STRUCTURAL LEARNING
II. Issues and Approaches

JOSEPH M. SCANDURA, EDITOR

1976

GORDON AND BREACH, SCIENCE PUBLISHERS

NEW YORK · LONDON · PARIS

Editorial office for the United Kingdom Gordon and Breach Science Publishers Ltd., 42 William IV Street, London W. C. 2, England

Editorial office for France Gordon & Breach, 7-9 rue Emile Dubois, 75014 Paris, France

Library of Congress catalog card number 75-34846 ISBN 0-677-15110-1.
 Printed in Great Britain by Bell & Bain Ltd., Glasgow

from Structural Learning II Issues and Approaches
Jos Scandura (ed.)
New York London Paris 1976
Gordon and Breach Science Publishers

TWO THEORIES OF PROOF

JOHN CORCORAN

> There was, until very lately, a special difficulty in the principles of mathematics. It seemed plain that mathematics consists of deductions, and yet the orthodox accounts of deduction were largely or wholly inapplicable to existing mathematics. Not only the Aristotelian syllogistic theory, but also the modern doctrines of Symbolic Logic, were either theoretically inadequate to mathematical reasoning, or at any rate required such artificial forms of statement that they could not be practically applied. --Russell

This part of the series has a dual purpose. In the first place we will discuss two kinds of theories of proof. The first kind will be called a theory of linear proof. The second has been called a theory of suppositional proof. The term "natural deduction" has often and correctly been used to refer to the second kind of theory, but I shall not do so here because many of the theories so-called are not of the second kind--they must be thought of either as disguised linear theories or theories of a third kind (see postscript below). The second purpose of this part is to develop some of the main ideas needed in constructing a comprehensive[25] theory of proof. The reason for choosing the linear and suppositional theories for this purpose is because the linear theory includes only rules of a very simple nature, and the suppositional theory can be seen as the result of making the linear theory more comprehensive.

1. THEORIES OF LINEAR PROOF

A theory of linear proof is a theory of proof which holds that proofs have a certain simple structure which can be metaphorically called linear.[26] As will be obvious shortly, such theories can be quite plausible a priori but, of course, the comprehensiveness of a theory of proof is an empirical

[25] See par. 2 of the second article in this series.

[26] A system of logic need not contain a theory of proof. There are other purposes for constructing such systems. For example, a logician may be concerned to codify consequences of sets of premises without even considering the problem of describing proofs per se. A system designed for such a purpose was called a consequence system in Corcoran (1969). Many consequence systems are systems of formal deductions which would be theories of linear proof were they put forth as theories of proof--which they usually are not.

3-84 e.g. Quine says in Ph, in effect, that he is not interested in actual proofs but only in procedures that produce consequences of premises. Cf also my remarks about Beeke in Conceptual Structures

matter. As usual, what is plausible a priori turns out to be a gross oversimplification of reality.

Because theories of linear proof are simple and (often) plausible it is perhaps remarkable that the first theory of proof was actually suppositional in nature (cf. my "A Mathematical Model of Aristotle's Logic"). However, theories of linear proof are quite old, tracing their common history at least as far back as Boole. Indeed, Boole (pp. 142, 143) was so clear about his own theory that his description of it can still serve as a concise introduction to the general topic.

> All demonstration essentially consists of the deduction of conclusions from premises,--a conclusion once deduced being itself admissible as a premiss. And it is in this order that reasoning usually proceeds. Certain premises are laid down, either from experience or from testimony, or from some other extralogical source; from these are deduced conclusions which simply or combined with other premises derived from the same class of sources as those first given, serve as bases for further inference, until the chain of argument is completed. At any stage of the process we may find ourselves dealing with two sorts of data, viz., such as have been deduced in the previous course of argument from given data, and such as have not before appeared. A very slight examination of any actual specimen of demonstrative reasoning will show that such are the materials of its composition and such the order of its progress.

In more modern terminology we can say that a linear proof of a conclusion c from a set P of premises is a sequence of lines, beginning with a list of all or some of the premises and such that each subsequent line is derived immediately from premises and/or previously proved lines and, finally, ending with c. In other words a linear proof of c from P is written linearly in a column, say, beginning with the premises P at the top and proceeding step-by-step through intermediate conclusions all derived from P and ending with c the final conclusion. This is the idea; in practice things are a little more complicated, but the following general statement always holds--in a linear proof from premises P to conclusion c each sentence in the proof is a logical consequence of P. (The reader should note that the concept of logical consequence as defined above is not relative to any system of proof.)

There are three minor modifications to be made to the above loose account of linear proofs. The first is that for clarity the premises shall be marked as such to make it clear that they are not asserted to follow from any sentences which they may happen to follow. The second is that in some systems premises may be written at any place in the proof, not just at the top.[27] Finally, in addition to assumptions and inferences,

---

[27] The motivation for allowing new premises to be introduced in the course of a linear proof may be the observation that one often tries to get a conclusion from only some of the available premises but then discovers in the course of constructing the proof that others are needed. From the present point of view this observation is irrelevant and taking cognizance of it may lead to an incorrect theory. Our purpose is not to describe how paths of reasoning emerge in thought but rather to describe how they are described once found. It seems to be a general

properly so-called, all linear systems of proof permit the writing of so-called logical axioms at any point in a proof. For example, in writing proofs in algebra we often have occasion to write logical identities, t = t, in proofs. Corresponding to this we would have a logical axiom rule which permits any proof to be lengthened by addition of a logical identity. For another example, in setting up a "proof by cases" we often write in proof lines of the form 'p or not-p' where p is a sentential formula. Corresponding to this we would have a logical axiom rule which permits any proof to be lengthened by addition of "excluded middle formulas." These two are probably the most prominent logical axioms rules.[28]

Below we will mark premises with a plus sign. Thus '+p' will be read "assume p as a premise" or simply "assume p."

