10. Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability

On several occasions it has been proposed that modal terms ('necessary', 'possible', and the like) be treated as predicates of expressions rather than as sentential operators.[1] According to this proposal, we should abandon such sentences as 'Necessarily man is rational' in favor of "'Man is rational' is necessary' (or "'Man is rational' is a necessary truth'). The proposal thus amounts to the following: to generate a meaningful context a modal term should be prefixed not to a sentence or formula but to a *name* of a sentence or formula (or perhaps a variable whose values are understood as including sentences).

The advantage of such a treatment is obvious: if modal terms becomes predicates, they will no longer give rise to non-extensional contexts, and the customary laws of predicate calculus with identity may be employed. The main purpose of the present paper is to consider to what extent within such a treatment the customary laws of modal logic can be maintained.

These considerations form the content of Section 2. Sections 3 and 4 use identical methods, but contain results unrelated to modal logic; Section 3 concerns the non-provability of certain

Originally published in *Acta Philosophica Fennica* 16:153–67 (1963). Reprinted with permission. This paper was written while the author held a United States National Science Foundation Grant, NSF G–19830. I wish to express gratitude also to my student, Mr. David Kaplan, and to Professor R. L. Vaught for helpful discussion and correspondence.

[1] Examples of such proposals may be found in [6], in which necessary truth is identified with provability in a certain system, as well as Carnap [1, pp. 233–60] and Quine [14].

arithmetical 'reflexion principles', and Section 4 contains general theorems on non-finite axiomatizability.[2]

1. PRELIMINARIES

The terminology of Tarski, Mostowski, and Robinson [17] will be adopted. A few additional notions will be employed, for instance, that of a *logical axiom* (for the first-order predicate calculus with identity). The logical axioms can be chosen in various ways; we impose only the following requirements: (1) the set of logical axioms is to be recursive; (2) all logical axioms are to be sentences (that is, formulas without free variables); (3) the logical axioms are to be complete under the rule of detachment (or *modus ponens*), in the sense that, for any theory T, the set of logically valid sentences of T is to be the smallest set closed under detachment and containing all logical axioms which are sentences of T.

Let φ be a formula whose only free variable is u. Then $\psi^{(\varphi)}$, or the *relativization* of ψ to φ, can be defined for an arbitrary formula ψ by the following recursion: if ψ is atomic, $\psi^{(\varphi)}$ is ψ; $(\sim\psi)^{(\varphi)}$ is $\sim\psi^{(\varphi)}$, $(\psi \rightarrow \chi)^{(\varphi)}$ is $\psi^{(\varphi)} \rightarrow \chi^{(\varphi)}$, and analogously for the other sentential connectives; $(\Lambda\alpha\,\psi)^{(\varphi)}$ is $\Lambda\alpha\,(\varphi(\alpha) \rightarrow \psi^{(\varphi)})$, and $(\mathrm{V}\alpha\,\psi)^{(\varphi)}$ is $\mathrm{V}\alpha\,(\varphi(\alpha) \wedge \psi^{(\varphi)})$.[3] If T is a theory, then $T^{(\varphi)}$, or the *relativization* of T to φ, is that theory whose constants are those of T together with those occurring in φ, and whose valid sentences are the logical consequences within this vocabulary of the set of sentences $\psi^{(\varphi)}$, where ψ is a valid sentence of T.

We shall be particularly interested in the theories P and Q of Tarski, Mostowski, and Robinson [17] (*Peano's arithmetic* and

[2] The results of this paper may be considered as applications of the Paradox of the Hangman, or rather the related Paradox of the Knower, in the sense that the proof of the basic lemma of the paper, Lemma 3, was partly suggested by the latter paradox. Both paradoxes were first exactly formulated in Kaplan and Montague [7], which contained the conjecture that they might, like the Liar, lend themselves to some sort of technical application.

[3] Thus the notation '$x(y)$' (as well as the notation '$x(y,z)$' in later passages) is used, as in Tarski [17], for proper substitution in a formula for the variable u (or the variables u, v). The same notation will also be used for the value of a function, but the context will always suffice to determine the intended sense.

