FASC. 2

## GÖDEL'S DIAGONALIZATION TECHNIQUE AND RELATED PROPERTIES OF THEORIES

 $\mathbf{BY}$ 

## A. MACINTYRE AND H. SIMMONS (ABERDEEN)

**0.** Introduction. Let T be any theory for which the usual proof of Gödel's (second) theorem works. In particular, we assume that for each sentence  $\sigma$  there is a term  $\lceil \sigma \rceil$  which denotes  $\sigma$ , and that there is a formal provability predicate  $Prv(\cdot)$  and a formal consistency statement CON(T). Gödel's theorem is then

(G) 
$$\operatorname{Not}[T \vdash \operatorname{CON}(T)].$$

Another theorem along these lines is that of Löb:

(L) For each sentence  $\sigma$ ,

$$T_{\cup}\{\Pr(\lceil \sigma \rceil)\} \vdash \sigma \Rightarrow T \vdash \sigma.$$

The motivation for this paper is the following theorem of Kreisel:

(K) For each universal sentence  $\sigma$ ,

$$T_{\cup}\{\ | \text{CON}(T)\} \vdash \sigma \Rightarrow T \vdash \sigma.$$

All the results of this paper derive from our proof of (K) and the comparison of this proof with those of (G) and (L). (One of us announced (K) is the abstract [4] and, subsequently, Kreisel's proof of (K) (see [2], bottom of p. 157) was brought to our attention. Our proof of (K) is not the same as that of Kreisel and appears to be more general.)

To prove (K) we found it convenient to prove

(C) For each sentence  $\sigma$ ,

$$T_{\cup}\{\ ]\operatorname{CON}(T)\} \vdash \sigma \Rightarrow T_{\cup}\{\operatorname{Prv}(\ |\ ]\sigma\ |\ )\} \vdash \sigma,$$

and to prove this we proved

(Q) For each sentence  $\sigma$ ,

$$T_{\cup}\{\Pr(\lceil \sigma \rceil), \Pr(\lceil \lceil \sigma \rceil)\} \vdash \sigma \Rightarrow T_{\cup}\{\Pr(\lceil \lceil \sigma \rceil)\} \vdash \sigma.$$

The method of proof of (L) generalizes to a proof of (Q). In fact, it turns out that (Q) is a special case of (L) since (Q) is equivalent to

(P) For each sentence  $\sigma$ ,

$$T_{\cup}\{\Pr(\lceil \sigma^- \rceil)\} \vdash \sigma^- \Rightarrow T \vdash \sigma^-,$$

where  $\sigma^-$  is  $Prv(\lceil \rceil \sigma \rceil) \rightarrow \sigma$ .

Each of these theorems are consequences of the diagonalization technique first used by Gödel. Moreover, each of (L), (P), (Q) and (C) can be deduced from their formatizations in T, and these are special cases of the diagonalization technique. In Section 2 we show that the formal version of (L), namely

(FL) For each sentence  $\sigma$ ,

$$T \vdash \operatorname{Prv}(\lceil \sigma^+ \rceil) \to \operatorname{Prv}(\lceil \sigma \rceil),$$

where  $\sigma^+$  is  $\Pr(\lceil \sigma \rceil) \to \sigma$ , gives precisely the instances of the diagonalization technique needed to prove (L). In Sections 3 and 4 we consider the relationships between properties (L), (P), (Q), (C), (K) and (G) and their formatizations.

To obtain these results we require that T satisfies some basic requirements. We call such theories *suitable*. These theories are discussed in Section 1.

In Section 5 we make some remarks about the Lindenbaum algebra of suitable theories, and, finally, in Section 6 we give some further remarks and open questions.

Throughout the paper we assume that T is some theory in a first order language. We are thinking of languages in which number theory can be formulated. Our notations are standard. We remark here that  $\exists_1$  is the set of formulas in prenex normal form having at most one block of quantifiers, these being existential. Similarly, we have sets  $\forall_1, \exists_2, \forall_2, \dots$ 

1. Suitable theories. Let L be any language with means of self reference. Thus for each sentence  $\sigma$  (of L) there is a term  $\lceil \sigma \rceil$  (of L) which denotes  $\sigma$ . Using this self reference, we can formalize within certain L-theories T the provability predicate and the consistency statement of T. We say such theories are *suitable* (a precise definition is given below).

For a theory T to be suitable, we require the formalized version of the provability predicate and the consistency statement to have certain properties. Consider first the formal provability predicate  $Prv(\cdot)$ , which is a formula in just one free variable. We require  $Prv(\cdot)$  to be both adequate and sound:

(ADQ) For each sentence  $\sigma$ ,

$$T \vdash \sigma \Rightarrow T \vdash \Pr(\lceil \sigma \rceil).$$

(SND) For each pair of sentences  $\sigma_1$ ,  $\sigma_2$ ,

$$T_{\cup}\{\Pr(\lceil \sigma_1 \rightarrow \sigma_2 \rceil)\} \vdash \Pr(\lceil \sigma_1 \rceil) \rightarrow \Pr(\lceil \sigma_2 \rceil).$$

(ADQ) is concerned with the extensional correctness of  $Prv(\cdot)$  and (SND) is concerned with the intensional correctness of  $Prv(\cdot)$ . Notice that we do not assume the converse of (ADQ), namely the following

For each sentence  $\sigma$ ,

$$T \vdash \Pr( \lceil \sigma \rceil) \Rightarrow T \vdash \sigma.$$

In general, to obtain this property we need T to be  $\omega$ -consistent or we must use some fake provability predicate (e.g. one constructed using Rosser's trick). The first of these is too restrictive and the second is out because of (SND).

For some sentences  $\sigma$ , (ADQ) can be strengthened to

$$T \vdash \sigma \rightarrow \Pr(\lceil \sigma \rceil).$$

We say such sentences are *nice*. In general,  $\exists_1$ -sentences will be nice, i. e.

(EN) Each sentence  $\sigma \in \mathbf{I}_1$  is nice.

