lost, nor gained. There is always a way back. By composing further with the inverses, we return to the original deductions.

6. IS THERE A SEMANTICS OF DEDUCTION?

In theoretical linguistics syntactical theory was much more prosperous than semantical theory. Not so in logic, where semantics, i.e., model theory, has for a long time been preponderant over syntax.

Logicians are concerned with language much more than other mathematicians, and the old name of mathematical logic, *symbolic logic*, rightly stressed that. It is true that linguistic preoccupations are not foreign to some other branches of mathematics – in particular, algebra – but their involvement with language rarely matches that of logic.

Proof theory is entirely within the sphere of language, and, with many good reasons, *syntactical* is usually taken as synonymous with *proof-theoretical*. It is also pretty secure to consider that the set-theoretic models of classical model theory give the *semantics* of mathematical theories based on classical logic. But to call "semantics" the production of any kind of models for other sorts of systems, like the lambda calculus, or systems of nonclassical logics, may well be abusive, if we understand "semantics" *à la lettre*, as the theory giving an explanation of meaning.

Did the untyped lambda calculus really acquire meaning only when, at a rather late date, some sorts of models were found for it?

Do the extant models of intuitionistic logic, or various substructural logics, give meaning to these logics, which are otherwise motivated mostly by proof theory? And what to say about various uses of the word "semantics" in theoretical computer science, or its borderlines, where some rather syntactical activities, like coding natural-deduction proofs with typed lambda terms, or just translating one formal language into another, are deemed a matter of semantics?

Completeness proofs are the glory of logic (though incompleteness proofs are even more glorious), but they should not serve as an excuse for the cheap question, often posed irresponsibly after colloquium talks, or in referee's reports: "What can you tell us about the semantics of your system? What about its models?" And this question should not receive a cheap answer, which consists in producing anything resembling models, or even not resembling them, as a semantics.

Classical model-theoretical semantics gives meaning to referential expressions like terms through models, and propositions acquire truth values through these models, but these models can hardly serve to give meaning to deductions. A consequence relation may be defined with respect to models, but we said that we need rather a consequence graph, where between the same premise and conclusion there may be several deductions. From the point of view of classical model theory, deductions are not bound to the models, but only to the language.

The fact that there is no room for deductions in the classical semantical framework, whose spirit is Platonistic, should be significant for the philosophy of mathematics. That part of mathematics which is bound to deduction – namely, logic – could be understood in a formalistic vein, whereas in the rest of mathematics we would have Platonism. This sort of formalistic conception would resemble Hilbert's formalism in so far as it is not purely formalistic – it understands formalistically just one part of mathematics. However, it differs very much from Hilbert's conception by finding formalism in logic, whereas Hilbert looked for it in those parts of mathematics transcending the finite. Moreover, Hilbert understood the foundational, finitistic, part of mathematics in a constructivist vein. Logic need not coincide with the finitistic part of mathematics, but it should presumably be found in the foundations. So Hilbert's formalism would indeed be turned upside down: formalism is in the foundations, and Platonism above, whereas with Hilbert, constructivism is in the foundations, and formalism above.

The fact that we are not prone to speak about models of deduction, and that this topic has not received much attention up to now, is

in accordance with a formalistic understanding of deduction. When we encounter different reconstructions of the same deductions, as happens when we have sequents on the one hand and natural deduction on the other hand, the usual inclination is not to speak about one reconstruction being a model of the other, but both are taken as alternative syntaxes.

Still, isn't there something model-theoretical in passing from the calculus of sequents to natural deduction? Couldn't one take natural deduction not just as an alternative syntax, but as a model giving meaning to the sequent calculus?

And what about modelling natural deduction itself? We can code natural-deduction proofs by typed lambda terms, according to the Curry–Howard correspondence, but this seems to be rather a matter of finding a suitable syntax to describe natural-deduction proofs, though there are authors who speak about the typed lambda calculus as providing a semantics of deductions.

Another kind of coding of natural deduction is obtained in categories, by proceeding as Lambek (see Lambek and Scott 1986, and references therein). The possibility of this coding is not fortuitous: one can prove rigorously that if we want to represent deductive systems set-theoretically by identifying, in the style of intuitionism, a proposition with deductions leading to it, or deductions starting from it, we must end up with categories. In this set-theoretical representation, one can also exhibit effectively the duality between composition of deductions, i.e., cut, and the identity deduction. Composition leads us from the deductive system to the representing category, and the identity deduction brings us back. This representation, which is summarized in the theorem that every small category is isomorphic to a concrete category, i.e., a subcategory of the category of sets with functions, is an elementary aspect of the Yoneda representation, and is related to some aspects of Stone's representation of lattice-orders and to Cayley's representation of monoids (see Došen 1998 or 1999; § 1.9).

We can always take as a model of a category the skeleton of this category, i.e., the category obtained by identifying isomorphic objects, but there are also more "dynamic" kinds of models. ("Dynamic" is often used in theoretical computer science and borderline areas just as a commending expression. It means roughly "okay", while "static" is a pejorative expression.)

Lambek, who first realized in the sixties that categorial equality of arrows coincides with proof-theoretical equivalence induced by normalization, conjectured also that the same equivalence relation

between deductions could be characterized by saying that the deductions have the same *generality*, by which he meant that by generalizing the deductions, by diversifying schematic letters as much as possible, while keeping the same rules, we shall end up with the same thing. Lambek's way of making precise the notion of generality of deductions was not successful. At roughly the same time, under the influence of Gentzen and Lambek, the matter was approached with more success in a geometrical vein by Eilenberg, Kelly and MacLane, in connection with so-called *coherence* problems in category theory (see Eilenberg and Kelly 1966, Kelly 1971). These are roughly decidability problems for the commuting of various classes of diagrams, i.e., decidability problems in an equational calculus of algebraic partial operations.

Lambek's conjecture that generality characterizes Gentzenian equivalence of deductions is not true in general; in particular, it is not true for intuitionistic implication (as it was conclusively shown in Petrić 1997). But what appears from these studies of coherence problems is that for categories interesting for logic, which codify deductions in various fragments of intuitionistic and substructural logics, we can find interesting geometrical models. That is, these categories can be faithfully embedded in some categories of geometrical morphisms. The matter was rediscovered two decades later with the proof nets of linear logic, and there is also a more recent rediscovery in Buss (1991) and Carbone (1997). However, proof nets are officially presented as a new kind of syntax, while Buss and Carbone disregard categories and don't deal explicitly with identity of deductions.

Lambek remarks in (1999) that this geometrization of algebraic matters goes against the direction given to mathematics by Descartes, but, concerning the matter at hand, this may nevertheless be the right direction.

Do these geometrical models give a semantics of deduction? I would refrain from answering the question. What is certain is that they give models of deduction, and these models are appealing and useful. To corroborate that, I shall present in the last section of this talk a geometrical model for the general notion of adjunction. (This model is explored in detail in Došen 1999.)

7. A GEOMETRICAL MODEL OF ADJUNCTION

Let us first briefly review one of the standard equational definitions of adjunction. We have two categories $\mathcal{A}$ and $\mathcal{B}$, and two functors, F

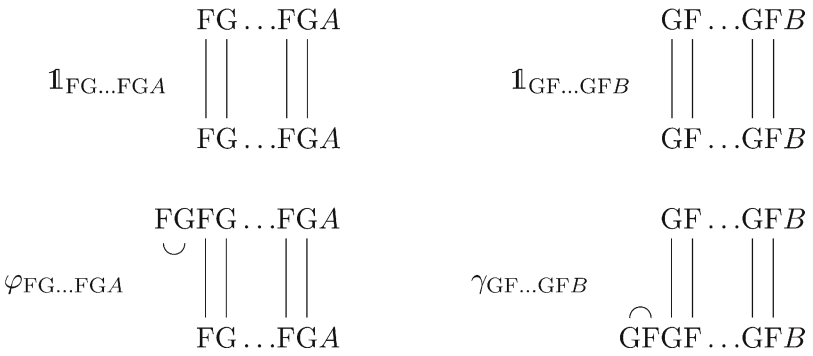
from $\mathcal{B}$ to $\mathcal{A}$ and G from $\mathcal{A}$ to $\mathcal{B}$. The former functor is called *left adjoint* and the latter *right adjoint*. Next, we have a natural transformation φ in $\mathcal{A}$, called the *counit* of the adjunction, whose components are $\varphi_A : \mathrm{FG}A \vdash A$ for every object A of $\mathcal{A}$, and a natural transformation γ in $\mathcal{B}$, called the *unit* of the adjunction, whose components are $\gamma_B : B \vdash \mathrm{GF}B$ for every object B of $\mathcal{B}$. Finally, the following *triangular equalities* must be satisfied:

$$\varphi_{\mathrm{F}B} \circ \mathrm{F}\gamma_B = \mathbb{1}_{\mathrm{F}B},$$
$$\mathrm{G}\varphi_A \circ \gamma_{\mathrm{G}A} = \mathbb{1}_{\mathrm{G}A}.$$

In logical situations we should imagine that one of the adjoint functors F and G is structural, and hence "invisible". Then the unit and counit correspond to rules for introducing and eliminating a connective.

Since all the assumptions in this definition are equational (the equalities in question are the categorial axioms of composing with identity arrows and the associativity of composition, the equalities of the functoriality of F and G, the equalities of naturalness of φ and γ, and the triangular equalities), we can take that we have here an equational calculus. Out of the linguistic material of this calculus we can build the *free adjunction* generated by a set of objects. The details of this construction, as well as other technical details concerning matters in this section, are exposed in Došen (1999).

To every arrow term in the free adjunction we assign a *graph*, which is made of links between occurrences of F and G in the source and target of the arrow term (these graphs should not be confused with the graphs of arrows underlying a category). Identity arrows and the components of the counit and unit have graphs like the following:



while given the graphs for $g : B_1 \vdash B_2$ and $f : A_1 \vdash A_2$ we obtain the graphs of $\mathrm{F}g$ and $\mathrm{G}f$ as follows:

Composition of graphs is defined in the obvious manner; for example, for the first triangular equality we have

It is clear that the composed graph on the left-hand side is equal to the graph on the right-hand side.

One can give a reformulation of the notion of adjunction where composition can be eliminated, in the style of cut elimination. For this reformulation one should replace the families of arrows, i.e., families of components, making the counit and unit of the adjunction by operations on arrows, as Gentzen replaced axioms like $A \wedge B \vdash A$ by rules like

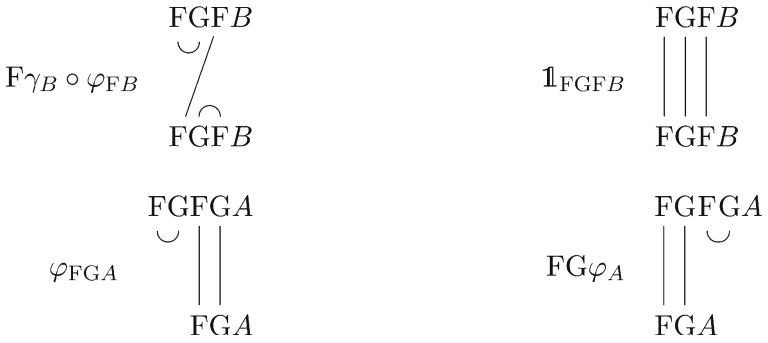
$$\frac{A, \Gamma \vdash C}{A \wedge B, \Gamma \vdash C}$$

One can then obtain a composition-free normal form for arrow terms, which is unique for every arrow.

With the help of this composition-free formulation of adjunction it can be proved that for all arrow terms h_1 and h_2 we have $h_1 = h_2$ in the free adjunction iff the graphs of h_1 and h_2 are equal. This result is of the kind called "coherence theorems" in category theory.

So graphs yield a very simple decision procedure for commuting of diagrams in free adjunctions. They enable us also to reduce to normal form arrow terms in the composition-free formulation without syntactical reduction steps. Uniqueness of normal form can also be demonstrated with the help of these graphs without involving anything like the Church–Rosser property of some reduction steps.

Consider, in the free adjunction, the following two pairs of arrow terms of the same type with different graphs:



There are infinitely many such pairs. It can be shown that if we extend the notion of adjunction by equating *any* such pair, we would trivialize the notion. Namely, the resulting free adjunction would be a preorder: any two arrow terms of the same type would be equal. This means that all these equalities of arrows with different graphs are equivalent with each other.

So the notion of adjunction is Post complete in some sense. Graphs are not absolutely needed to demonstrate this result, but they help to shorten calculations of a rather lengthy inductive argument.

Our coherence result for graphs in free adjunctions guarantees that there are faithful functors from the categories involved in the free adjunction to categories whose objects are finite sequences of alternating F's and G's, and whose arrows are the graphs. The faithfulness of these functors guarantees that we can speak of completeness with respect to the graph models (soundness amounts to functoriality). Our coherence result is exactly like a completeness theorem.

These categories of graphs are subcategories of categories of *tangles*, which have played recently a prominent role in the theory of quantum groups, in low-dimensional topology and in knot theory (see Kassel 1995, Chapter XII; Kauffman and Lins 1994, and references therein). Equality between our graphs covers planar ambient isotopies of tangles without crossings.

Since every logical constant is characterized by an adjunction, we can expect to find in the geometrical models of deductions involving these constants various avatars of our graphs of adjunction.

ACKNOWLEDGEMENTS

I would like to express my warmest thanks to Peter Schroeder-Heister for inviting me to deliver this talk at the workshop on *Proof-Theoretic*

Semantics in Tübingen in January 1999. I am also grateful to the Alexander von Humboldt Foundation for supporting my participation in this conference. The preparation of the written version of the talk was financed by the Ministry of Science, Technology and Development of Serbia through grant 1630 (Representation of proofs with applications, classification of structures and infinite combinatorics).

NOTE

* Since the text of this talk was written in 1999, the author has published several papers about related matters (see 'Identity of proofs based on normalization and generality', *The Bulletin of Symbolic Logic* **9** (2003), 477–503, corrected version available at: http://arXiv.org/math.LO/0208094; other titles are available in the same archive).

REFERENCES

Buss, S. R.: 1991, 'The Undecidability of k-Provability', *Annals of Pure and Applied Logic* **53**, 75–102.

Carbone, A.: 1997, 'Interpolants, Cut Elimination and Flow Graphs for the Propositional Calculus', *Annals of Pure and Applied Logic* **83**, 249–299.

Došen, K.: 1989, 'Logical Constants as Punctuation Marks', *Notre Dame Journal of Formal Logic* **30**, 362–381 (slightly amended version in D.M. Gabbay (ed.), *What is a Logical System?*, Oxford University Press, Oxford, 1994, pp. 273–296).

Došen, K.: 1996, 'Deductive Completeness', *The Bulletin of Symbolic Logic* **2**, 243–283, 523.

Došen, K.: 1998, 'Deductive Systems and Categories', *Publications de l'Institut Mathématique* **64**(78), 21–35.

Došen, K.: 1999, *Cut-Elimination in Categories*, Kluwer, Dordrecht.

Došen, K.: 2001, 'Abstraction and Application in Adjunction', in Z. Kadelburg (ed.), *Proceedings of the Tenth Congress of Yugoslav Mathematicians*, Faculty of Mathematics, University of Belgrade, pp. 33–46 (available at: `http://xxx.lanl.gov/math.CT.0111061`).

Dummett, M.: 1973, *Frege: Philosophy of Language*, Duckworth, London.

Eilenberg, S. and G. M. Kelly: 1966, 'A Generalization of the Functorial Calculus', *Journal of Algebra* **3**, 366–375.

Ishiguro, H.: 1990, *Leibniz's Philosophy of Logic and Language*, 2nd edn., Cambridge University Press, Cambridge.

Kassel, C.: 1995, *Quantum Groups*, Springer, Berlin.

Kauffman, L. H. and S. L. Lins: 1994, *Temperley–Lieb Recoupling Theory and Invariants of 3-Manifolds*, Princeton University Press, Princeton.

Kelly, G. M. and S. MacLane: 1971, 'Coherence in Closed Categories', *Journal of Pure and Applied Algebra* **1**, 97–140, 219.

Lambek, J.: 1974, 'Functional Completeness of Cartesian Categories', *Annals of Mathematical Logic* **6**, 259–292.
Lambek, J.: 1999, 'Type Grammars Revisited', in A. Lecomte et al. (eds.), *Logical Aspects of Computational Linguistics*, Lecture Notes in Artificial Intelligence 1582, Springer, Berlin, pp. 1–27.
Lambek, J. and P. J. Scott: 1986, *Introduction to Higher-Order Categorical Logic*, Cambridge University Press, Cambridge.
Lawvere, F. W.: 1969, 'Adjointness in Foundations', *Dialectica* **23**, 281–296.
Petrić, Z.: 1997, *Equalities of Deductions in Categorial Proof Theory* (in Serbian), Doctoral Dissertation, University of Belgrade.
Prawitz, D.: 1965, *Natural Deduction: A Proof-Theoretical Study*, Almqvist and Wiksell, Stockholm.

Mathematical Institute, SANU
Knez Mihailova 35, p.f. 367
11001 Belgrade
Serbia
E-mail: kosta@mi.sanu.ac.yu

REINHARD KAHLE

A PROOF-THEORETIC VIEW OF NECESSITY

ABSTRACT. We give a reading of binary necessity statements of the form "ϕ is necessary for ψ" in terms of proofs. This reading is based on the idea of interpreting such statements as "Every proof of ψ uses ϕ".

1. INTRODUCTION

In this paper, we give a proof-theoretic approach to necessity. From a philosophical point of view, we even argue that the *semantics* of necessity can be based on a proof-theoretic view.[1] This is in direct opposition to the usual model-theoretic view which interprets necessity by the use of the well-known *semantics of possible worlds*. However, we focus on necessity as a *binary relation*, in contrast to the traditional view of necessity as a unary operator. As a binary relation, it can be expressed by the schema

$$\phi \text{ is necessary for } \psi.$$

where ϕ and ψ are sentences.[2] The idea of the proof-theoretic approach is to read such a statement as

$$\text{Every proof of } \psi \text{ uses } \phi.$$

Obviously, on an informal level, this reading should be very plausible. From a technical point of view, there are two main difficulties: the notion of *use* and the quantifier over proofs. We will propose that for a meaningful necessity statement of this form, ϕ must be chosen from a list of what we call *potential axioms*. This list must be implicitly given by the context of the necessity statement. With this assumption, we can reduce the use of an arbitrary formula to the question of whether an axiom is used.

In Section 2, we will give a short discussion of the possible interpretations of binary necessity statements. In Section 3, we will present the technical preliminaries needed for the formal proof-theoretic reading, defined in Section 4. In Section 5, we will give examples

Synthese (2006) 148: 659–673
DOI 10.1007/s11229-004-6293-4

which illustrates our definition at work. The final section is devoted to a discussion of this approach.

2. UNDERSTANDING OF BINARY NECESSITY

The traditional treatment of necessity, even going back to Aristotle, is the view of necessity as a *modus* of a proposition. With the rise of modern logic, *modal logic* was chosen as an appropriate framework to deal with necessity (cf. Lewis and Langford 1932). In particular, because of the famous semantics of possible worlds (Kripke 1963), this approach is considered to be the standard formalization of necessity. There are several standard references for modal logic (e.g., Hughes and Cresswell 1968; Chellas 1980; Bull and Segerberg 1984). For a historical survey of possible worlds semantics, we refer to Copeland (2002). A comprehensive exposition of the technical aspects of modal logic is given by Kracht (1999).

The most prominent problem of this approach is certainly the *logical omniscience*. It is wellknown and welldiscussed (cf. e.g., Fagin et al. 1995), but, by no means, solved. We would like to criticize the modal approach from an even more general direction. It might be adequate for a notion of *logical necessity*, but it seems to be unsuited for the treatment of necessity in natural language use.

Our main claim is that necessity primarily occur as a *binary relation*. In principle, such a claim would require an empirical study of natural language use, which we cannot provide here. But, let us consider a typical example: Look at the sentence "I must hurry"[3] uttered when I am on my way to the train station. Obviously, there "exists" a possible world in which I do not hurry – but probably miss the train. Therefore, normally, such a sentence has to be understood as: "I must hurry *to catch my train*". Except for logical (and mathematical) statements, we think that necessity occurs in general in binary statements where the succedent may be omitted. However, it can be deduced from the context.

In fact, our analysis requires a pre-knowledge of the context in which a necessity statement is given. The pre-knowledge determine, so-to-speak, the search space for possible antecedents and succedents of necessity statements. As for the example, we would like to stress that necessity statements speak primarily about antecedents and succedents *in the future*.

The usual modal approach to necessity provides an analysis of binary necessity, by reading "ϕ is necessary for ψ" as "$\Box(\psi \rightarrow \phi)$".

However, this reading suffers the problem of logical omniscience: Every tautology is necessary for everything else; everything is necessary for any contradiction.[4]

In particular, pure modal logic seems to be highly unsuitable for the treatment of (our understanding of) necessity: the necessitation rule $\phi/\Box\phi$ makes sense, if the formal system is restricted to capture *only* necessity. However, the challenge is to speak of necessity *in connection with contingency*. Nevertheless, that this is (to some extent) possible in the modal framework by incorporating an operator for the *actual world*, we follow another approach, where necessity is reconstructed on the meta level of an axiom system for the actual world.[5]

3. TECHNICAL PRELIMINARIES

As discussed in the previous section, our approach is based on a deductive system. We will restrict ourselves to the propositional case.[6]

Starting with a set of atomic formulae, the language should be closed under the usual propositional connectives $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction) and $\rightarrow$ (implication). As metavariables for arbitrary formulae we will use Greek letters $\phi, \psi, \ldots$.

We need a standard derivability relation $\{\phi_1, \ldots, \phi_n\} \vdash \psi$ expressing that ψ can be derived from ϕ_1 to ϕ_n. It is not necessary to fix a specific type of calculus, such as *Hilbert-style calculus*, *sequent calculus* or *natural deduction*. However, it is required that the logical axioms include a complete axiomatization of *classical* propositional logic.

As metavariable for sets of axioms we use $\mathcal{D}$. $\mathcal{D} + \{\phi_1, \ldots, \phi_n\}$ and $\mathcal{D} + \psi$ have to be understood in the obvious way that ϕ_1 to ϕ_n or ψ are added as additional axioms to $\mathcal{D}$.

DEFINITION 1.

(1) $\mathcal{D}$ is *consistent*, iff there is a formula ϕ such that $\mathcal{D} \nvdash \phi$. $\mathcal{D}$ is *inconsistent*, iff $\mathcal{D}$ is not consistent.
(2) ϕ is *independent with respect to* $\mathcal{D}$, iff both, $\mathcal{D} + \phi$ and $\mathcal{D} + \neg\phi$, are consistent.[7]
(3) Formulae $\phi_1, \ldots, \phi_n$ *exclude each other* with respect to $\mathcal{D}$, iff for any i and j with $1 \leq i,j \leq n$ and $i \neq j$ the system $\mathcal{D} + \{\phi_i, \phi_j\}$ is inconsistent.

Based on the last definition we introduce the crucial notion of *variety of alternatives*:

DEFINITION 2.

(1) A set $\{\phi_1, \ldots, \phi_n\}$, $n \geq 2$, is called *a set of alternatives* with respect to $\mathcal{D}$, iff $\phi_1, \ldots, \phi_n$ are independent with respect to $\mathcal{D}$ and exclude each other with respect to $\mathcal{D}$.[8]
(2) A set of sets

$$\mathcal{A} = \{\{\phi_{11}, \ldots, \phi_{1m_1}\}, \{\phi_{21}, \ldots, \phi_{2m_2}\}, \ldots, \{\phi_{n1}, \ldots, \phi_{nm_n}\}\}$$

is called a *variety of alternatives* with respect to $\mathcal{D}$, iff the sets $\{\phi_{k1}, \ldots, \phi_{km_k}\}$ are sets of alternatives with respect to $\mathcal{D}$, for all $1 \leq k \leq n$ and any combination of single formulae of each set of alternatives is consistent with $\mathcal{D}$, i.e., $\mathcal{D} + \{\phi_{1l_1}, \phi_{2l_2}, \ldots, \phi_{nl_n}\}$ is consistent for arbitrary $1 \leq l_i \leq m_i$ and $1 \leq i \leq n$.

Finally, we can define the notions of *potential proof* and *use* of a *potential axiom* for a given axiom system $\mathcal{D}$ and a variety of alternatives $\mathcal{A}$:

DEFINITION 3. Let an axiom system $\mathcal{D}$ and a variety of alternatives

$$\mathcal{A} = \{\{\phi_{11}, \ldots, \phi_{1m_1}\}, \{\phi_{21}, \ldots, \phi_{2m_2}\}, \ldots, \{\phi_{n1}, \ldots, \phi_{nm_n}\}\}$$

be given.

(1) The formulae ϕ_{ij_i}, for $1 \leq j_i \leq m_i$ and $1 \leq i \leq n$, are called *potential axioms* for $\mathcal{D}$.
(2) A *potential proof* of a formula ψ is a proof of ψ in $\mathcal{D} + \{\phi_{1l_1}, \phi_{2l_2}, \ldots, \phi_{nl_n}\}$ for arbitrary $l_1, l_2, \ldots, l_n$.
(3) If $\mathcal{B}$ is a proof of ψ in $\mathcal{D} + \{\phi_{1l_1}, \phi_{2l_2}, \ldots, \phi_{nl_n}\}$ and ψ is not provable in $\mathcal{D} + \{\phi_{1l_1}, \phi_{2l_2}, \ldots, \phi_{k-1l_{k-1}}, \phi_{k+1l_{k+1}}, \ldots, \phi_{nl_n}\}$, we say that ϕ_{kl_k} is *used* in the proof $\mathcal{B}$.

Remark 4. In the last definition of the use of a potential axiom we do not refer to any appearance of ϕ in $\mathcal{B}$, but argue from a metatheoretical point of view: If ψ is not provable in the absence of ϕ, but in the presence of ϕ, ϕ was "used". This allows us to give a definition of use which does not depend on the underlying calculus (and it should contain every reasonable definition of use of an axiom in a proof given for a particular calculus). But, now the question arises,

how one can actually prove that ϕ was used, given a potential proof of ψ where ϕ is one of the potential axioms. We have to show that ψ is not provable in $\mathcal{D}$ in the presence of the potential axioms minus ϕ. This can be done by use of a counter-model where the potential axioms without ϕ hold and, in addition, $\neg\psi$. But we would like to stress that this use of a *model-theoretic* argument does not change the *proof-theoretic* character of our approach. Here, model theory is used only as a tool.

4. THE PROOF-THEORETIC READING

In the following we presuppose that a consistent axiom system $\mathcal{D}$ is given. Then, for the analysis of a statement of the form

($\star$) "ϕ is necessary for ψ"

we assume the following:

(1) ϕ and ψ are adequately formalizable in the language of the axiom system $\mathcal{D}$.
(2) ψ is independent of $\mathcal{D}$.
(3) There exists a variety of alternatives with respect to $\mathcal{D}$:

$$\mathcal{A} = \{\{\phi_{11}, \ldots, \phi_{1m_1}\}, \{\phi_{21}, \ldots, \phi_{2m_2}\}, \ldots, \{\phi_{n1}, \ldots, \phi_{nm_n}\}\}$$

such that ϕ is one ϕ_{kl_k}, $1 \leq k \leq n$, $1 \leq l_k \leq m_k$.

In principle, in the second condition we should also demand that ϕ is independent of $\mathcal{D}$. However, this follows from the third point, since all formulae of a set of alternatives in a variety have to be independent of $\mathcal{D}$, and ϕ has to belong to them.

With these conditions we can define:

DEFINITION 5. The necessity statement ($\star$) holds, iff

(1) there is a potential proof of ψ and
(2) every potential proof of ψ uses ϕ.