Robinson's arithmetic respectively).[4] Both theories have as non-logical constants the symbols, 0, S, $+$, $\cdot$; their possible realizations will consequently have the form $\langle A, z, s, p, t\rangle$, where $z \in A$, s is a function mapping A into A, and p, t are functions mapping the set of ordered pairs of elements of A into A. By the *standard realization* (of P or Q) is understood that realization $\langle A, z, s, p, t\rangle$ in which A is the set of natural numbers, z is the number zero, s is the successor function, and p, t are the respective operations of addition and multiplication of natural numbers. When we speak simply of *truth* or *definability* we mean truth or definability in the standard realization $\mathfrak{A}_0$. In other words, a sentence is said to be *true* if it is a sentence of Q and true in $\mathfrak{A}_0$, and a formula φ is said to *define* a set X of natural numbers if φ is a formula of Q whose only free variable is u, and φ is satisfied in $\mathfrak{A}_0$ by those and only those natural numbers which are members of X. Derivatively, we call a *set* of sentences *true* if all its members are true, and a *theory true* if its set of valid sentences is true; and we say that φ *defines* a set X of *expressions* if φ defines the set of Gödel numbers of members of X.[5]

If X is a set of natural numbers, T a theory, and φ a formula of T whose only free variable is u, then we say that φ *super-numerates* X in T if $\vdash_T \varphi(\Delta_n)$ whenever $n \in X$, and that φ *numerates* X in T if, for every natural number n, $n \in X$ if and only if $\vdash_T \varphi(\Delta_n)$. On the other hand, if X is a set of *expressions*, we say that φ *numerates* or *super-numerates* X in T if φ respectively numerates or super-numerates in T the set of Gödel numbers of members of X.

A function f mapping the set of natural numbers into itself is said to be *functionally numerable* in a theory T if there is a formula φ of T with free variables u, v such that, for all natural numbers n,

$$\vdash_T \varphi(\Delta_n, v) \leftrightarrow v = \Delta_{f(n)}.$$

[4] The only properties of Q used in this paper are that (i) the constants of Q are 0, S, $+$, $\cdot$, (ii) Q is true (in the sense about to be explained), (iii) Q is finitely axiomatizable, and (iv) all one-place recursive functions of natural numbers are functionally numerable in Q (again in a sense which is about to be explained). Thus Q could be replaced everywhere by any other theory with these properties.

[5] To make these definitions (as well as those in the next paragraph) perfectly unambiguous we should have to make certain disjointness assumptions, for instance, that no natural number is an expression. Without such assumptions, however, the intended sense of truth and definability will in what follows be clearly determined by the context.

We can associate with each expression a term of Q which can be regarded as the standard name of that expression; to be specific, we associate with an expression σ the name $\Delta_{nr(\sigma)}$ (that is, the result of prefixing $nr(\sigma)$ occurrences of S to the individual constant 0, where $nr(\sigma)$ is the Gödel number of σ). Keeping this interpretation in mind, we can regard the following lemma as a principle of self-reference; it is implicit in many earlier publications, and proved in essentially the present form in Montague [12].

Lemma 1. If T is a theory in which all one-place recursive functions of natural numbers are functionally numerable, ψ is a formula of T whose only free variable is u, and E is a one-place recursive function of expressions, then there is a sentence φ of T such that

$$\vdash_T \varphi \leftrightarrow \psi(\Delta_{nr(E(\varphi))}).$$

Lemma 2. If T is a theory which is an extension of $Q^{(\beta)}$, for some formula β whose only free variable is u, then all one-place recursive functions of natural numbers are functionally numerable in T.

Proof. This is easily derived from the fact, proved in [17], that all recursive functions are functionally numerable in Q.

The next statement is the central lemma of the present paper, and forms the basis of most of the results of the following sections.

Lemma 3. Suppose that T is a theory and α a formula whose only free variable is u, and, for all sentences φ, ψ of T,

(i) $\vdash_T \alpha(\Delta_{nr(\varphi)}) \rightarrow \varphi$,
(ii) $\vdash_T \alpha(\Delta_{nr(\varphi)})$, if φ is $\alpha(\Delta_{nr(\psi)}) \rightarrow \psi$,
(iii) $\vdash_T \alpha(\Delta_{nr(\varphi)})$, if φ is a logical axiom,
(iv) if $\vdash_T \alpha(\Delta_{nr(\varphi\rightarrow\psi)})$ and $\vdash_T \alpha(\Delta_{nr(\varphi)})$, then $\vdash_T \alpha(\Delta_{nr(\psi)})$,
(v) $Q^{(\beta)}$ is a subtheory of T, for some formula β whose only free variable is u.

Then T is inconsistent.