As our example of a theory of linear proof we will give a (non-comprehensive) theory of the proofs found in the abstract algebra of equations--the so-called equational algebras wherein all sentences are either equations or universal generalizations of equations. Following the statement of the rules we will give a proof of the theorem $(x)(x=x^{-1-1})$ [every element is identical to the inverse of its own inverse] from the group axioms.

Rule Set A

Initial String Rule (Kernel Rule)

(1) Premises: A finite sequence of sentences each affixed by + is a proof.[29]

Production Rules

(2) Identity Law: any proof may be lengthened by addition of any logical identity, (t = t) where t is a constant term.
(3) Substitution of "Equals": any proof containing (t = s) and also p may be lengthened by adding p' where p' is the result of replacing occurrences of t in p by s and/or vice versa.
(4) Instances: any proof containing (v)p(v) may be lengthened by adding p(t)--where v is a variable and t is a term composed of constants.
(5) Generalizations: any proof containing p(d), d a dummy[30] constant, may be lengthened by addition of (v)p(v) provided that

---

fact that in actual proofs where premises are made explicit at all they are put down at the beginning. If this is so then any theory which fails to take account of it is, strictly speaking, incorrect regardless of how valid it may be on other grounds.

[28] One way of characterizing the difference between normal reasoning and the so-called Hilbert-type systems of deduction is to note that in the former there are few logical axioms rules but many inference rules whereas in the latter there are commonly many logical axioms rules but few (usually one or two) inference rules. (Cf. Thomason, chapters III, IV, V, esp. pp. 62-63).

[29] Note that for purely heuristic reasons we have tacitly been using the term "proof" in such a way that a partial proof is counted as a proof, thus a finished proof will be a "proof" which satisfies some additional conditions. This issue will be discussed in more detail below. See especially, the discussion of "developments" of axiomatic theories in Section 5.

9-96 Somewhere I believe I said that there are only three axiomatic logical rules (I don't remember the Rand) but now looking at disjunction reasoning it seems clear that whenever P/Q is an instance of an immediate inference rule P/Q is an instance of a corresponding axiomatic rule. Cf Gaps & Remarks on Strict Deduction

no assumptions concern d (i.e., provided d is "arbitrary").

(6) Repetition: any proof may be lengthened by repeating any previous line dropping a '+' if it occurs.

Obviously, each of the above rules corresponds exactly to a rule commonly used in proofs in algebra. Notice however that there are commonly used rules which do not appear in the list. For example, the only way of instantiating here is by rule 4 and this permits the elimination of quantifiers only one per application. This will be an annoying deficiency. Similarly for generalizations. Another deficiency is that substitutions can be done using only one equation at a time. In the proof below we have starred the lines that would remain were the deficiencies eliminated.

$+ (x)(y)(z)(x.(y.z) = (x.y).z)$ *

$+ (x)(x.1 = x)$ *

$+ (x)(1.x = x)$ *

$+ (x)((x.x^{-1}) = 1)$ *

$+ (x)((x^{-1}.x) = 1)$ *

$(y)(z)(a.(y.z) = (a.y).z)$

$(z)(a.(a^{-1}.z) = (a.a^{-1}).z)$

$(a.(a^{-1}.a^{-1-1}) = (a.a^{-1}).a^{-1-1})$ *

$(a^{-1}.a^{-1-1}) = 1$ *

$a.1 = (a.a^{-1}).a^{-1-1}$

$a.a^{-1} = 1$ *

$a.1 = 1.a^{-1-1}$ *

$a.1 = a$ *

$a = 1.a^{-1-1}$

$1.a^{-1-1} = a^{-1-1}$ *

$a = a^{-1-1}$ *

$(x)(x = x^{-1-1})$ *

Having a more powerful instantiating rule would permit going from the associative law directly to the first unquantified line--skipping two lines. The other two unstarred lines would be skipped by doing two substitutions at a time.

Incidentally, the above rule set (or discourse grammar) describes proofs--

[30] Use of the term "dummy" is redundant here; a dummy constant is simply one which does not occur in the premises. Usually the constants

but it does not make explicit what "a proof of c from P" is. Naturally, we define a proof to be a proof of c from P if c is the last line of the proof and all premises in the proof are in P. The above example is a proof of $(x)(x = x^{-1-1})$ from the group axioms.