However, we require only

(PN) For each sentence  $\sigma$ ,  $Prv(\lceil \sigma \rceil)$  is nice.

Usually  $Prv(\cdot)$  will be an  $\exists_1$ -formula, so (PN) will be a special case of (EN).

Note that (ADQ), (SND) and (PN) are exactly conditions (IV), (I) and (V) of Löb [3].

Now, consider the formal consistency statement CON(T), which is a sentence. We require the following

(CNS) For each sentence  $\sigma$ ,

$$T \vdash \text{CON}(T) \rightarrow [ | \text{Prv}( | \sigma |) | | \text{Prv}( | \sigma |)].$$

Some theories also satisfy the following

(SNC) For each sentence  $\sigma$ ,

In general, CON(T) will be a sentence of the form

For all sentences  $\sigma$ , either  $\sigma$  is not provable or  $]\sigma$  is not provable,

so the verification of (CNS) is simply a matter of replacing a universal quantifier in CON(T) by a particular instance. However, (SNC) is concerned with the intensional correctness of  $Prv(\cdot)$  and CON(T).

The following is the diagonalization property first used by Gödel:

( $\Delta$ ) For each formula  $\varphi(\cdot)$  (having just one free variable), there is a sentence  $\delta$  such that

$$T \vdash \delta \leftrightarrow \varphi(\lceil \delta \rceil).$$

Definition. A theory T together with a formula  $Prv(\cdot)$  and a sentence CON(T) satisfying (ADQ), (SND), (PN) and (CNS) is *suitable*.

Notice that we do not assume that a suitable theory has (EN), (SNC) nor  $(\Delta)$ .

For each set of sentences Z, let

$$Prv[Z] = \{Prv(\lceil \sigma \rceil) : \sigma \in Z\}.$$

The following lemma is an important property of suitable theories: LEMMA 1.1. For any suitable theory T, set of sentences X, set of nice sentences Y, sentence  $\sigma$ ,

$$T \cup X \cup Y \vdash \sigma \Rightarrow T \cup \Pr(X) \cup Y \vdash \Pr(\neg \sigma \cap).$$

**Proof.** Let  $Z = X \cup Y$  so that  $Prv[Z] = Prv[X] \cup Prv[Y]$  and  $T \cup Y \vdash \tau$  for all  $\tau \in Prv[Y]$ . Thus it is sufficient to prove

$$T_{\cup}Z \vdash \sigma \Rightarrow T_{\cup}\Pr[Z] \vdash \Pr[C \cap \sigma]$$
.

Also it is sufficient to prove this implication for finite Z; we do this by induction on the size of Z.

For  $Z = \emptyset$  the implication is just (ADQ).

We now suppose the implication holds for Z and deduce the implication for  $Z' = Z_{\cup} \{\tau\}$  as follows:

Suppose that  $T_{\cup}Z' \vdash \sigma$ , so that  $T_{\cup}Z \vdash \tau \rightarrow \sigma$ . Thus the induction hypothesis gives

$$T_{\cup}\operatorname{Prv}[Z] \vdash \operatorname{Prv}(\lceil \tau \rightarrow \sigma \rceil),$$

and (SND) gives

$$T_{\cup}\operatorname{Prv}[Z] \vdash \operatorname{Prv}(\lceil \tau \rceil) \rightarrow \operatorname{Prv}(\lceil \sigma \rceil).$$

Hence we get /

$$T_{\cup}\operatorname{Prv}[Z'] \vdash \operatorname{Prv}(\lceil \sigma \rceil)$$

as required.

The following lemma contains three typical uses of Lemma 1.1. It follows immediately from Lemma 1.1 and (PN).

LEMMA 1.2. For any suitable theory T and sentences  $\sigma_1$ ,  $\sigma_2$ , the following hold:

- (i)  $T_{\cup}\{\Pr(\lceil \sigma_1 \rceil)\} \vdash \sigma_2 \Rightarrow T_{\cup}\{\Pr(\lceil \sigma_1 \rceil)\} \vdash \Pr(\lceil \sigma_1 \rceil).$
- (ii)  $T \vdash \sigma_1 \rightarrow \sigma_2 \Rightarrow T \vdash \Pr( \lceil \sigma_1 \rceil) \rightarrow \Pr( \lceil \sigma_2 \rceil).$

(iii) 
$$T_{\cup}\{\Pr(\lceil \sigma_1 \rceil)\} \vdash \sigma_1 \rightarrow \sigma_2 \Rightarrow T_{\cup}\{\Pr(\lceil \sigma_1 \rceil)\} \vdash \Pr(\lceil \sigma_2 \rceil).$$

There are many known suitable theories. Feferman in [1] gives an explicit construction for suitable theories. Let P be a Peano number theory and let Q be the theory of R. M. Robinson. Let B be the theory axiomatized by  $P \cap \bigvee_2$ . Thus  $Q \subseteq B \subseteq P$ , and Q is finitely axiomatized. We have elsewhere suggested that B is the basic number theory. Feferman's construction shows the following

THEOREM 1.3. Let T be any r. e. extension of B. Then T can be made suitable and also satisfies (EN), (SNC) and  $(\Delta)$ .

Proof. We make use of the results and notations of [1], which we refer to as F.

We have  $\exists_1 \subseteq BPF$  (see F3.8). Also the formalized version of Matijasevič's result [5] shows that, for each formula  $\psi \in BPF$ , there is a formula  $\theta \in \exists_1$  such that  $P \vdash \psi \leftrightarrow \theta$ . This together with F3.9 shows that for each formula  $\varphi \in RE$  there is a formula  $\theta \in \exists_1$  such that  $P \vdash \varphi \leftrightarrow \theta$ .

Let A be any r. e. set of axioms of T and let  $a(\cdot)$  be any RE formula which numerates A in Q. (The existence of such a formula follows from F3.10.) Let  $\Pr_a(\cdot)$  be the formula constructed in F4.3. Using F4.5 and the above-mentioned remarks, we have a formula  $\Pr(\cdot)$  in  $\exists_1$  such that  $P \vdash \Pr_a \leftrightarrow \Pr$ .