Proof. Since Q is finitely axiomatizable, so is $Q^{(\beta)}$. Let χ be a valid sentence of $Q^{(\beta)}$ from which all valid sentences of $Q^{(\beta)}$ are logically derivable, and let T' be that theory whose valid sentences are the sentences of T derivable from χ. By Lemmas 1 and 2, there is a sentence φ of T such that

(1) $\vdash_{T'} \varphi \leftrightarrow \alpha(\Delta_{nr(\chi \to \sim\varphi)})$.

If we let L be the theory whose valid sentences are the logically valid sentences of T, it follows by the Deduction Theorem that

$$\vdash_L \chi \to (\varphi \leftrightarrow \alpha(\Delta_{nr(\chi \to \sim\varphi)})).$$

Hence, by sentential logic,

(2) $\vdash_L (\alpha(\Delta_{nr(\chi \to \sim\varphi)}) \to (\chi \to \sim\varphi)) \to (\chi \to \sim\varphi)$.

Let γ be the sentence $\alpha(\Delta_{nr(\chi \to \sim\varphi)}) \to (\chi \to \sim\varphi)$. It follows from (iii) and (iv) that if ψ is any logically valid sentence of T, then $\vdash_T \alpha(\Delta_{nr(\psi)})$. In particular, we have by (2) that

(3) $\vdash_T \alpha(\Delta_{nr(\gamma \to (\chi \to \sim\varphi))})$.

Therefore, by (iv) and (ii),

(4) $\vdash_T \alpha(\Delta_{nr(\chi \to \sim\varphi)})$,

and hence, by (1),

$$\vdash_T \varphi.$$

By (4) and (i),

$$\vdash_T \chi \to \sim\varphi,$$

and hence, by (v),

$$\vdash_T \sim\varphi,$$

and T is inconsistent.

Tarski's theorem on the undefinability of truth (proved in [16], using the Paradox of the Liar) can be formulated as follows: if T is a theory, α is a formula whose only free variable is u, condition (v) of Lemma 3 is satisfied, and in addition

(i′) $\vdash_T \alpha(\Delta_{nr(\varphi)}) \leftrightarrow \varphi$, for all sentences φ of T,

then T is inconsistent.

This theorem is an immediate consequence of Lemma 3. There seems, however, to be no direct implication in the opposite

direction; indeed, Tarski's assumption (i′) appears intuitively much stronger than the combination of (i)–(iv).

The following lemma, closely related to Lemma 3, will also be used.

Lemma 4. Suppose that T is a theory and α a formula whose only free variable is u, and that, for all sentences φ, ψ of T, conditions (i), (iii), (v) of Lemma 3 hold, together with the following:

(iv′) if $\vdash_T \alpha(\Delta_{nr(\varphi \to \psi)})$, then $\vdash_T \alpha(\Delta_{nr(\varphi)}) \to \alpha(\Delta_{nr(\psi)})$.

Then $\vdash_T \sim \alpha(\Delta_{nr(\gamma)})$, for some sentence γ of T of the form $\alpha(\Delta_{nr(\delta)}) \to \delta$.

Proof. We form χ, φ, γ as in the proof of Lemma 3, and carry through that proof up to and including step (3). From (3) and (iv′) we conclude that

$$\vdash_T \alpha(\Delta_{nr(\gamma)}) \to \alpha(\Delta_{nr(\chi \to \sim\varphi)}).$$

From this we conclude by (1) that

$$\vdash_T \alpha(\Delta_{nr(\gamma)}) \to \varphi,$$

and by (i) that

$$\vdash_T \alpha(\Delta_{nr(\gamma)}) \to (\chi \to \sim\varphi).$$

But $\vdash_T \chi$ by (v); hence, by the two assertions above,

$$\vdash_T \sim \alpha(\Delta_{nr(\gamma)}),$$

and γ has by its construction the required form.

2. MODALITY

According to the proposal mentioned at the beginning of this paper, modal statements will occur in a syntax language (or metalanguage). As was intimated in connection with Lemma 1, we suppose syntax to be arithmetized and use as syntax languages those languages which contain the symbols of arithmetic.[6] Thus

[6] This approach is by no means essential, and is adopted only to allow us to build on terminology and results already present in the literature. An equivalent and perhaps more natural approach would employ a syntax language (such as the one introduced in Tarski [16]) which speaks directly about expressions.

we assume throughout the present section that N is a formula whose only free variable is u; and, for any sentence φ, we set $N[\varphi]$ equal to the sentence

$$N(\Delta_{nr(\varphi)}),$$

that is, the proper substitution of what might be regarded as the name of φ for all free occurrences of u in N. We think of $N[\varphi]$ as expressing the assertion that φ is necessarily true, but we impose no special conditions on N. For example, N may be a complex formula of set theory, or simply an atomic formula πu, where π is a one-place predicate. In particular, we do not assume that N is in any sense equivalent to a formula of P.