As the rule set is being used here, the (metalinguistic) symbols p, p(t), p(v), and p(d) refer to formulas in the language of groups. Thus this set of rules presupposed a sentential grammar for the language of groups.[31] However, if we interpreted the symbols as referring to formulas in the arithmetic language, then we could use Rule Set A for the theory of proof needed to complete the Partial Grammar of the Arithmetic Language given at the end of the first article. This would actually be a bit silly for two reasons: first, the Partial Grammar has no quantifiers so rules 6 and 7 would never apply; second, the Partial Grammar does have the logical connectives whereas none of the rules permit any inferences involving connectives. The point, therefore, is not that the Partial Grammar would be finished but rather that the reader can now see what a finished grammar would be like. The respective natures of an alphabet, a rule set for words, a rule set for phrases, and a rule set for sentences are already clear from the Partial Grammar. Now we have also seen a discourse grammar which describes or produces a certain set of proofs. This discourse grammar, Rule Set A, is a theory of proof.

Rule Set A is obviously a correct theory of proof--each of its rules corresponds exactly to (or is) an actual rule of inference that we have all used when doing proofs in elementary group theory. Rule Set A is obviously not comprehensive in the sense that I have defined the term because, e.g., it lacks the complex rules alluded to above which permit the unstarred lines to be omitted. However, it is complete in a certain sense.[32]

---

occurring in premises are given special symbolization: '0,' '1,' 'n,' 'e,' etc.; whereas dummies are indicated by 'a,' 'b,' 'c,' 'd,' or by variables subscripted with a '0,' e.g., x_0. Incidentally, Thomason (chapter IX, esp. p. 183) does not class his rule of generalization with the immediate inference rules. His rule of generalization is sound but, in my opinion, it does not correspond to actual reasoning as closely as does the present rule.

[31] The possibility of obtaining a correct theory of (symbolic) proof depends on having a "correct" symbolic sentential grammar to begin with. Indeed, finding "natural reasoning" blocked by restrictions dictated by peculiarities of the sentential grammar can indicate need for revision of the latter. For example, in the otherwise correct theory of symbolic proof given by Resnik (1970), every proof of Fyy from (x)(y)Fxy involves getting a generalization of Fyy as an intermediate step because of the need to avoid "capturing." Similar situations are common. However, it is possible to design the symbolic language in such a way as to make "capturing" grammatically impossible. This makes it unnecessary to add special restrictions on the rules. Once the symbolic language is thus revised, as in Lemmon (1965), as an unexpected advantage one finds that intrinsically awkward symbolic sentences are eliminated without loss of expressive power.

[32] A theory of proof for a particular language is called equationally complete when the following holds: given any set of equational sentences (either equations properly so-called or universal generalizations thereof) and any single equational sentence c, if c is a logical consequence of P, then there is a proof of c from P constructible by the rules of the theory. Rule Set A is equationally complete. This fact will be plausible to any reader who understands it. To the other readers the following remarks are addressed. Let P be the axioms for groups. Let

In any theory of proof which describes or produces only linear proofs, it is possible to give a very simple description of all proofs from a particular set P of premises to a particular conclusion, c. Given a definition of the logical axioms and the rules one can then say: a proof of P from c is a finite sequence of lines ending with c, each subsequent line of which either is an assumption in P or is a logical axiom or is obtained from previous lines by a rule.

The underlined expression (or rather an even simpler version of it) has become a slogan and, sometimes, a battlecry. One eminent logician related to me that when he first heard this slogan presented he was struck by its simplicity and truth and was moved to say to himself, "By God, that is what proofs are!"

If one takes the slogan as a rough description of all proofs, then one is led (1) to distinguish three kinds of rules of inference and (2) to believe that all rules of inference must be of one of the three kinds. The first kind contains only the rule of assumption--essentially to the effect that an assumption may be written to start (or to lengthen) any proof provided that it is marked as an assumption. The second kind contains all logical axiom rules--to the effect that a logical axiom may be written to lengthen any proof. The third kind contains all immediate inference rules; rules which state that any proof containing one or two (or some fixed finite number of) sentences of certain specified forms may be lengthened by adding a sentence in another form.

2. IMMEDIATE RULES AND SUBSIDIARY PROOF RULES

It so happens that by surveying the proofs in the mathematical literature (or by looking at our own proofs) we find many rules that are not of any of the above three kinds. Indeed, if all rules were of the three above kinds then there would be no room in mathematical reasoning for making subsidiary assumptions. Much of the most elegant and enlightening reasoning in mathematics turns on the ability to imagine good subsidiary assumptions. Below are some examples. (1) In proving that the square root of two is not rational, we assume, in addition to the axioms of arithmetic, the subsidiary assumption that the square root of two is rational.[33] (2) In proving the right cancellation law $[(x)(y)(z)((x.z = y.z) \supset x = y)]$ from the group axioms, we assume, in addition to the group axioms, that $a.d = b.d$ where a, b and d are arbitrarily chosen but fixed elements of the group. (3) Whenever we give proofs by cases after we have proved that there are two cases, say, we assume that the first case holds and then prove our theorem in that case, then we assume the second case and prove our theorem in that case--finally we conclude that the theorem holds in general.... In each of these three examples the proof involves making subsidiary assumptions, assumptions other than those from which the conclusion is shown to follow.

---

c be any equational sentence written in the language of groups and which is true in all groups. c, then, is a logical consequence of P; since (1) a group is by definition any mathematical system in which the axioms of groups are true and (2) to say that c is a logical consequence of P is to say that c is true in any mathematical system which makes all of the sentences in P true. The above-mentioned completeness condition implies, then, that by using Rule Set A one can construct a proof starting with P as assumptions (as in the example) and ending with c. In fact, such a proof can be gotten by lengthening the one given as a sample.

