OUTLINE OF AN INTRODUCTION
TO MATHEMATICAL LOGIC I

Abraham Robinson

1. _Preliminary_ Remarks. This is the first of (presumably) three articles on the subject mentioned in the title. The exposition is based on a course of fifteen lectures which formed part of the Edmonton (1957) Seminar of the Canadian Mathematical Congress. Limitations of space (and, originally, of time) compel us to be selective in two directions. First, while we shall refer to other branches of logic in passing, we shall be concerned principally with the two fundamental calculi - of propositions and of predicates (of the first order). Thus, except for a number of modern developments which are included here, our exposition will be similar in scope, though not in detail, to the first and third chapters of the well-known "Principles of Mathematical Logic" by D. Hilbert and W. Ackermann (English translation, Chelsea, New York, 1950) and this was in fact the recommended text for the Edmonton course. However, there exists a growing number of other good introductions to the field and some of these will be listed later.

Secondly, we shall not prove all the stated theorems in detail. However, by proving a selected number of our assertions, we shall endeavour to ensure (in most cases) that the reader will have no difficulty in working out detailed proofs for himself.

Modern Mathematical Logic (to give just one of its possible definitions) is the study of the laws of thought by means of the methods, including the symbolism, of Mathematics. As such it is not in any way in conflict with Classical (Aristotelian) Logic, but it is demonstrably more comprehensive than the latter. However, the historical aspects of the subject will not be considered here.

2. The Calculus of Propositions. Truth functions. Whatever the absolute character of the laws of thought, we cannot close our eyes to the fact that they appear to us in the first instance in our own subjective thinking. It is therefore only natural that we include, to begin with, some concrete examples from

ordinary life. Thus, consider the following sentences.

> A. All bears are tame.
> B. Not all bears are tame.
> C. Mountain climbing is more enjoyable than attending lectures.
> D. Some of the members of a recent mathematical convention spent more time outdoors than indoors.
> E. Mountain climbing, etc. (see C) and some of the members, etc. (see D).

Comparing B with A, we see that the truth or falsehood of B depends only on the truth or falsehood of A (if A is true then B is false; if A is false then B is true). Similarly, the truth or falsehood (briefly, the truth value) of E depends only on the truth values of C and D. In neither case is this connection affected by the correctness, or otherwise, of the sentences in question.

The Calculus of Propositions is concerned with the inter-relation of the truth values of propositions (sentences, statements) which may be taken from everyday life, or from Mathematics, or from any other subject. A moment's reflection shows that, in view of the fact that this interrelation is supposed to involve only the truth values of the sentences concerned, we may develop our theory simply in terms of functions whose domain and range are truth values, i.e. the two values T (truth) or F (falsehood). This leads to the

DEFINITION. A truth function

$$q = f(p_1, \ldots, p_n), \quad n \geq 1$$

is a function whose arguments range over the elements of the two-element set $V = \{T, F\}$ and which takes the values in the same set.

Thus, a truth function of n variables $p_1, \ldots, p_n$, is defined completely by assigning truth values, T or F, to the function for given truth values of the arguments. Since there are just 2^n different n - tuples of truth values for the arguments, it follows that there are just 2^{2^n} different truth functions of n variables. In particular, there are just four different truth functions for n = 1. They are given by the following tables.

p	T	F
f(p)	T	T

p	T	F
f(p)	F	F

p	T	F
f(p)	T	F

p	T	F
f(p)	F	T

In the third case we adopt a convention which is customary in Mathematics, and write $f(p) = p$, even though there is strictly speaking a difference between a variable and a function which takes the same values as the variable throughout. The first and second case are written as $f(p) = T$ and $f(p) = F$, with the qualification "for all p" whenever there is room for a misunderstanding. Finally the fourth function is denoted by $\sim p$ (read "non-p" or "not-p"). It expresses the connection between A and B in the set of concrete examples given above.