The first condition is needed to avoid pathological cases, since otherwise the quantification over every potential proof would be vacuous.[9]

Let us briefly discuss the three assumption we have made. The first one is trivial. For the independence of ϕ and ψ, we argue that in the case of formulae ϕ and/or ψ which are already proven (or disproven) language offers other, more adequate statements:

(1) ϕ must usually be a statement about the future and therefore independent. Otherwise, if $\mathcal{D} \vdash \phi$, one would use a necessity statement in the past tense: ϕ *was* necessary for ψ.[10] If $\mathcal{D} \vdash \neg\phi$ holds, the necessity statement has a subjunctive, or even a counterfactual meaning: "ϕ would have been necessary for ψ" or "If ϕ were the case, it would be necessary for ψ".[11]
(2) If ψ is already derivable from $\mathcal{D}$, there is no need for a necessity statement with respect to a antecedence in the future: the claim "ϕ is necessary for ψ" is rejected by the argument that ψ is already the case. This argument, of course, does not apply to necessity statements in the past tense. But, such statements must be distinguished from the ones under consideration here (nevertheless that there should be an analysis along the lines we present in this paper).[12]
(3) If $\mathcal{D} \vdash \neg\psi$, it is already to late for a necessity statement; there is no possibility of proving the succedent in any way. Here, one could at best think of a statement in the subjunctive mood: "ϕ would be necessary for ψ", probably understood in a counterfactual context: "ϕ would be necessary for ψ, if *something* would not have been the case."[13]

The reasons for the conditions on the variety of alternatives, stated in Definition 2, are the following:

(4) Clearly, there should be at least two alternatives in a set of alternatives. Often, a set of alternatives will consist only of a formula and its negation. However, the main example, presented below, illustrates a case where we can use more alternatives, cf. Remark 6 below.
(5) The independence of the elements of a set of alternatives is required to ensure that the antecedence of a necessity statement is independent, cf. the discussion above, item (1).
(6) The condition that the formulae in a set of alternatives exclude each other ensures that there is no overlap: If two potential axioms ϕ and ψ did not exclude each other, there could be different potential proofs for a formula χ of which one uses ϕ and another

uses ψ, i.e., neither ϕ nor ψ would be necessary according to our reading.[14]

(7) The consistency of combinations of elements of different sets of alternatives exclude potential proofs with an inconsistent set of potential axioms.

Of course, the variety of alternatives is the main parameter of our approach. This parameter is not uniquely determined. Probably, most of the disagreements about necessity statements can be explained with respect to disagreements about the variety of alternatives.[15]

5. AN EXAMPLE

As an example illustrating our approach, let us consider the following ranking of a soccer group before the last round.

Team	Points	Goals	Goal difference
A	4	2:0	+2
B	4	2:1	+1
C	3	5:2	+3
D	0	1:7	−6

In the last round, the following matches will be played:

A against D,

B against C.

We assume the reader is familiar with the usual rules for soccer rankings.[16] Now we investigate the following two necessity statements:

(a) "To win the group, *C* must win against *B*".
(b) "To finish second, *C* must win against *B*".

Let us give an informal analysis. For that, we would have to rephrase these statements according to our definition as:

– Every *potential* proof of "*C* wins the group" uses "*C* wins against *B*".
– Every *potential* proof of "*C* finishes second" uses "*C* wins against *B*".

We have to use *potential* proofs because neither "*C* wins the group" nor "*C* finishes second" is provable in the given situation. But it should

be obvious what the additional, *potential* axioms, which are allowed to be used in the potential proofs, are. The possible results of the remaining matches: "*A* wins against *D*", "*A* and *D* draw" and "*A* loses to *D*" build a *set of alternatives*.[17] So, the set of varieties is given by the two sets of alternatives of the two matches.

Remark 6. This example shows why we allow more then two alternatives. To consider only "win" and "not win" is not adequate for soccer results, which allows for two forms of "not win". In addition, we see from this example why a set of alternatives does not need to be logically complete. There is the forth alternative that the match will not be played at all.[18] But it is clear that the usual reading of the necessity statements will not consider this additional alternative.

For our reading of the necessity statement we have first to check the assumptions we made. Obviously, the rules of soccer and the given group can be formalized within an axiomatic framework.[19]

It is easy to observe that the last matches still allow *B* to finish first, second, or third. Therefore the succedents of both statements are independent. The independence of the potential axioms in the variety of alternatives follows from the fact that the games were not played yet.

Now, let us check whether (a) and/or (b) holds with respect to our reading of necessity.

For (a) it is indeed the case that every potential proof of "*C* wins the group" uses the potential axiom "*C* wins against *B*". First, one sees immediately that in the case of the two alternatives where *C* draws or loses to *B*, *B* will clearly stay ahead of *C*, and therefore, *C* cannot be the winner of the group. To complete the argument that (a) is correct, we must, however, check that there is at least one potential proof of "*C* wins the group" using "*C* wins against *B*". But that is, indeed, the case (e.g., if *A* draws against *D*).

If we turn to (b), it could be tempting to use an argument along the same lines, i.e., to argue that in both other alternatives *B* will stay ahead of *C*, and using the fact that *B* is second. In addition, there is a proof of "*C* finishes second" which uses "*C* wins against *B*" (if *A* wins against *D*). But, *all* proofs have to use this antecedent. And there is a proof of "*C* finishes second" which uses, instead, "*C* ties against *B*" – if *A* loses to *D*. In this case, *C* and *A* will both have four points, but the goal difference of *C* is better than that of *A*. Thus, (b) is not correct.

The example also allows for an illustration of the assumptions we made for the definitions.

(1) As claimed earlier, the antecedence of a necessity statement should be one about the future. If we consider, for example, results of matches already played, we would use sentences like "It *was* necessary that *C* won against *D* for *C* to win the group." or "It *would have been* necessary for *D* to win against *B* to win the group".[20]

(2) A statement of the form "To stay ahead of *D*, *A* must win against *D*" would be rejected, since it is already provable that *A* will stay ahead of *D* in any case, i.e., independent of the results of the remaining matches. In fact, we could consider this as a limit case of our analysis: if the succedent is already provable, there is obviously a (potential) proof not using the antecedence. Therefore, the necessity statement is trivially false.

(3) If the succedent is false, a necessity statement would be meaningless: "To win the group, *D* must win against *A*" is rejected since *D* cannot win the group in any case. It is rejected in a strong sense as meaningless, since its negation would not be considered as true, either: "It is not the case that to win the group, *D* must win against *A*". Both statements are meaningless, with the justification that *D* can not win the group no matter what.

(4/5) That there are at least two alternatives and that they should independent, is clear from the general considerations.

(6) For the sixth condition, let us consider another example: the statement "It is necessary that *C* does not lose to *B* to finish second" should be a true necessity statement. However, if we reconstruct it in a context with a set of alternatives containing "*C* does not lose to *B*" and "*C* draws against *B*", we would find a proof of *C* finishes second without using the first alternative. Of course, this would be rejected as an argument against the necessity statement, since the draw is included as a possibility in the first alternative. Formally, we ensure this by the requirement that the elements of a set of alternatives exclude each other.

(7) The consistence of the combinations avoids varieties of alternatives like the one where we have in one set "*B* wins" and in another "*C* wins". Since *B* will play against *C* a combination of these two potential axioms would result in a contradiction and we would have potential proofs for every formula without the use of any other formula. This, clearly, would result in contra-intuitive consequences.

6. DISCUSSION

6.1 *The Variety of Alternatives*

Obviously, the appropriate choice of the variety of alternatives is highly relevant for our approach. This is a question which depends on the context of the statement. We do not claim that there is any general method to fix a variety of alternatives. However, for the given example, the choice seems to be clear. Given a background theory, the question of finding an appropriate definition of the variety of alternatives is more a consequence than a presupposition of our analysis. Of course, the antecedence has to show up in a set of alternatives. In the easiest case, it just comes together with its negation. However, the other possible axioms might be able to be found by searching for a potential proof of the succedent. With this observation, we can even link the analysis of necessity statements to the field of *proof search* and the general field of *abduction*. This applies in particular, if we put the background theory itself in question: we do not search for a proof of the succedent only, but ask under which circumstances the necessity statement would be true.[21] Besides the antecedence, other formulae might be needed to prove the succedent. These are then assumed to be part of the background theory.

6.2. *Possible Worlds Semantics Revisited*

The variety of alternatives can be seen as a formal counterpart of the variety of possible worlds needed in the standard semantic approach to necessity. Given an axiom system, the combinations of it with the potential axioms from the sets of alternatives should give – at least, partial – axiomatizations of the different possible worlds. Then, the reading of "ϕ is necessary for ψ" as "ψ implies ϕ in all possible worlds" will probably coincide with our reading. From this perspective, the variety of alternatives focus only on the formulae which could be considered in necessity statements. Therefore, the problem of logical omniscience is rather "defined away" than directly solved: tautologies will not appear as elements of a variety of alternatives. Also, the consequence that everything is necessary for a contradiction is excluded by the requirement of at least one possible proof of the succedent, which means that there is at least one possible world in which it holds. We claim that this is not a bad solution, since it draws its support from the plausible idea of the *use* of a formula.

As discussed above, a proof-theoretic approach can even give some guidance on how to find the (or one) appropriate variety of alternatives. In contrast, possible world semantics itself does not provide any information on how to determine the appropriate variety of possible worlds.

6.3. *Modal Logic*

Modal logic, as an axiomatic counterpart of possible worlds semantics, has its limitations in relating contingency and necessity. Even if we add an operator to distinguish the *contingency* of a formula, necessity is stated only for given axioms and is then closed under logical and necessity operations. This is perfect for a notion of *logical necessity*. However, the proof-theoretic approach gives a direct relation of contingent formulae and its role in binary necessity statements as used in natural languages. The price we pay for this is that necessity statements must be read as *meta-statements*. Obviously, in our reading, they are not incorporated into the original axiom system. However, we consider this rather as a feature than as a "bug" of the analysis.

6.4. *Necessary Condition*

There is a relation of our approach to the usual definition of a *necessary condition*. In fact, from our definition it follows that, if ϕ is necessary for ϕ, $\neg\phi$ implies $\neg\psi$ (or ψ implies ϕ). (It holds $\psi \vee \neg\psi$. Since every proof of ψ uses ϕ, ψ can hold only in the presence of ϕ. Hence, $\phi \vee \neg\psi$, i.e., $\neg\phi \rightarrow \neg\psi$.)

However, the converse can be established only, if we take the variety of alternatives into account. Otherwise, on the one hand, we would have the problem of logical omniscience (tautologies would be necessary for everything). On the other hand, we might need extra potential axioms to prove the required implication.

As for the possible world semantics, we think that it is possible to define binary necessity in terms of necessary condition using the variety of alternatives. This would probably lead to an equivalent characterization. But, as for the possible world semantics, we claim that the approach given here is closer to the actual reasoning we use to justify (binary) necessity statements. Therefore, the given approach distinguishes itself as the primary definition. The characterization via necessary condition (as well as via possible worlds) is in this sense secondary and is rather a corollary about necessity than a definition of it.[22]

6.5. *Intensionality in General*

Necessity is paradigmatic for an intensional phenomena. Intensionality can be characterized as being incompatible with substitution of logical equivalent terms or formulas. Therefore, a standard set-theoretic semantics which is, by definition, extensional, fails in an intensional context. Such as for necessity, we claim that intensional phenomena can be explained by the use of additional information given within a *proof-theoretic* framework. One motivation for the proof-theoretic treatment can be found in *completeness*. As long as we have a complete axiomatization of a situation, models and proofs are of the same value with respect to validity and provability. But we have more structure on the proof-theoretic side. For instance, a statement can have several proofs, but should have only one truth value. This additional structure seems to be crucial for the analysis of intensional phenomena. This should be directly applicable to the notion of *relevance*. Here, the well-developed field of *relevance logic* (cf. e.g., Anderson and Belnap 1975; Anderson et al. 1992), investigates mainly the question of relevance as a property of an axiom system as a whole. The question whether "ϕ is relevant for ψ" could probably answered in a way similar to the way in which we treated necessity.

For the case of *belief revision*, *syntax based* approaches can be seen in the line of a proof-theoretic view (cf. Nebel 1992; Hansson 1998). From our point of view, they still often involve too many model-theoretic components, cf. e.g., the case of *safe contraction on belief bases* (Fuhrmann 1991; Nayak 1994). An example of a closer proof-theoretic view is given in (Kahle 2002). This approach is actually very close to the framework of *Truth maintenance systems* (cf. Doyle 1979; de Kleer 1986) proposed in the field of *artificial intelligence* (Russell and Norvig 1995). Indeed, in this field the advantage of proof-theoretic components over purely model-theoretic methods is clearly realized, cf. e.g., the problems of *model update* (or *interpretation update*) for *deductive databases* (or *knowledge bases*), (cf. Alferes et al., 2000).

As the most promising alternative to our approach we consider Moschovakis's *recursion-theoretic* approach, reading Frege's notion of *sense* and *denotation* as *algorithm* and *value* (Moschovakis 1994). We would link, alternatively, the *sense* of a declarative sentence, which Frege characterizes as *the mode of presentation* (Frege 1892; Frege 1952), with the *proofs* of it.

ACKNOWLEDGEMENT

I would like to thank Michael Arndt, Bertram Fronhöfer, Peter Schroeder-Heister, Greg Wheeler, and Bartosz Wieckowski for helpful comments on this paper.

NOTES

[1] The term"proof-theoretic" might be a little bit misleading, since we do not use any specific technique from (mathematical or structural) proof theory. A more modest term could be "proof-based". However, in differentiation from the modal approach and possible worlds semantics we think that the term "proof-theoretic" is justified. From a broader point of view, this paper also contributes to the general program of "proof-theoretic semantics" to which this volume is devoted.

[2] We avoid here the term *propositions* since propositions involve often a reference to possible worlds, at least to certain semantical presuppositions (cf. e.g., Anderson 1995).

[3] At the conference *Modality in Contemporary English*, Verona, 6–8 September 2001, (Facchinetti et al., 2003), we learned that "must" is being increasingly replaced by "have to" in contemporary English. However, here we are not focusing on this aspect of natural language use.

[4] See also the discussion about *necessary condition* at the end of the paper.

[5] A criticism of possible worlds semantics is given from a different perspective by Forster (200x).

[6] There are no principle reasons as to why the approach could not be extended to first or even second order logic. However, from a technical point of view, it would require a clear understanding of the independence of axioms. This is not obvious, if we think of instances with respect to their generalizations.

[7] With this definition independency coincides with a natural understanding of *possibility*. That means, in this case we could also say that ϕ *is possible with respect to* $\mathcal{D}$.

[8] We do not demand that a set of alternatives is *complete*, i.e., that $\mathcal{D}$ proves $\phi_1 \vee \ldots \vee \phi_n$. The reason for this is explained below in Remark 6.

[9] Therefore, we have already avoided the consequence that every single formula would be necessary for a contradictory statement.

[10] Exceptions are the *mathematical* necessity statement. However, in mathematics there is a well-defined notion of the *necessary condition*, as A is a necessary condition for B, if B implies A.

[11] We do not discuss here whether these two sentences might express the same thing.

[12] One can think of incorporating an explicit temporal structure into the axioms. In fact, our presupposition that potential axioms can become true only in the future, already contains such a temporal aspect.

[13] For an interesting analysis of counterfactuals which criticize the standard modal approach (cf. Wehmeier 200x).

[14] We are not sure whether this condition is absolutely necessary. There might be cases where one would like to allow such overlaps. But so far, we have not found any.
[15] Cf. also the discussion below in Section 6.
[16] Three points for a win, one for a draw, goal difference as first tie-breaker.
[17] For our example we can restrict ourselves to the "qualitative" results: win, tie, lose. The score does not matter.
[18] From another perspective, the disjunction is not complete since A has neither won, drawn, or lost "now", when the game has not yet been played.
[19] We have to ensure that this axiomatization can be expressed in propositional logic. In the example, this is probably not the most convenient way of formalization. However, it can be done by the use of appropriate propositional variables such as $p_1 \equiv$ "*A wins against D*", $p_2 \equiv$ "*A has now* 7 *points*", $p_3 \equiv$ "*A will have* 10 *points*" and axioms of the form $p_1 \wedge p_2 \rightarrow p_3$.
[20] The first statement should be considered as true, however, the truth of the second one is debatable.
[21] We are convinced by the fact that in a rigid reading of necessity – in particular, in terms of possible worlds – nearly every non-logical necessity statement uttered in a natural language conversation will turn out to be false.
[22] In particular, contra (Stalnaker 1995), we deny that possible worlds are the appropriate framework for articulating and sharpening the problem of the nature of modal truth, especially if it is considered as a presuppostion to speak about modalities.

REFERENCES

Anderson, A. and N. Belnap: 1975, *Entailment*, Volume I. Princeton University Press.
Anderson, A., N. Belnap, and M. Dunn: 1992, *Entailment*, Volume II. Princeton University Press.
Anderson, C. A.: 1995, 'Proposition, State of Affairs', in J. Kim and E. Sosa (eds.), *A Companion to Metaphysics*, Blackwell, pp. 419–421.
Alferes, J., J. Leite, L.M. Pereira, H. Przymusinska, and T. Przymusinski: 2000, 'Dynamic Updates of Non-monotonic Knowledge Bases', *Journal of Logic Programming* **45**(1–3), 43–70.
Bull, R. and K. Segerberg: 1984, 'Basic Modal Logic', in D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, Volume II, Kluwer, pp. 1–88.
Chellas, B.: 1980, *Modal Logic: An Introduction*, Cambridge University Press.
Copeland, B.J.: 2002, 'The Genesis of Possible Worlds Semantics', *Journal of Philosophical Logic* **31**(2), 99–137.
de Kleer, J.: 1986, 'An Assumption-based TMS', *Artificial Intelligence* **28**, 127–162.
Doyle, J.: 1979, 'A Truth Maintenance System', *Artificial Intelligence* **12**, 231–272.
Facchinetti, R., M. Krug, and F. Palmer (eds.): 2003, *Modality in Contemporary English*, de Gruyter.
Fagin, R., J. Halpern, Y. Moses, and M. Vardi: 1995, *Reasoning about Knowledge*, MIT Press.

Forster, Th.: 200x, 'The Modal Aether', in R. Kahle (ed.), *Intensionality*, A K Peters.
Frege, G.: 1892, 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* **(NF 100)**, 25–50.
Frege, G.: 1952, 'Sense and Meaning', in P. Geach and M. Black (eds.), *Translations from the Philosophical Writings of Gottlob Frege*. Basil Blackwell, 1952. English translation of (Frege, 1892).
Fuhrmann, A.: 1991, 'Theory Contraction Through Base Contraction', *Journal of Philosophical Logic* **20**, 175–203.
Hansson, S.O.: 1998, 'Revision of Belief Sets and Belief Bases', in D. Dubois and H. Prade (eds.), *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Volume 3: Belief Change*, Kluwer pp. 16–75.
Hughes, G. and M. Cresswell: 1968, *An Introduction to Modal Logic*, Methuen, London.
Kahle, R.: 2002, 'Structured Belief Bases', *Logical and Logical Philosophy* **10**, 45–58. Special issue of the Workshop *LLP* held spring 2001 at the TU Dresden.
Kracht, M.: 1999, *Tools and Techniques in Modal Logic*, Volume 142 of *Studies in Logic and the Foundations of Mathematics*, Elsevier.
Kripke, S.: 1963, 'Semantical Analysis of Modal Logic I, Normal Propositional Calculi', *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* **9**, 67–96.
Lewis, C. and C. Langford: 1932, *Symbolic logic*, The Century Co., New York.
Moschovakis, Y.: 1994, 'Sense and Denotation as Algorithm and Value', in J. Oikkonen and J. Väänänen (eds), *Logic Colloquium '90*, Volume 2 of *Lecture Notes in Logic*, Springer, pp. 210–249.
Nayak, A.: 1994, 'Foundational Belief Change', *Journal of Philosophical Logic* **23**, 495–533.
Nebel, B.: 1992, 'Syntax-based Approaches to Belief Revision', in P. Gärdenfors (ed.), *Belief Revision*, Cambridge University Press, pp. 52–88.
Russell, S. and P. Norvig: 1995, *Artificial Intelligence – A Modern Approach*, Prentice Hall.
Stalnaker, R.: 1995, 'Modalities and Possible Worlds', in J. Kim and E. Sosa (eds.), *A Companion to Metaphysics*, Blackwell, pp. 333–337.
Wehmeier, K.: 200x, 'Descriptions in the Mood', to appear in R. Kahle (ed.), *Intensionality*, AK Peters.

CENTRIA, Universidade Nova de Lisboa and
Departamento de Matemática
Universidade de Coimbra
Apartado 3008
P-3001-454 Coimbra
Portugal
E-mail: kahle@mat.uc.pt

GABRIELE USBERTI

TOWARDS A SEMANTICS BASED ON THE NOTION OF JUSTIFICATION

ABSTRACT. Suppose we want to take seriously the neoverificationist idea that an intuitionistic theory of meaning can be generalized in such a way as to be applicable not only to mathematical but also to empirical sentences. The paper explores some consequences of this attitude and takes some steps towards the realization of this program. The general idea is to develop a meaning theory, and consequently a formal semantics, based on the idea that knowing the meaning of a sentence is tantamount to having a criterion for establishing what is a justification for it. Section 1 motivates a requirement of epistemic transparency imposed onto justifications conceived as mental states. In Section 2, the formal notion of justification for an atomic formula is defined, in terms of the notion of cognitive state. In Section 3, the definition is extended to logically complex formulas. In Section 4, the notion of truth-ground is introduced and is used to give a definition of logical validity.

1. EPISTEMIC TRANSPARENCY

The gist of Dummett's anti-realistic argument is that if we equate the meaning of a sentence with its truth-conditions, and we conceive, realistically, truth as transcending our recognitional capacities, we are not able to give an account of our knowledge of the meaning of certain sentences satisfying a *manifestability* requirement, according to which

> [t]here must be an observable difference between the behaviour or capacities of someone who is said to have that knowledge and someone who is said to lack it (Dummett 1975, p. 7).

Suppose now you are a mentalist, so that you are naturally inclined to conceive knowledge of meaning as a sort of mental state, and that you think that behavior is no criterion of knowledge (though it can of course provide evidence for the possession of knowledge); in that case you will not be prepared to accept the manifestability requirement, since you are willing to admit that someone could be

Synthese (2006) 148: 675–699
DOI 10.1007/s11229-004-6272-9

in the mental state commonly called "knowing the meaning of A" even if he is not capable of exhibiting an appropriate behavior, perhaps because of some neurophysiological impairment.[1] But suppose also that you think that, to be justified in attributing a mental state to someone, you must satisfy a *specifiability* requirement, according to which you must be capable of specifying that state in terms of certain algorithms implemented in some way in his cognitive structure and of a finite amount of information. Under these conditions the anti-realistic conclusion of Dummett's argument follows, if it follows in the original argument;[2] for, if we equate the meaning of a sentence with its truth-conditions, and we conceive, realistically, truth as transcending our recognitional capacities, then we are not able to give an account of our knowledge of the meaning of certain sentences that satisfies the specifiability requirement: in the case of an undecidable true sentence of the form $\forall xA$ it seems impossible to see how knowledge of its truth condition, i.e., knowledge that infinitely many facts $A(t_1), A(t_2), \ldots$ subsist, could be specifiable as a mental state.

These remarks point out what is problematic in the truth-conditional theories of meaning: Knowledge of meaning is explained in term of a notion – truth – that is recognition transcendent. It is therefore natural to require that, if a theory is to provide a good account of knowledge of meaning, the key notion in terms of which it explains the meaning of sentences is not recognition transcendent, but recognition *immanent*. The problem is how to construe exactly this requirement of epistemic immanence. Let me distinguish between two interpretations: according to the first, or weak one, the key property must be such that it is *atemporally possible* for an appropriately idealized subject to recognize that an arbitrary sentence has this property, if it has it, and that it does not have this property, if it does not have it; according to the second, or strong one, the key property must be such that such a recognition is *presently possible*. What "atemporally possible" means is explained by Prawitz in the following passage, in which he speaks about the possibility not of recognition but of proof:

> [A] mathematical sentence is true if there exists a proof of it, in a tenseless or abstract sense of exists [...]. Or we may express the same idea by saying that a sentence A is true if 'we can prove A' [...]. That we can prove A is not to be understood as meaning that it is within our practical reach to prove A, but only that it is possible in principle to prove A [...]. Similarly, that there exists a proof of A does not mean that a proof of A will be constructed but only that the possibility is there for constructing a proof

of A. [...] I see no objection to conceiving the possibility that there is a specific method for curing cancer, which we may discover one day, but which may also remain undiscovered (Prawitz 1987, pp. 153–154).

On the other hand, "presently possible" alludes to a possibility which subsists or not according to our having or not having presently a procedure, or an algorithm, at our disposal, which is recognized as being such that its application to an arbitrary sentence would give as a result "Yes" if the sentence has the key property, "No" if it doesn't.

All theories using assertibility as the key notion adopt the weak interpretation, for it would be silly to say that even an idealized subject has presently at its disposal an algorithm answering "Yes" to exactly the assertible sentences. But if we interpret the immanence requirement in this way, it is not clear in which sense we can give a good account of knowledge of meaning. The possibility to which Prawitz alludes is not characterized as an epistemic state, but as merely *factual* accessibility to an epistemic state, so that its subsistence is completely independent of the subject's cognitive states; as a consequence it seems impossible to specify a cognitive state as the one of which knowledge of the meaning of a specific sentence consists.

On the contrary, if we interpret the immanence requirement in the strong way, we require that a subject who knows the meaning of A has an algorithm at his disposal, and therefore be in a specific mental state. Of course, in this case, we cannot choose the assertibility property as the key notion; but we can choose the relation "x is a canonical proof of A" (when A is a mathematical sentence). From this point of view, to know the meaning of a (mathematical) sentence A is tantamount to having a decision procedure for the relation "x is a canonical proof of A".

One might wonder whether it is intuitively plausible to require that such a relation is decidable. Prawitz has argued, for instance in Prawitz (1977), that it is not. The reason, as far as I can see, is the following:

> In the cases when A is an implication or a universal sentence [...] we must require not only a construction or description of an appropriate procedure but also an understanding of this procedure (p. 27)

in the sense that

> it cannot be enough that the person is just able to name or describe an operation which in fact always yields a certain kind of result when applied to objects within its domain; [...] he must also understand that the described procedure, when applied to an object within its domain, always yields a result of the stipulated kind (*Ibid.*);

but

it is doubtful in what sense, if any, one could decide the question whether this condition obtains in a certain situation (p. 29).

This reason is in fact the statement of a question, and it seems to me that there is an answer. If we reflect that the notion of proof at stake here is from the start an *intuitive* notion, not the notion of proof of some formal system, the sense in which one can decide the question whether something is a canonical proof or not becomes clear if we adopt the intuitionistic idea that proofs

> are mental constructions, that is, objects of thought not merely in the sense that they are thought about, but in the sense that, for them, *esse est concipi*. They exist only in virtue of our mathematical activity, which consists in mental operations, and have only those properties which they can be recognized by us as having (Dummett 1977, p. 7).

Let me restate this idea in the form of a requirement of epistemic transparency imposed onto proofs conceived as mental constructions:

(1) *Epistemic Transparency,* A mental construction is not a proof of A unless it is recognized as such by an *idealized* knowing subject.

I take this principle as a way of making explicit an essential characteristic of the intuitive notion of proof, or better: of *one* intuitive notion of proof. A basic intuition we have about proofs in this sense is that a proof of A is essentially what is recognized as such by an *idealized* knowing subject: there is not a point of view from which a construction can be judged *to be* a proof of A in spite of the fact that no idealized subject who is capable of recognizing it is aware of its being a proof of A, or from which a construction can be judged *not to be* a proof of A in spite of the fact that an idealized subject who is capable of recognizing it believes that it is. *To be* a proof of A is *to be conceived* as such by an idealized subject. Of course it may happen that an *empirical* subject mistakes something for a proof, or that he doesn't realize that something is a proof; but this is a consequence of limitations of memory, attention, and so on, from which we abstract when we appeal to an *idealized* subject.

Let me emphasize an essential point. Someone might deny that epistemic transparency is a characteristic of the *intuitive* notion of proof, or even of *one* intuitive notion; he might maintain that it is a characteristic of the notion of proof as *the anti-realist* conceives it; and – he might continue – it is scarcely plausible to call such a notion "intuitive"; for example, it seems to be in perfect agreement with intuition to say that something is a proof of A even if a subject who is

considering it does not realize that it is. My answer is, first, that intuition, when we consult it about fundamental notions, does not give us as definite answers as the objector implies. What is, for instance, the 'right' notion of possibility, of cause, of set, of probability, or of truth? If there were a unique answer there would simply be no space for as many philosophical discussions as there still are. Often, the only thing a philosophical discussion can do is bring to light the fact that what seemed to be one intuitive notion is in fact a cluster of notions. Sometimes it is sufficient to make them explicit in order to extricate notions that are simply different; it is the case of possibility: there is not 'the' intuitive notion of possibility, but there is logical possibility, deontic possibility, epistemic possibility, etc.; and none is 'more intuitive' than the others. Sometimes things are more complicated; in the case of truth, for example, after having extricated the metaphysical notion from the epistemic one we have to do with two notions that are not simply different: they vie for the role of the 'correct' notion of truth, or for the role of fundamental notion of the theory of meaning. In such cases (and I hold that the case of the notion of justification is similar), it is not by appealing to intuition that we can settle the question; what is decisive are considerations concerning the nature of the theories we can build on the basis of each notion, their coherence, their explicative power, and so on. In this connection it is not unacceptable, nor surprising, that some notion is both intuitive and congenial to a specific philosophical conception. What is important is, on the one hand, that the notion is intelligible to anyone who has different philosophical views, and, on the other hand, that the choice of it as a fundamental notion has clear motivations. In the present case, the choice of an epistemically transparent notion of proof has a very clear reason: it is the only choice compatible with the idea that knowing the sense of A is tantamount to having a criterion for establishing whether something is or is not a proof of A. For, if we admit that in some possible situation a construction *is* a proof of A, in spite of the fact that an (idealized) subject s is not aware of it, or that a construction *is not* a proof of A, in spite of the fact that s believes that it is, then we must also admit that s *has not* a criterion for establishing whether a construction is a proof of A, and we cannot equate his knowledge of the meaning of A to having such a criterion. On the other hand, the notion of an epistemically transparent proof is perfectly intelligible to the realist: it is the *internalist* notion of proof. In fact, I do not see how even a platonist might dispense with such a notion: even if mathematics is an

activity of discovery of a realm of entities *an sich* subsisting, the mathematician has to be absolutely confident in the reliability of proofs, the very tools by means of which he discovers mathematical truths, and there seems to be no other way of obtaining such a confidence than postulating the epistemic transparency of proofs.