At some point in each of these examples an inference is made not from certain previous lines in a proof but rather from (or on the basis of) a certain part of the proof. In other words, there are rules which can be stated as follows: any proof containing a subsidiary proof of a certain form may be extended by adding p. For example, in reductio reasoning we are following the rule: any proof containing a subsidiary proof beginning with p and containing a contradiction may be extended by adding ~p (not-p).

A subsidiary proof begins with a subsidiary assumption, a "new" assumption made for purposes of reasoning. The subsidiary assumption is marked with a "beginning" corner bracket '⌜'. Thus '⌜p' may be read "for purposes of reasoning suppose p" or simply "suppose p." When the subsidiary reasoning is completed one adds a "closing" or "ending" corner bracket '⌟' to the last line. Each time an ending bracket is added it is matched with the last beginning bracket not yet matched. The latter is always on the line containing the supposition which begins the subsidiary proof in question. Thus a subsidiary proof may be defined as a section of a proof enclosed in matching brackets. The details, if not already clear, will be so after considering a couple of examples.

Two paragraphs back we stated the reductio rule. We now give as an example an indirect (reductio) proof of $\sim(x) \sim (x = x^{-1})$ [not every element is different from its own inverse] from the group axioms.

$+ (x)(y)(z)((x.(y.z)) = ((x.y).z))$

$+ (x)(x.1 = x)$

$+ (x)(1.x = x)$

$+ (x)(x.x^{-1} = 1)$

$+ (x)(x^{-1}.x = 1)$

$\ulcorner (x) \sim (x = x^{-1})$

$\sim(1 = 1^{-1})$ *

$1.1^{-1} = 1$

$1.1^{-1} = 1^{-1}$

$\llcorner 1 = 1^{-1}$ *

subsidiary proof

$\sim(x) \sim (x = x^{-1})$

The subsidiary proof is enclosed in matching brackets. The contradiction in question is "between" the starred lines. Notice that the conclusion is inferred to follow from the group axioms (not from all assumptions) on

---

The equational completeness of Rule Set A was proved several years ago by Jan Kalicki and Dana Scott (1955).

[33]For a wide-ranging discussion of this particular proof in the general context of a concern with the history and the soundness of indirect reasoning see Cauman (1966).

the basis of the subsidiary proof. Once a subsidiary proof is marked off by an ending bracket (⌊), it must be regarded as an isolated, separate unit in the proof. In particular, one may no longer apply any of the immediate inference rules to lines inside of the subsidiary proof. For example, we could not write down as a next line $\sim(1 = 1^{-1})$ by repetition because this does not follow from only the group axioms.

Let us use the phrase 'subsidiary proof rule' to refer to rules which permit the lengthening of a proof on the basis of a subsidiary proof. Of course, the most notorious of subsidiary proof rules is the rule of conditionalization which permits inference of 'if p then q' on the basis of a subsidiary proof beginning with p and ending with q. We will give a proof of the right cancellation law from the group axioms to illustrate this. (In the proofs below we do not necessarily follow Rule Set A but use other commonly known rules as well.)

$+ (x)(y)(z)((x.(y.z)) = ((x.y).z))$

$+ (x)(x.1 = x)$

$+ (x)(1.x = x)$

$+ (x)(x.x^{-1} = 1)$

$+ (x)(x^{-1}.x = 1)$

$a.d = b.d$	
$(a.d).d^{-1} = (b.d).d^{-1}$	Subsidiary Proof
$a.(d.d^{-1}) = b.(d.d^{-1})$	
$a = b$	

$(a.d = b.d) \supset (a = b)$

$(x)(y)(z)((x.z = y.z) \supset x = y)$

It will be valuable to notice that in proofs by cases more than one subsidiary proof is needed--one for each case. Actually, all proofs-by-cases-rules are "combinations" of the two-case rule stated as follows: any proof containing 'c_1 or c_2', together with two subsidiary proofs, one beginning with c_1 the other beginning with c_2 both ending with c, can be extended by adding c. To illustrate this we will give a proof of the two-sided cancellation law. The proof will involve one application of the two-case rule inside of a subsidiary proof on which conditionalization is used.

$+ (x)(y)(z)((x.(y.z)) = ((x.y).z))$

$+ (x)\ (x.1 = x)$

$+ (x)\ (1.x = x)$

$+ (x)\ (x.x^{-1} = 1)$

$+ (x)\ (x^{-1}.x = 1)$

$\ulcorner(a.d = b.d) \vee (d.a = d.b)$	$-c_1$ or c_2
$\ulcorner a.d = b.d$	
$(a.d).d^{-1} = (b.d)d^{-1}$	first subsidiary proof
$a.(d.d^{-1}) = b.(d.d^{-1})$	
$\llcorner \quad a = b$	
$\ulcorner d.a = d.b$	
$d^{-1}.(d.a) = d^{-1}.(d.b)$	secondary subsidiary proof
$(d^{-1}.d).a = (d^{-1}.d).b$	
$\llcorner \quad a = b$	
$\llcorner \quad a = b$	-cases rule*
$((a.d = b.d) \vee (d.a = d.b)) \supset a = b$	-conditionalization**
$(x)(y)(z)(((x.z=y.z)\vee(z.x=z.y)) \supset x=y)$	

The notations on the right are designed to help the reader see exactly where and how the two subsidiary proof rules are applied (* and **).