The number of truth functions increases rapidly with increasing n. For $n = 2$, it is customary to specify the truth functions by a standard type of <u>truth</u> <u>tables</u> as follows. The truth table
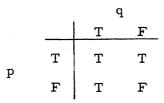
		q	
		T	F
p	T	V_{11}	V_{12}
	F	V_{21}	V_{22}

where Vik stands for T or F

refers to the truth function

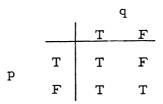$$f(T, T) = V_{11} \quad f(T, F) = V_{12} \quad f(F, T) = V_{21} \quad f(F, F) = V_{22} .$$

A number of important truth functions are given below. The first one expresses the connection between C and D on one hand, and E on the other, in our set of concrete examples.

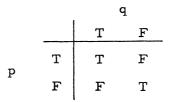2.1. <u>Conjunction</u>, symbol $p \wedge q$, read "p and q".

		q	
		T	F
p	T	T	F
	F	F	F

43

2.2. <u>Disjunction</u>, symbol $p \lor q$, read "p or q"

	q: T	q: F
p: T	T	T
p: F	T	F

2.3. <u>Implication</u>, symbol $p \supset q$, read "p ply q"

	q: T	q: F
p: T	T	F
p: F	T	T

2.4. <u>Biconditional</u> or equivalence, symbol $p \equiv q$, read "p equivalent to q"

	q: T	q: F
p: T	T	F
p: F	F	T

2.5. <u>Incompatibility</u> or Sheffer stroke, symbol $p \mid q$, read "p incompatible with q".

	q: T	q: F
p: T	F	T
p: F	T	T

The symbols which we have introduced in connection with the above truth functions, i.e. $\sim , \land , \lor , \supset , \equiv , \mid$ are called propositional connectives. It will be seen that the nomenclature adopted for 2.1, 2.4, 2.5, above is in agreement with everyday language. While the conjunction "or" has more than one possible interpretation, that given in 2.2 above is at least as acceptable as any other. However, whether the term "implication" in its common meaning is interpreted correctly by truth table 2.3 is open to question. Suffice it to say here that, in any case, these

44

terms are inessential for the development of the theory as such. Also, it should be noted that the "implication" discussed here is by no means a consequence relation, although we shall see later it is connected with such a relation.

3. <u>Formulae</u>. If in any truth function of n variables, $q = f(p_1, \ldots p_n)$ we substitute truth functions, $g_1, \ldots, g_n$ of $k_1, \ldots, k_n$ variables respectively, then we obtain again a truth function,

$$q = f(g_1 (p_1^1, \ldots, p_{k_1}^1), \ldots g_n(p_1^n, \ldots p_{k_n}^n))$$

where it is understood that some of the g_i's and some of the arguments may coincide. Strictly speaking, the operation of substitution which is familiar in Algebra and Analysis is an operation from functions to functions, which can itself be made the subject of a logical analysis. However, we shall assume here that the meaning of this operation is intuitively clear and shall call the result a compound truth function, or briefly, a <u>formula</u>. Thus the reader will have no difficulty in interpreting the meaning of the formulae $(\sim p) \lor q, \sim (p \lor q), (\sim p) \supset (q \lor (r \land s))$, etc. Moreover, it is customary to replace $(\sim p) \lor q$ by $\sim p \lor q$ on the understanding that the connective of negation, $\sim$, always applies only to the variable or bracketed expression following it immediately. Other bracket conventions will be given later. A formula which is identically true, i.e. which takes the value T for all assignments of truth values to the variables contained in it is called a <u>tauto-logy</u>.

We introduce a binary relation between formulae, called equivalence, as follows. We write $X \approx Y$, where X and Y are formulae (read X equivalent to Y and do not confuse with the biconditional truth function introduced under 2.4 above), if X and Y take equal truth values for equal truth values of their arguments, i.e. if X and Y yield identical truth functions. It is understood that X and Y need not contain the same variables (e.g. $p \approx p \land (q \lor \sim q)$). If X is a tautology then we write $X \approx T$, while if X is identically false we write $X \approx F$. This is in keeping with our previous notation if we interpret T(F) as the truth function (of an arbitrary number of variables) which is identically true (false). The relation $\approx$ is transitive, reflexive and symmetrical and is thus rightly called equivalence. The substitution of equivalent formulae for the variables of a given formula yields equivalent formulae. It is not difficult to compute the truth function which

45

is represented by a given formula. A convenient layout for the required computation is given below.

We wish to calculate the truth function represented by the formula $(p \supset q) \supset ((r \lor p) \supset (r \lor q))$.