2. JUSTIFICATIONS FOR ATOMIC SENTENCES

When we try to extend a verificationist theory of meaning to empirical sentences, the key notion of the theory must be non-factive and defeasible; it will therefore be more appropriate to use for such a key notion the word "justification" instead of "proof" or "verification". Justifications are *defeasible* in the sense that a justification for a sentence A can cease to be a justification for A as new information is received; they are *non-factive* in the sense that it may happen that a subject s has a justification to believe a sentence A that in fact is, intuitively, not true. I will call here *non-conclusiveness* the logical disjunction of defeasibility and non-factivity. That most empirical sentences can be justified only in a non-conclusive way will be assumed here without discussion.

An important consequence of the use of the non-conclusive notion of justification as the key notion of the theory of meaning is that the verificationist thesis according to which

(2) The truth of a mathematical proposition A can be defined as the existence of a (direct) verification of A

cannot be generalized in the obvious way resulting in the thesis that the truth of an empirical proposition A can be equated to the existence of a justification for A; the reason is simple: since the notion of justification is non-conclusive, if the truth of A were defined as the existence of a justification for A, we would have that some sentences which are true on one occasion are not true on another occasion[3] – a very counterintuitive consequence. This fact is considered as a serious difficulty by any verificationist who makes the further assumption that

(3) The meaning of a sentence is given by its (direct) verifiability conditions.

The reason is the following. If we make this assumption, and put it together with (2), we can immediately derive that

(4) The meaning of a sentence is given by its truth-conditions;

and many verificationists are willing to accept this familiar principle as expressing an essential link between meaning and truth – a link that also realists postulate, the only difference being the way truth is understood.

But this link is broken if we replace "verify" with "justify", since the verificationist thesis (2) cannot be generalized to empirical propositions.

I will call "verificationist" whoever believes that it is possible and interesting to develop a theory of the meaning of mathematical propositions based on the notion of verification, and "justificationist" whoever believes that it is possible and interesting to develop a more general theory of the meaning of mathematical and empirical propositions based on the notion of justification. Suppose now that a verificationist already has some reasons not to subscribe to (3), for example of the kind illustrated in the first section; then the link between meaning and truth would be broken from the start (namely also in the more restricted domain of mathematical sentences), and there would be no *new* problem for the justificationist project.

If we renounce the link between meaning and truth from the start, how can we conceive knowledge of the meaning of an empirical sentence A? My suggestion is to equate it with the capacity to recognize a canonical justification for A, i.e., to decide the relation "x is a canonical justification for A". Of course this idea remains empty unless I can define a plausible theoretical notion of canonical justification for A which turns out to be epistemically transparent, and therefore decidable. This is what I shall try to do in what follows. The strategy I shall adopt consists in inductively defining the set of justifications for A, and in isolating the set of canonical justifications by means of a decidable property.

The problem of atomic sentences has been ignored since Heyting's inductive definition of the notion "x is a proof of A", as far as I know; but Heyting had a good reason to neglect it: he was exclusively interested in a theory of the meaning of the logical constants; while a contemporary verificationist or justificationist has a more ambitious program: he aims at an explanation for the meaning of all (mathematical or empirical) sentences, so the problem cannot be avoided. Obviously, an answer of the sort: "A justification for an atomic sentence is whatever authorizes us to believe A" would be unexplanatory: it is the intuitive relation of warranting that we are trying to explain, at least partially, through the notion of justification, not the other way round.

My idea is to define the notion of justification for an atomic sentence in terms of two other notions: the notion of *authorization to*

use a name to refer to a given entity, and the notion of *authorization to concatenate* a predicate with a name.[4]

I will give an idea of what an *authorization to use* the name *n* to refer to a given object is through an example. Consider the name "Chomsky". According to the model for the sense of a proper name sketched in Dummett (1973), to know the sense of this name is to have a criterion of identification of an object as the referent of the name. The problem is how to explain the notion of an object's being 'given' or 'presented' to a subject. Dummett explains it by appealing to the general assumption that there is a 'fundamental' way of presenting physical objects, i.e., demonstrative identification; as a consequence, the understanding of the sense of a name amounts to an ability to determine the truth-value – more properly, to know what would determine the truth-value – of what he calls a "recognition statement": a sentence of the form "This is *X*", where "*X*" is the name in question and the "is" occurs as the sign of identity. I find this explanation inadequate for two reasons. First, the assumption that there is some privileged way in which persons, for instance, are given to us seems quite unplausible: we can know a person through perception but also through testimony, or by interpreting the traces of some of her/his actions (a footprint, a book, a statue, and so on), or by locating her/him within a net of parental relations some of whose elements we already know, and so on; and the same is true of a great variety of objects. Second, the reason why Dummett considers recognition sentences as basic is simply that they are construed by him as assuring recognition; in other terms, the ability to determine the truth-value of the recognition statement "This is Chomsky" can be plausibly suggested as an explanation of the knowledge of the meaning of the name "Chomsky" only if the statement is interpreted as expressing the identity of an unknown entity named "Chomsky" with the known man that is being indicated, and which is known just because he is perceptively present. But it is not always the case that a statement of the form "This is *X*" can be interpreted in this way. Suppose, for example, that both I and my audience know of a certain boy we have never met, that he got lost in the park, that his name is "John", that he is four years old, that he wears a red shirt, and that he is fair-haired. Suppose I walk through the park and I meet a boy who corresponds to what I know about him; in this situation I am entitled to say to a friend of mine who is walking with me: "This is John". It is true that I use a sentence of the form "This is *X*"', but I use it in a completely different way from the one assumed by

Dummett as normal when he speaks of recognition statements: what we know is the referent of "John", and the referent of "this" becomes known only by virtue of his identity with the referent of "John". We seem therefore to lack a non-circular criterion for distinguishing the 'good' uses of the recognition statements.

It may seem that, if we withdraw the assumption of a 'fundamental' way of identifying physical objects, the notion of authorization to use a name to refer to a given entity becomes irremediably vague, subjective and context-dependent, so that it cannot be used to explain the sense of names. I think this is not the case. What is necessary is to extract from the intuitive notion of context the aspects that are relevant in determining a subject's being or not being authorized to use a name. To this effect I will introduce the notion of cognitive state.

For the purposes of the present paper I will identify an (atomic) *cognitive state* with a triple $c = \langle e, \langle t, f_t \rangle, i \rangle$, where e is a specification, for every name n and every predicate P, of the *epistemic contents* e_n and e_P associated with it; t is the activated term of the internal representational system, and f_t the associated identity criterion; and i is a function mapping (i) the activated term t to the information i_t encoded into or associated with it; (ii) every name n to a (possibly empty) set i_n of auxiliary terms activated in connection with n, together with information concerning them; (iii) every primitive predicate P to some supplementary information i_P from perception, memory, etc. associated with P. Let me explain this terminology.

The epistemic content e associated with a name or to a predicate is a certain amount of verbal or non-verbal information attached to them. For instance, the name "Chomsky" might be associated with the pieces of verbal information "Chomsky is a linguist" and "Chomsky teaches at M.I.T.", but also to such non-verbal information as the mental representation of a face, or of a voice, stored in some mental catalogue; and the predicate "x is square" might be associated with the piece of verbal information "x has four equal sides", but also to the mental representation of a square.[5]

Two restrictions are imposed on the epistemic content associated with a name n. (i) Verbal information must be *atomic*, in the sense that it has to be expressible by means of atomic sentences. The reason is the following: as I said at the beginning, in order to explain the notion of justification for (and therefore the meaning of) an atomic sentence, I appeal to the notion of authorization to use a name; if the explanation of this notion made reference to logically complex

sentences, the whole approach could not satisfy a molecularity requirement, since the explanation of the meaning of an atomic sentence would presuppose the understanding of the meaning of sentences of unlimited logical complexity. (ii) The atomic sentences expressing verbal information associated with n must contain occurrences of n; and non-verbal information associated with n must be explicitly associated with n. The reason for this requirement is the necessity of ensuring epistemic transparency for the formal notion of authorization: any subject who associates with n the epistemic content e must *know*[6] that he associates e with n.

I assume that the epistemic content associated with a name n or to a predicate P is articulated into a *fixed part* $\mathrm{fp}e_n$ or $\mathrm{fp}e_P$, which must be contained in every possible epistemic content associated with n or with P, from a *variable part* $\mathrm{vp}e_n$ or $\mathrm{vp}e_P$, which is not subject to this restriction. Intuitively, the information contained in $\mathrm{fp}e_n$ or $\mathrm{fp}e_P$ is the information without which a subject cannot be said to know the meaning of n or P, or to have semantic competence about them. For example, if n is a name for an ostensible object, it seems intuitively correct to require that $\mathrm{fp}e_n$ contain nothing more than a sortal indicating the sort of object n is intended to denote; if n is a name for a number, it seems necessary to require that $\mathrm{fp}e_n$ contain more: all that is necessary in order to identify the referent of n within the sequence of natural numbers; if P is the predicate "square", it seems necessary to require that $\mathrm{fp}e_P$ contain some sort of mental model of a square; if P is the predicate "kill" it seems necessary to require that $\mathrm{fp}e_P$ contains some 'meaning postulate' to the effect that if x killed y, then y is dead; and so on.[7] Which pieces of information are to be put into $\mathrm{fp}e_n$ or $\mathrm{fp}e_P$ is presumably an empirical question pertaining to lexical semantics; it seems plausible to me that information in $\mathrm{fp}e_n$ or $\mathrm{fp}e_P$ comes from the lexicon, while information in $\mathrm{vp}e_n$ or $\mathrm{vp}e_P$ comes from the belief system.

Imagine now that on a certain occasion a subject s meets a certain person, so that his visual apparatus generates a visual representation t_1, for example of a face: t_1 is an instance of a term of the internal representational system[8] activated on that occasion. When, on another occasion, s hears someone speaking of a common friend, s's memory activates some other representation t_2, which is another instance. When his teacher tells him: "Think of a number with such and such properties", s elaborates through attention another representation t_3. And so on. I assume that with the activated term t an identity criterion is associated, i.e., a function f_t such that,

for every term t' of the internal representational system, $f_t(t') = 1$ iff t' is a mental representation *of the same object* as t; so a class naturally corresponds to f_t: the class of terms t' such that $f_t(t') = 1$. The specific nature of such a criterion may be extremely variable; I simply assume that, in a given cognitive state, we associate to the activated term *some* function of this kind. Plausibly, each term belonging to the same class corresponding to f_t is experienced by the subject as a representation *of* the same entity, as a *mode of giving* it. In a nutshell, this is how I suggest to explain the notion of an object's being given to a subject.[9]

The need of auxiliary terms and supplementary information concerning them is illustrated by the following example. Suppose that s associates with "Chomsky" the epistemic content "Chomsky is a linguist" and "Chomsky teaches at M.I.T.", and imagine that on a certain occasion someone tells s something about *two* linguists t_1 and t_2 who teach at M.I.T.; clearly in this situation s is not intuitively authorized to use "Chomsky" to refer to both; nor is s intuitively authorized to use "Chomsky" to refer to either one of them. A cognitive state is defined by choosing one of the two terms t_1 and t_2 as the activated term, and the other as an auxiliary term activated in connection with "Chomsky".

Once a cognitive state $c = \langle e, \langle t, f_t \rangle, i \rangle$ is specified, the question "Does e_n authorize an idealized subject to use n to refer to the entity given by t, in the presence of i_n?" has a definite meaning, in the sense that the answer does not depend on other hidden features of the context. It depends on two questions: (i) (*the matching question*) whether there is an appropriate *matching* between the information contained in e_n and the information i_t associated with t;[10] and (ii) (*the uniqueness question*) whether, for every auxiliary term t' activated in connection with n and which matches e_n, the information $i_{t'}$ and the identity criterion f_t permit putting t' into the same class as t.[11] If the answers to the two questions are affirmative, any subject in the cognitive state c is authorized to use n to refer to the entity given by t, otherwise he is not.

We can therefore define the *meaning* of a name n as a function M_n such that, for every cognitive state $c = \langle e, \langle t, f_t \rangle, i \rangle$, $M_n(c) = 1$ iff the answer to both the matching question and to the uniqueness question is "Yes". An *authorization to use n to refer to the object given by t* is a cognitive state $c = \langle e, \langle t, f_t \rangle, i \rangle$ such that $M_n(c) = 1$.

Let us see now how the notion of denotation can be defined. Given the meaning M_n of a name n and a cognitive state $c = \langle e, \langle t, f_t \rangle, i \rangle$, the

following relation can be defined on the class IR of the terms of the internal representation system:

$$tRt' =_{\text{def}} M_n(\langle e, \langle t, f_t \rangle, i \rangle) = 1 \quad \text{iff}$$
$$M_n(\langle e, \langle t', f_{t'} \rangle, i \rangle) = 1;$$

obviously it is an equivalence relation, so it induces a partition on IR.[12] If t is an element of IR, I will call "$|t|_c$" its equivalence class.

When $M_n(\langle e, \langle t, f_t \rangle, i \rangle) = 1$, an idealized subject s in the state $c = \langle e, \langle t, f_t \rangle, i \rangle$ is intuitively authorized to use n to refer to $|t|_c$; so it is natural to identify $|t|_c$ with the denotation of n, in the state c. This is correct, in fact, but it cannot be the whole story. The reason is that it is possible either (1) that afterwards, i.e., in a subsequent cognitive state c', s realizes that in c he associated with n an intuitively *incorrect* epistemic content; or (2) that in a subsequent cognitive state c' s realizes that the uniqueness condition is not satisfied. When this happens, and s has reasons to believe that his new cognitive state is intuitively 'better' than c, then s is no longer authorized to use n to refer to $|t|_c$, and we can no longer say that n denotes $|t|_c$. It is therefore necessary to define the denotation of n relatively to cognitive states intuitively 'better' than c.

How is a correct cognitive state to be characterized? Let us consider an example; I will articulate the short story I am going to tell into different cognitive states of the subject s.[13] Suppose s associates with "Chomsky" the epistemic content e_1: "Chomsky is a postman in Brooklyn"; this is his initial cognitive state c_1. Later on (c_2), in a bookshop, he finds a book with the name "Noam Chomsky" on its cover; it is probable that at this point he glances through the book. Why? Because it is not common, but it is possible, that a postman in Brooklyn writes a book; in that case it is probable that it is an autobiography or something similar; turning the pages of the book, s makes a test, and the outcome is negative: the book deals with linguistics. Now s has several options, i.e., several possible explanations of the data at his disposal: (1) Chomsky is a postman who does linguistics in his spare time; (2) There are two persons named Noam Chomsky, a linguist and a postman; (3) The friend who told s that Chomsky is a postman in Brooklyn pulled s's leg; (4) the book s has in his hands is an April fool's joke; and so on. To make a choice s needs some selection criteria, and perhaps to acquire more information. Suppose that, after this work, he selects (3) and therefore associates with "Chomsky" the new epistemic content e_2: "Chomsky

is a linguist'' (c_3); at this point s can legitimately assert that e_1 was incorrect.

From this example, we can extract the following definitions:

DEFINITION 2.1. A cognitive state $c' = \langle e', \langle t', f_{t'} \rangle, i' \rangle$ is *better than* $c = \langle e, \langle t, f_t \rangle, i \rangle$ *with respect to n* (in symbols $c' \geq_n c$) iff the following condition (a) and one of the conditions (b) or (c) are satisfied:

(a) the amount I_n of information contained in i_n is part of the amount I'_n of information contained in i'_n;
(b) the amount E_n of information contained in e_n is part of the amount E'_n of information contained in e'_n;
(c) E_n is not part of E'_n, and the association of e'_n with n yields a better explanation of the data contained in i'_n than the association of e_n.

DEFINITION 2.2. The cognitive state c is *n-correct relative to the cognitive state* c' iff conditions (a) and (b) are satisfied. It is *n-incorrect relative to* c' iff conditions (a) and (c) are satisfied.

DEFINITION 2.3. The cognitive state $\langle e, \langle t, f_t \rangle, i \rangle$ is *n-correct* and *n-complete relative to the cognitive state* $\langle e', \langle t', f_{t'} \rangle, i' \rangle$ iff conditions (a) and (b) are satisfied and, for every auxiliary term t'' in i'_n which matches e'_n, the information present in i'_n permits putting t'' into the same equivalence class as t.

We can now define the notion of denotation:

DEFINITION 2.4. Given two cognitive states $c = \langle e, \langle t, f_t \rangle, i \rangle$ and $c' = \langle e', \langle t', f_{t'} \rangle, i' \rangle$, *n denotes the object o relative to* c' iff $o = |t|_c$ and c is n-correct and n-complete relative to c'.

Let us consider predicates. An important feature of their behavior is that if a subject is authorized to concatenate a predicate with a name, then he is authorized to concatenate it with any other name of the same object, provided he is authorized to believe that it is a name of the same object. If I am justified in asserting, for instance, that the boy in front of me is running, then I am thereby justified in asserting that Matthew is running, and that the elder son of my brother is running, provided I am justified in believing that the boy in front of me is Matthew, the elder son of my brother. In more elaborate terms, we might say that predication, the operation of concatenating a predicate with a name, has an implicit *modal* aspect, in the sense that we do not simply ask ourselves whether we are authorized to concatenate a predicate with a given name, but with any other name we

could use to refer to the same object. This seems to me the main reason why we cannot content ourselves with the notion of authorization to use a name to refer to a given entity, but we need the notion of *denotation* of a name.

As in the case of names, a subject's being or not being authorized to concatenate a predicate with a name depends essentially, although in a different measure,[14] on the epistemic content he associates with the predicate. For instance, if a subject associates with the predicate "*x* is square" the epistemic content "*x* has four equal sides", and the activated term is a mental representation of a book with four equal sides, then, since the activated term matches the epistemic content he associates with the predicate, there is shape recognition, and *s* is authorized to concatenate the predicate with that name; on the other hand, if the subject associates the same epistemic content with the predicate, and the activated term is a mental representation of a book with four sides such that only opposite sides are equal, then, since the activated term does not match the epistemic content associated with the predicate, there is no shape recognition, and *s* is not authorized to concatenate the predicate with that name.

But the intuitions we must account for are more complex. Consider for example a situation t_1 in which two subjects s_1 and s_2 sitting in positions p_1 and p_2, respectively, look at a round disk placed on a table. Through a suitable location of the subjects and in appropriate lighting conditions it is possible to make s_1 see the disk as round and s_2 see it as elliptical. Under the hypothesis that neither of the subjects can move, it seems quite legitimate to say that in the situation described, s_1 is intuitively authorized to concatenate the predicate "*x* is round" with the term "that disk", while s_2 is authorized to concatenate the predicate "*x* is elliptical" with the same term. Imagine now that at t_2 the two subjects switch positions, so that s_1 now sees the disk as elliptical and s_2 sees it as round. It is not intuitively legitimate to say that at t_2 both subjects are authorized to concatenate both the predicate "*x* is round" and the predicate "*x* is elliptical" with the term "that disk"; the subjects will probably be uncertain about the shape of the disk, and under normal conditions they will try to acquire new relevant information, for example by touching the disk, or by changing its position, etc. – a clear indication of the fact that *neither* of them is authorized to concatenate either of the two predicates with the term at t_2. It seems plausible to say that, in order to arrive at a cognitive state in which he is again authorized to con-

catenate one of the predicates with a name of that disk, each subject engages in a process whose goal is the selection of one representation of that disk, among the ones to which he has access through perception, memory, attention and so on, as *the best* one – the best from the point of view of its capacity to inform the subject about the properties of the disk. For instance, they will select the visual representation that is 'in accord' with the tactile representation; they will select the representation which, together with some general laws, permits them to account for the others; and so on.[15] From an abstract point of view, we might say that they select the representation which offers the best explanation of the relevant data.

If a cognitive state $c = \langle e, \langle t, f_t \rangle, i \rangle$ is specified, then, for every name n, the question "Does e_P authorize an idealized subject to concatenate P with n, in presence of i_P?" has a definite meaning, in the sense that the answer does not depend on other hidden features of the context. It depends on another question (the *matching question*): whether the hypothesis that there is an appropriate *matching* between the information contained in e_P and the information encoded into the $t' \in |n|_c$ that is selected as the best on the basis of the information contained in i_P is the best explanation of the data contained in e_P, i_P, e_n and i_n. If the answer to the matching question is affirmative, any subject in the cognitive state c is authorized to concatenate P with n, otherwise he is not. Since this has to hold for every name n, we can identify an *authorization to concatenate P with a name* with a cognitive state c in which a function f is accessible such that, for every name n, $f(n) = 1$ iff the answer to the matching question is "Yes". I will call f "the *concept* expressed by P in c". The *meaning* of a (unary) predicate P is therefore the function M_P such that, for every cognitive state c, $M_P(c)$ is the *concept* expressed by P in c.

On the basis of this definition, we can explain the fact that neither of the two subjects s_1 or s_2 of the preceding example is authorized to concatenate either of the two predicates "is round" and "is elliptical" with the term "that disk" at t_2 by saying that, at t_2, both subjects have access (through memory or perception) to both the representation of a round disk and to the representation of an elliptical disk, but neither representation can be selected by either subject as the best one.

In order to deal with predicates of variable arity, I shall henceforth re-interpret the definition of an atomic cognitive state given above in the following way: it is a triple $c = \langle e, \langle t, f_t \rangle, i \rangle$, where all is as before, save that t is a k-tuple $\langle t_1, \ldots, t_k \rangle$ of activated terms. An analogous generalization of the matching question for predicates yields the

definition of authorization to concatenate a k-ary predicate P with a k-tuple of names.

Let me illustrate this generalization by considering the binary predicate "=". It seems reasonable to require that $f_p e_=$ contains the piece of information "$x = y$ iff x and y are the same thing";[16] so, when a cognitive state $c = \langle e, \langle t, f_t \rangle, i \rangle$ is specified, then, for every pair of names $\langle n_1, n_2 \rangle$, the question of whether a subject s is authorized to concatenate "=" with $\langle n_1, n_2 \rangle$ boils down to the question of whether s is authorized to use n_1 and n_2 to refer to the same equivalence class of terms. Suppose that n_1 is "Hesperus", n_2 is "Phosphorus", that e_{n_1} contains the piece of information "Hesperus is the first celestial body shining in the evening sky", and that e_{n_2} contains "Phosphorus is the last celestial body shining in the morning sky". According to our general formulation of the matching question, we must perform a (mental or real) experiment: we must look for the 'best' terms (not necessarily activated in c) belonging to $|\text{Hesperus}|_c$ and $|\text{Phosphorus}|_c$ and ask whether the hypothesis that they belong to the same equivalence class of terms is the best explanation of the data available in c.

Of course, for predicates as well as for names, it is possible that the epistemic content associated with P is intuitively incorrect. Therefore, for reasons analogous to the ones explained above concerning names, the notion of extension of a predicate at c must be relativized to cognitive states. So we define the notions "$c' \geq_P c$"and "c is *P-correct relative to* c'" in a way perfectly analogous to our definitions of "$c' \geq_n c$" and "c is n-correct relative to c'", and then we can propose the following

DEFINITION 2.5. Given two cognitive states c and c', $M_P(c)$ is a *satisfaction-ground of P in c, relative to* c', iff c is P-correct relative to c'.

We can now define the notion of justification for atomic sentences. A *justification* for a sentence of the form "$P(n_1, \ldots, n_k)$" is a cognitive state c such that $M_P(c)(\langle n_1, \ldots, n_k \rangle) = 1$, i.e., intuitively, a cognitive state in which an idealized subject *is* authorized to concatenate P with $n_1, \ldots, n_k$. The *meaning* of "$P(n_1, \ldots, n_k)$" can therefore be identified with $\lambda c.\, M_P(c)(\langle n_1, \ldots, n_k \rangle)$, i.e., intuitively, with a *classification criterion* of cognitive states, associating 1 with c iff an idealized subject in c is justified to believe "$P(n_1, \ldots, n_k)$".

Consider a justification j for $P(n_1, \ldots, n_k)$, and a cognitive state c': j is a *truth-ground of* $P(n_1, \ldots, n_k)$*, relative to* c', iff j is n_1-n_k-complete and P, n_1-n_k-correct, relative to c'.

3. THE LOGICAL CONSTANTS AND CANONICAL JUSTIFICATIONS

The introduction of the logical constants amounts to the assumption that the class of cognitive states is closed under certain operations. We will see now which ones.

I will assume without discussion that $f_p e_\wedge$ contains the piece of information that a justification for $A \wedge B$ is a pair $\langle j_1, j_2 \rangle$, where j_1 is a justification for A and j_2 a justification for B; so, when a cognitive state c is specified, then, for every pair of sentences $\langle A, B \rangle$, the question as to whether a subject s is justified in believing $A \wedge B$ boils down to the question of whether s has access both to a cognitive state j_1 which is a justification for A and to a cognitive state j_2 which is a justification for B; a justification for $A \wedge B$ can therefore be identified with the pair $\langle j_1, j_2 \rangle$.

Analogously, a justification for $A \vee B$ can be identified with either a justification for A or a justification for B; and a justification for $\exists x A$ with a pair $\langle n, j \rangle$, where n is a name, and j a justification for $A[n/x]$.

Let us consider implication. Imagine the following situation: John knows that the phone numbers in Milan have been changed according to the following rule: dial 48 plus the result of adding 53 to the old number. He wants to call Charles, a friend of his who lives in Milan; it seems to him that his old number was 341951, but he is not sure. In this situation John has neither a justification for the sentence

(5) Charles' old number was 341951

nor a justification for the sentence

(6) Charles' current number is 48342004;

but he does have, and knows that he has, a justification for the conditional

(7) If Charles' old number was 341951, then his current number is 48342004.

What, exactly, does his justification consist of? Of his knowledge of the rule described above. We can therefore suggest that $f_p e_\supset$ contains the piece of information that a justification for $A \supset B$ is a general method m that is recognized as transforming every justification j for A into a justification $m(j)$ for B. As a consequence, when a cognitive state c is specified, then, for every pair of sentences $\langle A, B \rangle$, the question of whether a subject s is justified in believing $A \supset B$ boils down to the question of whether s has access to a cognitive state in

which he knows such a general method; a justification for $A \supset B$ can therefore be identified with the method m. Analogously, a justification for $\forall xA$ can be identified with a method m that is recognized as associating with each name n a justification for $A[n/x]$.

These definitions may seem inadequate for (at least) two reasons. First, I have required that justifications be epistemically transparent, and one may wonder whether such general methods as the ones introduced by the definitions *are* epistemically transparent. But it should be kept in mind that we are making reference to an idealized subject, i.e., a subject having no limits of memory, attention, and so on: a subject whose cognitive capacities and performances in any given occasion are taken as representative of the ones of an arbitrary member of the same species. Well: if such an idealized subject is not able, when acquainted with m, to recognize it as a method with such and such characteristics, the sole conclusion it is natural to draw is that m is *not* such a method. How could it be a method with such and such characteristics if *nobody* were capable of acknowledging that it is? Of course, it is possible that *I* am not capable of realizing that something is such a method, because of the limits of my IQ, memory, attention, and so on; but these are precisely the factors from which we abstract when we make reference to an idealized subject.