Using F4.4(i) and the actual construction of  $Pr_a$ , we have, for each sentence  $\sigma$ ,

$$T \vdash \sigma \Rightarrow P \vdash \Pr_{a}(\lceil \sigma \rceil)$$

$$\Rightarrow P \vdash \Pr(\lceil \sigma \rceil)$$

$$\Rightarrow B \vdash \Pr(\lceil \sigma \rceil)$$

$$\Rightarrow T \vdash \Pr(\lceil \sigma \rceil),$$

and so T has (ADQ).

Using F4.6(iv), we see that T has (SND).

Let  $\gamma$  be the sole axiom of Q so that  $T \vdash \gamma$ , and hence  $T \vdash \Pr_a(\lceil \gamma \rceil)$ . Also [Q](v) is the formula  $v = \lceil \gamma \rceil$ . Consider any formula  $\pi(\cdot)$ . A particular case of F4.7(i) is

$$P \hspace{0.2em}\vdash\hspace{0.9em} \pi(\hspace{0.2em}\lceil\hspace{0.9em}\gamma\hspace{0.9em}\rceil) \hspace{-0.9em}\rightarrow\hspace{-0.9em} (\forall v) \hspace{0.2em} [\operatorname{Pr}_{[Q]}(v) \hspace{-0.9em}\rightarrow\hspace{-0.9em} \operatorname{Pr}_{\pi}(v)] \hspace{0.5em}.$$

Hence, letting  $\pi(\cdot) = \Pr_{\alpha}(\cdot)$ , we have

$$P \vdash (\forall v) [\Pr_{[Q]}(v) \rightarrow \Pr_{\pi}(v)],$$

and so F4.7(ii) gives

$$P \vdash (\forall v)[\Pr_{[O]}(v) \rightarrow \Pr_{[V]}(v)].$$

Now consider any sentence  $\sigma \in \mathbf{I}_1$ . From F5.5 we have

$$P \vdash \sigma \rightarrow \Pr_{Q}(\lceil \sigma \rceil)$$

so that

$$P \vdash \sigma \rightarrow \Pr(\lceil \sigma \rceil).$$

Thus going via B we see that T has (EN). Also, since  $Prv \in \mathbb{I}_1$ , T has (PN).

Now let  $\operatorname{Con}_a$  be the sentence constructed in F4.9(ii). Since  $\operatorname{Fm}_K$  and  $\operatorname{Pr}_a$  are both RE, we have some sentence  $\operatorname{CON}(T) \in \bigvee_1$  such that

$$T \vdash \operatorname{Con}_a \leftrightarrow \operatorname{CON}(T)$$
.

Immediately from F4.9(ii) we have, for each sentence  $\sigma$ ,

$$P \vdash \text{CON}(T) \rightarrow []\text{Prv}(\lceil \sigma \rceil) \vee ]\text{Prv}(\lceil \rceil \sigma \rceil)],$$

hence going  $via\ B$  we see that T has (CNS). This shows that T is suitable. F4.10(i) gives

$$P \vdash \text{CON}(T) \leftrightarrow \text{Prv}(\lceil \bot \rceil),$$

where  $\perp$  is some logically false sentence. Thus we have

$$T \vdash \text{CON}(T) \leftrightarrow \text{Prv}(\lceil \bot \rceil)$$
.

Also, for any sentence  $\sigma$ , (ADQ) followed by (SND) gives

$$T \vdash \Pr(\lceil \bot \rceil) \rightarrow \Pr(\lceil \sigma \rceil)$$

so that T has (SNC).

Finally, F5.1 shows that Q and hence T has  $(\Delta)$ .

2. A version of the diagonalization property. Most applications of  $(\Delta)$  appear to be those using formulas

$$\varphi(\cdot) = \operatorname{Prv}(\cdot) \rightarrow \sigma$$

for some sentence  $\sigma$ , i. e. ( $\Delta$ ) is used to produce a sentence  $\delta$  satisfying

(D) 
$$T \vdash \delta \leftrightarrow [\Pr(\lceil \delta \rceil) \to \sigma].$$

The following theorem gives a useful equivalent to the existence of such a sentence  $\delta$ :

THEOREM 2.1. Let T be any suitable theory and  $\sigma$  any sentence. If there is a sentence  $\delta$  satisfying (D), then the following hold, where  $\sigma^+$  is the sentence  $\text{Prv}(\lceil \sigma \rceil) \rightarrow \sigma$ :

- (i)  $T \vdash \sigma^+ \leftrightarrow \delta$ .
- (ii)  $T \vdash \operatorname{Prv}(\lceil \sigma \rceil) \leftrightarrow \operatorname{Prv}(\lceil \sigma^+ \rceil)$ .
- (iii)  $T \vdash \Pr(\lceil \sigma \rceil) \leftrightarrow \Pr(\lceil \delta \rceil)$ .
- (iv)  $T \vdash \Pr(\lceil \sigma^+ \rceil) \leftrightarrow \Pr(\lceil \sigma \rceil)$ .

Conversely, if

$$T \vdash \Pr( \lceil \sigma^+ \rceil) \rightarrow \Pr( \lceil \sigma \rceil),$$

then such a sentence  $\delta$  exists, namely  $\delta = \sigma^+$ .

Proof. Suppose first that  $\delta$  satisfied (D) so that

$$T_{\cup}\{\operatorname{Prv}(\lceil \delta \rceil)\} \vdash \delta \rightarrow \sigma$$

from which Lemma 1.2(iii) gives

$$T \vdash \Pr(\lceil \delta \rceil) \rightarrow \Pr(\lceil \sigma \rceil).$$

We also have  $T \vdash \sigma \rightarrow \delta$  so that Lemma 1.2(ii) gives

$$T \vdash \Pr(\lceil \sigma \rceil) \rightarrow \Pr(\lceil \delta \rceil).$$

Hence we have (iii). Equivalence (i) now follows from (D) and (iii), and (iv) follows from (i) using Lemma 1.2(ii). Finally, (ii) follows from (iii) and (iv).