The following remark concerning the limitations of syntactical interpretations of modal logic is due to Gödel [6]: if N is the standard formula of P defining the set of valid sentences of P, then $N[N[\varphi] \rightarrow \varphi]$ is untrue for some sentences φ of P.

Gödel's proof of this remark employs his earlier theorem (in [5]) to the effect that the consistency of P is not provable in P, and like the proof of that earlier theorem depends essentially on the exact structure of the formula N. Now it is easy to find formulas of P (other than the standard formula) which define the set of valid sentences of P and with respect to which the consistency of P is provable in P.[7] We might hope to use one of these more complex formulas to express necessity and still maintain the characteristic modal law

$$N[N[\varphi] \rightarrow \varphi].$$

But this hope is vain, as our investigations (in particular, Theorem 4, which is a generalization of Gödel's remark) will show. Stronger negative results are also obtainable, like the following.

Theorem 1. Suppose that T is any theory such that

(i) T is an extension of Q (or of $Q^{(\alpha)}$, for some formula α whose only free variable is u),

[7] Examples may be found in Feferman [3], but simpler examples may be constructed in a rather obvious way. (Feferman's formulas were devised so as to satisfy certain additional conditions as well.)

and, for all sentences φ, ψ of T,

(ii) $\vdash_T N[\varphi] \rightarrow \varphi$,
(iii) $\vdash_T N[N[\varphi] \rightarrow \varphi]$,
(iv) $\vdash_T N[\varphi \rightarrow \psi] \rightarrow (N[\varphi] \rightarrow N[\psi])$,
(v) $\vdash_T N[\varphi]$, if φ is a logical axiom.

Then T is inconsistent.

Proof. By Lemma 3, taking N for α.

Notice that the assumptions of Theorem 1 concern only those expressions $N[\varphi]$ in which φ is without free variables. Thus the difficulties exhibited both here and below are seen to be independent of the familiar problems resulting from quantification into modal contexts.

It has been shown in joint work of Mr. David Kaplan and the author that none of the hypotheses (ii)–(v) can be dropped. Theorem 1 can, however, be slightly strengthened in view of Lemma 4:

Theorem 2. Suppose that T satisfies conditions (i), (ii), (iv), (v) above. Then there is a single sentence φ of T such that

$$\vdash_T \sim N[N[\varphi] \rightarrow \varphi].$$

The following assertion, which overlaps Theorem 1, can be obtained directly from Lemma 3.

Theorem 3. Suppose that T satisfies conditions (i) and (ii) of Theorem 1, and that

(vi) $\vdash_T N[\varphi]$ whenever φ is a sentence such that $\vdash_T \varphi$.

Then T is inconsistent.

The following generalization of Gödel's remark is a simple consequence of Theorem 1.

Theorem 4. Suppose that U is any true theory with the same constants as P, and that N is a formula of P defining the set of valid sentences of U. Then $N[N[\varphi] \rightarrow \varphi]$ is untrue for some sentences φ of P.

Proof. Assume the hypothesis, and that $N[N[\varphi] \to \varphi]$ is true for all sentences φ of P. Let T be the theory whose valid sentences are the true sentences of P. Then T and N will satisfy conditions (i)–(v) of Theorem 1. Hence T is inconsistent, which is absurd.

Now what general conclusions can be drawn from Theorems 1–4? In the first place, observe that the schemata in conditions (ii)–(v) of Theorem 1 are provable in the well-known systems of first-order modal logic with identity, that is, the systems of Carnap [2] and Kripke [8]. These schemata would, moreover, be provable in any reasonable extension to predicate logic of S1, the weakest of the Lewis modal calculi.[8] Further, it is not unnatural to impose condition (i): modal logic, like ordinary logic, ought to be applicable to an arbitrary subject matter, including arithmetic. Condition (vi), though not needed in Theorem 1, is rather natural, and appears as an inference rule in many familiar systems.

Thus if necessity is to be treated syntactically, that is, as a predicate of sentences, as Carnap and Quine have urged, then virtually all of modal logic, even the weak system S1, must be sacrificed.