Before we proceed to a discussion of theories of suppositional proof (theories involving subsidiary proof rules), the reader should note that the above three proofs are not linear because the subsidiary assumptions are not among the premises from which the proof proceeds and neither are they consequences of the premises. That is, for example, in the proof of the cancellation law from the group axioms there are sentences which are <u>not</u> logical consequences of the group axioms. Thus in these proofs we do not reason in a linear fashion--we take "side trips."

3. THEORIES OF SUPPOSITIONAL PROOF

The defining characteristic of a theory of suppositional proof is that the rules permit the use of subsidiary assumptions which are later "discharged" and are not among the assumptions from which the final conclusion is shown to follow. These rules are subsidiary proof rules which countenance an inference <u>not</u> from previous lines but rather on the basis of a subsidiary proof. Such rules are not unusual but rather they comprise the essence of clear, elegant mathematical reasoning. Indeed, I think the mathematically experienced reader will agree that linear proofs have a very computational flavor to them, whereas suppositional proofs seem to embody more creative and enlightening reasoning.

There are a few questions concerning the formulation of suppositional rules which might have been annoying some readers. I will digress slightly at this point to take up some of them.

In the first place we must give an explicit rule for adding subsidiary

assumptions: rule of supposition--any proof can be lengthened by addition of a formula prefixed by a beginning bracket. Secondly the following rule explicitly accounts for introduction of closing brackets: closing rule--any proof containing more beginning brackets than ending brackets can be modified by affixing a closing bracket to its last line. The idea is that each supposition line, ⌜p, starts a subsidiary proof and that each subsidiary proof must start with the last supposition line not already a part of another subsidiary proof. Each time an ending bracket is put into a proof there is exactly one beginning bracket with which to match it--namely the last one not already matched.[34]

A subproof of a proof is a sequence of lines beginning and ending with matched brackets. An occurrence of a sentence is inactive in a proof if it occurs within a subproof. An occurrence is active if not inactive. A sentence is active in a proof if it has an active occurrence therein. A subproof occurrence in a proof is inactive if it occurs within another subproof. An occurrence of a subproof is active if not inactive. A subproof is active in a proof if it has an active occurrence in the proof.

All rules must be stated so that they apply only to sentences and subproofs which are active.

A given proof is a proof of its last line (if active) from its set of active assumptions (premises plus active suppositions). Now we can state two important general principles for suppositional proofs.

Let p be a given line of a suppositional proof. (1) The sequence of lines up to and including p is itself a proof. Let us call this the partial proof ending with p. (2) In any suppositional proof, each given line p is a logical consequence of the active assumptions of the partial proof ending with p. (If p is prefixed by a ⌜, the ⌜ counts as in the partial proof--if by ⌞ the ⌞ does not count as in the subproof).

Now we can define a finished proof to be one which satisfies the following two conditions: first, it contains no active subsidiary assumptions; second, it ends with an active sentence. The first condition guarantees that any reasoning for purposes of which a supposition has been made is completed. The second condition allows a subsidiary proof rule to be applied after the last subsidiary proof has been completed. This definition includes every proof that one would want to count as finished and excludes most unfinished proofs, but it still counts as "finished" certain proofs which one would not wish to consider as such. A more adequate definition would involve intricacies undesirable in an article of this sort (but see below for an easy improvement).

It is obvious that the framework of a suppositional theory is much more adequate for characterizing mathematical proofs[35] than is the framework of a linear theory--even though anything that can be proved in a given suppositional theory will also admit of proof in some linear theory. In other words, we are not contrasting the abstract power of such theories but rather their relative adequacies in characterizing the proofs which we actually write. Given the advantage of suppositional theories we can ask: are there other kinds of rules of proof which could be added and

[34]Proofs in suppositional theories have the abstract form of nest structures in the sense of Smullyan (1965).

[35]In particular, there are many proofs which cannot be accounted for except in a suppositional theory.

which would constitute an even more adequate framework? Let us put this question another way. Besides the premises rule and the closing rule we have seen four kinds of rules of inference: (1) assumption rules, (2) logical axiom rules, (3) immediate inference rules and (4) subsidiary proof rules. Are there other kinds of rules which are actually employed in writing of proofs?

The most obvious kind of rule to suggest adding is a rule that permits the writing of "goals." Frequently when we are writing a proof, after some assumptions (premises and/or suppositions) have been entered, we indicate our goal by writing, for example, "we want to show p." This is actually a very handy device which helps convey the reasoning to be expressed in the proof. Since the purpose of proofs is to express reasoning we should certainly consider such a rule. We could state it: Any proof may be lengthened by adding ?p. The question mark in this context could be read "to prove," say.[36] We would then have to define all occurrences of ?p as inactive because otherwise we would be applying immediate rules to what we were trying to prove--thus begging the question.