(1)	(2)	(3)	(4)=(1)⊃(2)	(5)=(3)∨(1)	(6)=(3)∨(2)	(7)=(5)⊃(6)	(8)=(4)⊃(7)
p	q	r	p ⊃ q	r ∨ p	r ∨ q	(r∨p)⊃(r∨q)	(p⊃q)⊃ ((r∨p)⊃ (r∨q))
T	T	T	T	T	T	T	T
T	T	F	T	T	T	T	T
T	F	T	F	T	T	T	T
T	F	F	F	T	F	F	T
F	T	T	T	T	T	T	T
F	T	F	T	F	T	T	T
F	F	T	T	T	T	T	T
F	F	F	T	F	F	T	T

Thus, the formula in question is a tautology.

The reader will have no difficulty in verifying that the following equivalences hold for arbitrary formulae X, Y, Z.

3.1.	$X \land Y \approx Y \land X$
3.2.	$X \lor Y \approx Y \lor X$
3.3.	$(X \land Y) \land Z \approx X \land (Y \land Z)$
3.4.	$(X \lor Y) \lor Z \approx X \lor (Y \lor Z)$
3.5.	$X \lor (Y \land Z) \approx (X \lor Y) \land (X \lor Z)$
3.6.	$X \land (Y \lor Z) \approx (X \land Y) \lor (X \land Z)$
3.7.	$\sim (\sim X) \approx X$
3.8.	$\sim (X \land Y) \approx \sim X \lor \sim Y$
3.9.	$\sim (X \lor Y) \approx \sim X \land \sim Y$.

3.1 and 3.2 above show that the conjunction and disjunction obey the commutative law and 3.3 and 3.4 establish that they satisfy also the associative law. It is therefore customary to omit brackets in consecutive conjunctions or disjunctions (e.g. $p \lor q \lor r$ stands for $(p \lor q) \lor r$ as well as for $p \lor (q \lor r)$). 3.5 and 3.6 represent distributive laws while 3.8 and 3.9 are known as "de Morgan's formulae". The latter permit us to interchange the order of negation and conjunction (disjunction) subject to the

46

change of the original conjunction (disjunction) into a disjunction (conjunction).

Normal forms. A formula is said to be in disjunctive normal form (or, simply, a disjunctive normal formula), if it is a disjunction or conjunctions of variables, and (or) of their negations. Thus

$$(p \land q \land \sim r) \lor (p \land \sim p \land q) \lor \sim r$$

is in disjunctive normal form, while

$$p \land (q \lor (p \land \sim r)) \qquad , \qquad \sim (r \lor p)$$

are not in disjunctive normal form.

Similarly a formula is said to be in conjunctive normal form if it is a conjunction of disjunctions of variables and (or) of their negations, e.g. $(p \lor q \lor \sim r) \land (p \lor \sim p \lor q) \land \sim r$.

The parts of a disjunctive normal formula which are separated by the sign of disjunction are called its disjuncts. There is a corresponding definition for the conjuncts of a conjunctive normal formula.

4.1. THEOREM. Let X be a conjunctive normal formula. In order that X be a tautology, it is necessary and sufficient that in each of the conjuncts of X at least one variable appear both with and without the sign of negation prefixed to it.

The condition is necessary. For if a conjunction is a tautology, then each of its conjuncts must be a tautology. Let X^1 be a conjunct of X, and suppose that X^1 does not contain any variable together with its negation. Then to every variable p which occurs in X^1 without the sign of negation, we assign the truth value F, while to every variable which appears in X^1 with the sign of negation we assign the truth value T. This yields the truth value F for X^1 and hence for X, which is contrary to the assumption that X is a tautology. This proves that the condition is necessary.

The condition is also sufficient. For suppose that it is satisfied and consider any assignments of truth values to the variables of X. Let X^1 be any conjunct of X. By assumption, X^1 contains at least one variable, p say, together with its negation, $\sim p$. It follows that if T is assigned to p then T is assigned also

47

to $p \vee \sim p$ and hence to X^1; while if F is assigned to p, then T is assigned to $\sim p$ and hence also to $p \vee \sim p$ and, further, to X^1. It follows that T is assigned also to the entire formula X. This completes the proof of the theorem.