This is not to say that the Lewis systems have no natural interpretation. Indeed, if necessity is regarded as a sentential operator, then perfectly natural model-theoretic interpretations may be found, for instance in Kripke [8] and Montague [10], which satisfy all the Lewis systems S1–S5. It should be observed, however, that the *natural* model-theoretic interpretations (as opposed to *ad hoc* interpretations) provide no justification for any of S1–S4; for, though they satisfy the theorems of these systems, they satisfy additional modal principles as well, and indeed give all of S5, the strongest of the Lewis systems and the system whose quantified version was in a sense proved complete in Kripke [8].

Thus it seems at present doubtful that any philosophical interest can be attached to S1–S4. The natural model-theoretic treatment gives a system stronger than all of them, and no satisfactory syntactical treatment can be provided for any of them.

[8] For the Lewis calculi see Lewis and Langford [9, pp. 492 ff.]

3. ARITHMETICAL REFLEXION PRINCIPLES

Feferman has in [4] constructed 'recursive progressions of theories' in which the passage from a theory T to its successor consists in adding to T all instances of the schema

$$\alpha(\Delta_{nr(\varphi)}) \to \varphi,$$

where α is a rather special formula defining the set of valid sentences of T. In view of Gödel's theorem on non-demonstrable consistency (or rather the version of it given in Feferman [3]), it is clear that in the cases considered by Feferman the successor of a theory is always stronger than the theory itself. The theorems of the present section show this to be the case even in more general situations, when no limitation is placed on the structure of the formula α.

Lemma 5. Suppose that A, T are theories and α, β formulas whose only free variable is u, and that

(i) T is an extension of A,
(ii) T is an extension of $Q^{(\beta)}$,
(iii) the constants of A include those of $Q^{(\beta)}$,
(iv) α numerates in T the set of valid sentences of A,
(v) $\vdash_A \alpha(\Delta_{nr(\varphi)}) \to \varphi$, for each sentence φ of A.

Then T is inconsistent.

Proof. Let T' be the theory whose valid sentences are those sentences of A which are valid in T. Then T' and α satisfy the hypothesis of Lemma 3. Hence T' and therefore T are inconsistent.

Theorem 5. If A is a true theory with the same constants as Q, and α defines the set of valid sentences of A, then there is a sentence φ of A such that

$$\text{not } \vdash_A \alpha(\Delta_{nr(\varphi)}) \to \varphi.$$

Proof. Assume the hypothesis and that, for each sentence φ of A, $\vdash_A \alpha(\Delta_{nr(\varphi)}) \to \varphi$. Let T be the theory whose valid sentences are the true sentences of Q, and β the formula $u = u$. Then (i)–(v) of Lemma 5 hold, and hence T is inconsistent, which is absurd.

Theorem 5 can be strengthened; we may permit A to contain additional constants beyond those of Q.

Theorem 6. Suppose that

(i) A is a theory whose constants include those of Q,
(ii) all sentences of Q which are valid in A are true,
(iii) α defines the set of valid sentences of A.

Then there is a sentence φ of A such that

$$\text{not } \vdash_A \alpha(\Delta_{nr(\varphi)}) \to \varphi.$$

Proof. By Theorem 5, taking for A the theory whose valid sentences are those sentences of Q which are valid in A, and for α the formula $\alpha \wedge \beta$, where β is a formula of Q defining the set of sentences of Q.

A further extension is possible. Instead of requiring, as in Theorem 6, that the arithmetical part of A be true, we may require only that it be true in a certain relativized sense.

Theorem 7. Suppose that A is a theory and α, β formulas whose only free variable is u, and that

(i) the constants of A include those of $Q^{(\beta)}$,
(ii) φ is true, whenever φ is a sentence of Q and $\vdash_A \varphi^{(\beta)}$,
(iii) α defines the set of valid sentences of A.

Then there is a sentence φ of A such that

$$\text{not } \vdash_A \alpha^{(\beta)}(\Delta_{nr(\varphi)}) \to \varphi.$$

Proof. Assume (i)–(iii), and that, for every sentence φ of A, $\vdash_A \alpha^{(\varphi)}(\Delta_{nr(\varphi)}) \to \varphi$. Let T be the theory whose valid sentences are the sentences of A logically derivable from the valid sentences of A together with the sentences $\varphi^{(\beta)}$, where φ is a true sentence of Q.