Now let us consider another important kind of rule. We have actually given an example of this kind of rule, but we did not classify it. Notice that all of the above kinds of rules apply only to a part of a proof to which they apply, i.e., it is usually unnecessary to look at each line in the whole proof in order to apply any of the above four kinds of rules--supposition does not require looking at any lines, the same for logical axiom rules, immediate inference rules involve only fixed finite numbers of lines, subsidiary proof rules involve perhaps a few subsidiary proofs plus perhaps a few active lines. The rule of generalization, however, requires looking at a particular line p(d) and then checking through the whole proof to determine that nothing has been assumed about d--i.e., that d is indeed arbitrary ('d' is dummy). Such rules we call <u>global immediate rules</u>. Thus, the classification of linear rules above was inadequate.

In addition there are subsidiary rules which involve reference to the entire proof to which they are applied. The most prominent example of a <u>global subsidiary rule</u> is the rule that is generally used in reasoning from an existentially quantified statement. For example, suppose that we have assumed the right cancellation law in a proof and we are aiming to prove $(\exists x)(y)(y.x=x) \supset (x)(x=x^{-1})$. We assume the antecedent $(\exists x)(y)(y.x=x)$ and we say "let x_0 be such an object." ("Let" is a sure sign of an assumption.) We are assuming that x_0 is an "arbitrary object" satisfying the condition $(y)(y.x_0 = x_0)$. We reason then of an (genuinely) arbitrary b that $b.x_0 = x_0$ and that $b^{-1}.x_0 = x_0$. Then, using the cancellation law, infer $b = b^{-1}$. Since b is arbitrary, $(x)(x = x^{-1})$. Now we say: "Since x_0 was arbitrary and $(x)(x = x^{-1})$ does not depend on x_0, the

---

[36] This rule may profitably be compared with a similar device of Kalish and Montague (pp. 14ff) which involves writing 'show p' to indicate a goal and which requires the 'show' to be crossed out once "the goal has been reached." As useful and valid as this device surely is, it is not correct in our sense because it violates the principle that every subproof of a (partial) proof is itself a (partial) proof. The latter is a rough statement which corresponds to the apparent facts that we do not alter previously written (partial) proofs and that we read them "top to bottom" checking each line as encountered. The Kalish-Montague device may correspond better to a description of how proofs "emerge in thought" which, of course, is not our goal.

conclusion follows from the original assumption." This corresponds, in the below formalized version, to taking $(x)(x = x^{-1})$ out of the subsidiary proof and making it active [starred line].

$+ (x)(y)(z)((x.z=y.z) \supset x=y)$

$?(\exists x)(y)(y.x=x) \supset (x)(x=x^{-1})$

$\ulcorner(\exists x)(y)(y.x=x)$

$\quad\ulcorner(y)(y.x_0=x_0)$ "let x_0 be such an object"

$\quad\quad b.x_0 = x_0$

$\quad\quad b^{-1}.x_0 = x_0$

$\quad\quad b.x_0 = b^{-1}.x_0$

$\quad\quad (b.x_0 = b^{-1}.x_0) \supset b = b^{-1}$ (cancellation law)

$\quad\quad b = b^{-1}$

$\quad\llcorner(x)(x=x^{-1})$

$\llcorner\ (x)(x=x^{-1})$ *

$(\exists x)(y)(y.x=x) \supset (x)(x=x^{-1})$

It might be worthwhile to do another example using the above rule. We will prove $(y)((\exists x)(Dx\&Hyx) \supset (\exists z)(Az\&Hyz))$ from $(x)(Dx \supset Ax)$.

$+ (x)(Dx \supset Ax)$

$\ulcorner(\exists x)(Dx\&Hbx)$

$\quad\ulcorner(Da\&Hba)$ "let a be such an object"

$\quad\quad Da$

$\quad\quad Da \supset Aa$

$\quad\quad Aa$

$\quad\quad Hba$

$\quad\quad Aa\ \&\ Hba$

$\quad\llcorner(\exists z)(Az\&Hbz)$

$\llcorner(\exists z)(Az\&Hbz)$ *

$(\exists x)(Dx\&Hbx) \supset (\exists z)(Az\&Hbz)$

$(y)((\exists x)(Dx\&Hyx) \supset (\exists z)(Az\&Hyz))$

The rule just exemplified could be called "existential instantiation" because it involves "instantiating" as existential statement to begin the

subsidiary proof.

Often in writing a proof after a pair of contradictions have been proved (made active) we write 'a contradiction' and it is on the basis of that notation that we apply the reductio rule. Thus it is necessary (for comprehensiveness) to add a special symbol, say X, to the language of proofs.[37] "X" can be read 'A contradition.' The rule of contradiction introduction is the following: any proof which contains active sentences p and not-p may be lengthened by addition of X. Given this we can now state two new reductio rules: any proof which ends in a subsidiary proof beginning p (respectively ~p) and ending with X can be lengthened by addition of ~p (respectively p).

The usual proof of Russell's theorem [no set contains exactly the sets not containing themselves] involves all three of the rules just mentioned together with the subsidiary proof rule of "existential instantiation." It should be mentioned that Russell's theorem is proved without the use of premises--it is proved using logic alone. For this reason it is often counted as a "law of logic"--indeed, its denial implies a contradiction.