Now suppose that

$$T \vdash \Pr(\lceil \sigma^+ \rceil) \rightarrow \Pr(\lceil \sigma \rceil).$$

Thus, since  $T \vdash \sigma \rightarrow \sigma^+$ , Lemma 1.2(ii) gives us (ii), and so

$$T \vdash \sigma^+ \leftrightarrow [\Pr(\lceil \sigma^+ \rceil) \rightarrow \sigma].$$

Hence (D) is satisfied by  $\delta = \sigma^+$ .

**3. Properties** (L), (P), (Q) and (C). Given some fixed suitable theory T we can associate with each sentence  $\sigma$  four other sentences  $\sigma^+$ ,  $\sigma^-$ ,  $\sigma^\vee$  and  $\sigma^c$  as follows:

$$egin{aligned} \sigma^+ &= \operatorname{Prv}(\, \lceil \sigma \, \rceil) \! o \sigma, & \sigma^- &= \operatorname{Prv}(\, \lceil \, ]\sigma \, \rceil) \! o \sigma, \ \sigma^{arphi} &= \sigma^+ \ _{ee} \ \sigma^-, & \sigma^{\operatorname{c}} &= \ |\operatorname{CON}(T) \! o \sigma. \end{aligned}$$

The theory T can have one or more of the following four properties:

(L) For each sentence  $\sigma$ ,

$$T \vdash \sigma^+ \Rightarrow T \vdash \sigma$$
.

(P) For each sentence  $\sigma$ ,

$$T \vdash \sigma^{\vee} \Rightarrow T \vdash \sigma^{-}$$
.

(Q) For each sentence  $\sigma$ ,

$$T_{\cup}\{\Pr(\lceil \rceil \sigma \rceil)\} \vdash \sigma^+ \Rightarrow T_{\cup}\{\Pr(\lceil \rceil \sigma \rceil)\} \vdash \sigma.$$

(C) For each sentence  $\sigma$ ,

$$T \vdash \sigma^{c} \Rightarrow T \vdash \sigma^{-}$$
.

Of course, (L) is the property considered by Löb in [3].

In this section we consider the relationships between these four and various other properties. Each of the above-mentioned four properties are implied by the corresponding formalized versions. These are as follows:

(FL) For each sentence  $\sigma$ ,

$$T \vdash \Pr(\lceil \sigma^+ \rceil) \rightarrow \Pr(\lceil \sigma \rceil).$$

(FP) For each sentence  $\sigma$ ,

$$T \vdash \operatorname{Prv}(\lceil \sigma^{\vee} \rceil) \rightarrow \operatorname{Prv}(\lceil \sigma^{-} \rceil).$$

(FQ) For each sentence  $\sigma$ ,

$$T_{\cup}\{\Pr(\lceil \sigma \rceil)\} \models \Pr(\lceil \sigma^+ \rceil) \rightarrow \Pr(\lceil \sigma \rceil).$$

(FC) For each sentence  $\sigma$ ,

$$T \vdash \Pr(\lceil \sigma^{c} \rceil) \rightarrow \Pr(\lceil \sigma^{-} \rceil).$$

Note that (FQ) is not a genuine formalization of (Q).

To deduce a property from its formalized version it is convenient to go through a weak formalized version. These are as follows:

(WFL) For each sentence  $\sigma$ ,

$$T \vdash \Pr(\lceil \sigma^+ \rceil) \Rightarrow T \vdash \Pr(\lceil \sigma \rceil).$$

(WFP) For each sentence  $\sigma$ ,

$$T \vdash \operatorname{Prv}(\lceil \sigma^{\vee} \rceil) \Rightarrow T \vdash \operatorname{Prv}(\lceil \sigma^{-} \rceil).$$

(WFQ) For each sentence  $\sigma$ ,

$$T_{\cup}\{\Pr(\lceil \rceil \sigma \rceil)\} \vdash \Pr(\lceil \sigma^+ \rceil) \Rightarrow T_{\cup}\{\Pr(\lceil \rceil \sigma \rceil)\} \vdash \Pr(\lceil \sigma \rceil).$$

(WFC) For each sentence  $\sigma$ ,

$$T \vdash \Pr(\lceil \sigma^{c} \rceil) \Rightarrow T \vdash \Pr(\lceil \sigma^{-} \rceil).$$

In this section we prove the following theorem:

THEOREM 3.1. For any suitable theory T the following implications hold:

(a) 
$$(\Delta) \Rightarrow (\mathbf{FL})$$
.

(b) 
$$(FL) \Rightarrow (FP) \Rightarrow (FC)$$

$$\psi \qquad \psi \qquad \psi$$

$$(WFL) \Rightarrow (WFP) \Rightarrow (WFC)$$

$$\psi \qquad \psi \qquad \psi$$

$$(L) \Rightarrow (P) \Rightarrow (C).$$

(c) 
$$(\mathbf{FQ}) \Leftrightarrow (\mathbf{FP})$$

$$\downarrow \qquad \qquad \downarrow \qquad \qquad \downarrow$$
 $(\mathbf{WFQ}) \Rightarrow (\mathbf{WFP})$ 

$$\downarrow \qquad \qquad \downarrow \qquad \qquad \downarrow$$
 $(\mathbf{Q}) \Leftrightarrow (\mathbf{P}).$ 

This theorem will be proved in a series of lemmas. Several of these lemmas will be stated without proofs since these proofs are very easy. Throughout we have the blanket assumption that T is a suitable theory.