Then T is consistent. For suppose otherwise. Then there would exist true sentences $\varphi_0, \ldots, \varphi_n$ such that $\vdash_A \sim (\varphi_0^{(\beta)} \wedge \ldots \wedge \varphi_n^{(\beta)})$. Hence $\vdash_A [\sim(\varphi_0 \wedge \cdots \wedge \varphi_n)]^{(\beta)}$, and thus, by (ii), $\sim(\varphi_0 \wedge \cdots \wedge \varphi_n)$ is true, which is impossible.

To show that $\alpha^{(\beta)}$ numerates in T the set of valid sentences of A, assume that $\vdash_A \varphi$. Then, by (iii), $\alpha(\Delta_{nr(\varphi)})$ is true, and hence $\vdash_T [\alpha(\Delta_{nr(\varphi)})]^{(\beta)}$; that is, $\vdash_T \alpha^{(\beta)} (\Delta_{nr(\varphi)})$. For the converse, assume that $\vdash_T \alpha^{(\beta)}(\Delta_{nr(\varphi)})$ and it is not the case that $\vdash_A \varphi$. Then, by (iii), $\sim\alpha(\Delta_{nr(\varphi)})$ is true, and hence $\vdash_T \sim\alpha^{(\beta)}(\Delta_{nr(\varphi)})$, which is impossible by the consistency of T.

Thus A, T, $\alpha^{(\beta)}$ satisfy condition (iv) of Lemma 5. Condition (v) has been assumed, and it is easily seen that the remaining hypotheses of that lemma are also satisfied. It follows that T is inconsistent, contrary to what has been shown.

If we consider super-numerations rather than defining formulas, then the assumption of truth involved in Theorems 5–7 can be replaced by an assumption of consistency.

Theorem 8. Supose that A is a consistent theory and α a formula such that

(i) A is an extension of $Q^{(\beta)}$, for some formula β whose only free variable is u,
(ii) α super-numerates in A the set of valid sentences of A.

Then there is a sentence φ of A such that

$$\text{not } \vdash_A \alpha(\Delta_{nr(\varphi)}) \to \varphi.$$

Proof. Assume the contrary, and apply Lemma 5, taking A for both A and T. (The verification of condition (iv) of Lemma 5 requires a small argument.) We conclude that A is inconsistent, contradicting the hypothesis of the theorem.

4. NON-FINITE AXIOMATIZABILITY

Theorem 8 yields some general theorems on non-finite axiomatizability. Observe that in these theorems we speak of formulas super-numerating the set of *logically* valid sentences of a theory, not as before the set of *all* valid sentences of the theory.

Theorem 9. Suppose that A is a theory such that, for some formula α,

(i) A is an extension of $Q^{(\beta)}$, for some formula β whose only free variable is u,
(ii) variable α super-numerates in A the set of logically valid sentences of A,

(iii) $\vdash_A \alpha(\Delta_{nr(\varphi)}) \to \varphi$, for each sentence φ of A.

Then A, if consistent, is not finitely axiomatizable.

Proof. Assume the contrary, and let χ be a valid sentence of A from which all valid sentences of A are logically derivable. By (i) and Lemma 2, there is a formula γ of A whose only free variables are u, v, and such that, for each sentence φ of A,

(1) $\vdash_A \gamma(\Delta_{nr(\varphi)}, v) \leftrightarrow v = \Delta_{nr(\chi \to \varphi)}$.

Let δ be the formula $\bigvee v(\gamma(u, v) \land \alpha(v))$. Then it is easily deduced from (ii) that δ super-numerates in A the set of valid sentences of A. Hence, by Theorem 8, there is a sentence φ of A such that

not $\vdash_A \delta(\Delta_{nr(\varphi)}) \to \varphi$.

But, by (1),

$\vdash_A \delta(\Delta_{nr(\varphi)}) \to \bigvee v(v = \Delta_{nr(\chi \to \varphi)} \land \alpha(v))$; hence
$\vdash_A \delta(\Delta_{nr(\varphi)}) \to \alpha(\Delta_{nr(\chi \to \varphi)})$; thus, by (iii),
$\vdash_A \delta(\Delta_{nr(\varphi)}) \to (\chi \to \varphi)$; hence
$\vdash_A \delta(\Delta_{nr(\varphi)}) \to \varphi$;

and we have arrived at contradiction.