The reader will have no difficulty in formulating a similar test which decides whether or not a given disjunctive normal formula is identically false.

Next we prove:

4.2. THEOREM. To every truth function $f(p_1, \ldots, p_n)$ there exists a conjunctive normal formula X of the variables $p_1, \ldots p_n$ such that

$$f(p_1, \ldots, p_n) \approx X$$

Proof. We first dispose of the possibility that f is identically true. In that case we define X simply by

$$(p_1 \vee \sim p_1) \wedge (p_2 \vee \sim p_2) \wedge \ldots \wedge (p_n \vee \sim p_n)$$

Now any truth function $f(p_1, \ldots, p_n)$ is determined by the specification of the truth value F or T for given truth values of $p_1, \ldots, p_n$. This is exemplified by the following scheme:

p_1	p_2	p_3	$\cdots$	p_n	$f(p_1, \ldots, p_n)$
T	F	F	$\cdots$	T	F

There are just 2^n rows such as this, which we may arrange in an arbitrary but definite order. Consider the rth row, $1 \leq r \leq 2^n$. Corresponding to it, we define the disjunction X_r, which is of the form

$$(\sim) p_1 \vee (\sim) p_2 \vee \ldots \vee (\sim) p_n$$

In this expression, the terms $(\sim)$ are understood to mean that $\sim$ appears before any p_k if T is assigned to p_k in the rth row, while $\sim$ does not appear before p_k if F is assigned to p_k in that row. We now define X as the conjunction of all X_r (taken in an arbitrary order) such that F is assigned to $f(p_1, \ldots, p_n)$ in the rth row. We have already disposed of the truth functions which are identically true so that we may assume that X is not empty. Now consider any row in which the truth value F is

48

assigned to $f(p_1, \ldots, p_n)$. Then F is assigned also to the corresponding X_r for this distribution of truth values, and the same therefore applies to X. On the other hand, suppose that in the rth row, $f(p_1, \ldots, p_n)$ obtains the truth value T. Consider any row other than the rth row, the sth row say. Then for some p_k the truth value of p_k in the sth row is different from the truth value of p_k in the rth row. Hence p_k appears in X_s without (with) the sign of negation according as it appears in X_r with (without) that sign. Now the truth value assigned to X_r for the distribution of truth values given in the rth row is F. It follows that the truth value assigned to p_k in the rth row makes X_s true and that X_r is not included in X. It follows that the entire X obtains the value T for the assignment of truth values of the rth row. This completes the proof of the theorem.

We have the corollary that every truth function can be represented by (i.e., is equivalent to) a formula which contains only the connectives of negation, disjunction, and conjunction. Moreover by one of de Morgan's laws one may eliminate the conjunction by representing it in terms of disjunction and negation. Indeed, by 3.7 and 3.8

4.3. $X \wedge Y \approx \; \sim(\sim X \vee \sim Y)$

for all formulae X and Y. Similarly, we obtain the equivalence

4.4. $X \vee Y \approx \; \sim(\sim X \wedge \sim Y)$

by means of which we may eliminate the disjunction while retaining conjunction and negation. But even a single connective suffices to represent all truth functions in view of the equivalences

4.5. $\sim X \approx X \mid X$

4.6. $X \vee Y \approx (X \mid X) \mid (Y \mid Y)$

which can be verified without difficulty.

We may also take note of the following equivalences for future reference

4.7. $X \supset Y \approx \; \sim X \vee Y$

4.8. $X \equiv Y \approx \; (X \supset Y) \wedge (Y \supset X)$

4.9. $X \mid Y \approx \; \sim(X \wedge Y).$

49

We have shown that to every formula, there exists an equivalent formula in conjunctive normal form. We shall now give a procedure for obtaining such a normal formula from the given formula without the use of truth tables. Instead of expounding the procedure in abstract we give a single example.