The second reason for perplexity may come from having noticed that, on the one hand, I have said that justifications for empirical sentences are non-conclusive, on the other hand, I have suggested to define justifications for $A \supset B$ and for $\forall xA$ in the same way *proofs* of the same sentences are defined, where proofs of mathematical sentences are indefeasible; in other words, how might a method associating with each name n a justification for $A[n/x]$ be non-conclusive? My provisory answer[17] is very simple: remember that non-conclusiveness is the logical disjunction of defeasibility and non-factivity; although it must be conceded that a justification for $\forall xA$, as I have defined it, is indefeasible, it is equally clear that it is not factive: a justification for $A[n/x]$, for some n, might not be a sufficient condition of the truth of $A[n/x]$; of course, the notion of truth involved here needs further clarification.[18]

Let us consider negation. The first, obvious, idea is to apply the intuitionistic definition of the negation of a mathematical sentence to empirical sentences, and to define a justification for the negation of A as a general method transforming every alleged justification for A

into a justification for the absurdity $\perp$, where $\perp$ is defined as the sentence for which there is no justification.

Indeed, in the case of many empirical sentences, their negations cannot be construed except as intuitionistic. Consider for instance the sentence

(8) Not all prehistoric men were black-eyed;

probably we will never be able to say, of a specific prehistoric man, that he was black-eyed; at the same time, it is quite plausible to say that we have justifications to believe that (8) is true; and if we reflect on the nature of these justifications, we realize that each of them can be verbalized as a *reductio ad absurdum* of the assumption that all prehistoric men were black-eyed, as the intuitionistic explanation of negation requires.

But there are important classes of empirical sentences whose negations cannot plausibly be conceived as intuitionistic. Let us return to the example of the two subjects looking at a round disk from different positions; at t_1 s_1, who sees the disk as round, has an obvious intuitive justification to believe the sentence

(9) That disk is not elliptical.

Suppose now that we define a justification for the negation of A as a justification for $A \supset \perp$, and ask ourselves whether, with this definition, we can allow s_1 to have a justification for (9). In order to have a justification for (9) $\supset \perp$, s_1 should know a general method m he recognizes as transforming every justification j for

(10) That disk is elliptical

into a justification for $\perp$. We know that a justification for (10) is a cognitive state c such that $M_{\text{elliptical}}(c)(\text{that disk}) = 1$. Imagine now that at t_2 the two subjects switch positions, so that s_1 now sees the disk as elliptical and s_2 sees it as round; if s_1 had known m at t_1, at t_2 he should be able to transform c into a justification for $\perp$; but this is just what s_1 is *not* able to do at t_2: what does happen at t_2 is that s_1 feels himself no longer justified in believing that the disk is round, precisely because he considers c as a possible justification for the belief that the disk is elliptical. The answer to our question is therefore negative.

The preceding remarks suggest that the negation of an empirical sentence is in many cases an operation that differs greatly from implication. Another suggestion in the same direction comes from observing that, in empirical contexts, a very natural way of justifying the negation of a sentence is to exhibit a counterexample to the sentence. For instance, the most natural way to justify "Not all men are good" is to

present a bad man; by the way, to derive a contradiction from the assumption that all men are good would be much more complicated, in spite of the fact that the justified proposition would be, in a sense, weaker. Analogously, the most natural way to justify "It is not the case that John is away and Mary is at home" is to justify either "John is not away" or "Mary is not at home".

The last example shows another feature of empirical negated sentences to which intuitionistic negation is not faithful. Intuitively, a justification j for "It is not the case that John is away and Mary is at home" is as defeasible as a justification j' for "John is away and Mary is at home"; but if we construe the negated conjunction intuitionistically, it becomes a case of implication, and justifications for implications are normally indefeasible.

My suggestion of counterexamples as justifications for negated sentences might give the impression of an implicit appeal to the classical, or realist, meaning of negation, since it depends on a tacit use of de Morgan's laws and similar laws for quantifiers. But this impression is false. Classical negation is not the sole negation satisfying those laws: it is – I conjecture – the sole *explicitly definable* negation satisfying those laws; but if we reject the idea that negation must be explicitly definable, and embrace the idea of an inductively defined operation, we can look for a constructive negation. In fact, a very good candidate is already at hand: Nelson's constructible falsity.[19] I will give below the inductive clauses.

As concerns atomic sentences, let us come back to the two subjects s_1 and s_2. I have said that at t_1 s_1, who sees the disk as round, has an obvious intuitive justification for sentence (9); what does this justification consist of? One might be tempted to answer that it consists of two things: the justification j_1 he has to believe

(11) That disk is round,

and the justification j_2 he has to believe that being round and being elliptical are two *incompatible* properties of physical objects: it is because he knows that being round is incompatible with being elliptical that he may legitimately choose the former property as a token for the absence of the latter.[20] Notice that if we defined a justification for (14) in this way, we would not run into the difficulty that a justification for (14) would simultaneously be also a justification for

(12) That disk is not square

and for several other sentences which, intuitively, are not synonymous to (9); for j_2, whatever it is, will be different from a justifi-

cation for believing that being *square* and being elliptical are incompatible. But we would run into another difficulty: the choice of j_1 as the first component of a justification for (9) is not more motivated than the choice of a justification for any other sentence of the form "That disk is P", where P is a property incompatible with being elliptical; and, since there seems to be no way of regimenting the class of the properties incompatible with being elliptical, the choice of that class as the first component would entail the loss of the decidability of the relation "x is a justification for A" when A is a negated predication.

Consider the cognitive state $c = \langle e, \langle t, f_t \rangle, i \rangle$, where $e_{\text{elliptical}}$ contains a stored model of an ellipse, $t_{\text{that disk}}$ is a newly derived mental representation of a round disk and $i_{\text{that disk}}$ is empty; of course $M_{\text{elliptical}}(c)(\text{that disk}) = 0$, since there is no matching between $e_{\text{elliptical}}$ and the information encoded into the term $t_{\text{that disk}}$ which, since s_1 has not yet seen the disk from position p_2, is selected as the best among the terms belonging to $|\text{that disk}|_c$. My suggestion is simply to take c as a justification for (9) and, in general, to define a justification for $\neg P(n_1, \ldots, n_k)$ as a cognitive state c such that $M_P(c)(\langle n_1, \ldots, n_k \rangle) = 0$.

Finally, I will define a *canonical* justification for a sentence A as a justification for A such that in order to establish that it is a justification for A, only information from $f_p e_E$, for every constituent expression E of A, is needed. Notice that, according to this definition, the following argument for $A \wedge B$

$$(13)\quad \frac{\begin{array}{cc}\Pi_1 & \Pi_2 \\ C \supset A \wedge B & C\end{array}}{A \wedge B}$$

is a justification for $A \wedge B$, since from it a justification for A and a justification for B can be extracted; but it is not a canonical justification for $A \wedge B$, since in order to establish that it is such a justification one needs more than information from $f_p e_\wedge$ and from the epistemic contents associated with the constituents of A and B.

4. TRUTH-GROUNDS AND CONSTRUCTIVE VALIDITY

To conclude, let us see how the ideas illustrated above can give rise to a definition of constructive validity.

Given a first-order language $\boldsymbol{L}$ with a set N of names and identity predicate, a *cognitive structure* for $\boldsymbol{L}$ is a pair $\boldsymbol{C} = \langle C, M\rangle$, where C is a (nonempty) set of temporally ordered cognitive states $c = \langle e, \langle t, f_t\rangle, i\rangle$ and M a meaning-assignment, i.e., a function such that

- for every name n, M_n is a function such that, for every $c \in C$, $M_n(c) \in \{0, 1\}$;
- for every predicate P^k, M_{P^k} is a function such that, for every $c \in C$, $M_{P^k}(c)$ is a function such that, for every $n_1 \ldots n_k \in C$, $M_{P^k}(c)(\langle n_1 \ldots n_k\rangle) \in \{0, 1\}$.

Given a cognitive structure $\boldsymbol{C}$, and an atomic sentence $P^k(n_1, \ldots, n_k)$, the sets $\{c | M_{P^k}(c)(\langle n_1 \ldots n_k\rangle) = 1\}$ and $\{c | M_{P^k}(c)(\langle n_1 \ldots n_k\rangle) = 0\}$ are determined; in other words, for every atomic sentence A, the set $\boldsymbol{J}_{\boldsymbol{C}}(A)$ of its $\boldsymbol{C}$-justifications and the set $\boldsymbol{J}_{\boldsymbol{C}}(\neg A)$ of $\boldsymbol{C}$-justifications for its negation are determined. The set of the $\boldsymbol{C}$-justifications of every sentence can be inductively defined in the way sketched in Section 3; let me sum up the inductive clauses:

1. $\boldsymbol{J}_{\boldsymbol{C}}(\bot) = \emptyset$
2. $\boldsymbol{J}_{\boldsymbol{C}}(B \wedge C) = \boldsymbol{J}_{\boldsymbol{C}}(B) \times F_{\boldsymbol{J}}(C)$
   $\boldsymbol{J}_{\boldsymbol{C}}(\neg(B \wedge C)) = \boldsymbol{J}_{\boldsymbol{C}}(\neg B) \cup \boldsymbol{J}_{\boldsymbol{C}}(\neg C)$
3. $\boldsymbol{J}_{\boldsymbol{C}}(B \vee C) = \boldsymbol{J}_{\boldsymbol{C}}(B) \cup \boldsymbol{J}_{\boldsymbol{C}}(C)$
   $\boldsymbol{J}_{\boldsymbol{C}}(\neg(B \vee C)) = \boldsymbol{J}_{\boldsymbol{C}}(\neg B) \times \boldsymbol{J}_{\boldsymbol{C}}(\neg C)$
4. $\boldsymbol{J}_{\boldsymbol{C}}(B \supset C) = \boldsymbol{J}_{\boldsymbol{C}}(C)^{\boldsymbol{J}_{\boldsymbol{C}}(B)}$
   $\boldsymbol{J}_{\boldsymbol{C}}(\neg(B \supset C)) = \boldsymbol{J}_{\boldsymbol{C}}(B) \times \boldsymbol{J}_{\boldsymbol{C}}(\neg C)$
5. $\boldsymbol{J}_{\boldsymbol{C}}(\forall x B) = \prod_{n \in N} \boldsymbol{J}_{\boldsymbol{C}}(B[n/x])$
   $\boldsymbol{J}_{\boldsymbol{C}}(\neg(\forall x B)) = \sum_{n \in N} \boldsymbol{J}_{\boldsymbol{C}}(\neg B[n/x])$
6. $\boldsymbol{J}_{\boldsymbol{C}}(\exists x B) = \sum_{n \in N} \boldsymbol{J}_{\boldsymbol{C}}(B[n/x])$
   $\boldsymbol{J}_{\boldsymbol{C}}(\neg(\exists x B)) = \prod_{n \in N} \boldsymbol{J}_{\boldsymbol{C}}(\neg B[n/x])$

We must now define the notion "j is a $\boldsymbol{C}$-truth-ground of A, relative to c" (in symbols $j \models_c^{\boldsymbol{C}} A$).

1. If A is $P^k(n_1, \ldots, n_k)$ or $\neg P^k(n_1, \ldots, n_k)$, and $j \in \boldsymbol{J}_{\boldsymbol{C}}(A)$, then $j \models_c^{\boldsymbol{C}} A$, iff j is n_1-n_k-complete and P, n_1-n_k-correct, relative to c.
2. If A is $B \wedge C$, and $j = \langle j_1, j_2\rangle \in \boldsymbol{J}_{\boldsymbol{C}}(A)$, then $j \models_c^{\boldsymbol{C}} A$ iff $j_1 \models_c^{\boldsymbol{C}} B$ and $j_2 \models_c^{\boldsymbol{C}} C$.

If A is $\neg(B \wedge C)$, and $j \in \boldsymbol{J}_{\boldsymbol{C}}(A)$, then: if $j = \langle j_1, 0\rangle$, $j \models_c^{\boldsymbol{C}} A$ iff $j_1 \models_c^{\boldsymbol{C}} \neg B$; if $j = \langle j_2, 1\rangle$, $j \models_c^{\boldsymbol{C}} A$ iff $j_2 \models_c^{\boldsymbol{C}} \neg C$.

3. If A is $B \vee C$, and $j \in \boldsymbol{J}_{\boldsymbol{C}}(A)$, then: if $j = \langle j_1, 0\rangle$, $j \models_c^{\boldsymbol{C}} A$ iff $j_1 \models_c^{\boldsymbol{C}} B$; if $j = \langle j_2, 1\rangle$, $j \models_c^{\boldsymbol{C}} A$ iff $j_2 \models_c^{\boldsymbol{C}} C$.
   If A is $\neg(B \vee C)$, and $j = \langle j_1, j_2\rangle \in \boldsymbol{J}_{\boldsymbol{C}}(A)$, then $j \models_c^{\boldsymbol{C}} A$ iff $j_1 \models_c^{\boldsymbol{C}} \neg B$ and $j_2 \models_c^{\boldsymbol{C}} \neg C$.
4. If A is $B \supset C$, and $j \in \boldsymbol{J}_{\boldsymbol{C}}(A)$, then $j \models_c^{\boldsymbol{C}} A$ iff, for every $j' \in \boldsymbol{J}_{\boldsymbol{C}}(B)$, if $j' \models_c^{\boldsymbol{C}} B$, then $j(j') \models_c^{\boldsymbol{C}} C$.
   If A is $\neg(B \supset C)$, and $j = \langle j_1, j_2\rangle \in \boldsymbol{J}_{\boldsymbol{C}}(A)$, then $j \models_c^{\boldsymbol{C}} A$ iff $j_1 \models_c^{\boldsymbol{C}} B$ and $j_2 \models_c^{\boldsymbol{C}} \neg C$.
5. If A is $\forall x B$, and $j \in \boldsymbol{J}_{\boldsymbol{C}}(A)$, then $j \models_c^{\boldsymbol{C}} A$ iff, for every $n \in N$, $j(n) \models_c^{\boldsymbol{C}} B[n/x]$. If A is $\neg(\forall x B)$, and $j = \langle n, j'\rangle \in \boldsymbol{J}_{\boldsymbol{C}}(A)$, then $j \models_c^{\boldsymbol{C}} A$ iff $j' \models_c^{\boldsymbol{C}} [1]\neg B[n/x]$.
6. If A is $\exists x B$, and $j = \langle n, j'\rangle \in \boldsymbol{J}_{\boldsymbol{C}}(A)$, then $j \models_c^{\boldsymbol{C}} A$ iff $j' \models_c^{\boldsymbol{C}} B[n/x]$. If A is $\neg(\exists x B)$, and $j \in \boldsymbol{J}_{\boldsymbol{C}}(A)$, then $j \models_c^{\boldsymbol{C}} A$ iff, for every $n \in N$, $j(n) \models_c^{\boldsymbol{C}} \neg B[n/x]$.

Given a cognitive structure $\boldsymbol{C} = \langle C, M\rangle$, a sentence A is *$\boldsymbol{C}$-constructively valid* iff there is a $j \in \boldsymbol{J}_{\boldsymbol{C}}(A)$ such that, for every $c \in C$, $j \models_c^{\boldsymbol{C}} A$.

A sentence A is *constructively valid* iff, for every cognitive structure $\boldsymbol{C}$, A is $\boldsymbol{C}$-constructively valid.

We can also define the notion "A is $\boldsymbol{C}$-true relative to c" (in symbols $\models_c^{\boldsymbol{C}} A$), in the obvious way: $\models_c^{\boldsymbol{C}} A$ iff there is a j such that $j \models_c^{\boldsymbol{C}} A$. Notice that this notion of truth distributes over $\vee$; for $\models_c^{\boldsymbol{C}} (B \vee C)$ iff there is a j such that $j \models_c^{\boldsymbol{C}} B \vee C$; then either $j = \langle j_1, 0\rangle$, and $j_1 \models_c^{\boldsymbol{C}} B$, and therefore there is a j such that $j \models_c^{\boldsymbol{C}} B$, i.e. $\models_c^{\boldsymbol{C}} B$; or $j = \langle j_2, 1\rangle$, and therefore there is a j such that $j \models_c^{\boldsymbol{C}} C$, i.e. $\models_c^{\boldsymbol{C}} C$.

NOTES

[1] See Chomsky (1980, pp. 51 ff.) for an illustration of such a position.

[2] I am not implying that Dummett's argument is valid; I am only saying that the modified argument is not less compelling than the original one for someone who finds the specifiability requirement more acceptable than the manifestability requirement. For a more detailed justification of this thesis see Usberti (1995, IV.2).

[3] I am presupposing here that the existence of a justification is (anti-realistically) a temporal notion, in the sense that a justification for A may exist at t and not exist before or after t. Of course a constructivist might tend to embrace an atemporal notion of existence, and therefore of truth; in this case the reason explained in the text should be restated by saying that for some sentence A it might happen that there is a justification for both A and $\neg A$.

[4] First, I will concentrate on unary predicates.

[5] It is not necessary here to enter into a more detailed analysis of the nature of such mental representations. Some interesting suggestions may be found in Marr (1982).
[6] Not necessarily consciously.
[7] The difference emphasized in footnote 14 below could be expressed by saying that the fixed part of the epistemic content associated with a name – i.e. the amount of information a language user must associate to the name in order to be credited with knowledge of its meaning – is minimal, while the fixed part of the epistemic content associated with an adjectival predicate and to a verb is maximal.
[8] As a matter of fact, there will be several representational systems: visual, auditory, and so on; and also memory and attention may be considered, for our present purposes, as such systems. For simplicity I consider them as one.
[9] For a more detailed account see Usberti (2002).
[10] Of course, the specific nature of this matching may vary according to the nature of information to be matched, and the devices that verify it may be very different from one another.
[11] The relevant notion of equivalence class is defined two paragraphs down.
[12] The classes belonging to this partition should not be confused with the ones corresponding to the identity criteria associated with activated terms: two terms of the internal representational system may be in the same class independently of their matching the epistemic content associated with any name.
[13] Of course, a much more fine-grained analysis would be possible; but the one suggested in the text is sufficient to explain my point.
[14] There is a clear intuitive difference between proper names, on the one hand, and adjectival predicates and verbs, on the other hand: while we should not say that someone who does not know that Chomsky is a linguist ignores the sense of "Chomsky", we do say that someone who does not know that a square field has four equal sides ignores the meaning of "square", and that someone who does not know that whoever has been killed is dead ignores the meaning of "kill".
[15] For an account along these lines of our construction of empirical reality see Musatti (1926).
[16] Of course we are not forced to make this choice; if the epistemic content associated with "=" were not the same in every cognitive state, then $M_{=}(c)$ would simply vary according to c.
[17] Provisory in the sense that a thorough analysis of empirical implications should be much more fine-grained and distinguish between several types of implication.
[18] On this point cp. Usberti (1995).
[19] Cp. Nelson (1949).
[20] "[I]t might be maintained that every significant observation must be an observation of some property, and further that the absence of a property P if it may be established empirically at all, must be established by the observation of (another) property N which is taken as a token for the absence of P." (Nelson 1959, p. 208.).

REFERENCES

Chomsky, N.: 1980, *Rules and Representations*, Columbia University Press, New York.
Dummett, M.: 1973, *Frege. Philosophy of Language*, Duckworth, London.

Dummett, M.: 1975, 'The Philosophical Basis of Intuitionistic Logic', in H. E. Rose and J. C. Sheperdson (eds.), *Logic Colloquium '73*, North Holland, Amsterdam, pp. 5–40.
Dummett, M.: 1977, *Elements of Intuitionism*, Clarendon Press, Oxford.
Marr, D.: 1982, *Vision*, W. H. Freeman and Company, New York.
Musatti, C.: 1926, 'Analisi del Concetto di Realtà Empirica', now in C. Musatti, *Condizioni dell'esperienza e Fondazione Della Psicologia*, Giunti-Barbera, Firenze, 1964, pp. 13–175.
Nelson, D.: 1949, 'Constructible Falsity', *The Journal of Symbolic Logic* **14**, 16–26.
Nelson, D.: 1959, 'Negation and Separation of Concepts in Constructive Systems', in A. Heyting (ed.), *Constructivity in Mathematics*, North-Holland, Amsterdam, pp. 208–225.
Prawitz, D.: 1977, 'Meaning and Proofs: On the Conflict Between Classical and Intuitionistic Logic', *Theoria* **43**, 2–40.
Prawitz, D.: 1987, 'Dummett on a Theory of Meaning and Its Impact on Logic', in B. Taylor (ed.), *Michael Dummett: Contributions to Philosophy*, The Hague, Nijhoff, pp. 117–165.
Usberti, G.: 1995, *Significto e Conoscenza*, Guerini e Associati, Milano.
Usberti, G.: 2002, 'Names: Sense, Epistemic Content, and Denotation', *Topoi* **21**, 99–112.

Dipartimento di Filosofia e Scienze sociali
via Roma 47
53100 Siena
Italia
E-mail: usberti@unisi.it

GRIGORI MINTS

NOTES ON CONSTRUCTIVE NEGATION

ABSTRACT. We put together several observations on constructive negation. First, Russell anticipated intuitionistic logic by clearly distinguishing propositional principles implying the law of the excluded middle from remaining valid principles. He stated what was later called Peirce's law. This is important in connection with the method used later by Heyting for developing his axiomatization of intuitionistic logic. Second, a work by Dragalin and his students provides easy embeddings of classical arithmetic and analysis into intuitionistic negationless systems. In the last section, we present in some detail a stepwise construction of negation which essentially concluded the formation of the logical base of the Russian constructivist school. Markov's own proof of Markov's principle (different from later proofs by Friedman and Dragalin) is described.

1. INTRODUCTION

We put together several little-known observations on constructive negation. Section 2 contains a description of a passage in (Russell 1903) where Russell anticipated intuitionistic logic by clearly distinguishing propositional principles implying the law of the excluded middle from remaining valid principles. In fact, he states what was later called Peirce's law. This is important in connection with the method used later by Heyting (see Troelstra 1990, Section 3.1) for developing his axiomatization of intuitionistic logic.

Section 3 presents some little-known Russian work on negationless mathematics. It turns out that classical arithmetic and analysis can be embedded into intuitionistic negationless systems. Section 4 presents in some detail a stepwise construction of negation which essentially concluded the formation of the logical base of the Russian constructivist school.

2. RUSSELL'S ANTICIPATION OF INTUITIONISTIC LOGIC

In Section 18 (Chapter II) of (Russell 1903), Russell lists 10 propositional axioms. The first nine (Section 2.1 below) are intuitionisti-

Synthese (2006) 148: 701–717
DOI 10.1007/s11229-004-6294-3

cally valid. The next requires some comments. Its statement takes into account Russell's definition of a proposition in Section 16 of (Russell 1903): "To say 'p is a proposition' is equivalent to saying 'p implies p' ". We quote from Section 18:

(10) If p implies p and q implies q then " 'p implies q' implies p" implies p. This is called the principle of *reduction*; it has less self-evidence than the previous principles, but is equivalent to many propositions that are self-evident. I prefer it to these, because it is explicitly concerned, like its predecessors, with implication, and has the same kind of logical character as they have. If we remember that "p implies q" is equivalent to "q or not-p," we can easily convince ourselves that the above principle is true; for " 'p implies q' implies p" is equivalent to "p or the denial of 'q or not-p,'" i.e. to "p or 'p and not-q,'" i.e. to p. But this way of persuading ourselves that the principle of reduction is true involves many logical principles which have not yet been demonstrated, and cannot be demonstrated except by reduction or something equivalent. The principle is especially useful in connection with negation. Without its help, by means of the first nine principles, we can prove the law of contradiction; we can prove, if p and q be propositions, that p implies not-not-p; that "p implies not-q" is equivalent to "q implies not-p" and to not-pq; that "p implies q" implies "not-q implies not-p"; that p implies that not-p implies p; that not-p is equivalent to "p implies not-p"; and that "p implies not-q" is equivalent to "not-not-p implies not-q." But we cannot prove without reduction or some equivalent (so far at least as I have been able to discover) that p or not-p must be true (the law of excluded middle); that every proposition is equivalent to the negation of some other proposition; that not-not-p implies p; that "not-q implies not-p" implies "p implies q"; that "not p implies p" implies p, or that "p implies q" implies "q or not-p." Each of these assumptions is equivalent to the principle of reduction, and may, if we choose, be substituted for it. Some of them – especially excluded middle and double negation – appear to have far more self-evidence. But when we have seen how to define disjunction and negation in terms of implication, we shall see that the supposed simplicity vanishes, and that, for formal purposes at any rate, reduction is simpler than any of the possible alternatives. For this reason I retain it among my premisses in preference to more usual and more superficially obvious propositions.

Let us repeat the principles which Russell claims follow from the first nine principles (conjunction of p and q is expressed here by pq):

$$\neg(p(\neg p)) \quad ; p \to \neg\neg p; \quad (p \to \neg q) \iff (q \to \neg p)$$

$$(p \to \neg q) \iff \neg(pq); \quad (p \to q) \to (\neg q \to \neg p); \quad p \to (\neg p \to p)$$

$$\neg p \iff (p \to \neg p); \quad (p \to \neg q) \iff (\neg\neg p \to \neg q).$$

All these principles are provable in intuitionistic and even in minimal logic.

Next we list the principles Russell claims cannot be proved without the reduction principle. None of these principles is provable intuitionistically, since each of them has an instance which

intuitionistically implies $p \vee \neg p$. Such an instantiation is given in square brackets below:

$p \vee \neg p$	
$\exists q(p \iff \neg q)$	$[p/p \vee \neg p]$
$\neg\neg p \rightarrow p$	$[p/p \vee \neg p]$
$(\neg q \rightarrow \neg p) \rightarrow (p \rightarrow q)$	$[p/\top;\ q/q \vee \neg q]$
$(\neg p \rightarrow p) \rightarrow p$	$[p/p \vee \neg p]$
$(p \rightarrow q) \rightarrow (q \vee \neg p)$	$[q/p]$

So Russell was right in both the positive and negative part. This is especially interesting in view of the evidence (van Dalen (1999)) that Brouwer was studying Russell's work at the time when the principles of intuitionism were first developed.

2.1. *Russell's Original Propositional Principles*

(1) $(p \rightarrow q) \rightarrow (p \rightarrow q)$

(2) $(p \rightarrow q) \rightarrow (p \rightarrow p)$

(3) $(p \rightarrow q) \rightarrow (q \rightarrow q)$

(4) A true hypothesis in an implication may be dropped, and the consequent asserted

(5) $(p \rightarrow p)$ and $(q \rightarrow q) \rightarrow (pq \rightarrow p)$

(6) $(p \rightarrow q)$ and $(q \rightarrow r) \rightarrow (p \rightarrow r)$

(7) $(q \rightarrow q)$ and $(r \rightarrow r)$ and $(p \rightarrow (q \rightarrow r)) \rightarrow (pq \rightarrow r)$

(8) $(p \rightarrow p)$ and $(q \rightarrow q) \rightarrow (pq \rightarrow r) \rightarrow (p \rightarrow (q \rightarrow r))$

(9) $(p \rightarrow q)$ and $(p \rightarrow r) \rightarrow (p \rightarrow qr)$

3. NEGATIONLESS MATHEMATICS

Philosophical difficulties connected with the use of negation and related notions such as the empty set were noticed very early, and stressed in modern time by Griss (1955), who suggested constructing intuitionistic mathematics without negation. To avoid trivial solutions like smuggling negation $\neg A$ back as $A \rightarrow 0 = 1$ it was proposed to eliminate implication in favor of a new connective $A \rightarrow_{\mathbf{x}} B$ where $\mathbf{x} \equiv x_1 \ldots x_n$. Formula $A \rightarrow_{\mathbf{x}} B$ is to be understood as $\forall \mathbf{x}(A \rightarrow B)\ \&\ \exists \mathbf{x} A$. In other

words, implication is allowed only if its premise is non-void. Positive suggestions by Griss caused some doubts: it was not clear how to develop significant parts of mathematics in a negationless way. Several systems of negationless logic and arithmetic were proposed (see Nelson 1966), but their strength was not obvious. It turns out they have the same strength as the corresponding intuitionistic systems. The simplest proof was given recently by Krivtsov (2000a). It uses the fact that traditional implication

$$A \to B$$

is intuitionistically equivalent to

$$\forall x((A \vee x = 0) \to (B \vee x = 0)).$$

The latter implication can be written as $(A \vee x = 0) \to_{\mathbf{x}} (B \vee x = 0)$ and has a non-void premise since x can be instantiated by 0. The actual modeling of intuitionistic arithmetic, analysis and higher type systems in negationless terms given in (Krivtsov 2000a,b) is slightly more complicated.