Theorem 9 applies only to theories which explicitly contain arithmetic. In many situations, however, we consider theories which, like certain well-known systems of set theory, contain arithmetic only implicitly, by way of interpretability; the next theorem concerns non-finite axiomatizability in such situations.

We call a theory T' a *definitional extension* of a theory T if T' is an extension of T and there is a set D such that (i) D is a set of possible definitions[9] in T of non-logical constants of T' which are not constants of T, (ii) each non-logical constant of T' which is not a constant of T occurs in exactly one member of D, and (iii) a sentence of T' is valid in T' if and only if it is logically derivable from the union of D and the set of valid sentences of T.

[9] For the notion of a possible definition of a constant in a theory see Tarski, Mostowski, and Robinson [17].

Theorem 10. Suppose that A is a theory such that, for some A' and α,

(i) A' is a definitional extension of A,
(ii) A' is an extension of $Q^{(\beta)}$, for some formula β whose only free variable is u,
(iii) α super-numerates in A' the set of logically valid sentences of A,
(iv) $\vdash_{A'} \alpha(\Delta_{nr(\varphi)}) \to \varphi$, for each sentence φ of A.

Then A, if consistent, is not finitely axiomatizable.

Proof. Assume the contrary. Then there will clearly be a theory A' which satisfies (i)–(iv), as well as the following additional conditions:

(1) A' has only finitely many constants which are not constants of A,
(2) A' is finitely axiomatizable.

Using (i) and (1), it is not difficult to show the existence of a recursive function f of expressions such that, for each sentence φ of A',

(3) $f(\varphi)$ is a sentence of A,
(4) $\vdash_{A'} \varphi \leftrightarrow f(\varphi)$,
(5) if φ is logically valid, then so is $f(\varphi)$.

(Loosely speaking, we construct $f(\varphi)$ by first eliminating defined constants from φ, and then, if φ contains operation symbols, appending to the result an antecedent containing the existence and uniqueness conditions for the formulas used to define the operation symbols.)

By (ii) and Lemma 2, there is a formula γ of A' whose only free variables are u, v, and such that, for each sentence φ of A',

(6) $\vdash_{A'} \gamma(\Delta_{nr(\varphi)}, v) \leftrightarrow v = \Delta_{nr(f(\varphi))}$.

Let δ be the formula $\bigvee v(\gamma(u, v) \land \alpha(v))$. By (iii), (3), and (5),

(7) δ super-numerates in A' the set of logically valid sentences of A'.

We show that, for each sentence φ of A',

(8) $\vdash_{A'} \delta(\Delta_{nr(\varphi)}) \rightarrow \varphi.$

Assume that φ is a sentence of A'. By (6),

$\vdash_{A'} \delta(\Delta_{nr(\varphi)}) \rightarrow \alpha(\Delta_{nr(f(\varphi))})$; hence, by (3) and (iv),
$\vdash_{A'} \delta(\Delta_{nr(\varphi)}) \rightarrow f(\varphi)$;

and (8) follows by (4).

Now A' is consistent, by (i) and the fact that A is consistent. Therefore, by (7), (8), (ii), and Theorem 9, A' is not finitely axiomatizable, which contradicts (2).

The conclusions of the last two theorems can be strengthened.

Theorem 11. (i) If A satisfies the hypothesis of Theorem 9, then A is essentially non-finitizable (that is, no consistent extension of A having the same constants as A is finitely axiomatizable). (ii) If A satisfies the hypothesis of Theorem 10, then A is again essentially non-finitizable.

Proof. It is easily seen that if A satisfies the hypothesis of one of the two theorems in question, then so does every consistent extension of A with the same constants as A.

Now Theorem 11 (i) can be applied to the theory P, and 11 (ii) to what in Montague [11] is called *general set theory*. Indeed, any theory which satisfies the condition, introduced in Definition 9 of [11], of being *strongly semantically closed* will also satisfy the hypothesis of Theorem 10 (or at least will be isomorphic to a theory satisfying the hypothesis of Theorem 10). This follows from Theorem 1 of [11]; we take for α (in Theorem 10) the standard formula of Q defining the set of logically valid sentences (in terms of derivability by detachment from the logical axioms).

These applications are not new; it was already shown in [11] that every strongly semantically closed theory is essentially non-finitizable. Theorem 11 (ii), however, leads to a much simpler proof of this conclusion than was previously available. In the first place, previous methods required as a lemma Gödel's theorem to the effect that the consistency of certain theories is not provable within the theories themselves; this is no longer required.