Let the formula X be given by

$$(\sim(p \wedge (q \vee \sim r)))) \vee t.$$

We first shift the signs of negation until they apply only to the individual variables, by the use of 3.7, 3.8, and 3.9, so

$$X \approx (\sim p \vee (\sim q \wedge r))) \vee t.$$

Next, in view of the commutative and associative laws of disjunction

$$X \approx \sim p \vee t \vee (\sim q \wedge r)$$

and finally, applying one of the distributive laws,

$$X \approx (\sim p \vee t \vee \sim q) \wedge (\sim p \vee t \vee r).$$

We note that the conjunctive normal formula which represents X is not unique since we have equally well in the case just considered, e.g.

$$X \approx (\sim p \vee t \vee \sim q) \wedge (\sim p \vee t \vee r) \wedge (s \vee \sim s).$$

There is a corresponding procedure for reducing any given formula to one in disjunctive normal form.

Combining the procedure just described with Theorem 4.1, we obtain an alternative test for deciding whether a given formula is a tautology. Given X we first find a conjunctive normal formula X^1 such that $X \approx X^1$ and we then check whether the conditions of 4.1 are satisfied. The difference between this and the earlier procedure (by truth values) is that we now argue directly from the external "shape" of X^1 while at the same time X^1 is obtained from X by a purely formal transformation.

5. <u>Review</u> of <u>the</u> <u>concept</u> <u>of</u> <u>a</u> <u>formula</u>. Earlier, we looked upon a formula as a compound truth function obtained by applying operations of substitution to single truth functions. This point of view can be justified with perfect rigour, but the following alternative approach throws some light on other aspects of our calculus.

We begin with a set P of objects called "propositional variables", $P = \{p, q, r, \ldots\}$ where P is supposed infinite and may or may not be countable. Next we define a set F of objects called "formulae" inductively as follows:

5.1. All variables are formulae.
5.2. If X is a formula then $\sim (X)$ is a formula.
5.3. If X, Y are formulae, then $(X) \vee (Y)$, $(X) \wedge (Y)$,
    $(X) \supset (Y)$, $(X) \equiv (Y)$, and $(X) \mid (Y)$ are formulae.

It is then understood that F contains only elements which can be shown to be such of the (repeated) applications of 5.1 - 5.3. The brackets are inserted automatically but in some cases they are unnecessary and in others they can be omitted by agreement, as before. However, in its fully bracketed form, every formula is obtained by a unique sequence of operations corresponding to 5.1, 5.2, or 5.3. The variables which are used in the definition of a formula are said to be contained in the formula. Only a finite number of variables are contained in each formula.

Consider the set of all formulae which contain no variables other than $p_1, \ldots, p_n$. For a given assignment of truth values to $p_1 \ldots, p_n$, we then define truth values for the remaining formulae of the set inductively as follows:

5.4. $p_1, \ldots, p_n$ take the given truth values.

5.5. If X has already been assigned the truth value T (or alternatively F) then $\sim (X)$ is assigned the truth value F (or, in the alternative case, T).

5.6. If truth values have already been assigned to formulae X and Y then the truth values of $(X) \vee (Y)$, $(X) \wedge (Y)$, etc. are determined from the corresponding truth tables in Section 2.

In this way, every formula X which contains only the variables $p_1, \ldots, p_n$ is made to correspond to a truth function, and this is the same truth function that is obtained from X if we regard this formula as the result of successive substitutions.

51

It follows that the notions of equivalence and of tautology apply equally to the formulae just defined. For this reason we may, if we so desire, dispense with the connectives of conjunction, implication, biconditional, and incompatibility, and consider only formulae which are built up by means of negation and disjunction.

At this point we ought to investigate in more detail the notion of a symbol, e.g. of the symbol for a propositional variable. Is the symbol for the propositional variable p (i.e. 'p') to be regarded as an entity which is distinct from p? We leave it to the reader to ponder this question.

6. <u>The calculus of deduction</u>. So far the reader will have discovered no trace of the traditional approach to Logic, which is to some extent also that of the working mathematician and which regards Logic as "the science of drawing necessary conclusions". The mathematician is in fact accustomed to a procedure by which he derives conclusions, or "theorems", from initial postulates or axioms. Thus, informally speaking, the theorems in question are satisfied in a particular mathematical structure only if the given axioms are satisfied in it. Since in Logic we do not wish to confine ourselves immediately to the consideration of particular (mathematical) structures, it is natural to introduce a deductive system in which the "theorems" to be derived are such as to apply to <u>all</u> structures. In other words, we shall consider a deductive system in which the provable theorems are precisely the tautologies. Although this would appear to differ considerably from the approach used in Mathematics, the gap can be bridged without difficulty as we shall show in due course.