An earlier work (Mezhlumbekova 1975) by V. Mezhlumbekova used similar ideas in a more mathematical setting. It is contained in her Ph.D. thesis advised by Dragalin. Consider the Dialectica interpretation (cf. Avigad and Feferman 1998) of first order intuitionistic arithmetic HA into the language of functionals of finite types. It transforms an arbitrary HA-formula A into a formula $\exists \mathbf{y} \forall \mathbf{x} \phi$, where $\mathbf{x}, \mathbf{y}$ are finite sequences of variables of finite types, and ϕ is a quantifier-free formula. We assume that ϕ has a form $t = 0$ where t is a term in the language of primitive recursive functionals of finite type. Every quantifier free formula ϕ can be put into a form $t = 0$ by familiar transformations of primitive recursive formulas. Let ϕ^* be obtained from ϕ by "untangling" higher order application using Kleene's normal form for partial recursive functions:

$$\psi[Ap(t,s)] := \exists z(T_1(t,s,z) \ \& \ \psi[Uz]).$$

Recall that the hereditarily recursive operations (HRO) of type 0 are natural numbers for $\theta = 0$. HRO of type $\sigma \to \tau$ are Gödelnumbers of partial recursive functions that define total functionals from HRO of type σ to HRO of type τ. Let V_σ denote the definition of HRO of type σ with $V_0(x) := (x = x)$.

For an arbitrary formula A of HA having Dialectica interpretation $\exists \mathbf{y}^\sigma \forall \mathbf{x}^\tau \phi$, consider a negationless formula of first-order arithmetic

$$\exists \mathbf{y}(\mathbf{y} \in V_\sigma \ \& \ (\mathbf{x} \in V_\tau \to_{\mathbf{x}} \phi^*)).$$

All sets V_σ are obviously non-void. Mezhlumbekova sketches a proof of the following statement:

If HA $\vdash A$, then for some natural numbers $\mathbf{k}$ it is provable in a weak system of negationless arithmetic HAG that

$$\mathbf{k} \in V_\sigma \ \& \ (\mathbf{x} \in V_\tau \to_{\mathbf{x}} \phi^*[\mathbf{y}/\mathbf{k}]).$$

In other words, HA is embedded into HAG. Combined with negative translation of first order classical arithmetic (Peano arithmetic) PA into HA, this shows that HAG has the same deductive power as classical arithmetic. More precisely, for negative formulas ϕ one has

$$\text{HA} \vdash (\exists \mathbf{y}(\mathbf{y} \in V_\sigma \ \& \ (\mathbf{x} \in V_\tau \to_{\mathbf{x}} \phi^*))) \iff \phi.$$

4. TREATMENT OF NEGATION IN THE RUSSIAN CONSTRUCTIVIST (MARKOV) SCHOOL

This section presents an account of a work which essentially concluded the construction of the logical base for the work of the Russian constructivist school.[1] This school, founded by A.A. Markov (1903–1979), insisted on developing mathematics using only effective, mechanizable means. Here we describe Markov's approach to the semantics of negative arithmetical formulas. The important requirement of Markov-style constructivism was that the objects should be coded by natural numbers. This directed attention to theories formalizable in the language of first order arithmetic.

The logic of Markov's school can be described using three basic principles: recursive realizability, Markov's principle and classical logic for sentences containing no constructive problems, i.e., $\exists, \vee$-free sentences (Mints 1983; Troelstra and van Dalen 1988). This is a complete description from the classical point of view. Markov presents his approach in quite a different way. He considers a formal language $L_{2\omega+1}$ (Markov 1976) having the same expressive power as the language of first-order arithmetic with all logical connectives $\&, \vee, \supset, \forall, \exists, \forall \leq, \exists \leq$ (the two latter quantifiers are bounded). Arbitrary formulas of this language are transformed into formulas of the form

$$\exists x A \tag{1}$$

where A is an *almost negative* formula: only decidable formulas can occur after $\exists, \vee$. Transformation into the form (1) is done by the algorithm SH (Markov 1976, §12) based on the ideas of Shanin (1958) and equivalent (under a suitable coding in the intuitionistic arithmetic

HA) to recursive realizability as introduced by Kleene (1958). This step and related logical principles were intensively discussed in the early stages of the Russian constructivist school, and by the mid-sixties they were sufficiently clear. Markov's discussion of the algorithm SH is relatively brief.

It is easy to transform the formula A in (1) into a *negative* formula A' which contains neither $\vee$ nor $\exists$-quantifiers (so that HA derives $\neg\neg A' \leftrightarrow A'$) using *Markov's principle*: it is enough to replace $\exists x M$ by $\neg\forall x\neg M$ and $A \vee B$ by $\neg(\neg A \,\&\, \neg B)$. But Markov's goal was to construct a semantics for almost negative formulas in a bottom-up fashion and simultaneously justify *the principle of constructive selection* (his name for *Markov's principle*). In the following, we consider only semantics of almost negative formulas sliced into stages (steps) according to implication nesting: recall that negation is defined via implication and $\perp$. It is known that in the presence of the universal quantifier $\forall$ implication complexity corresponds to quantifier complexity (number of quantifier alternations) used in the investigations of classical arithmetic. Using the notation $\perp = (0 \neq 0)$, $\neg A = A \supset \perp$, $\dot{\exists} x A = \neg\forall x\neg A$ a formula $\forall x_1 \dot{\exists} x_2 \ldots \forall x_k \dot{\exists} x_{k+1} M$ containing k alternations of quantifiers is translated into a formula $\forall x_1(\forall x_2(\ldots(\forall x_{k+1}\neg M \supset \perp)\ldots \supset \perp) \supset \perp)$ containing k "essential" implications. The importance of implication in constructive mathematics was observed early enough (Heyting 1930; Kolmogorov 1932) and by the beginning of the 1960s there were hopes of finding its interpretation in almost negative formulas based on a deductive understanding sketched in papers by Lorenzen (cf. Lorenzen 1954) and based on the notions of derivable and admissible rule. For a given formal system **C**, a rule

$$\frac{A}{B} \tag{2}$$

is *admissible* if $\mathbf{C} \vdash A$ implies $\mathbf{C} \vdash B$ where $\vdash$ is the derivability symbol. The rule (2) is *derivable* (in the system **C**) if there is a deduction of B from A by the rules of **C**. Every derivable rule is obviously admissible, but the converse is not true generally. A good example is the disjunction property of intuitionistic formal systems: the rule

$$\frac{\neg A \to B \vee C}{(\neg A \to B) \vee (\neg A \to C)}$$

for closed formulas A, B, C is admissible in the intuitionistic propositional calculus, but is not derivable there. Otherwise the formula

$$(\neg A \to B \vee C) \to ((\neg A \to B) \vee (\neg A \to C))$$

would be intuitionistically derivable. If **C** is a standard logical system, i.e., the set of derivable formulas is recursively enumerable, the derivability of the rule (2) can be expressed by an arithmetical formula

$$\exists x \text{ Proof } (x, \mathcal{G}(A)) \to \exists y \text{ Proof } (y, \mathcal{G}(B)), \tag{3}$$

where $\mathcal{G}(A), \mathcal{G}(B)$ are Gödelnumbers (numerical codes) of the formulas A, B and Proof is the familiar proof predicate. If the calculus **C** is complete for a class of formulas containing A and B, then (3) can be rewritten as

$$\text{True}(A) \to \text{True}(B),$$

where $\text{True}(F)$ means that F is true. This underlies Markov's definition of implication as admissibility of a rule for the formulas of a basic language containing essentially Σ_1^0-formulas (language L_1 in Markov 1976). By the principles codified by recursive realizability, formula (3) is interpreted as expressing the existence of an effective function transforming x into y:

$$(\exists f \in \text{Rec})(\forall x(\text{Proof}(x, \mathcal{G}(A)) \to \text{Proof}(f(x), \mathcal{G}(B)))),$$

where $f \in \text{Rec}$ means that a natural number f is a code of a unary total recursive function.

For an implication $F \supset G$ which is not immediately reducible to implications of Σ_1^0-formulas Markov suggested interpretation by derivability of the corresponding rule $F \vdash G$. He used such a definition mainly in a situation when there exists a logical system **C** (which is not recursively enumerable in general) which is sound for some previously defined notion of truth $\text{True}(H)$ for the formulas H of the same complexity as F and G. In such a case it was always understood that all true (in the sense of the predicate True) formulas H are added to the axioms of **C**). In other words, a formula $F \supset G$ is true if $\mathcal{F}, F \vdash G$ by the rules of the system **C** where $\mathcal{F} = \{F^{\frown} : \text{True}(F)\}$. Formulas $\forall x A[x]$ beginning with an unbounded universal quantifier are reduced to simpler formulas using the ω-rule (also called Carnap's rule):

$$\frac{\ldots A[n] \ldots \textit{all } n}{\forall x A[x]}$$

Markov requires that every application of this rule be constructive (there should be a uniform general method) and introduces by a generalized inductive definition a notion of truth based on this rule.

Such a treatment is equivalent to the notion of a recursive derivation with ω-rule if one admits a principle of generalized induction corresponding to this inductive definition.

The only remaining special rule of the stepwise semantics is Markov's principle. It can be stated as an implication

$$\neg\neg\exists x M \supset \exists x M, \tag{4}$$

where M is a formula containing no unbounded quantifiers. For the premise $\neg\neg\exists x M$ of the implication (4) the truth turns out to be equivalent to derivability in a system (to be denoted by $\mathbf{C}_2$) with effective ω-rule and *intuitionistic* logic. This made it possible to use the following important observation (Markov 1966b,1976):

Theorem 4.1 *Markov's rule*

$$\frac{\neg\neg\exists x M}{\exists x M} \tag{5}$$

where M is an arithmetical formula without unbounded quantifiers, is admissible in the system $\mathbf{C}_2$.

Similar results for other systems were known in the literature (Kreisel 1958). It was the possibility of justifying the rule at a low stage of the introduced hierarchy that was important to Markov. Another important feature was a method of the proof which anticipated later proofs by Dragalin (1980) and Friedman (1978). The latter two proofs use much weaker means (Kalmar elementary functions are sufficient) to transform the proofs than earlier proofs which used cut elimination. Markov sketched his proof only for systems with the ω-rule described below, but its specialization for the system $\mathbf{C}_2$ (Markov 1966b) would be elementary in the same sense as the proofs by Friedman and Dragalin.

After admissibility of the rule (5) is established, Markov postulates it at the next level as a basic rule, so that the implication $\neg\neg\exists x M \supset \exists x M$ turns out to be true. The final result is a stepwise definition of the constructive truth for almost negative arithmetical formulas which is equivalent to the standard classical definition if one uses classical logic (cf. Mints 1983).

4.1. *One-Quantifier Systems*

4.1.1 *The First Approximation*

Details of the definitions changed in the process of investigation. The first publication (Markov 1966a,b) contained only the two lowest levels. The language L_1 in (Markov 1966b) deals with strings in a

finite alphabet and has the same expressive power as an arithmetical language $\mathcal{L}_1$ including all Kalmar-elementary functions, equations, inequations (i.e., negated equations), $\&, \vee$, bounded quantifiers $(\forall \leq t), (\exists \leq t)$ and the unbounded existential quantifier $\exists$. Recall that Kalmar-elementary functions are constructed from the constants, projections, $+, \times$ and $[a/b]$ by bounded sums, bounded products and substitution.

Formulas of $\mathcal{L}_1$ are essentially Σ_1^0-formulas. The truth definition for closed L_1-formulas in (Markov 1966b) is standard. A calculus C_1 described in (Markov 1966b) is complete for L_1. We describe a similar calculus which is complete for the arithmetical language $\mathcal{L}_1$. Let $|t|$ stand for the numerical value of the closed term t, the letter n stand for an arbitrary numeral, i.e., an expression of the form $0 + 1 + \cdots + 1$, and $F[v/t]$ is the result of substituting the term t for all free occurrences of a variable v in F with standard precautions. $F[t]$ stands for $F[x/t]$.

Calculus C_1:
Axioms:
True closed equations $t = r$ and true closed inequations $t \neq r$
Inference rules:
Introduction rules for $\&, \vee, \exists$:

$$\frac{F\ G}{F\&G} \qquad \frac{F}{F \vee G} \qquad \frac{G}{F \vee G} \qquad \frac{F[t]}{\exists x A}$$

Introduction rules for bounded quantifiers:

$$\frac{F[n]}{(\exists x \leq n)F} \qquad \frac{(\exists x \leq n)F}{(\exists x \leq n+1)F} \qquad \frac{(\exists x \leq |t|)F}{(\exists x \leq t)F}$$

$$\frac{F[0]}{(\forall x \leq 0)F} \qquad \frac{(\forall x \leq n)F\ F[n+1]}{(\forall x \leq n+1)F} \qquad \frac{(\forall x \leq |t|)F}{(\forall x \leq t)F}$$

It is easy to see that these rules are sound for the (obvious understanding of the) language $\mathcal{L}_1$. By induction on the formula it is easy to show that these rules are complete: every true formula is derivable. The language L_2 in (Markov 1966b) has the same expressive power as the arithmetical language $\mathcal{L}_2$ which has as its formulas all formulas of $\mathcal{L}_1$, all implications $F \supset G$ where $F, G \in \mathcal{L}_1$ and arbitrary conjunctions of formulas in $\mathcal{L}_2$. Hence nested implications are not allowed. The truth of a closed implication $F \supset G$ is defined as admissibility of the rule

$$\frac{F}{G}$$

in the calculus $\mathbf{C}_1$, and conjunction is understood in a natural way. By the completeness of the calculus $\mathbf{C}_1$ this definition is equivalent to a standard classical definition. For the formulas of the language L_2 Markov defines (1966b) a sound calculus C_2. The main technical feature of that calculus is the reformulation of standard axioms and rules of the intuitionistic logic and arithmetic adapted to the restricted language considered here. Except for axioms corresponding to the standard defining equations for computable functions and usual properties of bounded and unbounded quantifiers, calculus C_2 contains two induction rules for proving properties of strings in a finite alphabet. The corresponding rule for the arithmetical language would have the form

$$(Ind)\ \frac{F[0] \supset G[0] \quad F[x+1] \supset (F[x] \vee G[x+1]) \quad F[x+1] \,\&\, G[x] \supset G[x+1]}{F[x] \supset G[x]}$$

where $F, G \in \mathcal{L}_1$. Let's note that both premises and the conclusion of the rule *Ind* are formulas of the language $\mathcal{L}_2$. On the other hand, in the traditional formalism the rule *Ind* is equivalent to the standard induction rule for the formula $F \supset G$, which contains implications more complicated than the formulas of the language $\mathcal{L}_2$:

$$(Ind+)\ \frac{F[0] \supset G[0] \quad (F[x] \supset G[x]) \supset (F[x+1] \supset G[x+1])}{F[x] \supset G[x]}$$

Indeed, the premises of the rules (*Ind*) and (*Ind*+) are equivalent in the intuitionistic predicate calculus with Markov's principle:

$$[(F[x] \supset G[x]) \supset (F[x+1] \supset G[x+1])] \leftrightarrow$$

$$[(F[x+1] \supset (F[x] \vee G[x+1])) \,\&\, (F[x+1] \& G[x] \supset G[x+1])].$$

This trick of reducing logical complexity by the implicit use of relations to be justified later is characteristic of Markov's constructive mathematics and its predecessors (intuitionism, finitism). Markov does not explain his choice of the induction rule, but notes in ([Markov 1966a) the impossibility of a complete calculus for the formulas of the language $\mathcal{L}_2$. He also states there that Markov's principle holds for refutability in C_2, which probably means the admissibility of Markov's rule. Although (Markov 1966a,b) contain no indication to the proof of the latter statement, it is plausible that Markov had in mind the same method that was later used to prove admissibility of Markov's rule in the systems with the ω-rule.

4.2. *Systems with ω-Rule*

4.2.1 *The Language* $\mathcal{L}_2$

As already pointed out, the set of the true formulas of the language L_2 is not recursively enumerable: true formulas of the form $\neg\exists x M(\leftrightarrow \forall x \neg M)$ with M containing at most bounded quantifiers form a complete Π_1^0-set. The truth definition for the implications $\exists x M \supset \exists x K$ appeals to the notion of a uniform general method or algorithm. This notion was used in (Markov 1967a,b) to construct a system of rules for the language L_2 containing the ω-rule. Informal justification of such an approach using (what Markov called) *intuition of generality* is given in (Markov 1972,§7, 1976, §2).

The rules for the language $\mathcal{L}_2$, similar to the rules from (Markov (1967a (1967b) (1976)) are the intuitionistic rules for $\supset, \&, \vee$ adapted to avoid nested implications and the ω-rule:

$$\text{(R1)}\ \frac{F \quad F \supset G}{G} \qquad \text{(R2)}\ \frac{F \supset G \quad G \supset H}{F \supset H} \qquad \text{(R3)}\ \frac{G}{F \supset G}$$

$$\text{(R4)}\ \frac{F \supset GF \supset H}{F \supset G \,\&\, H} \qquad \text{(R5)}\ \frac{F \supset H \quad G \supset H}{F \vee G \supset H} \qquad \text{(R6)}\ \frac{K \quad L}{K \,\&\, L}$$

$$\text{(R7)}\ \frac{K \,\&\, L}{K} \qquad \text{(R8)}\ \frac{K \,\&\, L}{L} \qquad \text{(R9)}\ \frac{\ldots I[n] \supset G \ldots n = 0, 1, \ldots}{\exists x I \supset G}$$

Here F, G, H are $\mathcal{L}_1$-formulas, i.e., formulas of the language $\mathcal{L}_1$, K, L are $\mathcal{L}_2$-formulas, $\exists x I$ is a closed $\mathcal{L}_1$-formula. Let's recall that the derivations according to these rules can use (as axioms) arbitrary true $\mathcal{L}_1$-formulas. It is obvious that the rules R1–R9 are sound (or semantically acceptable in Markov's terminology): from true formulas one can derive only true formulas.

4.2.2. *Languages* $\mathcal{L}_3$–$\mathcal{L}_\omega$

Formulas of the arithmetical language $\mathcal{L}_3$, which will have the same expressive power as the language L_3 in (Markov 1967a, b, 1976) are implications $J \supset K$ of $\mathcal{L}_2$-formulas and arbitrary conjunctions of $\mathcal{L}_3$-formulas. In other words, single nesting of implication is allowed. The truth of an implication $J \vdash K$ is defined as derivability $J \vdash K$ by the rules R1–R9 using arbitrary true $\mathcal{L}_2$-formulas. It turns out that the language $\mathcal{L}_3$ (as well as the wider languages $\mathcal{L}_n, \mathcal{L}_\omega$ introduced below) is equivalent to $\mathcal{L}_2$, and it is sufficient to have at most one ω-rule in every branch of the derivation. Justification of that statement is combined in (Markov 1967a, b) with the admissibility proof for Markov's rule. The proof is based on the following observation which is implicit in Markov's constructions.

Every propositional combination of $\mathcal{L}_1$-formulas can be transformed into a classically equivalent $\mathcal{L}_2$-formula, i.e., into a conjunction of implications of $\mathcal{L}_1$-formulas. Markov defines such a transformation $\mathcal{R}$. The main non-intuitionistic steps of $\mathcal{R}$ are

$$\mathcal{R}((I \supset F) \supset G) = (I \vee G)\&(F \supset G)$$

and

$$\mathcal{R}((I \supset F) \supset (H \supset G)) = (H \supset I \vee G)\&(F\&H \supset G).$$

Remaining steps like $\mathcal{R}(F \supset (G \supset H)) = F\&G \supset H$ are defined in a natural way.

Theorem 4.2 A formula $K \in \mathcal{L}_3$ is true iff the formula $\mathcal{R}(K)$ is true.

This proposition is stated in (Markov 1967a, b) only for the case when a justification of the truth of $K \supset L$ uses only a bounded number of ω-rules in every branch, but it is obviously true (together with its proof) without this restriction. The proof is done by induction on the derivation with ω-rule. The detailed proof by cases was never published, but none of these cases presents any difficulty. Theorem 2 amounts to completeness of the rule $\mathcal{R}(K) \vdash K$ for the language $\mathcal{L}_3$. An important instance

$$\mathcal{R}((\exists xM \supset \bot) \supset \bot) = ((\exists xM \vee \bot)\ \&\ (\bot \supset \bot)) \leftrightarrow \exists xM$$

justifies Markov's principle.

The language $\mathcal{L}_{n+1}$, $n+1 > 3$ where n times nested implications are allowed is defined in a natural way. The truth of $\mathcal{L}_{n+1}$-implication is defined as derivability by the rules R1–R9 plus the rules:

$$\text{(R10)}\ \frac{D}{\mathcal{R}(D)} \text{(R11)}\ \frac{\mathcal{R}(D)}{D}$$

Then a union is formed: $\mathcal{L}_\omega = \bigcup_{n>0} \mathcal{L}_n$, and $\mathcal{L}_\omega$ is reduced to $\mathcal{L}_2$ by the transformation $\mathcal{R}$. Hence the truth of an $\mathcal{L}_\omega$-formula in stepwise semantics is the same as the truth of its translation into $\mathcal{L}_2$, and so it coincides with usual arithmetic truth. In particular, familiar classical propositional logic is valid in $\mathcal{L}_2$.

4.3. *The Language* $\mathcal{L}_{\omega+1}$

The last step made in (Markov 1967a, b) is adding initial universal quantifiers to $\mathcal{L}_\omega$-formulas. These quantifiers are understood in a natural way: $\forall xK$ means that there exists a uniform general method allowing the truth of every formula $K[n]$ to be established. The language $\mathcal{L}_{\omega+1}$ having the same expressive power as the language $L_{\omega+1}$ from (Markov 1976) contains $\mathcal{L}_\omega$ and allows unrestricted use of $\forall$ and

&. The rules R1 (*modus ponens*), R6–R8 (introduction and elimination of &) in Section 5, which are denoted below by $\Pi(\omega+1)1-4$ as well as the rules

$$\Pi(\omega+1)5\,\frac{\forall x(F\supset B)}{\exists xF\supset B}\quad \Pi(\omega+1)6\,\frac{\forall xG}{G[n]}\quad \Pi(\omega+1)7\,\frac{\ldots G[n]\ldots}{\forall xG}$$

are sound for $\mathcal{L}_{\omega+1}$. It is easy to see that this system of rules denoted below by $\Pi(\omega+1)$ is also complete for deriving true $\mathcal{L}_{\omega+1}$-formulas from the true $\mathcal{L}_{\omega}$-formulas, and it is possible to have a finite bound for the number of ω-rules in any branch: the bound is the maximum nesting of universal quantifiers. Indeed, a true conjunction is obtained by the rule R8, a true $\forall$-formula is obtained by the ω-rule, and in this way every true $\mathcal{L}_{\omega+1}$-formula is reduced to true $\mathcal{L}_{\omega}$-formulas which are initial (axioms). Since this reduction preserves equivalence in a traditional sense, stepwise semantics coincides with traditional truth for $\mathcal{L}_{\omega+1}$-formulas.

4.4. *Languages* $\mathcal{L}_{\omega+n}, n\geq 2$

The language $L_{\omega+1}$ corresponds exactly to the level Π^0_2 of the arithmetical hierarchy. The next step increases the complexity by 1. Its iteration leads to formulas with arbitrary nesting of implications and universal quantifiers, so that all almost negative formulas are obtained.

The formulas of the language $\mathcal{L}_{\omega+n+1},\ n>0$ are constructed from $\mathcal{L}_{\omega+n}$-formulas by one application of implication and by unrestricted application of $\&,\forall$ to such implications. In fact, Markov used $\supset N$ to indicate the level-N implication, but we drop N. Semantics is defined as before: $\forall xF$ means availability of a uniform general method making it possible to establish the truth of every formula $F[n]$. An implication $A\supset B$ where $A,B\in\mathcal{L}_{\omega+n}$ means derivability $A\vdash B$ by the rules $\Pi_{\omega+n}$ described below (using arbitrary true $\mathcal{L}_{\omega+n}$-formulas). The rules $\Pi_{\omega+n}$ for $n>1$ include (modulo technical details) the rules R1–R4, R6–R8 from section 4.2.1 (*modus ponens*, the transitivity of implication, adding a redundant premise, introduction of conjunction into the conclusion of an implication, introduction and elimination of conjunction) as well as the following rules:

$$\frac{\ldots A\supset G[n]\ldots}{A\supset\forall xG}\ \forall^+\ \frac{\ldots H[n]\ldots}{\forall xH}\ \forall^+\ \frac{\forall xH}{H[n]}\ \forall^-\ \frac{A\&I\supset B}{A\supset\mathcal{R}(I,B)}\ (exp\ )$$

where $\mathcal{R}(I,B)$ for $I\in\mathcal{L}_{\omega+n-2}, B\in\mathcal{L}_{\omega+n-1}$ is the result of transforming an implication $A\supset(B\supset C)$ into a $\mathcal{L}_{\omega+n}$-formula by moving

universal quantifiers and conjunctions forward and replacing $A \supset (B \supset D)$ by $A \,\&\, B \supset D$. In other words, the rule (exp) replaces exportation $A \,\&\, I \supset B \vdash A \supset (I \supset B)$ which has a conclusion outside the language $\mathcal{L}_{\omega+n}$. In fact, the rules $\Pi_{\omega+n}$ were stated for the implication $\supset (N-1)$ and are admissible for implications $\supset M$ $(M < N-1)$. We do not stress these distinctions. It is easy to see that the rules $\Pi_{\omega+n}$ are sound both for $n = 1$ and for $n > 1$. It is possible to show that they are complete: every formula of the language $\mathcal{L}_{\omega+n+1}, n \geq 1$ which is true in the traditional sense is also true in stepwise semantics. We only sketch the argument to clarify the role of every rule. Let us note that an arbitrary $\mathcal{L}_{\omega+n+1}$-formula is reduced to implications of $\mathcal{L}_{\omega+n}$-formulas by analyzing it using the rules of &- and $\forall$-introduction. For example, a formula $K\&L$ is true iff both K and L are true. The completeness proof is concluded by the following assertion.

LEMMA 4.3. If $n \geq 1$ and the formula $A \supset B$ where $A, B \in \mathcal{L}_{\omega+n}$ is true in the traditional sense, then $A \vdash B$ by the rules $\Pi_{\omega+n}$.

Proof. Analysis using the rules of &- and $\forall$-elimination makes it possible to reduce our task (non-constructively) to deriving the relation below, which is true in the traditional sense:

$$E \supset F \vdash G \supset H, \tag{6}$$

where $E, F, G, H \in \mathcal{L}_{\omega+n-1}$. One can assume that the formula $G \supset H$ is false, since otherwise (6) is an axiom. Then $E \supset F$ is false too, and hence E is true and F is false. For example, if a formula $A = \forall x D$ is false, then for some n the formula $D[n]$ is false and $A \vdash D[n]$ by $\forall$-elimination. The formula $F \supset H \in \mathcal{L}_{\omega+n}$ is true and for $n \geq 2$ the required derivation is as follows:

$$\dfrac{\dfrac{\dfrac{E \supset F \quad E}{F} \quad F \supset H}{H}}{G \supset H}$$

For $n = 1$ the true formula $(E \supset F) \supset (G \supset H)$ is in $\mathcal{L}_{\omega}$, and hence (6) is obtained by the rule $\Pi(\omega+1)1$ (*modus ponens*), which concludes the argument.

A more constructive proof for the equivalence of stepwise and traditional semantics would use natural deduction calculus with ω-rule as an equivalent of traditional semantics, and induction on the normal (cut-free) proofs in this calculus (cf. Dragalin 1980; Mints 1983). Note that the rule $\Pi(\omega+1)9$ is not used in our proof, and as noted by Gimon (1973), Markov showed that this rule can be omitted.

4.5. *Languages* $\mathcal{L}_{2\omega}, \mathcal{L}_{2\omega+1}$

The language $L_{2\omega}$ from (Markov (1976)) is the union of all $L_{\omega+n}, n \geq 0$, and under the traditional interpretation it coincides with the language of classical arithmetic. In particular, for every formula $A \in L_{2\omega}$, the formula $\neg\neg A \supset A$ is true in stepwise semantics. Hence the classical propositional calculus is sound both for $L_{2\omega}$ and for the language $\mathcal{L}_{2\omega} = \bigcup_n \mathcal{L}_{\omega+n}$. The arithmetical language $\mathcal{L}_{2\omega+1}$ corresponding to the language $L_{2\omega+1}$ from (Markov (1976)) is simply the language of first order arithmetic. All logical connectives $\&, \vee, \supset, \forall, \exists$ can be applied here without restrictions. This language is reduced to the language $\mathcal{L}_{2\omega}$ (and hence to one of $\mathcal{L}_{\omega+n}, n \geq 0$) by "elucidation of the constructive problem". Arbitrary sentences A are transformed by an algorithm SH (Section 1) into a formula $\exists x A'(x)$ where $A'(x) \in \mathcal{L}_{2\omega}$, and then the truth of A is defined as the existence of a natural number n such that $A'(n)$ is true.