Secondly, the earlier methods required the verification of condition (iv) of Theorem 10 for one particular formula α, the 'standard' formula defining logical validity. But as Professor R. L. Vaught has observed and an inspection of the proof of Theorem 1 of [11] will reveal, the verification of (iv) becomes much easier if we take for α, as we may according to Theorem 11 (ii), the formula of Q which defines logical validity in terms of *genetic provability*[10] rather than in terms of derivability from the logical axioms; we thus avoid showing that the Herbrand theorem is, so to speak, 'provable in the object language'.

REFERENCES

1. Carnap, R. *The Logical Syntax of Language*. London, 1937.
2. —. Modalities and Quantification. *Journal of Symbolic Logic* 11:33–64 (1946).
3. Feferman, S. Arithmetization of Metamathematics in a General Setting. *Fundamenta Mathematicae* 49:35–92 (1960).
4. —. Transfinite Recursive Progressions of Axiomatic Theories. *Journal of Symbolic Logic* 27:259–316 (1962).
5. Gödel, K. Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I. *Monatshefte für Mathematik und Physik* 38:173–98 (1931).
6. —. Eine Interpretation des intuitionistischen Aussagenkalküls. *Ergebnisse eines mathematischen Kolloquiums* 4:39–40 (1933).
7. Kaplan, D. and R. Montague. A Paradox Regained. *Notre Dame Journal of Formal Logic* 1:79–90 (1960). Chap. 9 in this book.
8. Kripke, S. A Completeness Theorem in Modal Logic. *Journal of Symbolic Logic* 24:1–14 (1959).
9. Lewis, C. and C. Langford. *Symbolic Logic*. New York and London, 1932.
10. Montague, R. Logical Necessity, Physical Necessity, Ethics, and Quantifiers. *Inquiry* 3:259–69 (1960). Chap. 1 in this book.
11. —. Semantical Closure and Non-Finite Axiomatizability I. In *Infinitistic Methods*, Warsaw, 1961.
12. —. Theories Incomparable with Respect to Relative Interpretability. *Journal of Symbolic Logic* 27:195–211 (1962).

[10] For an informal description of genetic provability see Montague [11, p. 51], and for an exact definition see Schütte [15].

13. Mostowski, A. On Models of Axiomatic Systems. *Fundamenta Mathematicae* 39:133–58 (1952).
14. Quine, W. Three Grades of Modal Involvement. *Proceedings of the XI[th] International Congress of Philosophy* 14:65–81 (Brussels, 1953).
15. Schütte, K. Schlussweisen-Kalküle der Prädikatenlogik. *Mathematische Annalen* 22:47–65 (1950).
16. Tarski, A. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica* 1:261–405 (1935).
17. Tarski, A., A. Mostowski, and R. Robinson. *Undecidable Theories*. Amsterdam, 1953.

11. Deterministic Theories

1. INTRODUCTION

The purpose of this paper is to analyse the notions of a deterministic theory and of a deterministic history, and to investigate some of the properties of these notions. To Laplace is attributed the assertion that if we were given the positions and momenta of all particles at one particular time, we could deduce, using the laws of mechanics, the positions and momenta of all particles at any later time. Implicit in this assertion is an analysis of determinism as applied to theories. The analysis has been made more explicit by Nagel (in [20, p. 422]), who characterizes a theory as deterministic if it will "enable us given the state (of a system)... at one time to deduce the formulation of the state at any other time." (Derivatively, we might call a *history* deterministic if there is a deterministic *theory* which describes it.)

Now this definition cannot be taken quite literally. For let us suppose that classical mechanics is deterministic in the sense of Laplace and Nagel. It would follow, if we were to spell out the definition explicitly, that for any instants t_0 and t, there are sentences $\varphi(t_0)$ and $\varphi(t)$, expressing the state of the universe at

Originally published in *Decisions, Values and Groups* 2. Oxford, Pergamon Press, 1962; pp. 325–70. Reprinted with permission. This paper is based largely on work accomplished by the author while he was in attendance at the Interdisciplinary Behavioral Sciences Research Conference at the University of New Mexico, June–August, 1958. This work was supported by the Behavioral Sciences Division, Air Force Office of Scientific Research, under contract AF 49 (638)–33. The author wishes also to express his gratitude to Mr. David Kaplan, Professor Azriel Lévy and Mrs. Ruth Anna Mathers, who offered important suggestions and criticism.