The deductive system which we are going to describe is that of Whitehead and Russell (in Principia Mathematica) as modified by Bernays.

<u>Formulae</u> will be defined as in the preceding section, when only the connectives $\sim$ and $\vee$ are used. Other connectives will be regarded as abbreviations, as follows: $p \supset q$ is an abbreviation for $\sim p \vee q$, $p \wedge q$ is an abbreviation for $\sim (\sim p \vee \sim q)$, $p \equiv q$ is an abbreviation for $(p \supset q) \wedge (q \supset p)$ and $p \mid q$ is an abbreviation for $\sim (p \wedge q)$.

The <u>theorems</u> of our system are obtained by the use of the following rules.

52

6.1. The following are theorems:

(1) $\sim (p \lor p) \lor p$
(2) $\sim p \lor (p \lor q)$
(3) $\sim (p \lor q) \lor (q \lor p)$
(4) $\sim (\sim p \lor q) \lor (\sim (r \lor p) \lor (r \lor q))$.

The same formulae are, in abbreviated notation,

(1) $(p \lor p) \supset p$
(2) $p \supset (p \lor q)$
(3) $(p \lor q) \supset (q \lor p)$
(4) $(p \supset q) \supset ((r \lor p) \supset (r \lor q))$.

We use the same numbering as before, i.e. (1) - (4) - since from our point of view corresponding expressions, e.g. $\sim (p \lor p) \lor p$ and $(p \lor p) \supset p$ are supposed to denote the same formula.

(1) to (4) are known as the axioms of the system.

6.2. By substituting a formula for one of the variables in a given theorem (whenever the variable occurs in that theorem) we obtain a theorem.

6.3. If X and $\sim X \lor Y$ (in abbreviated notation, $X \supset Y$) are theorems, then Y is a theorem. This is the rule of <u>modus ponens</u>.

We now regard it as our task to produce theorems by the use of rules 6.1. - 6.3. As for 6.2., we shall use the symbol $\frac{X}{p}$ (Y) to indicate the result of the substitution of the formula X for the variable p in the formula Y. If Y does not contain p then we define $\frac{X}{p}$ (Y) as Y. Thus any substitution can be applied to any formula.

Similarly, we write MP(X, Z) for the result, (i.e. Y) of applying the rule 6.3. to two formulae X and Z where Z is of the form $X \supset Y$. If Z is not of this form then we agree to regard X itself as the result of the operation. In this way any application of 6.2. and 6.3. to theorems already obtained yields a theorem.

For example, let us substitute $\sim r$ for r in (4), so

$$\frac{\sim r}{r} (4) = (5) \quad (p \supset q) \supset ((\sim r \lor p) \supset (\sim r \lor q))$$

53

and this is, in abbreviated notation

$$(5) \qquad (p \supset q) \supset ((r \supset p) \supset (r \supset q)).$$

Similarly, the following are all theorems

$$\frac{p}{q}(2) \;=\; (6) \quad p \supset (p \lor p)$$

$$\frac{p \lor p}{p}\;(5) = (7) \quad ((p \lor p) \supset q) \supset ((r \supset (p \lor p)) \supset (r \supset q))$$

$$\frac{p}{q}\;(7) \;=\; (8) \quad ((p \lor p) \supset p) \supset ((r \supset (p \lor p)) \supset (r \supset p))$$

$$\frac{p}{r}\;(8) \;=\; (9) \quad ((p \lor p) \supset p) \supset ((p \supset (p \lor p)) \supset (p \supset p))$$

$$MP\,((1),\,(9)) = (10) \quad (p \supset (p \lor p)) \supset (p \supset p)$$

$$MP\,((6),\,(10)) = (11) \quad p \supset p$$

and this is, more explicitly (11) $\sim p \lor p$.

Of the theorems obtained so far, (5) and (11) are of particular interest.

Instead of continuing the derivation of our theorems indefinitely, we shall now ask ourselves some basic questions concerning the character of our deductive calculus.

(To be continued)

Hebrew University, Jerusalem, Israel

54