$?\sim(\exists x)(y)(x\epsilon y \equiv \sim(y\epsilon y))$

$\ulcorner(\exists x)(y)(x\epsilon y \equiv \sim(y\epsilon y))$

$\ulcorner(y)(x_0\epsilon y \equiv \sim(y\epsilon y))$ "let x_0 be such an object"

$x_0\epsilon x_0 \equiv \sim(x_0\epsilon x_0)$

$\ulcorner x_0\epsilon x_0$

$\sim x_0\epsilon x_0$

$\llcorner X$

$\sim x_0\epsilon x_0$

$x_0\epsilon x_0$

$\llcorner X$ "but x_0 was arbitrary"

$\llcorner X$

$\sim(\exists x)(y)(x\epsilon y \equiv \sim(y\epsilon y))$

Because of limitations of space we merely mention a class of rules called definitional rules which actually form a subclass of the global subsidiary rules and which, as can be surmised from the name, countenance the use of nominal definitions within proofs.

As a final question we consider the nature of an axiomatic development of a mathematical theory. An axiomatic development of a theory begins with the axioms. Subsequently the first theorem is proved, then the

---

[37] Linguistically, this may be a radical move. We are adding to the "sentences" used in discourses something that does not appear in the underlying language.

second, then the third, etc. However, after the first proof the axioms are not repeated. Moreover, in addition to the axioms, previously proved theorems are also used as new "axioms"--but these are generally not rewritten either. One way of characterizing such a development is to say that it is one long proof and that axioms and previously proved theorems can be used because they are already active above. There is something artificial about this characterization--we usually say that a development of a theory contains many proofs, here we say that it is just one long proof. It is obvious that there is a level above the level of proofs--a level containing "axiomatic developments" which, in a sense, are composed of proofs. This implies that in a development of a theory there is structure which is not reducible to the structure of proofs.[38] Thus there are at least two levels of language above the sentential level.

4. SUMMARY OF SUPPOSITIONAL THEORIES

We have seen that linear theories contain four kinds of rules: premises, logical axiom, immediate inference, and global immediate inference. Next, we noticed that suppositional theories contain two additional kinds of rules: subsidiary proof rules and global subsidiary proof rules. It is important to realize that relative to linear systems both kinds of subsidiary rules are radical innovations because they countenance inferences not based on previously proved sentences but rather on the basis of previously performed patterns of reasoning.[39] In addition, we pointed out that the definitional rules are merely a species of the global subsidiary proof rules.

We explained the concept of an active sentence[40] in a proof and we asserted that the general principle behind suppositional proofs has two parts: (1) that given a proof and a sentence occurrence p in the proof, the part of the proof ending with p is also a proof (called the partial proof ending with p) and (2) each such p is a logical consequence of the active assumptions of the partial proof ending with p. Given this principle, the notation for subsidiary proofs, and the classification of the rules,

---

[38] In a development of an axiomatic theory each theorem and each lemma is a "main goal" and within the course of deduction of a main goal one often choses "intermediate goals" in order to focus on the local direction of the reasoning. Several things follow. The first is that one needs at least two "goal indicators," one for main goals and one for intermediate goals. One way of handling this is to use a single question mark to indicate a main goal, two question marks to indicate a conclusion to be reached in proving a main goal, (perhaps) three question marks to indicate a conclusion to be reached in proving a "level-two" goal, etc. The second is that the notion of a "finished proof" must be modified in order that a proof is counted as finished only if all of its goals have been reached in the required order. As each subsequent theorem or lemma has been reached the entire proof up to that point should be finished and it may be necessary to have a special symbol to indicate the end of a finished proof. Indeed many current authors use such symbols. Kelley (1955) uses a small shaded rectangle which he attributes to Halmos; Suppes (1960) uses the traditional 'Q.E.D.;' and Dean (1966) uses a triple asterisk. For further discussion of the structure of a development of an axiomatic theory see my "A Mathematical Model of Aristotle's Syllogistic."

[39] For a more detailed discussion see my "Three Logical Theories."

[40] See Section 3 above.

anyone having a background in mathematics is prepared to formulate his own theory of proof.[41]

5. SUMMARY OF THE SERIES

In the interest of accuracy we must admit that the obvious heuristic value of the notion of a partial proof probably refutes the hypothesis that the class of discourses has a kernel/transformations structure. The proof discourses are clearly the "finished proofs" and it does not seem to be the case that these have the requisite structure: one does not build up finished proofs by applying "natural" transformations to other finished proofs. Indeed, it seems to be generally the case with discourses that the beginning of a discourse is not itself a discourse but rather it seems that the beginning of a discourse makes "promises" which must be fulfilled later in order for the discourse to be "finished." When we put down some axioms and "a goal" (see above), that proof is not finished until the goal has been reached. Likewise with discourses, generally. For example, if someone were to say, "I have called this meeting to give you my views on the latest crisis," and then sat down, he would not have uttered a complete discourse. There are innumerable similar examples. The conclusion that the class of discourses fails to have a kernel/transformations structure seems inescapable.

In part I we discussed some fundamental concepts involved in the analysis of mathematical reasoning. In addition, we introduced the concept of levels of language and pointed out that a grammar of an entire language should be composed of several grammars, one at each level. We also made the point that a proof is a certain kind of discourse which, in turn, suggested the possibility of a theory of proof--a discourse grammar which describes the proofs of a language.