This concludes our description of Markov's stepwise semantics.

4.6. *Further Work*

In his papers published in *Doklady Akad. Nauk SSSR* **214,** 1974 (English translation in *Soviet Mathematics, Doklady* **15,** 1974), Markov presented a new version of stepwise semantics equivalent to one in (Markov 1976).

The work of Markov on stepwise semantics was continued by his students. They studied extensions of the language of arithmetic contained in the language of ramified analysis, where the language of a rank r allows quantifiers over sets definable by formulas of lower rank. Semantics which were obtained in this way turned out to be equivalent to the traditional semantics.

Dragalin (1972) proved that the levels of stepwise semantics form a strict hierarchy. Kanovich (1975) extended stepwise semantics to languages with quantifiers over arithmetical sets. Burgina (1984) sketched an extension of stepwise semantics to ramified analysis of all finite ranks. The truth of the formula $\forall x A$ with a variable x of rank r is defined as the truth of all substitution instances $A[x/\mathbf{q}xB]$ where $\mathbf{q}xB$ is a closed abstract of a rank $\leq r$. M. Dombrowskii-Kabanchenko (1979) extends stepwise semantics to all constructive levels: he constructs languages L_x for all $x \in O$, so that every hyperarithmetical set is representable in one L_x.

ACKNOWLEDGEMENTS

Most of this material was presented in the seminar on negation taught by the author jointly with J. Moravcsik in the winter quarter 1999/2000 at Stanford University. The author is grateful to organizers of the Tübingen workshop, 1999, and to participants of the Stanford seminar on negation for support and discussions.

NOTE

[1]The Russian version of this section was written as a comment to several papers, most of them published by Nauka Publishers, Moscow in the Collected Works of Markov (Markov 2002).

REFERENCES

Avigad J. and S. Feferman: 1998, 'Gödel's Functional ("Dialectica") Interpretation', in S. Buss (eds.), *Handbook of Proof Theory*, Elsevier, pp. 337–405.

Burgina, E. S.: 1984, 'A Step Semantic System for Set Theory', *Mathematical Notes* **35** (6), 448–456 (Russian original: *Matematicheskie Zametki* **35** (6), 855–868.

Dombrowskii-Kabanchenko, M. N.: 1979, 'On a Transfinite Extension of the Stepwise Semantical System of A. A. Markov' (Russian), in B. Kushner and N. Nagorny (eds.), *Theory of algorithms and mathematical logic*, Computing Center of the Acad. of Sci., Moscow, pp. 18–26.

Dragalin, A. G.: 1980, 'New types of realizability and Markov's rule', *Soviet Math. Dokl.* (Russian Original: *Dokl. Akad. Nauk SSSR* **251** (3), 534–537).

Dragalin, A. G.: 1972, 'On a Stepwise Semantical System of A. A. Markov' (Russian), *Second Soviet Conference in Mathematical Logic*, abstracts, 15.

Friedman H.: 1978, 'Classically and Intuitionistically Provably Recursive Functions', *Springer Lecture Notes in Mathematics* **669**, 21–27.

Gimon, V. V.: 1973, 'The Dependence of the Second-Level Rules of Inference in Markov's Stratified Semantic System', *Soviet Math. Dokl.* **14**(5), 1504–1507, (Russian original: *Dokl. Akad. Nauk SSSR* **212**, (5), 1036–1038).

Griss G.: 1955, 'La Mathématique Intuitioniste Sans Negation', *Nieuw Archief voor Wiskunde (3)* **III**, 134–142.

Heyting A.: 1930, 'Die Formalen Regeln der Intuitionistischen Logic', *Sitzungsber. der Preus. Acad. der Wiss., Physik-math. Kl.*, 1930, pp. 42–56.

Kanovich M. I.: 1975, 'A Hierarchical Semantic System with Set Variables', *Soviet Math. Dokl.* **16**,(2), 504–508, (Russian original: *Dokl. Akad. Nauk SSSR* **221**(6), 1256–1259).

Kleene S. C.: 1958, 'On the Interpretation of Intuitionistic Number Theory', *J. Symbolic Logic* **23**, 155–182.

Kolmogorov A. N.: 1932, 'Zur Deutung der Intuitionistischen Logik', *Math. Z.* **35**, 58–65.

Kreisel G.: 1958, 'Mathematical Significance of Consistency Proofs', *JSL* **23**, 155–182.

Krivtsov V.: 2000, 'A Negationless Interpretation of Intuitionistic Theories', *I. Studia Logica* **65** (2), 155–179.

Krivtsov V.: 2000, 'A Negationless Interpretation of Intuitionistic Theories', *I. Studia Logica* **64** (3), 323–344.
Lorenzen P.: 1954, *Einführung in die Operative Logik und Mathematik*, Springer, Heidelberg.
Markov A. A.: 1966a, 'Lower Steps of the Constructive Mathematical Logic' (Russian), in *International Congress of Mathematicians, Abstracts of short talks*, **Section 1**, Moscow, 20.
Markov A. A.: 1966b, 'On Lower Degrees of Constructive Mathematical Logic', Computing Center of the Akad. of Sci., Moscow, 11pp. (mimeographed).
Markov A. A.: 1967a, 'An Approach to Constructive Mathematical Logic', Computing Center of the Akad. of Sci., Moscow, 15pp. (mimeographed).
Markov A. A.: 1967b, 'An Approach to Constructive Mathematical Logic', in B. van Rootselaar and J. F. Stahl (eds), *Logic, Methodology and Philosophy of Science, III*, North-Holland, Amsterdam, pp. 283–294.
Markov A. A.: 1971, 'Essai de construction d'une logique de la mathématique constructive', *Rev. Intern. Philos.* **25**, 477–507.
Markov A. A.: 1972, *On the logic of constructive mathematics* (Russian), Znanie, Moscow, 47p.
Markov A. A.: 1976, 'An attempt of construction of the logic of constructive mathematics' (Russian), in B. Kushner and N. Nagorny (eds), *Theory of algorithms and mathematical logic*, Computing Center of the Acad. of Sci., Moscow, pp. 3–31.
Markov A. A. and N. M. Nagorny: 1988, *The theory of algorithms*, Kluwer Academic Publishers, Dordrecht.
Markov A. A.: 2002, *Collected Works*, in N. M. Nargornyi (ed.). MZNMO Publishers, Moscow, XLVIII + 448p. (in Russian).
Mezhlumbekova, V.: 1975, 'Deductive Capabilities of Negationless Intuitionistic Arithmetic', *Moscow University Mathematical Bulletin* **Vol. 30** (2), Allerton Press Inc., New York.
Mints, G. E.: 1983, 'Stepwise Semantics of A. A. Markov' (Russian), a supplement to the Russian translation of the *Handbook of Mathematical Logic*, Nauka, Moscow, part IV, pp. 348–357.
Nelson, D.: 1966, 'Non-Null Implication', *Journal of Symbolic Logic* **31**(4).
Russell, B.: 1903, *The Principles of Mathematics*, Allen and Unwin, London.
Shanin, N. A.: 1958, 'On Constructive Understanding of Mathematical Judgments' (Russian), *Proceedings Steklov Institute of Mathematics* **52**, 226–231.
Troelstra, A. and D. van Dalen: 1988, *Constructivism in mathematics. An introduction*, Vol. I, II. North-Holland, Amsterdam.
Troelstra, A.: 1990, 'On Early History of Intuitionistic Logic', in P. Petkov (ed.), *Mathematical Logic*, Plenum Press, New York and London, pp. 3–18.
van Dalen, D.: 1999, *Mystic, geometer, and intuitionist: the life of L.E.J. Brouwer*, Clarendon Press, Oxford.

Department of Philosophy,
Building 90
Stanford University
Stanford,
CA 94305-2155
U.S.A.
E-mail: mints@csli.stanford.edu

MICHAEL RATHJEN

THEORIES AND ORDINALS IN PROOF THEORY

ABSTRACT. How do ordinals measure the strength and computational power of formal theories? This paper is concerned with the connection between ordinal representation systems and theories established in ordinal analyses. It focusses on results which explain the nature of this connection in terms of semantical and computational notions from model theory, set theory, and generalized recursion theory.

1. INTRODUCTION

A central theme running through proof theory is the classification of theories by means of ordinals. This is manifest in the assignment of 'proof theoretic ordinals' to theories, gauging their 'consistency strength' and 'computational power'. To put it roughly, such ordinal analyses attach ordinals in a given representation system to formal theories.

The present paper gathers together results which explain the nature of the connection between ordinal representation systems and theories established in ordinal analyses by a more semantical approach in that it characterizes these ordinals in terms of familiar notions from model theory, set theory, and generalized recursion theory.

2. MEASURES IN PROOF THEORY

2.1. *Gentzen's Result*

Gentzen showed that transfinite induction up to the ordinal

$$\varepsilon_0 \;=\; \sup\{\omega, \omega^{\omega}, \omega^{\omega^{\omega}}, \ldots\} \;=\; \text{least } \alpha.\ \omega^{\alpha} = \alpha$$

suffices to prove the consistency of Peano Arithmetic (PA). To appreciate Gentzen's result it is pivotal to note that he applied transfinite induction up to ε_0 solely to primitive recursive predicates and besides that his proof used only finitistically justified means. Hence, a more precise rendering of Gentzen's result is

Synthese (2006) 148: 719–743
DOI 10.1007/s11229-004-6297-0

$$\mathrm{F} + \text{PR-TI}(\varepsilon_0) \vdash \mathsf{Con}(\mathrm{PA}), \tag{1}$$

where F signifies a theory that is acceptable in finitism (e.g., F = PRA = Primitive Recursive Arithmetic) and PR-TI(ε_0) stands for transfinite induction up to ε_0 for primitive recursive predicates. Gentzen also showed that his result is best possible in that PA proves transfinite induction up to α for arithmetic predicates for any $\alpha < \varepsilon_0$. The compelling picture conjured up by the above is that the non-finitist part of PA is encapsulated in PR-TI(ε_0) and therefore "measured" by ε_0, thereby tempting one to adopt the following definition of *proof-theoretic ordinal* of a theory T:

$$|T|_{\mathsf{Con}} = \text{least } \alpha.\ \mathrm{PRA} + \text{PR-TI}(\alpha) \vdash \mathsf{Con}(T). \tag{2}$$

The foregoing definition of $|T|_{\mathsf{Con}}$ is, however, inherently vague because the following issues have not been addressed:

- How are ordinals to be represented in PRA?
- (2) is definitive only with regard to a prior choice of *ordinal representation system*.
- Different ordinal representation systems may provide different answers to (2).

Notwithstanding that, for 'natural' theories T and with regard to a 'natural' ordinal representation system, the ordinal $|T|_{\mathsf{Con}}$ encapsulates important information about the proof strength of T.

The next section will introduce a notion of proof-theoretic ordinal, $|T|_{\mathrm{sup}}$, which does not hinge on the choice of a particular ordinal representation system.

3. THE GENERAL FORM OF AN ORDINAL ANALYSIS

In this section I attempt to say something general about all ordinal analyses that have been carried out thus far. One has to bear in mind that these concern 'natural' theories. Also, to circumvent countless and rather boring counter examples, I will only address theories that have at least the strength of PRA.

3.1. *Theories*

Ordinal analysis is concerned with theories serving as frameworks for formalizing parts of mathematics. It is known that virtually all of ordinary mathematics can be formalized in Zermelo–Fraenkel set

theory with the axiom of choice (ZFC). Hilbert and Bernays (1938) showed that large chunks of mathematics can already be formalized in second order arithmetic. Owing to these observations, proof theory has been focusing on set theories and subsystems of second order arithmetic.

3.1.1. *Subsystems of Second-order Arithmetic*

The language $\mathcal{L}_2$ of second-order arithmetic contains (free and bound) number variables $a, b, c, \ldots, x, y, z, \ldots$, (free and bound) set variables $A, B, C, \ldots, X, Y, Z, \ldots$, the constant 0, function symbols $Suc\ , +, \cdot$, and relation symbols $=, <, \in$. Suc stands for the successor function.

Terms are built up as usual. For $n \in \mathbf{N}$, let $\bar{n}$ be the canonical term denoting n. Formulae are built from the prime formulae $s = t$, $s < t$, and $s \in A$ using $\wedge, \vee, \neg, \forall x, \exists x, \forall X$ and $\exists X$ where s, t are terms.

Note that equality in $\mathcal{L}_2$ is only a relation on numbers. However, equality of sets will be considered a defined notion, namely

$$A = B \text{ iff } \forall x[x \in A \leftrightarrow x \in B].$$

As usual, number quantifiers are called bounded if they occur in the context $\forall x(x < s \rightarrow \ldots)$ or $\exists x(x < s \wedge \ldots)$ for a term s which does not contain x. The Δ_0^0-formulae are those formulae in which all quantifiers are bounded number quantifiers, Σ_k^0-formulae are formulae of the form $\exists x_1 \forall x_2 \ldots Q x_k F$, where F is Δ_0^0, Π_k^0-formulae are those of the form $\forall x_1 \exists x_2 \ldots Q x_k F$. The union of all Π_k^0- and Σ_k^0-formulae for all $k \in \mathrm{N}$ is the class of *arithmetical* or Π_∞^0-*formulae*. The Σ_k^1-formulae (Π_k^1-formulae) are the formulae $\exists X_1 \forall X_2 \ldots Q X_k F$ (resp. $\forall X_1 \exists X_2 \ldots Q x_k F$) for arithmetical F.

The basic axioms in all theories of second-order arithmetic are the defining axioms of $0, 1, +, \cdot, <$ and the *induction axiom*

$$\forall X(0 \in X \wedge \forall x(x \in X \rightarrow x + 1 \in X) \rightarrow \forall x(x \in X)),$$

respectively the *schema of induction*

$$\text{IND} \qquad F(0) \wedge \forall x(F(x) \rightarrow F(x + 1)) \rightarrow \forall x F(x),$$

where F is an arbitrary $\mathcal{L}_2$-formula.

We consider the axiom schema of $\mathcal{C}$-*comprehension* for formula classes $\mathcal{C}$ which is given by

$$\mathcal{C} - \text{CA} \qquad \exists X \forall u(u \in X \leftrightarrow F(u))$$

for all formulae $F \in \mathcal{C}$ in which X does not occur.

For each axiom schema Ax we denote by (Ax) the theory consisting of the basic arithmetical axioms, the schema $\Pi^0_\infty - \mathrm{CA}$, the schema of induction and the schema Ax. If we replace the schema of induction by the induction axiom, we denote the resulting theory by $(\mathrm{Ax})\restriction$.

An example for these notations is the theory $(\Pi^1_1 - \mathrm{CA})$ which contains the induction schema, whereas $(\Pi^1_1 - \mathrm{CA})\restriction$ only contains the induction axiom in addition to the comprehension schema for Π^1_1-formulae.

In the framework of these theories one can introduce defined symbols for all primitive recursive functions. Especially, let $\langle,\rangle\colon \mathbb{N}\times\mathbb{N}\longrightarrow\mathbb{N}$ be a primitive recursive and bijective pairing function.

The *xth* section of U is defined by $U_x := \{y : \langle x, y\rangle \in U\}$. Observe that a set U is uniquely determined by its sections on account of $\langle,\rangle$'s bijectivity.

Any set R gives rise to a binary relation $\prec_R$ defined by $y \prec_R x := \langle y, x\rangle \in R$.

Using the latter coding, we can formulate the axiom of choice for formulae F in $\mathcal{C}$ by

$$\mathcal{C} - \mathrm{AC} \qquad \forall x \exists Y F(x, Y) \to \exists Y \forall x F(x, Y_x).$$

A special form of comprehension is Δ^1_n-comprehension, that is

$$\Delta^1_n - \mathrm{CA} \qquad \forall u\big[\phi(u) \leftrightarrow \vartheta(u)\big] \to \exists X \forall u(u \in X \leftrightarrow \phi(u))$$

for all Π^1_n-formulae ϕ and Σ^1_n-formulae ϑ.

Bar induction is the schema

$$\mathrm{BI} \qquad \forall X\big[\mathrm{WF}(\prec_X) \wedge \forall u\big(\forall v \prec_X u \phi(v) \to \phi(u)\big) \to \forall u \phi(u)\big]$$

for all formulae ϕ, where $\mathrm{WF}(\prec_X)$ expresses that $\prec_X$ is well-founded (see Definition 3.4).

3.1.2. *Subsystems of Set Theory*

The axiom systems for set theory considered in this paper are formulated in the usual language of set theory (called $\mathcal{L}_\in$ hereafter) containing $\in$ as the only non-logical symbol besides $=$. Formulae are built from prime formulae $a \in b$ and $a = b$ by use of propositional connectives and quantifiers $\forall x, \exists x$. Bounded quantifiers $\forall x \in a$, $\exists x \in a$ are defined as usual. Δ_0-formulae are the formulae wherein all quantifiers are bounded; Σ_1-formulae are those of the form $\exists x \varphi(x)$ where $\varphi(a)$ is a Δ_0-formula. For $n > 0$, Π_n-formulae (Σ_n-formulae)

are the formulae with a prefix of n alternating unbounded quantifiers starting with a universal (existential) one followed by a Δ_0-formula. The class of Σ-formulae is the smallest class of formulae containing the Δ_0-formulae which is closed under $\wedge$, $\vee$, bounded quantification and unbounded existential quantification.

One of the set theories which is amenable to ordinal analysis is Kripke–Platek set theory (KP). Its standard models are called *admissible sets*. One of the reasons that this is a truly remarkable theory is that a great deal of set theory requires only the axioms of KP. An even more important reason is that admissible sets have been a major source of interaction between model theory, recursion theory and set theory (cf. Barwise 1975). KP arises from ZF by completely omitting the power set axiom and restricting separation and collection to absolute predicates (cf. Barwise 1975), i.e., Δ_0 formulas. These alterations are suggested by the informal notion of 'predicative'.

DEFINITION 3.1. The axioms of KP are:

Extensionality:	$\forall x(x\in a \leftrightarrow x\in b) \rightarrow a = b.$
Foundation:	$\forall x[(\forall y\in x)G(y) \rightarrow G(x)] \rightarrow \forall x G(x)$ for all formulae G.
Pair:	$\exists\, x\,(x = \{a, b\}).$
Union:	$\exists x\,(x = \bigcup a).$
Infinity:	$\exists x\,[x \neq \emptyset \wedge (\forall y\in x)(\exists z\in x)(y\in z)]$.[1]
Δ_0*Separation:*	$\exists x\,\big(x = \{y\in a : F(y)\}\big)$[2] for all Δ_0-formulas F in which x does not occur free.
Δ_0*Collection:*	$(\forall x\in a)\exists y G(x,y) \rightarrow \exists z(\forall x\in a)(\exists y\in z)G(x,y)$ for all Δ_0-formulas G in which z does not occur free.

L_α, the αth level of Gödel's constructible hierarchy L, is defined by $L_0 = \emptyset$, $L_{\beta+1} = \big\{X : X \subseteq L_\beta;\ X$ definable over $\langle L_\beta, \in\rangle\big\}$ and $L_\lambda = \bigcup\{L_\beta : \beta < \lambda\}$ for limits λ. So any element of L of level α is definable from elements of L with levels $< \alpha$ and L_α.

A transitive set A such that $(A, \in)$ is a model of KP is called an *admissible set*. An ordinal α is *admissible* if the structure $(L_\alpha, \in)$ is a model of KP.

Some systems of set theories will be used later for illustrative purposes. KPi is an extension of KP via the axiom

$$(Lim) \qquad \forall x \exists y [x \in y \wedge y \text{ is an admissible set}].$$

KPl denotes the system KPi without Δ_0 Collection. KPl^r and KPi^r arise from KPl and KPi, respectively, by restricting the scheme of Foundation to Δ_0-formulae G.

KPi^w is obtained from KPi^r by adding the schema

$$\mathrm{IND}_\omega \quad \forall x \in \omega(\forall y \in x F(y) \to F(x)) \to \forall x \in \omega F(x)$$

of induction on ω for all formulae F.

The foregoing set theories are closely related to well-studied subsystems of second-order arithmetic. KPl^r, KPl, KPi^w, and KPi prove the same sentences of second-order arithmetic as $(\Pi^1_1 - \mathrm{CA})$, $(\Pi^1_1 - \mathrm{CA}) + \mathrm{BI}$, $(\Delta^1_2 - \mathrm{CA})$, and $(\Delta^1_2 - \mathrm{CA}) + \mathrm{BI}$, respectively.

3.2. *Proof-Theoretical Reductions*

Ordinal analyses of theories allow one to compare the strength of theories. This subsection defines the notions of *proof-theoretic reducibility* and *proof-theoretic strength* that will be used henceforth.

All theories T considered in the following are assumed to contain a modicum of arithmetic. For definiteness let this mean that the system PRA of Primitive Recursive Arithmetic is contained in T, either directly or by translation.

DEFINITION 3.2. Let T_1, T_2 be a pair of theories with languages $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively, and let Φ be a (primitive recursive) collection of formulae common to both languages. Furthermore, Φ should contain the closed equations of the language of PRA.

We then say that T_1 is *proof-theoretically* Φ-*reducible to* T_2, written $T_1 \leq_\Phi T_2$, if there exists a primitive recursive function f such that

$$\mathrm{PRA} \vdash \forall \phi \in \Phi \forall x [\mathtt{Proof}_{T_1}(x, \phi) \ \to \ \mathtt{Proof}_{T_2}(f(x), \phi)]. \tag{3}$$

T_1 and T_2 are said to be *proof-theoretically* Φ-*equivalent*, written $T_1 \equiv_\Phi T_2$, if $T_1 \leq_\Phi T_2$ and $T_2 \leq_\Phi T_1$.

The appropriate class Φ is revealed in the process of reduction itself, so that in the statement of theorems we simply say that T_1 is *proof-theoretically reducible* to T_2 (written $T_1 \leq T_2$) and T_1 and T_2 are *proof-theoretically equivalent* (written $T_1 \equiv T_2$), respectively.

Alternatively, we shall say that T_1 and T_2 have the *same proof-theoretic strength* when $T_1 \equiv T_2$.

REMARK 3.3. Feferman's notion of proof-theoretic reducibility in (Feferman 1988) is more relaxed in that he allows the reduction to be given by a T_2-recursive function f, i.e.,

$$T_2 \vdash \forall \phi \in \Phi \forall x [\mathtt{Proof}_{T_1}(x, \phi) \rightarrow \mathtt{Proof}_{T_2}(f(x), \phi)]. \quad (4)$$

The disadvantage of (4) is that one forfeits the transitivity of the relation $\leq_\Phi$. Furthermore, in practice, proof-theoretic reductions always come with a primitive recursive reduction, so nothing seems to be lost by using the stronger notion of reducibility.

3.3. $|T|_{\mathrm{sup}}$

Before delineating the general form of an ordinal analysis, we need several definitions.

DEFINITION 3.4. Let T be a framework for formalizing a certain part of mathematics. T should be a true theory (i.e., all its theorems are true) which contains a modicum of arithmetic.

Let A be a subset of $\mathbb{N}$ ordered by $\prec$ such that A and $\prec$ are both definable in the language of T. If the language of T allows for quantification over subsets of $\mathbb{N}$, like that of second order arithmetic or set theory, *well-foundedness* of $\langle A, \prec \rangle$ will be formally expressed by

$$\begin{aligned} \mathrm{WF}(A, \prec) := \forall X \subseteq \mathbb{N} [\forall u \in A (\forall v \prec u v \in X \rightarrow u \in X) \\ \rightarrow \forall u \in A u \in X], \end{aligned} \quad (5)$$

where $\forall v \prec u \ldots$ is short for $\forall v (v \prec u \rightarrow \ldots)$. If, however, the language of T does not provide for quantification over arbitrary subsets of $\mathbb{N}$, like e.g., that of PA, we shall assume that it contains a new unary predicate U. U acts like a free set variable, in that no special properties of it will ever be assumed. We will then resort to the following formalization of well-foundedness:

$$\mathrm{WF}(A, \prec) := \forall u \in A (\forall v \prec u \mathbf{U}(v) \rightarrow \mathbf{U}(u)) \rightarrow \forall u \in A \mathbf{U}(u). \quad (6)$$

We shall use $\mathrm{WF}(\prec)$ as an abbreviation for $\mathrm{WF}(\mathbb{N}, \prec)$. We also set

$$\mathrm{WO}(A, \prec) := \mathrm{LO}(A, \prec) \wedge \mathrm{WF}(A, \prec). \quad (7)$$

If $\langle A, \prec \rangle$ is well-founded, we use $|\prec|$ to signify its set-theoretic order-type. For $a \in A$, the ordering $\prec \restriction a$ denotes the restriction of $\prec$ to $\{x \in A : x \prec a\}$.

The ordering $\langle A, \prec \rangle$ is said to be *provably well-ordered in* T if

$$T \vdash WO(A, \prec). \tag{8}$$

The supremum of the provable well-orderings of T, $|T|_{\text{sup}}$, is defined as follows:

$$|T|_{\text{sup}} := \sup \{\alpha : \ \alpha \textit{ provably recursive in } T\} \tag{9}$$

where an ordinal α is said to be provably recursive in T if there is a recursive well-ordering $\langle A, \prec \rangle$ with order-type α such that

$$T \vdash \ WO(A, \prec)$$

with A and $\prec$ being provably recursive in T. Note that, by definition, $|T|_{\text{sup}} \leq \omega_1^{CK}$, where ω_1^{CK} is the supremum of the order-types of all recursive well-orderings on $\mathbb{N}$. Another characterization of ω_1^{CK} is that it is the least admissible ordinal $> \omega$.

AGREEMENT. From now on the *proof-theoretic ordinal* of a theory T is taken to be $|T|_{\text{sup}}$.

3.4. *The Robustness of* $|T|_{sup}$

This subsection gathers together several results which show that there is a lot of leeway in defining $|T|_{\text{sup}}$. Instead of recursive well-orderings we could have restricted ourselves to primitive recursive or even elementary recursive well-orderings. On the other hand it is also possible to go into the other direction by allowing for well-orderings of greater complexity.

The statements below involve certain well known subsystems of PA and second order arithmetic. $\mathrm{I}\Sigma_1$ denotes the fragment of PA obtained by restricting induction to Σ_1 formulas. WKL_0 is a fragment of second-order arithmetic whose main set existence axiom is a version of König's lemma restricted to binary trees. WKL_0 is proof-theoretically of the same strength as $\mathrm{I}\Sigma_1$, and thus weaker than PA.

For an exact definition and the role of these theories in the program of *Reverse Mathematics* see Simpson (1999).

PROPOSITION 3.5.

(i) Suppose that for every elementary well-ordering $\langle A, \prec \rangle$, whenever $T \vdash \ WO(A, \prec)$, then

$$T \vdash \forall u[A(u) \wedge \forall v \prec uP(v) \rightarrow P(u)] \rightarrow \forall u[A(u) \rightarrow P(u)]$$

holds for all provably recursive predicates P of T. Then

$$|T|_{\sup} = \sup\{\alpha : \alpha \textit{ is provably elementary in } T\} \quad (10)$$
$$= \sup\{\alpha : \alpha \textit{ is provably } \Sigma^0_1 \textit{ in } T\}.$$

Moreover, if $T \vdash WO(A, \prec)$ *and* $A, \prec$ are provably recursive in T, then one can find an elementary well-ordering $\langle B, \lessdot\rangle$ and a recursive function f such that $T \vdash WO(B, \lessdot)$, f is provably recursive in T, and T proves that f supplies an order isomorphism between $\langle B, \lessdot\rangle$ and $\langle A, \prec\rangle$.

Examples for (i) are the theories $I\Sigma_1$, WKL_0 and PA.

(ii) If T contains $(\Pi^0_\infty - CA)\restriction$, then

$$|T|_{\sup} = \sup\{\alpha : \alpha \textit{ is provably arithmetic in } T\}. \quad (11)$$

(iii) If T contains $(\Sigma^1_1 - AC)\restriction$, then

$$|T|_{\sup} = \sup\{\alpha : \alpha \textit{ is provably analytic in } T\}, \quad (12)$$

where a relation on $\mathbb{N}$ is called *analytic* if it is lightface Σ^1_1.

Proof. (Rathjen 1999), Proposition 2.19. □

A theory is said to be Π^1_1-faithful if all of its theorems of complexity Π^1_1 are true.

THEOREM 3.6. Let T be a Σ^1_1 axiomatizable theory.