LEMMA 3.2. For each sentence  $\sigma$ ,  $\sigma^{\vee}$  is logically equivalent to any of the following three sentences:

$$[\operatorname{Prv}(\lceil \sigma \rceil) \wedge \operatorname{Prv}(\lceil \rceil \sigma \rceil)] \to \sigma, \qquad \operatorname{Prv}(\lceil \rceil \sigma \rceil) \to \sigma^+, \qquad \operatorname{Prv}(\lceil \sigma \rceil) \to \sigma^-.$$

LEMMA 3.3. For any sentence  $\sigma$ , the following hold:

(i) 
$$T \vdash \sigma \rightarrow \sigma^+$$
.

(iv) 
$$T \vdash \sigma^+ \rightarrow \sigma^\vee$$
.

(ii) 
$$T \vdash \sigma \rightarrow \sigma^{c}$$
.

(v) 
$$T \vdash \sigma^{c} \rightarrow \sigma^{\vee}$$
.

(iii) 
$$T \vdash \sigma \rightarrow \sigma^-$$
.

(vi) 
$$T \vdash \sigma^- \rightarrow \sigma^{\vee}$$
.

Proof. (v) follows from (CNS) and the remaining implications are logical.

LEMMA 3.4. For any sentence  $\sigma$ , the following hold:

(i) 
$$T_{\cup}\{\Pr(\lceil \rceil \sigma \rceil) \to \Pr(\lceil \sigma \rceil)\} \vdash \sigma^{\vee} \to \sigma^{-}$$
.

(ii) 
$$T_{\vee}\{\text{Prv}(\lceil \sigma^{-} \rceil)\} \vdash \sigma^{\vee} \rightarrow \sigma^{-}$$
.

LEMMA 3.5. For any sentence  $\sigma$ , the following hold:

(i) 
$$T_{\cup}\{\Pr(\lceil \rceil \sigma \rceil)\} \vdash \sigma \leftrightarrow \sigma^{-}$$
.

(ii) 
$$T_{\cup}\{\Pr(\lceil \rceil \sigma \rceil)\} \vdash \Pr(\lceil \sigma \rceil) \leftrightarrow \Pr(\lceil \sigma \rceil)$$
.

(iii) 
$$T_{\cup}\{\Pr(\lceil \rceil \sigma \rceil)\} \vdash \sigma^+ \leftrightarrow \sigma^{\vee}$$
.

(iv) 
$$T_{\cup}\{\Pr(\lceil \neg \sigma \rceil)\} \vdash \Pr(\lceil \sigma^{+} \rceil) \leftrightarrow \Pr(\lceil \sigma^{\vee} \rceil).$$

LEMMA 3.6. For any sentence  $\sigma$ , the following holds:

$$T \vdash \sigma^{\vee} \leftrightarrow \sigma^{-+}$$
.

**Proof.** We know that  $\sigma^{\vee}$  is logically equivalent to

$$[\Pr(\lceil \sigma \rceil) \land \Pr(\lceil \rceil \sigma \rceil)] \rightarrow \sigma$$

and  $\sigma^{-+}$  is  $Prv(\lceil \sigma^{-} \rceil) \rightarrow \sigma^{-}$  which is logically equivalent to

$$[\operatorname{Prv}(\lceil \sigma^- \rceil) \wedge \operatorname{Prv}(\lceil \sigma \rceil)] \rightarrow \sigma.$$

But Lemma 3.5(ii) gives

$$T \vdash [\Pr(\lceil \sigma \rceil) \land \Pr(\lceil \lceil \sigma \rceil)] \leftrightarrow [\Pr(\lceil \sigma \rceil) \land \Pr(\lceil \lceil \sigma \rceil)]$$

which gives the required result.

Although the following lemma is not used in the proof of Theorem 3.1, it is of interest:

LEMMA 3.7. Suppose T has (SNC). For any sentence  $\sigma$ , the following holds:

$$T \vdash \sigma^{c} \leftrightarrow \sigma^{\vee}$$
.

In the next three lemmas properties (FL), (FP), (FQ) and (FC) are considered.

LEMMA 3.8. For any sentence  $\sigma$ , the following are equivalent:

- (i)  $T \vdash \Pr(\lceil \sigma^+ \rceil) \rightarrow \Pr(\lceil \sigma \rceil)$ .
- (ii)  $T_{\cup}\{\text{Prv}(\lceil \sigma^+ \rceil)\} \vdash \sigma^+ \rightarrow \sigma$ .

Proof. Implication (i)  $\Rightarrow$  (ii) is trivial and (ii)  $\Rightarrow$  (i) follows by an application of Lemma 1.2(iii).

LEMMA 3.9. For any sentence  $\sigma$ , the following are equivalent:

- (i)  $T \vdash \Pr(\lceil \sigma^{\vee} \rceil) \rightarrow \Pr(\lceil \sigma^{-} \rceil)$ .
- (ii)  $T_{\cup}\{\Pr(\lceil \rceil \sigma \rceil)\} \vdash \Pr(\lceil \sigma^+ \rceil) \rightarrow \Pr(\lceil \sigma \rceil).$
- (iii)  $T_{\cup}\{\Pr(\lceil \sigma^{\vee} \rceil)\} \vdash \sigma^{\vee} \rightarrow \sigma^{-}$ .

Proof. Equivalence (i)⇔(iii) follows from Lemmas 3.6 and 3.8. Implication (i)⇒(ii) follows from Lemma 3.5(ii), (iv).

From (ii) and Lemma 3.5(iv) we have

$$T_{\cup}\{\Pr(\lceil \sigma^{\vee} \rceil)\} \vdash \Pr(\lceil \lceil \sigma \rceil) \rightarrow \Pr(\lceil \sigma \rceil),$$

and so implication (ii) > (iii) follows from Lemma 3.4(i).

LEMMA 3.10. For any sentence  $\sigma$ , the following are equivalent:

- (i)  $T \vdash \Pr(\lceil \sigma^c \rceil) \rightarrow \Pr(\lceil \sigma^- \rceil)$ .
- (ii)  $T_{\cup}\{\Pr(\lceil \sigma^c \rceil)\} \vdash \Pr(\lceil \lceil \sigma \rceil) \rightarrow \Pr(\lceil \sigma \rceil).$
- (iii)  $T_{\cup}\{\operatorname{Prv}(\lceil \sigma^{\operatorname{c}} \rceil)\} \vdash \sigma^{\vee} \rightarrow \sigma^{-}$ .
- (iv)  $T_{\cup}\{\Pr(\lceil \sigma^{c} \rceil)\} \vdash \sigma^{c} \rightarrow \sigma^{-}$ .