In part II we outlined what a theory of proof would be like. We noted that the grammatical rules used in describing proofs are the rules of inference according to which we write proofs. We discussed the nature of our knowledge of rules of inference distinguishing weak and strong varieties of such knowledge. Finally, we speculated concerning the utility of a theory of proof vis-a-vis improvements in mathematical education.

In the course of Part III, we contrasted what has become the traditional theory with a newer and more adequate theory whose essential features were discovered in the 1920's (Jaskowski). The older theory holds that mathematical reasoning proceeds from axioms step-by-step to conclusions

---

[41] In mathematical logic one constructs a precisely defined mathematical analog (formal deductive system) of a system of proofs and a precise mathematical analog (formal semantic system) corresponding to the (actual or imagined) system of interpretations associated with the language. In this way the philosophical problem of the soundness of a system of proofs is replaced by a precise mathematical problem. The form of the main lemma in a soundness proof for a system of linear proofs is this: for every proof π the assumptions of π taken together imply each sentence in π. In my opinion the form of the corresponding lemma for any correct theory of suppositional proofs is this: for every proof π the active assumptions of π taken together imply each active sentence of π. This opinion, if correct, will account for the feeling of strangeness encountered in trying to construct proofs in the system of Quine's Methods (pp. 159-167).

in a strictly linear fashion; i.e., each step in a proof must be a logical consequence of the axioms. Apparently this view was first systematized by Boole in the nineteenth century. It became the commonly accepted view until the 1920's when Lukasiewicz pointed out in his seminar that the theory did not agree with mathematical practice. Jaskowski, who was a student in the seminar, accepted the project of developing the exact details of a theory of proof which would take into account the salient features of mathematical reasoning not accounted for by Boole's theory. The newer theory is largely the result of Jaskowski's effort. The older theory we called linear, the newer suppositional.

We gave several examples of rules and proofs with the intention of supplying enough detail so that the basic ideas can be grasped in a useful way.

6. POSTSCRIPT

The linguist and the logician will doubtless disagree with many of the above assertions. Several serious oversimplifications have been made--mostly concerning linguistics. My hope has been to show the overlap and possible cross-fertilization between, on the linguistic side, the ideas of Harris and Chomsky and, on the logical side, the ideas of Jaskowski. I have tried to do this in a way that would be of benefit to persons of diverse backgrounds. I was trying to write to an audience of mathematics educators, linguists, mathematicians, psychologists, and logicians.

One final technical point: the so-called natural deduction systems found in books by Suppes, Lemmon, and Mates are <u>not</u> theories of suppositional proof. By looking carefully at each of them, one notices that the lines of their proofs are <u>not</u> sentences, but rather ordered pairs (P,c) where P is a set of "premises" and c is a single sentence. Moreover, a grammar to generate their proofs takes the <u>form</u> of a linear theory without <u>any</u> assumptions. In particular, in each of these systems each proof is a finite sequence of lines (P_1, c_1), (P_2, c_2), ..., (P_n, c_n) where each subsequent line is either (axiomatically) of the form $(\{c\}, c)$ or else is the result of applying an immediate rule to a fixed, finite number of preceding lines. An example of such a rule would be: if (P_i, d) and $(P_j, d \supset c)$ are lines in a proof, then the proof can be lengthened by writing $(P_i + P_j, c)$. The idea behind constructing a proof of c from P in these systems is not to try to deduce c from P, but rather to construct the ordered pair (P,c) starting initially from ordered pairs $(\{x\}, x)$ using rules which when applied to "valid arguments" produce "valid arguments." In a word, these systems stack-up valid arguments starting with the simple and building to the complex. As far as either the characterization of normal reasoning or utility in teaching is concerned, it seems to me that none of these systems fares well in comparison to a suppositional system as found in the following: Anderson and Johnstone (1963), Kalish and Montague (1964), Leblanc (1966), or Thomason (1970).

Acknowledgements

This work originated as a talk given at the Conference on Mathematical and Structural Learning held at the University of Pennsylvania in April of 1968. I wish to thank Dr. Joseph Scandura for inviting me to what developed into a valuable conference. If the final version is a material improvement over the talk, then Professors James Greeno, Paul Rosenbloom

and Joseph Scandura deserve credit for their suggestions and criticism (not all of which I had wisdom to agree with). I also wish to acknowledge the fact that had I not been fortunate enough to receive a Summer Research Fellowship (NSF-1G-68-3) from the National Science Foundation through the auspices of the University of Pennsylvania, then I likely would not have written these pages. I gratefully acknowledge the helpful and sympathetic criticism that I have received from the students in my logic seminar. Especially significant in regard to this work were the ideas of John Herring, William Frank, Edward Keenan and George Weaver. Finally, I acknowledge ideas received in private communication separately from Mr. James Munz, Linguistics Project, University of Pennsylvania and Professor J. J. LeTourneau, Mathematics Department, Hampshire College. I wish to dedicate this work to the memory of Albert L. Hammond and Ludwig Edelstein; both late of Johns Hopkins University. These two men were instrumental not only in bringing me to appreciate the search for truth but also in making practical arrangements in order that I could pursue graduate studies. Both had tragic lives, one was harassed by philistines in American higher education, the other by Nazi's in Germany as well--but they both fought the good fight and neither lost faith in the ultimate value of truth and kindness.