(i) If T is Π^1_1-faithful, then $|T|_{\sup} < \omega_1^{CK}$.
(ii) If $(\Pi^0_\infty - CA) \subseteq T$ and $|T|_{\sup} < \omega_1^{CK}$, then T is Π^1_1-faithful.
(iii) There are consistent primitive recursive theories T such that $|T|_{\sup} = \omega_1^{CK}$.

Proof. See Rathjen (1999), Theorem 2.4. □

As Kreisel observed, another feature of $|T|_{\sup}$ is that this ordinal does not change when one augments T by true Σ^1_1 statements.

PROPOSITION 3.7. Let T be a primitive recursive, Π^1_1-faithful theory of second order arithmetic such that $PA \subseteq T$. Let $\lhd$ be a primitive recursive well-ordering such that $|T|_{\sup} = |\lhd|$ and

$$PA + TI(\lhd) \vdash \mathtt{Proof}_T(\ulcorner F \urcorner) \to F \quad (13)$$

holds for all arithmetic formulae F which may contain free second order set variables but no free number variables. Then, for any true Σ^1_1 statement B,

$$|T|_{\sup} = |T + B|_{\sup}.$$

Proof. See Rathjen (1999), Proposition 2.6.

REMARK 3.8. In all the examples I know, if T is a subsystem of classical second order arithmetic for which an ordinal analysis has been carried out via an ordinal representation system $(A, \lhd)$, (13) is satisfied.

4. MODEL-THEORETIC CHARACTERIZATIONS

This first part of this section shows that $|T|_{\mathrm{sup}}$ can be couched in terms of partial models in the constructible hierarchy. The second part presents Carlson's approach of obtaining ordinal representations from finite structures.

4.1. *Partial Models*

Recall that L_α, the αth level of Gödel's constructible hierarchy L, is defined by $L_0 = \emptyset$, $L_{\beta+1} = \{X : X \subseteq L_\beta;\ X \textit{ definable over } \langle L_\beta, \in\rangle\}$ and $L_\lambda = \bigcup\{L_\beta : \beta < \lambda\}$ for limits λ. So any element of L of level α is definable from elements of L with levels $< \alpha$ and L_α.

DEFINITION 4.1. For a collection of sentences (in the language of set theory), $\mathcal{F}$, we say that L_α is an $\mathcal{F}$-*model of* T if for all $B \in \mathcal{F}$, whenever $T \vdash B$, then $L_\alpha \models B$. Let

$$|T|_{\mathcal{F}} := L_\alpha \text{is an } \mathcal{F}\text{-model of} T\}.$$

Suppose an ordinal δ is definable by a formula $D(x)$ in T, that is $T \vdash \exists!\xi D(\xi)$ and $D(\delta)$ is true. Let $T \vdash B^{L_\delta}$ stand for $T \vdash \exists\xi[D(\xi) \wedge B^{L_\xi}]$. L_α is said to be an $\mathcal{F}(L_\delta)$ *model of* T if whenever $T \vdash B^{L_\delta}$ holds for $B \in \mathcal{F}$, then $L_\alpha \models B$. For interesting ordinals δ and theories T it is often fruitful to consider the following ordinal:

$$|T|_{\mathcal{F}(L_\delta)} := \min\{\alpha :\ L_\alpha \text{ is an } \mathcal{F}(L_\delta)\text{-model of T}\}.$$

DEFINITION 4.2. Let $\mathcal{F}$ be a collection of sentences. A set theory T is said to be $\mathcal{F}$-*sound with respect to* L if for every $\mathcal{F}$ theorem ϕ of T, $L \models \phi$ holds.[3] For the sake of brevity, in what follows I shall use the shorthand "$\mathcal{F}$-sound" rather than "$\mathcal{F}$-sound with respect to L".

The system PRST (for *Primitive Recursive Set Theory*) is formulated in the language of set theory augmented by symbols for all primitive recursive set functions in the sense of Jensen and Karp (Jensen and Karp 1971). The *axioms of* PRST are Extensionality, Pair, Union, Infinity, Δ_0-Separation, the Foundation Axiom (i.e., $x \neq \emptyset \rightarrow (\exists y{\in}x)(\forall z{\in}y)z \notin x$) and the defining equations for the primitive recursive set functions (see (Rathjen 1992) for a precise definition).

In the following we shall assume that all set theories contain PRST either directly or via interpretation.

The next theorem gives a characterization of the proof-theoretic ordinal of T in terms of $|T|_{\mathcal{F}}$ for two classes of formulae. It requires, however, that T proves the existence of ω_1^{CK}. Recall that ω_1^{CK} stands for the least admissible ordinal $> \omega$. There is a canonical Π_3-sentence θ of set theory such that for every $\alpha > 0$, $L_\alpha \models \theta$ iff L_α is an admissible set (cf. Richter and Aczel 1973). We will say that T proves the existence of ω_1^{CK} if $T \vdash \exists \alpha > \omega \theta^{L_\alpha}$.

THEOREM 4.3. If T is Π_2-sound and T proves the existence of ω_1^{CK}, then

$$|T|_{\sup} = |T|_{\Sigma_1(L(\omega_1^{CK}))} = |T|_{\Pi_2(L(\omega_1^{CK}))}.$$

Proof. The equality $|T|_{\sup} = |T|_{\Sigma_1(L(\omega_1^{CK}))}$ follows from (Rathjen 1991), Theorem 7.14.

$|T|_{\Sigma_1(L(\omega_1^{CK}))} = |T|_{\Pi_2(L(\omega_1^{CK}))}$ is an immediate consequence of the proof of (Rathjen (1992), Theorem 2.1. □

An ordinal analysis of T also allows one to determine the ordinals $|T|_{\Sigma_1}$ and $|T|_{\Pi_2}$. This will be addressed in more detail in the last section. In point of fact, these ordinals are the same if T satisfies some mild requirements.

PROPOSITION 4.4. Suppose T is Π_2 sound and contains Δ_0-collection. Furthermore, suppose that $T \vdash B$ implies $T \vdash \exists \alpha \exists x\ (x = L_\alpha \wedge B^x)$ for all Σ_1-sentences B. If T has a Σ_1-model then T has a Π_2-model and

$$|T|_{\Sigma_1} = |T|_{\Pi_2}. \tag{14}$$

Proof. (Rathjen (1992). Theorem 2.1. □

There are theories where $|T|_{\Sigma_1}$ and $|T|_{\sup}$ coincide. A prominent example is KP. The ordinal $\psi_\Omega(\varepsilon_{\Omega+1})$ is known as the *Bachmann–Howard ordinal.*

THEOREM 4.5. $|\mathrm{KP}|_{\sup} = |\mathrm{KP}|_{\Sigma_1} = |\mathrm{KP}|_{\Pi_2} = \psi_\Omega(\varepsilon_{\Omega+1})$.

Proof. See (Jäger 1982) and (Rathjen 1992). □

4.2. *Patterns of Resemblance*

An intriguing new way of defining ordinal representation systems has been pursued by Carlson (cf. Carlson 1999, 2001). In this approach

the class of ordinals gets furnished with a relation of $\exists_1$ elementary substructurehood and the ordinal representations correspond to finite substructures of this class structure.

DEFINITION 4.6. Suppose $\mathfrak{A} = (ORD, \ldots, \leq)$ is a (class) structure whose universe is the class of ordinals ORD, with the ordering $\leq$ of ordinals. A finite substructure $\mathfrak{F}$ of $\mathfrak{A}$ is said to be *isominimal* if there is no finite substructure $\mathfrak{F}'$ of $\mathfrak{A}$ such that

- $\mathfrak{F}' \equiv \mathfrak{F}$
- $\mathfrak{F}' \neq \mathfrak{F}$
- $\mathfrak{F}' \leq_{pw} \mathfrak{F}$,

where $\leq_{pw}$ denotes the *pointwise partial ordering* of finite sets of ordinals, i.e., $\mathfrak{F}' \leq_{pw} \mathfrak{F}$ iff both structures have the same number of elements and if $\alpha_0, \ldots, \alpha_{n-1}$ enumerates the elements of $\mathfrak{F}'$ in increasing order and $\beta_0, \ldots, \beta_{n-1}$ enumerates the elements of $\mathfrak{F}$ in increasing order then $\alpha_i \leq \beta_i$ for $i < n$.

Since the definition of Σ_1 formula in the usual set-theoretic sense allows arbitrary bounded quantifiers inside the initial existential quantifiers, we specify that a $\exists_1$ formula is a quantifier-free formula prefixed by a string of existential quantifications.

The *core* of $\mathfrak{A}$ is the union of the isominimal substructures of $\mathfrak{A}$.

Carlson (1999) introduces a structure $\mathcal{R}_0$ whose core turns out to be the ubiquitous ordinal ε_0.

DEFINITION 4.7. $\preceq_1^0$ is the partial ordering on the class of ordinals defined by induction so that

$$\alpha \preceq_1^0 \beta \text{ iff } (\alpha, 0, \leq, \preceq_1^0) \text{ is a } \exists_1\text{-elementary substructure of } (\beta, 0, \leq, \preceq_1^0).$$

To be more precise, by induction on β we define the set of α such that $\alpha \preceq_1^0 \beta$ (note that we have taken some liberty in writing $(\alpha, \leq, \preceq_1^0)$ where we should have restricted the relations to α).

THEOREM 4.8. (Carlson 1999). The core of $\mathcal{R}_0$ *is the ordinal* ε_0.

Augmenting the ordinals by the function of addition, Carlson (2001) introduces a richer structure $\mathcal{R}_1$ whose core turns out to be the proof-theoretic ordinal of $(\Pi_1^1 - \mathrm{CA})\restriction$.

DEFINITION 4.9. $\preceq_1^1$ is the partial ordering on the class of ordinals defined by induction so that

$$\alpha \preceq_1^1 \beta \text{ iff } (\alpha, 0, +, \leq, \preceq_1^1) \text{ is a } \exists_1\text{-elementary substructure of } (\beta, 0, +, \leq, \preceq_1^0).$$

It should be pointed out that contrary to standard practice, one allows structures to interpret $+$ as a partial operation on the universe, e.g., if $\beta, \gamma < \alpha$ but $\beta + \gamma \geq \alpha$ then $+$ is not defined for the arguments β, γ in the structure $(\alpha, 0, +, \leq, \preceq_1^1)$.

THEOREM 4.10 (Carlson 2001). The core of $\mathcal{R}_1$ is $\psi_{\Omega_1}\Omega_\omega$ (in the notation of Buchholz 1986), the proof-theoretic ordinal of $(\Pi_1^1 - \mathrm{CA})\restriction$.

To give an idea of how the core of $\mathcal{R}_1$ gives rise to a recursive ordinal representation system we need some notions. A substructure $\mathfrak{B}$ of $\mathcal{R}_1$ is *closed* if $0 \in \mathfrak{B}$ and whenever $\omega^{\alpha_1} + \cdots + \omega^{\alpha_m}$ is in $\mathfrak{B}$ with $\alpha_1 \geq \cdots \geq \alpha_m$ then $\omega^{\alpha_1}, \ldots, \omega^{\alpha_m} \in \mathfrak{B}$ and $\omega^{\alpha_1} + \cdots + \omega^{\alpha_i} \in \mathfrak{B}$ for $i = 1, \ldots, m$. Notice that every finite set of ordinals is contained in a finite set of ordinals which is closed.

It can be shown that for a fixed finite closed substructure $\mathfrak{F}$ of $\mathcal{R}_1$, there is a unique isominimal substructure $\mathfrak{F}^*$ of $\mathcal{R}_1$ which is isomorphic to $\mathfrak{F}$. Moreover, $\mathfrak{F}^*$ is closed. This provides a system of ordinal representations for the ordinals which occur in the core of $\mathcal{R}_1$: if α appears as the nth element of some closed isominimal substructure $\mathfrak{F}^*$ of $\mathcal{R}_1$ we can use the pair (τ, n) as a notation for α where τ is the isomorphism type of $\mathfrak{F}^*$. These notations allow one to show that the core of $\mathcal{R}_1$ is isomorphic to a recursive structure.

REMARK 4.11. At first glance, the difference between the structures $\mathcal{R}_0$ and $\mathcal{R}_1$ seems only tiny as the operation of addition on ordinals appears to be innocent enough. A similar effect, though, has also been unearthed in a paper by Schütte and Simpson (Schütte and Simpson 1985) wherein they showed that omitting the operation $+$ from the ordinal representation system for the proof-theoretic ordinal of $(\Pi_1^1 - \mathrm{CA})$ has a dramatic effect in that the order-type drops to ε_0.

REMARK 4.12. Carlson has also considered richer structures than $\mathcal{R}_1$ whose cores are conjectured to provide ordinal representations for very strong subsystems of second order arithmetic.

5. CHARACTERIZATIONS VIA *E*-RECURSION

A particularly interesting measure that can be assigned to a set theory T is $|T|^E$. Here the superscript E signifies *E-recursion*, also

termed *set recursion*. *E*-recursion theory extends the notion of computation from the natural numbers to arbitrary sets. For details see Sacks (1990).

DEFINITION 5.1. The intent is to assign meaning to $\{e\}(x)$ for every set x via an appropriate notion of computation. The definition of $\{e\}(x)$ is in terms of schemes introduced by Normann (1978), and subsequently and independently by Moschovakis (1976). The first three schemes are projection, difference and pairing. The fifth is composition, and the sixth is enumeration. Bounding with union, the fourth scheme, is the sole source of infinitely long computations. To be precise, *E*-recursion is defined by the following schemes:

1. $e = \langle 1, n, i\rangle$,
   $\{e\}(x_1, \ldots, x_n) = x_i$.
2. $e = \langle 2, n, i, j\rangle$,
   $\{e\}(x_1, \ldots, x_n) = x_i \backslash x_j$.
3. $e = \langle 3, n, i, j\rangle$,
   $\{e\}(x_1, \ldots, x_n) = \{x_i, x_j\}$.
4. $e = \langle 4, n, c\rangle$,
   $\{e\}(x_1, \ldots, x_n) \simeq \bigcup\{\{c\}(y, x_2, \ldots, x_n) : y \in x_1\}$.
   The left side is not defined unless $\{c\}(y, x_2, \ldots, x_n)$ is defined for all $y \in x_1$.
5. $e = \langle 5, n, m, e', e_1, \ldots, e_n\rangle$,
   $\{e\}(x_1, \ldots, x_n) \simeq \{e'\}(\{e_1\}(x_1, \ldots, x_n), \ldots, \{e_m\}(x_1, \ldots, x_n))$.
6. $e = \langle 6, n, m\rangle$,
   $\{e\}(e_1, x_1, \ldots, x_n, y_1, \ldots, y_m) \simeq \{e_1\}(x_1, \ldots, x_n)$.

$\simeq$ is Kleene's symbol for strong equality. If g and f are partial functions, then $f(x) \simeq g(x)$ iff neither $f(x)$ nor $g(x)$ is defined, or $f(x)$ and $g(x)$ are defined and equal.

To some, enumeration is a theorem, not a scheme. Casting it as a scheme makes it possible to omit the least number operator and primitive recursion, two schemes well abandoned when there is no underlying effective wellordering of the sets.

DEFINITION 5.2. A partial function from V, the class of all sets, into V is *partial E-recursive* if it belongs to the least class of partial functions closed under the Normann schemes. The graph of such a

function is Σ_1 (in the language of set theory), the converse, however, does not hold. An example is $O(x)$, Gödel's order of constructibility function

$$O(x) \simeq \mu\gamma[x \in L_{\gamma+1} \setminus L_\gamma].$$

$O(x)$ is Σ_1 but not partial E-recursive. If $x \in L$ then $O(x)$ is found by an unbounded search devoid of effective content.

A theorem of van de Wiele (1982) explains the gap between Σ_1 definability and partial E-recursiveness.

DEFINITION 5.3. Let $f: V \to V$ be a total function. F is *uniformly* Σ_1 *definable on every admissible set* if there is a Σ_1 formula $\phi(x, y)$ (which contains at most the free variables exhibited) such that for every admissible set A:

$-(\forall x \in A) f(x) \in A$;
$-f \restriction A = \{\langle a, b\rangle : \langle A, \in\rangle \models \phi(a, b)\}$.

THEOREM 5.4 (van de Wiele 1982). For every total function $f: V \to V$ the following are equivalent:

(i) f is E-recursive.
(ii) f is uniformly Σ_1 definable on every admissible set.

DEFINITION 5.5. The next notions are due to A. Schlüter (1993).

$$|T|^E_{\Sigma_1} := \min\{\alpha : \text{ for all } e \in \omega,\ T \vdash \{e\}(\omega) \downarrow \text{ implies } \{e\}(\omega) \in L_\alpha\}.$$

$|T|^E_{\Pi_2}$ denotes the ordinal

$$\min\{\alpha > \omega : \text{ for all } e \in \omega,\ T \vdash \forall x \{e\}(x) \downarrow \text{ implies } \forall x \in L_\alpha \{e\}(x) \in L_\alpha\}.$$

For the remainder of this subsection it is assumed that all set theories contain PRST.

THEOREM 5.6 (Schlüter 1993). If T is a Π_2 sound theory, then

$$|T|^E_{\Sigma_1} = |T|^E_{\Pi_2}. \tag{15}$$

Proof. (15) is stated and proved in Schlüter (1993, 6.14). □

THEOREM 5.7 (Schlüter 1993). If T is Π_2-sound, then

$$|T|_{\mathrm{sup}} = |T|^E_{\Sigma_1} = |T|^E_{\Pi_2}.$$

Proof. A detailed proof of $|T|^E_{\Sigma_1} = |T|_{\mathrm{sup}}$ can be found in Schlüter (1993), Satz 6.15. □

In point of fact, the proof of Theorem 5.6 also yields the following result.

THEOREM 5.8. If T is a Π_2 sound theory, then

$$|T|_{\mathrm{sup}} = \{\alpha : \exists e{\in}\omega[\alpha = \{e\}(\omega) \wedge T \vdash \{e\}(\omega)\downarrow]\}.$$

6. THE Σ_1 SPECTRUM OF A THEORY

The extraction of classifications of provable functions from ordinal analyses is not confined to recursive functions on natural numbers. In the case of fragments of second order arithmetic, one may also classify the provable hyperarithmetical as well as the provable Δ^1_2 functions on $\mathbb{N}$. Such results can be obtained by interpreting the ordinals of the representation system used in the pertaining ordinal analysis as large admissible ordinals (see Section 7). In the case of set theories one may classify several kinds of provable set functions and ordinals.

In the following we will be concerned with norms that can be assigned to set theories. In general, they can also be extracted from an ordinal analysis of a set theory T. Among other results, they lead to a classification of the provable set functions of T.

DEFINITION 6.1. Another notion that is closely related to the norm $|T|_{\Sigma_1}$ is the notion of *good* Σ_1*-definition* from admissible set theory (see Barwise 1975, II.5.13). Given a set theory T, we say that an ordinal α has a *good* Σ_1*-definition in* T if there is a Σ_1-formula $\phi(u)$ such that

$$L \models \phi[\alpha] \text{ and } T \vdash \exists!\xi\phi(\xi).$$

Let

$$\mathtt{spec}_{\Sigma_1}(T) := \{\alpha : \alpha \text{ has a good } \Sigma_1 \text{ definition in T}\}.$$

If T is Σ_1 sound one obviously has $\sup(\mathtt{spec}_{\Sigma_1}(T)) = |T|_{\Sigma_1}$. In many cases the set $\mathtt{spec}_{\Sigma_1}(T)$ bears interesting connections to the ordinals of the representation system that has been used to analyze T. Ordinal representation systems that have been developed via a detour through

large cardinals allow for an alternative interpretation wherein the large cardinals are replaced by their recursively large counterparts. The latter interpretation gives rise to a canonical interpretation of the ordinal terms of the representation system in $\mathtt{spec}_{\Sigma_1}(T)$. In general, however, the ordinals of $\mathtt{spec}_{\Sigma_1}(T)$ stemming from the ordinal representation form a proper subset of $\mathtt{spec}_{\Sigma_1}(T)$ with many 'holes' as will be shown in the last section. It would be very desirable to find a 'natural' property which could distinguish the ordinals of the representation system within $\mathtt{spec}_{\Sigma_1}(T)$ so as to illuminate their naturalness. I consider this to be one of the most important problems in the area of strong ordinal representation systems. A more thorough discussion will follow in Section 7.

We will show that under very weak assumptions on T that the spectrum of T is an initial segment of the ordinals. Let T be a theory such that $T \vdash \phi$ implies $L \models \phi$ for all Σ_1- and Π_1-sentences ϕ. We shall require that T contains a modicum of primitive recursive set theory in the sense that T contains the theory PRST of Definition 4.2. In particular, we assume that the function $\eta \mapsto L_\eta$ is provable in T. Moreover, we shall assume that $T \vdash \forall\alpha\exists\lambda \geq \alpha[\lambda \textit{ is a limit}]$. Then the following holds.

THEOREM 6.2 (Möllerfeld and Rathjen 2002). $\mathtt{spec}_{\Sigma_1}(T)$ is an ordinal, that is an initial segment of the ordinals.

Proof. The proof makes use of the notion of stable ordinal. For an introduction to stable ordinals the reader is referred to Barwise's textbook (1975). An ordinal η is β-*stable* if $\eta \leq \beta$ and L_η is a Σ_1-elementary substructure of L_β.

Set $\sigma_T := \sup(\mathtt{spec}_{\Sigma_1}(T))$. For each $\eta \in \mathtt{spec}_{\Sigma_1}(T)$ we pick a Σ_1-formula ϕ_η such that

$$T \vdash \exists!\xi\phi_\eta(\xi) \quad \text{and} \quad L \models \phi_\eta[\eta]. \tag{16}$$

We shall proof, by induction on $\alpha < \sigma_T$, that $\alpha \in \mathtt{spec}_{\Sigma_1}(T)$.

For this assume $\alpha \subseteq \mathtt{spec}_{\Sigma_1}(T)$.

For each limit ordinal $\lambda \in \mathtt{spec}_{\Sigma_1}(T)$ with $\lambda > \max(\omega, \alpha)$, define

$$A_\lambda := \{a \in L_\lambda \mid \text{there is a } \Sigma_1\text{-definition of } a \text{ in } L_\lambda \text{ using parameters } < \alpha.\}$$

Let ρ_λ be the least ordinal not in A_λ. By the proof of Theorem 7.8. in chap. V of Barwise (1975), we get

$$A_\lambda = L_{\rho_\lambda} \quad \text{and} \quad \rho_\lambda \text{ is the least } \lambda\text{-stable ordinal} \geq \alpha. \quad (17)$$

In actuality, this proof assumes that λ is admissible. However, ruminating on the proof, it turns out that all which is required is that the predicate $x \in L_\delta$ (of x and δ) and the constructible ordering $<_L$ are absolute for L_λ. Therefore it suffices to assume that λ is a limit $> \omega$ (for more details see Devlin (1984), II, Theorem 5.2).

Note that $\alpha \leq \rho_\lambda \leq \lambda$.

Case 1: $\alpha < \rho_\lambda$ for some $\lambda \in \mathtt{spec}_{\Sigma_1}(T)$, where λ is a limit $> \max(\omega, \alpha)$.

Then α has a Σ_1-definition in L_λ using parameters $\beta_1, \ldots, \beta_n < \alpha$. Let $\psi(x, \beta_1, \ldots, \beta_n)$ be the defining formula. Put

$$\theta(x) := \big[L_\lambda \models \exists!\xi\psi(\xi, \beta_1, \ldots, \beta_n) \wedge x \in L_\lambda \wedge L_\lambda \models \psi(x, \beta_1, \ldots, \beta_n)\big] \vee \big[L_\lambda \models \neg\exists!\xi\psi(\xi, \beta_1, \ldots, \beta_n) \wedge x = \lambda\big].$$

θ is a Σ_1-formula with parameters $\lambda, \beta_1, \ldots, \beta_n$ all of which are in $\mathtt{spec}_{\Sigma_1}(T)$. Using the formulae $\phi_\lambda, \phi_{\beta_1}, \ldots, \phi_{\beta_n}$ (see (16)), we can re-write θ to an equivalent Σ_1-formula such that $T \vdash \exists!\eta\theta(\eta)$. Moreover, $L \models \theta[\alpha]$. Therefore $\alpha \in \mathtt{spec}_{\Sigma_1}(T)$.

Case 2: For all limits $\lambda \in \mathtt{spec}_{\Sigma_1}(T)$ with $\lambda > \max(\alpha, \omega)$, it holds $\alpha = \rho_\lambda$.

By (17), α is λ-stable for all such λ. Therefore α is σ_T-stable, as σ_T is the limit of all these ordinals. However, from $T \vdash \exists!\xi\psi(\xi)$ with ψ Σ_1, we get $L_{\sigma_T} \models \exists\xi\psi(\xi)$ and hence $L_\alpha \models \exists\xi\psi(\xi)$, which yields $\mathtt{spec}_{\Sigma_1}(T) \subseteq \alpha$, contradicting $\alpha < \sigma_T$. □

A quick glance at the preceding proof reveals that the proof utilizes that T is a classical theory at every turn. Therefore one might expect a different behaviour for intuitionistic set theories. Up till now, however, no intuitionistic set theory int-T has been found where $\mathtt{spec}_{\Sigma_1}$(int-T) contains holes. On the other hand, for several intuitionistic theories it has been shown that their spectrum yields a segment like in the classical case. If T is a classical set theory let int-T be the theory with the same axioms but based on intuitionistic logic.

THEOREM 6.3. If T denotes any of the theories KPl, KPl^r, or KPi^w, then

$$\mathtt{spec}_{\Sigma_1}(T) = \mathtt{spec}_{\Sigma_1}(\mathit{int}{-}T).$$

Proof. This has been shown by my student Nikolaus Thiel. The results are presented in his Ph.D. thesis (Thiel 2003). □

Regarding the previous result a caveat should be entered here as it might hinge on what notion of ordinal one employs. Theorem 6.3 holds if ordinals are defined as transitive sets whose elements are transitive as well. Classically such ordinals are linearly ordered by $\in$ but intuitionistically this is not provable. Therefore it is conceivable that Theorem 6.3 does no longer hold when one takes a more restricted notion of ordinals which requires them to be linearly ordered by $\in$. I consider this a very interesting riddle.

7. RECURSIVELY LARGE ORDINALS AND ORDINAL REPRESENTATION SYSTEMS

It is probably not widely known that an ordinal analysis of a theory T not only characterizes the provably recursive ordinals of T but also provides information about provable sets of higher complexity. For instance the first T-stable ordinal, σ_T, that is the first ordinal ρ such that all Σ_1 definable ordinals of T are $< \rho$ can be obtained from an ordinal analysis of T as well. σ_T can also be characterized as the first ordinal which is closed under all ∞-functions on ordinals which are provably total in T (cf. Hinman 1978, VIII), or in the case of subsystems of second order arithmetic as the supremum of its provable Δ^1_2 ordinals. To illustrate these more subtle features by means of a simple example (which nevertheless encapsulates the generic case) the last section of this paper introduces an ordinal representation system that has been employed in the ordinal analysis of the subsystem of second order arithmetic based on Δ^1_2 comprehension and bar induction or equivalently the set theory KPi.

7.1. *Ordinal Functions Based on a Weakly Inaccessible Cardinal*

Recall that KPi is a set theory which originates from Kripke–Platek set theory and in addition has an axiom which says that any set is contained in an admissible set. Thus the standard models of KPi in L are the segments L_κ with κ being recursively inaccessible. The ordinal analysis for KPi (cf. Jäger and Pohlers 1982) used an ordinal representation system built from ordinal functions, so-called *collapsing functions*, which have originally been defined with the help of a weakly inaccessible cardinal. This subsection expounds on the development of this particular ordinal representation system with an eye towards the role of cardinals therein. Traditionally the cardinals $\aleph_\alpha$ have been named Ω_α in proof theory and this tradition will also be adhered to below. Let

$$\mathrm{I} := \text{“first weakly inaccessible cardinal”} \tag{18}$$

and let $(\alpha \mapsto \Omega_\alpha)_{\alpha<\mathrm{I}}$ enumerate the infinite cardinals below I in increasing order. Further let

$$\Re^{\mathrm{I}} := \{\mathrm{I}\} \cup \{\Omega_{\xi+1} : \xi < \mathrm{I}\}. \tag{19}$$

Variables κ, π will range over $\Re^{\mathrm{I}}$.