Proof. Implication (i)  $\Rightarrow$  (ii) follows by (SND), implication (ii)  $\Rightarrow$  (iii) by Lemma 3.4(i), implication (iii)  $\Rightarrow$  (iv) by Lemma 3.3(v), and, finally, (iv)  $\Rightarrow$  (i) follows by Lemma 1.2(iii).

In the next four lemmas properties (WFL), (WFQ), (WFP) and (WFC) are considered.

LEMMA 3.11. For any sentence  $\sigma$  such that  $T \vdash \text{Prv}(\lceil \sigma^+ \rceil)$ , the following are equivalent:

- (i)  $T \vdash \operatorname{Prv}(\lceil \sigma \rceil)$ .
- (ii)  $T \vdash \sigma^+ \rightarrow \sigma$ .

Proof. Implication (i)  $\Rightarrow$  (ii) is straightforward, and (ii)  $\Rightarrow$  (i) follows by an application of Lemma 1.2(ii).

In the next three lemmas we consider the equivalence of the following four conditions on the sentence  $\sigma$ :

- (a)  $T_{\cup}\{\Pr(\lceil \rceil \sigma \rceil)\} \vdash \Pr(\lceil \sigma \rceil)$ .
- (b)  $T_{\cup}\{\Pr(\lceil \rceil \sigma \rceil)\} \vdash \sigma^+ \rightarrow \sigma$ .
- (c)  $T \vdash \sigma^{\vee} \rightarrow \sigma^{-}$ .
- (d)  $T \vdash \Pr(\lceil \sigma^{\vee} \rceil) \rightarrow \Pr(\lceil \sigma^{-} \rceil)$ .

LEMMA 3.12. For any sentence  $\sigma$  such that

$$T_{\cup}\{\Pr(\lceil \rceil \sigma \rceil)\} \vdash \Pr(\lceil \sigma^+ \rceil),$$

conditions (a), (b), (c) and (d) are equivalent.

Proof. Equivalence (a) $\Rightarrow$ (b) follows by the proof of Lemma 3.11. Implication (a) $\Rightarrow$ (c) follows by Lemma 3.4(i), and (c) $\Rightarrow$ (d) follows by Lemma 1.2(ii). It remains to prove (d) $\Rightarrow$ (a).

Since

$$T_{\cup}\{\Pr(\lceil \rceil \sigma \rceil)\} \vdash \Pr(\lceil \sigma^+ \rceil),$$

we have from Lemma 3.5(iv) and (d)

$$T_{\cup}\{\Pr(\lceil \rceil \sigma \rceil)\} \vdash \Pr(\lceil \sigma^- \rceil),$$

and so Lemma 3.5(ii) gives (a).

LEMMA 3.13. For any sentence  $\sigma$  such that  $T \vdash \Pr(\lceil \sigma^{\vee} \rceil)$ , conditions (a), (b), (c) and (d) are each equivalent to

(e) 
$$T \vdash \Pr(\lceil \sigma^- \rceil)$$
.

**Proof.** The hypothesis  $T \vdash \operatorname{Prv}(\lceil \sigma^{\vee} \rceil)$  gives

$$T_{\cup}\{\Pr(\lceil \rceil \sigma \rceil)\} \vdash \Pr(\lceil \sigma^+ \rceil),$$

and so (a), (b), (c) and (d) are equivalent by Lemma 3.12. The equivalence of (e) follows using (d).

LEMMA 3.14. For any sentence  $\sigma$  such that  $T \vdash \Pr(\lceil \sigma^c \rceil)$ , conditions (a), (b), (c), (d) and (e) are each equivalent to

(f) 
$$T \vdash \sigma^{c} \rightarrow \sigma^{-}$$
.

We are now in a position to prove Theorem 3.1.

Proof of Theorem 3.1. Implication  $(\Delta) \Rightarrow (FL)$  follows from Theorem 2.1.

Implications (FL)  $\Rightarrow$  (FP), (WFL)  $\Rightarrow$  (WFP) and (L)  $\Rightarrow$  (P) follow from Lemma 3.6, and implications (FP)  $\Rightarrow$  (FC), (WFP)  $\Rightarrow$  (WFC) and (P)  $\Rightarrow$  (C) follow from Lemma 3.3(v). Each of implications (FL)  $\Rightarrow$  (WFL), (FP)  $\Rightarrow$  (WFP) and (FC)  $\Rightarrow$  (WFC) is trivial, and implications (WFL)  $\Rightarrow$  (L), (WFP)  $\Rightarrow$  (P) and (WFC)  $\Rightarrow$  (C) follow by (ADQ) and Lemmas 3.11 or 3.13 or 3.14, respectively.

Equivalence  $(Q)\Leftrightarrow(P)$  follows from Lemma 3.2, and equivalence  $(FQ)\Leftrightarrow(FP)$  follows from Lemma 3.9. Implication  $(WFQ)\Rightarrow(WFP)$  follows from Lemma 3.13. Implication  $(FQ)\Rightarrow(WFQ)$  is trivial, and implication  $(WFQ)\Rightarrow(Q)$  follows from Lemmas 1.3(i) and 3.12.

**4. Properties** (K) and (G). We could now consider the properties of Section 3 restricted to  $\forall_1$ -sentences. The negation of an  $\forall_1$ -sentence is  $\exists_1$ , and hence (assuming (EN)) is nice. Thus, for each property (X), we could consider the restriction (X') of (X) obtained by replacing the

quantifier "For each sentence  $\sigma$ " be the quantifier "For each sentence  $\sigma$  such that  $|\sigma|$  is nice". The properties so obtained can be dealt with by noting the following lemma:

LEMMA 4.1. For each sentence  $\sigma$  the following are equivalent,

- (i)  $]\sigma$  is nice,
- (ii)  $T \vdash \sigma^- \rightarrow \sigma$ ,

and the following are equivalent,

- (iii)  $T \vdash \sigma^{\vee} \rightarrow \sigma^{+}$ ,
- (iv)  $T \vdash \sigma^- \rightarrow \sigma^+$ .

Also (i)  $\Rightarrow$  (iii).

Proof. Equivalence (i) $\Leftrightarrow$ (ii) follows since  $\neg \sigma \to \Pr(\neg \sigma \neg)$  and  $\sigma^- \to \sigma$  are logically equivalent. Implication (iii)  $\Rightarrow$  (iv) follows from Lemma 3.3(vi), and implication (iv)  $\Rightarrow$  (iii) follows since Lemma 3.2 gives

$$T_{\cup}\{\sigma^{\vee}, \operatorname{Prv}(\lceil \sigma \rceil)\} \vdash \sigma^{-},$$

so that (iv) gives

$$T_{\cup}\{\sigma^{\vee}, \operatorname{Prv}(\lceil \sigma \rceil)\} \vdash \sigma^{+}$$

which gives (iii). Clearly, (ii)  $\Rightarrow$  (iv), and so (i)  $\Rightarrow$  (iii).

We are particularly interested in (C'), i.e. the following property:

(K) For each sentence  $\sigma$  such that  $]\sigma$  is nice

$$T \vdash \sigma^{c} \Rightarrow T \vdash \sigma$$
.

This is related to Gödel's property

(G) 
$$\operatorname{Not}[T \vdash \operatorname{CON}(T)]$$

and its formalization

$$(\mathbf{FG}) \qquad \qquad T \vdash \mathbf{CON}(T) \to \mathsf{Prv}(\mathsf{CON}(T) \mathsf{D}).$$

We prove the following theorem:

THEOREM 4.2. For any suitable theory T, the following implication hold:

$$(C) \Rightarrow (FG)$$

$$\psi \qquad \psi \\
(\mathbf{K}) \Rightarrow (\mathbf{G}).$$

Proof. We have already seen that  $(C) \Rightarrow (K)$ . To see that  $(C) \Rightarrow (FG)$ , we note that  $[CON(T)]^{\circ}$  is logically valid, and so (C) gives

$$T \vdash []CON(T)]^{-}$$

which is

(FG) now follows from (SND).

Implication  $(FG) \Rightarrow (G)$  holds since (ADQ) gives

$$T \vdash \text{CON}(T) \Rightarrow T \vdash \text{Prv}(\lceil \text{CON}(T) \rceil),$$

and so (FG) gives

$$T \vdash CON(T) \Rightarrow T \vdash CON(T)$$

which gives (G).

To see that  $(K) \Rightarrow (G)$  we suppose  $T \vdash CON(T)$  so that  $T \vdash \sigma^c$  for all sentence  $\sigma$ . Thus (K) gives  $T \vdash \sigma$  for all sentence  $\sigma$  such that  $|\sigma|$  is nice. However, there is at least one such sentence which is not provable (e.g.,  $0 \neq 0$ ). Thus we have (G).

From this theorem we obtain the following precise version of Kreisel's remark [2]:

THEOREM 4.3. Let T be any suitable theory having  $(\Delta)$  and (EN). For each sentence  $\sigma \in \bigvee_{1}$ ,

$$T_{\cup}\{|CON(T)\} \vdash \sigma \Rightarrow T \vdash \sigma.$$

5. The Lindenbaum algebra of a suitable theory. Let B(T) be the Lindenbaum algebra of the suitable theory T. The elements of B(T) are equivalence classes of sentences where two sentences  $\sigma_1$  and  $\sigma_2$  are equivalent if  $T \vdash \sigma_1 \leftrightarrow \sigma_2$ . We will confuse the sentences with the equivalence classes in which they lie. B(T) is a boolean algebra under the natural operations and the ordering of B(T) is given by

$$\sigma_1 \leqslant \sigma_2 \Leftrightarrow T \vdash \sigma_1 \rightarrow \sigma_1.$$

From Lemma 1.2(ii) we see that  $Prv(\cdot)$  can be considered as an order preserving operation on B(T). Similarly,  $(\cdot)^+$  and  $(\cdot)^-$  can be considered as operations on B(T). In this section we consider some of the properties of these two operations.

THEOREM 5.1. For any suitable theory T, the operation  $(\cdot)^-$  is a closure operation, i.e. it is increasing, idempotent and order preserving.

Proof. Lemma 3.3(iii) shows that (·) is increasing.

Consider any sentence  $\sigma$ . Since  $|\sigma^- \leqslant |\sigma$ , Lemma 1.2(ii) gives  $\Pr(\lceil |\sigma^- |) \leqslant \Pr(\lceil |\sigma^- |)$ . Also we have

$$T_{\cup}\{\Pr(\lceil \rceil \sigma \rceil)\} \vdash \lceil \sigma \rightarrow \rceil \sigma^{-},$$

so that Lemma 1.2(iii) gives  $\Pr(\lceil \rceil \sigma \rceil) \leq \Pr(\lceil \rceil \sigma^- \rceil)$ . Hence we have  $\Pr(\lceil \rceil \sigma \rceil) = \Pr(\lceil \rceil \sigma^- \rceil)$ . Using this, we have

$$\sigma^{--} = \Pr(\lceil \rceil \sigma^{-} \rceil) \rightarrow \sigma 
= [\Pr(\lceil \rceil \sigma^{-} \rceil)_{\land} \Pr(\lceil \rceil \sigma^{-} \rceil)] \rightarrow \sigma 
= \Pr(\lceil \rceil \sigma^{-} \rceil) \rightarrow \sigma 
= \sigma^{-},$$

so that  $(\cdot)^-$  is idempotent.