DEFINITION 7.1. An ordinal representation system for the analysis of KPi can be derived from the following functions and Skolem hulls of ordinals defined by recursion on α:

$$C^{\mathrm{I}}(\alpha,\beta) = \begin{cases} \textit{closure of } \beta \cup \{0,\mathrm{I}\} \\ \textit{under}: \\ +, (\xi \mapsto \omega^\xi) \\ (\xi \mapsto \Omega_\xi)_{\xi<\mathrm{I}} \\ (\xi\pi \mapsto \psi_\pi(\xi))_{\xi<\alpha} \end{cases} \tag{20}$$

$$\psi_\pi(\alpha) \simeq \min\{\rho < \pi :\ C^{\mathrm{I}}(\alpha,\rho) \cap \pi = \rho \ \wedge\ \pi \in C^{\mathrm{I}}(\alpha,\rho)\}. \tag{21}$$

Note that if $\rho = \psi_\pi(\alpha)$, then $\psi_\pi(\alpha) < \pi$ and $[\rho,\pi) \cap C^{\mathrm{I}}(\alpha,\rho) = \emptyset$, thus the order-type of the ordinals below π which belong to the Skolem hull $C^{\mathrm{I}}(\alpha,\rho)$ is ρ. In more pictorial terms, ρ is the α^{th} *collapse* of π.

LEMMA 7.2. If $\pi \in C^{\mathrm{I}}(\alpha,\pi)$, then
$\psi_\pi(\alpha)$ is defined; in particular $\psi_\pi(\alpha) < \pi$.

Proof. Note first that for a limit ordinal λ,

$$C^{\mathrm{I}}(\alpha,\lambda) = \bigcup_{\xi<\lambda} C^{\mathrm{I}}(\alpha,\xi)$$

since the right hand side is easily shown to be closed under the clauses that define $C^{\mathrm{I}}(\alpha,\lambda)$. We can pick $\omega \le \eta < \pi$ such that $\pi \in C^{\mathrm{I}}(\alpha,\eta)$. Now define

$$\begin{aligned} \eta_0 &= \sup C^{\mathrm{I}}(\alpha,\eta) \cap \pi \\ \eta_{n+1} &= \sup C^{\mathrm{I}}(\alpha,\eta_n) \cap \pi \\ \eta^* &= \sup_{n<\omega} \eta_n. \end{aligned} \tag{22}$$

Since the cardinality of $C^{\mathrm{I}}(\alpha,\eta)$ is the same as that of η and therefore less than π, the regularity of π implies that $\eta_0 < \pi$. By repetition of this argument one obtains $\eta_n < \pi$, and consequently $\eta^* < \pi$. The definition of η^* then ensures

$$C^{\mathrm{I}}(\alpha, \eta^*) \cap \pi = \bigcup_n C^{\mathrm{I}}(\alpha, \eta_n) \cap \pi \;=\; \eta^* < \pi.$$

Therefore, $\psi_\pi(\alpha) < \pi$. □

Let $\varepsilon_{\mathrm{I}+1}$ be the least ordinal $\alpha > \mathrm{I}$ such that $\omega^\alpha = \alpha$. The next definition singles out a subset $\mathcal{T}(\mathrm{I})$ of $C^{\mathrm{I}}(\varepsilon_{\mathrm{I}+1}, 0)$ which gives rise to an ordinal representation system, i.e., there is an elementary ordinal representation system $\langle \mathcal{OR}, \lhd, \hat{\aleph}, \hat{\psi}, \ldots \rangle$, so that

$$\langle \mathcal{T}(\mathrm{I}), <, \aleph, \psi, \ldots \rangle \;\cong\; \langle \mathcal{OR}, \lhd, \hat{\aleph}, \hat{\psi}, \ldots \rangle. \tag{23}$$

"..." is supposed to indicate that more structure carries over to the ordinal representation system.

DEFINITION 7.3 $\mathcal{T}(\mathrm{I})$ is defined inductively as follows:

1. $0, \mathrm{I} \in \mathcal{T}(\mathrm{I})$.
2. If $\alpha_1, \ldots, \alpha_n \in \mathcal{T}(\mathrm{I})$ and $\alpha_1 \geq \cdots \geq \alpha_n$, then $\omega^{\alpha_1} + \cdots + \omega^{\alpha_n} \in \mathcal{T}(\mathrm{I})$.
3. If $\alpha \in \mathcal{T}(\mathrm{I})$, $0 < \alpha < \mathrm{I}$ and $\alpha < \Omega_\alpha$, then $\Omega_\alpha \in \mathcal{T}(\mathrm{I})$.
4. If $\alpha, \pi \in \mathcal{T}(\mathrm{I})$, $\pi \in C^{\mathrm{I}}(\alpha, \pi)$ and $\alpha \in C^{\mathrm{I}}(\alpha, \psi_\pi(\alpha))$, then $\psi_\pi(\alpha) \in \mathcal{T}(\mathrm{I})$.

The side conditions in 7.3.2, 7.3.3 are easily explained by the desire to have unique representations in $\mathcal{T}(\mathrm{I})$. The requirement $\alpha \in C^{\mathrm{I}}(\alpha, \psi_\pi(\alpha))$ in 7.3.4 also serves the purpose of unique representations (and more) but is probably a bit harder to explain. The idea here is that from $\psi_\pi(\alpha)$ one should be able to retrieve the stage (namely α) where it was generated. This is reflected by $\alpha \in C^{\mathrm{I}}(\alpha, \psi_\pi(\alpha))$.

It can be shown that the foregoing definition of $\mathcal{T}(\mathrm{I})$ is deterministic, that is to say every ordinal in $\mathcal{T}(\mathrm{I})$ is generated by the inductive clauses of 7.3 in exactly one way. As a result, every $\gamma \in \mathcal{T}(\mathrm{I})$ has a unique representation in terms of symbols for $0, \mathrm{I}$ and function symbols for $+, (\alpha \mapsto \Omega_\alpha), (\alpha, \pi \mapsto \psi_\pi(\alpha))$. Thus, by taking some primitive recursive (injective) coding function $\lceil \cdots \rceil$ on finite sequences of natural numbers, we can code $\mathcal{T}(\mathbf{I})$ as a set of natural numbers as follows:

$$\ell(\alpha) = \begin{cases} \lceil 0, 0 \rceil & \text{if } \alpha = 0 \\ \lceil 1, 0 \rceil & \text{if } \alpha = \mathrm{I} \\ \lceil 2, \ell(\alpha_1), \cdots, \ell(\alpha_n) \rceil & \text{if } \alpha = \omega^{\alpha_1} + \cdots + \omega^{\alpha_n} \\ \lceil 3, \ell(\beta) \rceil & \text{if } \alpha = \Omega_\beta \\ \lceil 4, \ell(\beta), \ell(\pi) \rceil & \text{if } \alpha = \psi_\pi(\beta), \end{cases}$$

where the distinction by cases refers to the unique representation of 7.3. With the aid of ℓ, the ordinal representation system of (23) can

be defined by letting $\mathcal{OR}$ be the image of ℓ and setting $\lhd := \{(\ell(\gamma), \ell(\delta)) : \gamma < \delta \wedge \delta, \gamma \in \mathcal{T}(\mathrm{I})\}$ etc. However, for a proof that this definition of $\langle \mathcal{OR}, \lhd, \hat{\Re}, \hat{\psi}, \ldots \rangle$ in point of fact furnishes an elementary ordinal representation system, we have to refer to the literature (cf. Buchholz 1986; Buchholz and Schütte 1988; Rathjen 1994).

7.1.1. *Recursively Large Ordinals*

The large cardinal hypothesis that I is the first weakly inaccessible cardinal is outrageously strong when compared with the strength of KPi. However, it enters the definition procedure of the collapsing function ψ_{I}, which is then employed in the shape of terms to 'name' a countable set of ordinals. As one succeeds in establishing recursion relations for the ordering between those terms, the set of terms gives rise to an ordinal representation system. It has long been suggested that, instead, one should be able to interpret the collapsing functions as operating directly on the recursively large counterparts of those cardinals. To give an example, the ordinal notations used in the determination of the ordinals $|\mathrm{ID}_n|$ for theories of n-iterated inductive definitions (cf. Buchholz et al. 1981) embody collapsing functions $\psi_{\Omega_1}, \ldots, \psi_{\Omega_n}$, which are contingent upon the cardinals $\aleph_1, \ldots, \aleph_n$. The conceptual problem here is that the definition procedure of these functions makes essential use of the set-theoretical universe, whilst the resulting notation system corresponds to a countable, indeed recursive ordinal. Feferman wrote (cf. Feferman 1987, p. 436):

> It has been suggested that, instead, one should be able to interpret the long hierarchies as operating directly on the (Kripke–Platek) admissible number classes τ_α, where $\tau_1 = \omega_1^{rec}$. However, no theory of such classes currently available allows one to 'name' higher admissibles in the definition of a function and have a given admissible such as τ_1 closed under it.

For example, taking such an approach in Definition 7.1 would consist in letting

$$\mathrm{I} := \text{first recursively inaccessible ordinal}$$

and setting $\Re^{\mathrm{I}} := \{\pi < \mathrm{I} : \pi \text{ admissible}, \pi > \omega\}$. The difficulties with this approach arise with the proof of Lemma 7.2. One wants to show that, for all α, $\psi_{\mathrm{I}}(\alpha) < \mathrm{I}$, but the arguments of the cardinal setting no longer work here. To get a similar result for a recursively inaccessible ordinal κ one would have to work solely with κ-recursive operations.

In addition, the functions $\psi_\pi : \varepsilon_{\kappa+1} \to \pi$ would have to be defined for admissible ordinals π with $\omega < \pi < \kappa$. In the cardinal setting this comes down to a simple cardinality argument. To get a similar result for an admissible π one would have to work exclusively with π-recursive operations. How this can be accomplished is far from being clear as the definition of $C^{\mathrm{I}}(\alpha, \rho)$ for $\rho < \pi$ usually refers to higher admissibles than just π. Notwithstanding that, the admissible approach is workable as was shown in Rathjen (1993, 1994), Schlüter (1995). A key idea therein is that the higher admissibles which figure in the definition of $\psi_\pi(\alpha)$ can be mimicked via names within the structure L_π in a π-recursive manner.

The drawback of the admissible approach is that it involves quite horrendous definition procedures and computations, which when taken as the first approach tend to be at the limit of human tolerance.

On the other hand, the admissible approach provides a natural semantics for the terms in the ordinal representation system s. Recalling the notion of *good* Σ_1*-definition* from Definition 6.1, it turns out that all the ordinals of $\mathcal{T}(\mathrm{I}) \cap \mathrm{I}$ possess a good Σ_1-definition in KPi (cf. Rathjen 1994) under the interpretation which takes I to be the first recursively inaccessible ordinal and lets the functions ψ_π operate on admissible ordinals π instead of regular cardinals.

Unlike in the case of KP, $\mathcal{T}(\mathrm{I}) \cap \mathrm{I}$ only forms a proper subset of $\mathtt{spec}_{\Sigma_1}(\mathrm{KPi})$ with many 'holes'.[4] To illuminate the nature of the ordinals in $\mathcal{T}(\mathrm{I}) \cap \mathrm{I}$, it would be desirable to find another property which distinguishes them among the ordinals of $\mathtt{spec}_{\Sigma_1}(\mathrm{KPi})$.

ACKNOWLEDGEMENTS

I am grateful for a stay at the Mittag-Leffler Institute during which I wrote this paper.

NOTES

[1] This contrasts with (Barwise 1975) where Infinity is not included in KP.

[2] $x = \{y \in a : F(y)\}$ stands for the Δ_0-formula $(\forall y \in x)[y \in a \wedge F(y)] \wedge (\forall y \in a)[F(y) \to y \in x]$.

[3] This notion is definable in Zermelo–Fraenkel set theory as long as $\mathcal{F} \subseteq \Pi_n$ for some n. However, if e.g., $\mathcal{F}$ contains all sentences of set theory, then one has to go beyond Zermelo–Fraenkel set theory.

[4] The ordinals of $\mathcal{T}(\mathrm{I}) \cap \mathrm{I}$ are cofinal in $\mathtt{spec}_{\Sigma_1}(\mathrm{KPi})$, though. Letting $\pi_0 := \psi_{\mathrm{I}}\varepsilon_{\mathrm{I}+1}$, one has $\sup(\mathtt{spec}_{\Sigma_1}(\mathrm{KPi})) = \pi_0$.

REFERENCES

Bachmann, H.: 1950, 'Die Normalfunktionen und das Problem der ausgezeichneten Folgen von Ordinalzahlen', *Vierteljahresschrift Naturforsch. Ges. Zürich* **95**, 115–147.

Barwise, J.: 1975, *Admissible Sets and Structures*, Springer, Berlin.

Buchholz, W.: 1986, 'A New System of Proof-Theoretic Ordinal Functions', *Annals of Pure and Applied Logic* **32**, 195–207.

Buchholz, W., S. Feferman, W. Pohlers, and W. Sieg: 1981, *Iterated Inductive Definitions and Subsystems of Analysis*, Springer, Berlin.

Buchholz, W. and K. Schütte: 1988, *Proof Theory of Impredicative Subsystems of Analysis*, Bibliopolis, Naples.

Carlson, T.: 1999, 'Ordinal Arithmetic and Σ_1 Elementarity', *Archive for Mathematical Logic* **38**, 449–460.

Carlson, T.: 2001, 'Elementary Patterns of Resemblance', *Annals of Pure and Applied Logic* **108**, 19–77.

Devlin, K.: 1984, *Constructibility*, Springer, Berlin.

Feferman, S.: 1987, 'Proof Theory: A Personal Report', in G. Takeuti (ed.), *Proof Theory*, 2nd ed., North-Holland, Amsterdam, pp. 445–485.

Feferman, S.: 1988, 'Hilbert's Program Relativized: Proof-theoretical and Foundational Reductions', *The Journal of Symbolic Logic* **53**, 364–384.

Gentzen, G.: 1936, 'Die Widerspruchsfreiheit der reinen Zahlentheorie', *Mathematische Annalen* **112**, 493–565.

Hilbert, D. and P. Bernays: 1938, *Grundlagen der Mathematik II*, Springer, Berlin.

Hinman, P.: 1978, *Recursion-Theoretic Hierarchies*, Springer, Berlin.

Jäger G.: 1982, 'Zur Beweistheorie der Kripke–Platek Mengenlehre über den natürlichen Zahlen', *Archiv für Mathematische Logik* **22**, 121–139.

Jäger, G. and W. Pohlers: 1982, 'Eine beweistheoretische Untersuchung von $\Delta^1_2 - \mathrm{CA} + \mathrm{BI}$ und verwandter Systeme', *Sitzungsberichte der Bayerischen Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse* **(1982)**.

Jensen, R.B. and C. Karp: 1971, 'The Primitive Recursive Set Functions' in D. Scott (ed.): *Axiomatic Set Theory*, *Proc. Symp. Pure Math* **13**, American Mathematical Society, Providence, pp. 143–167.

Möllerfeld, M. and M. Rathjen: 2002, 'A Note on the Σ_1 Spectrum of a Theory', *Archive for Mathematical Logic* **41**, 33–34.

Moschovakis, Y.N.: 1976, *Recursion in the Universe of Sets*, mimeographed note.

Normann, D.: 1978, 'Set Recursion', in Fenstad et al. (eds.), *Generalized Recursion Theory II*, North-Holland, Amsterdam, pp. 303–320.

Rathjen, M.: 1991, 'Proof-Theoretic Analysis of KPM', *Archive for Mathematical Logic* **30**, 377–403.

Rathjen, M.: 1992, 'A Proof-Theoretic Characterization of the Primitive Recursive Set Functions', *Journal of Symbolic Logic* **57**, 954–969.

Rathjen, M.: 1992, 'Fragments of Kripke–Platek Set Theory with Infinity', in P. Aczel, H. Simmons and S. Wainer (eds.), *Proof Theory*, Cambridge University Press, pp. 251–273.

Rathjen, M.: 1993, 'How to Develop Proof-Theoretic Ordinal Functions on the Basis of Admissible Sets', *Mathematical Quarterly* **39**, 47–54.

Rathjen, M.: 1994, 'Collapsing Functions Based on Recursively Large Ordinals: A Well-ordering Proof for KPM', *Archive for Mathematical Logic* **33**, 35–55.
Rathjen, M.: 1994, 'Proof Theory of Reflection', *Annals of Pure and Applied Logic* **68**, 181–224.
Rathjen, M.: 1999, 'The Realm of Ordinal Analysis', in S. Cooper and J. Truss (eds.), *Sets and Proofs*, Cambridge University Press, pp. 219–279.
Richter, W. and P. Aczel: 1973, 'Inductive Definitions and Reflecting Properties of Admissible Ordinals', in J. E. Fenstad and P. Hinman (eds.), *Generalized Recursion Theory*, North Holland, Amsterdam, pp. 301–381.
Sacks, G.E.: 1990, *Higher Recursion Theory*, Springer, Berlin.
Schlüter, A.: 1993, *Zur Mengenexistenz in formalen Theorien der Mengenlehre*, Thesis, University of Münster.
Schlüter, A.: 1995, 'Provability in Set Theories with Reflection', preprint.
Schütte, K. and S. Simpson: 1985, 'Ein in der Zahlentheorie unbeweisbarer Satz über endliche Folgen von natürlichen Zahlen', *Archiv für Mathematische Logik und Grundlagenforschung* **25**, 75–89.
Simpson, S.: 1999, *Subsystems of Second-Order Arithmetic*, Springer, Berlin.
Thiel, N.: 2003, *Metapredicative Set Theories and Provable Ordinals*, Ph.D. thesis, University of Leeds.
van de Wiele, J.: 1982, 'Recursive Dilators and Generalized Recursion', in *Proceedings of Herbrand Symposium*, North-Holland, Amsterdam, pp. 325–332.

Department of Pure Mathematics
University of Leeds
Leeds LS2 9JT
Great Britain
E-mail: rathjen@math.ohio-state.edu

CONTENTS OF VOLUME 148

SYNTHESE / *Volume 148 No. 1 January I 2006*

SYNTHESE / *Volume 148 No. 2 January II 2006*

SYNTHESE / *Volume 148 No. 3 February 2006*

PROOF-THEORETIC SEMANTICS

Editors:
Reinhard Kahle and Peter Schroeder-Heister

AUTHOR INDEX

INSTRUCTIONS FOR AUTHORS

Online Manuscript Submission

Springer now offers authors, editors and reviewers of *Synthese* the option of using our fully web-enabled online manuscript submission and review system. To keep the review time as short as possible (no postal delays!), we encourage authors to submit manuscripts online to the journal's editorial office. Our online manuscript submission and review system offers authors the option to track the progress of the review process of manuscripts in real time.

The online manuscript submission and review system for *Synthese* offers easy and straightforward log-in and submission procedures. This system supports a wide range of submission file formats: for manuscripts – Word, WordPerfect, RTF, TXT and LaTex; for figures – TIFF, GIF, JPEG, EPS, PPT, and Postscript.

Papers for the special *Synthese* section *Knowledge, Rationality & Action* can also be submitted via Editorial Manager. Authors who want to submit an article to *Knowledge, Rationality & Action* can also make use of the online submission and review system. They should indicate that the manuscript is intended for the special section by selecting the article type: "*Knowledge, Rationality & Action* submission" from the drop down menu.

In general, *Synthese* does not publish short book reviews. However, we are interested in receiving longer critical notes on recent publications. We are especially interested in essays in which a number of related publications are critically examined. Please send these directly to the Editor-in-Chief of the journal, Dr. John Symons: jsymons@utep.edu.

Note:
By using the online manuscript submission and review system, it is NOT necessary to submit the manuscript also in printout + disk. In case you encounter any difficulties while submitting your manuscript online, please get in touch with the responsible Editorial Assistant by clicking on "CONTACT US" from the tool bar.

Manuscripts should be submitted to:
www.editorialmanager.com/synt

LaTex

For submission in LaTeX, Springer have developed a Kluwer LaTeX class file, which can be downloaded from the link below. Use of this class file is highly recommended. Do not use versions downloaded from other sites. Technical support is available at: texhelp@springer.com. If you are not familiar with TeX/LaTeX, the class file will be

of no use to you. In that case, submit your article in a common word processor format.
www.springeronline.com/authors/jrnlstylefiles

Reviewing Procedure

Synthese follows a double-blind reviewing procedure. Authors are therefore requested not to include their name or affiliation in their submitted papers. Self-identifying citations and references in the article text should be avoided. Authors should thus make sure that their names and/or affiliations are NOT mentioned on any of the manuscript pages. If authors do include their names on submitted papers, anonymous reviewing cannot be guaranteed.

Manuscript Presentation

The journals language is English. British English or American English spelling and terminology may be used, but either one should be followed consistently throughout the article. Manuscripts should be printed or typewritten on A4 or US Letter bond paper, one side only, leaving adequate margins on all sides to allow reviewers remarks. Please double-space all material, including notes and references. Quotations of more than 40 words should be set off clearly, either by indenting the left-hand margin or by using a smaller typeface. Use double quotation marks for direct quotations and single quotation marks for quotations within quotations and for words or phrases used in a special sense.

Number the pages consecutively with the first page containing:

- running head (shortened title)
- title

Abstract

- Please provide a short abstract of 100 to 250 words. The abstract should not contain any undefined abbreviations or unspecified references.

Figures and Tables

- *Submission of electronic figures*
  In addition to hard-copy printouts of figures, authors are requested to supply the electronic versions of figures in either Encapsulated PostScript (EPS) or TIFF format. Many other formats, e.g., Microsoft Postscript, PiCT (Macintosh) and WMF (Windows), cannot be used and the hard copy will be scanned instead.

  Figures should be saved in separate files without their captions, which should be included with the text of the article. Files should be named according to DOS conventions, e.g., figure1.eps. For vector graphics, EPS is the preferred format. Lines should not be thinner than 0.25pts and in-fill patterns and screens should have a density of at least 10%. Font-related problems can be avoided by using standard fonts such as Times Roman and Helvetica. For bitmapped graphics,

TIFF is the preferred format but EPS is also acceptable. The following resolutions are optimal: black-and-white line figures – 600–1200 dpi; line figures with some grey or coloured lines – 600 dpi; photographs – 300 dpi; screen dumps – leave as is. Higher resolutions will not improve output quality but will only increase file size, which may cause problems with printing; lower resolutions may compromise output quality. Please try to provide artwork that approximately fits within the typeset area of the journal. Especially screened originals, i.e. originals with grey areas, may suffer badly from reduction by more than 10–15%.

– *Avoiding problems with EPS graphics*
Please always check whether the figures print correctly to a PostScript printer in a reasonable amount of time. If they do not, simplify your figures or use a different graphics program.

If EPS export does not produce acceptable output, try to create an EPS file with the printer driver (see below). This option is unavailable with the Microsoft driver for Windows NT, so if you run Windows NT, get the Adobe driver from the Adobe site (www.adobe.com).

If EPS export is not an option, e.g., because you rely on OLE and cannot create separate files for your graphics, it may help us if you simply provide a PostScript dump of the entire document.

– *How to set up for EPS and Postscript dumps under windows*
Create a printer entry specifically for this purpose: install the printer Apple Laserwriter Plus and specify FILE: as printer port. Each time you send something to the printer you will be asked for a filename. This file will be the EPS file or PostScript dump that we can use.

The EPS export option can be found under the PostScript tab. EPS export should be used only for single-page documents. For printing a document of several pages, select Optimise for portability instead. The option Download header with each job should be checked.

– *Submission of hard-copy figures*
If no electronic versions of figures are available, submit only high-quality artwork that can be reproduced as is, i.e., without any part having to be redrawn or retypeset. The letter size of any text in the figures must be large enough to allow for reduction. Photographs should be in black-and-white on glossy paper. If a figure contains colour, make absolutely clear whether it should be printed in black-and-white or in colour. Figures that are to be printed in black-and-white should not be submitted in colour. Authors will be charged for reproducing figures in colour.

Each figure and table should be numbered and mentioned in the text. The approximate position of figures and tables should be indicated in the margin of the manuscript. On the reverse side of each figure, the name of the (first) author and the figure number should be written in pencil; the top of the figure should be clearly indicated. Figures and tables should be placed at the end of the manuscript following the Reference section. Each figure and table should be accompanied by an explanatory legend. The figure legends should be grouped and placed on a separate page. Figures are not returned to the author unless specifically requested.

In tables, footnotes are preferable to long explanatory material in either the heading or body of the table. Such explanatory footnotes, identified by superscript letters, should be placed immediately below the table.

Section Headings

– First-, second-, third-, and fourth-order headings should be clearly distinguishable and numbered. (e.g., 1., 1.1, 1.1.1, 2., 2.1, etc.).

Appendices

– Supplementary material should be collected in an Appendix and placed before the Notes and Reference sections.

Notes

– Please use endnotes rather than footnotes. Notes should be indicated by consecutive superscript numbers in the text and listed at the end of the article before the References. A source reference note should be indicated by means of an asterisk after the title. This note should be placed at the bottom of the first page.

Cross-Referencing

– In the text, a reference identified by means of an authors name should be followed by the date of the reference in parentheses and page number(s) where appropriate. When there are more than two authors, only the first authors name should be mentioned, followed by et al.. In the event that an author cited has had two or more works published during the same year, the reference, both in the text and in the reference list, should be identified by a lower case letter like a and b after the date to distinguish the works.

– Examples:
Winograd (1986, 204)
(Winograd 1986a, b)
(Winograd 1986; Flores et al. 1988)
(Bullen and Bennett 1990)

Acknowledgements

– Acknowledgements of people, grants, funds, etc. should be placed in a separate section before the References.

References

– References to books, journal articles, articles in collections and conference or workshop proceedings, and technical reports should be listed at the end of the article in alphabetical order (see examples below). Articles in preparation or

articles submitted for publication, unpublished observations, personal communications, etc. should not be included in the reference list but should only be mentioned in the article text (e.g., T. Moore, personal communication).

– References to books should include the authors name; year of publication; title; page numbers where appropriate; publisher; place of publication, in the order given in the example below.

  Krantz, D.H., R.D. Luce, P. Suppes, and A. Tversky: 1971, Foundations of Measurement, Vol. 2. Academic Press, New York.

– References to articles in an edited collection should include the authors name; year of publication; article title; editors name; title of collection; first and last page numbers; publisher; place of publication, in the order given in the example below.

  Hintikka, J.: 1966, A Two-Dimensional Continuum of Inductive Methods, in J. Hintikka and P. Suppes (eds), Aspects of Inductive Logic, North-Holland, Amsterdam, pp. 113–32.

– References to articles in periodicals should include the authors name; year of publication; article title full title of periodical; volume number (issue number where appropriate); first and last page numbers in the order given in the example below.

  Lewis, D.: 1984, Putnams Paradox, Australasian Journal of Philosophy 62, 221–236.

– References to technical reports or doctoral dissertations should include the authors name; year of publication; title of report or dissertation; institution; location of institution, in the order given in the example below.

  Sprites, P., R. Scheines, C. Glymour and C. Meek: 1993, TETRAD II. Documentation for Version 2.2.

  Technical Report, Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA.

Proofs

Proofs will be sent to the corresponding author by e-mail (if no e-mail address is available or appears to be out of order, proofs will be sent by regular mail). Your response, with or without corrections, should be sent within 72 hours. Please do not make any changes to the PDF file. Minor corrections (+/− 10) should be sent as an e-mail attachment to: proofscorrection@springer.com.

Always quote the four-letter journal code and article number and the PIPS No. from your proof in the subject field of your e-mail. Extensive corrections must be clearly marked on a printout of the PDF file and should be sent by first-class mail (airmail overseas).

Offprints

25 offprints of each article will be provided free of charge. Additional offprints (both hard copies and PDF files) can be ordered by means of an offprint order form supplied with the proofs.

Page Charges and Colour Figures

No page charges are levied on authors or their institutions. Colour figures are published at the authors expense only.

Copyright

Authors will be asked, upon acceptance of an article, to transfer copyright of the article to the Publisher. This will ensure the widest possible dissemination of information under copyright laws.

Permissions

It is the responsibility of the author to obtain written permission for a quotation from unpublished material, or for all quotations in excess of 250 words in one extract or 500 words in total from any work still in copyright, and for the reprinting of figures, tables or poems from unpublished or copyrighted material.

Springer Open Choice

In addition to the normal publication process (whereby an article is submitted to the journal and access to that article is granted to customers who have purchased a subscription), Springer now provides an alternative publishing option: Springer Open Choice. A Springer Open Choice article receives all the benefits of a regular subscription-based article, but in addition is made available publicly through Springers online platform SpringerLink. To publish via Springer Open Choice, upon acceptance please click on the link below to complete the relevant order form and provide the required payment information. Payment must be received in full before publication or articles will publish as regular subscription-model articles. We regret that Springer Open Choice cannot be ordered for published articles.
www.springeronline.com/openchoice.

Additional Information

Additional Information can be obtained from:

Springer
SYNTHESE
P.O. Box 990
3300 AZ Dordrecht
The Netherlands
Fax: 00 31 (0) 78 6576254