Finally, consider any two sentences  $\sigma_1 \leqslant \sigma_2$ , so that  $\Pr(\lceil |\sigma_2| \rceil) \leqslant \Pr(\lceil |\sigma_1| \rceil)$ . Then

$$\begin{array}{ll} \sigma_1^- &= \operatorname{Prv}(\, | \, \neg \sigma_1 \, \neg) \rightarrow \sigma_1 \\ &\leqslant \operatorname{Prv}(\, | \, \neg \sigma_2 \, \neg) \rightarrow \sigma_1 \\ &\leqslant \operatorname{Prv}(\, | \, \neg \sigma_2 \, \neg) \rightarrow \sigma_2^- \\ &= \sigma_2^- \,, \end{array}$$

and so  $(\cdot)^-$  is order preserving.

THEOREM 5.2. For any suitable theory T, the operation  $(\cdot)^+$  is both increasing and idempotent.

**Proof.** Lemma 3.3(i) shows that  $(\cdot)^+$  is increasing.

Consider any sentence  $\sigma$ . Since  $\sigma \leqslant \sigma^+$ , Lemma 1.2(ii) gives  $\Pr(\lceil \sigma \rceil) \leqslant \Pr(\lceil \sigma^+ \rceil)$ , so that

$$\sigma^{++} = \Pr(\lceil \sigma^{+} \rceil) \rightarrow \sigma^{+}$$

$$= [\Pr(\lceil \sigma^{+} \rceil) \land \Pr(\lceil \sigma \rceil)] \rightarrow \sigma$$

$$= \Pr(\lceil \sigma \rceil) \rightarrow \sigma$$

$$= \sigma^{+}.$$

Thus  $(\cdot)^+$  is idempotent.

The order preserving properties  $(\cdot)^+$  are more involved.

THEOREM 5.3. Let T be a suitable theory having (FL). For any two sentences  $\sigma_1$  and  $\sigma_2$ ,

$$\sigma_1 \leqslant \sigma_2 \leqslant \sigma_1^+ \Rightarrow \sigma_1^+ = \sigma_2^+$$
.

The operation  $(\cdot)^+$  is never order inverting.

**Proof.** From  $\sigma_1 \leqslant \sigma_2 \leqslant \sigma_1^+$  we have

$$\operatorname{Prv}(\lceil \sigma_1 \rceil) \leqslant \operatorname{Prv}(\lceil \sigma_2 \rceil) \leqslant \operatorname{Prv}(\lceil \sigma_1^+ \rceil),$$

and since (FL) gives  $Prv(\lceil \sigma_1^+ \rceil) \leq Prv(\lceil \sigma_1 \rceil)$ , we have

$$\Pr(\lceil \sigma_1 \rceil) = \Pr(\lceil \sigma_2 \rceil) = \Pr(\lceil \sigma_1^+ \rceil) = \pi$$

for some sentence  $\pi$ . Thus

$$egin{array}{ll} \sigma_1^+ &= \pi \! 
ightarrow \! \sigma_1 \ &\leqslant \pi \! 
ightarrow \! \sigma_2 &= \sigma_2^+ \ &\leqslant \pi \! 
ightarrow \! \sigma_1^+ &= \sigma_1^{++}, \end{array}$$

and so Theorem 5.2 gives  $\sigma_1^+ = \sigma_2^+$ , as required.

The operation  $(\cdot)^+$  is never order inverting for if  $\sigma_1 \leqslant \sigma_2$  and  $\sigma_2^+ \leqslant \sigma_1^+$ , then (since  $\sigma_2 \leqslant \sigma_2^+$ ) we have  $\sigma_1^+ = \sigma_2^+$ .

THEOREM 5.4. Let T be a suitable theory having (L) and  $(\cdot)^+$  order preserving. Then the following hold:

- (i) For each nice sentence  $\sigma$ ,  $T \vdash \Pr(\lceil \sigma \rceil)$ .
- (ii)  $T \vdash |CON(T)|$ .

Proof. Let  $\sigma$  be any nice sentence. Thus  $\sigma \leq \Pr(\lceil \sigma \rceil)$ , and so  $\sigma^+ \leq \Pr(\lceil \sigma \rceil)^+$ . But the sentence  $\sigma^+ \to \Pr(\lceil \sigma \rceil)^+$  is logically equivalent to  $\Pr(\lceil \sigma \rceil)^+$ , and so  $T \vdash \Pr(\lceil \sigma \rceil)^+$ . An application of (L) now gives (i).

Now, consider any sentence  $\sigma$  such that both  $\sigma$  and  $\sigma$  are nice (e.g.  $\sigma$  = 0). Part (i) gives

$$T \vdash \operatorname{Prv}(\lceil \sigma \rceil) \land \operatorname{Prv}(\lceil \lceil \sigma \rceil),$$

and so (ii) follows from (CNS).

Proof. By the theorem it is sufficient to show that  $(\cdot)^+$  is order preserving whenever  $T \vdash |CON(T)|$ . However,  $T \vdash |CON(T)|$  together with (SNC) give  $T \vdash Prv(\lceil \sigma \rceil)$  for each sentence  $\sigma$ . Thus  $\sigma^+ = \sigma$  and  $(\cdot)^+$  is trivially order preserving.

THEOREM 5.6. Let T be a suitable theory having (FL). For any two sentences  $\sigma_1$  and  $\sigma_2$ ,  $(\sigma_1 \wedge \sigma_2)^+ = (\sigma_1^+ \wedge \sigma_2^+)^+$ .

Proof. Since  $(\sigma_{1} \wedge \sigma_{2}) \leqslant \sigma_{1}$ , we have  $Prv(\lceil \sigma_{1} \wedge \sigma_{2} \rceil) \leqslant Prv(\lceil \sigma_{1} \rceil)$ , so that

$$\begin{aligned}
\sigma_1^+ &= \Pr(\lceil \sigma_1 \rceil) \to \sigma_1 \\
&\leqslant \Pr(\lceil \sigma_{1 \wedge} \sigma_2 \rceil) \to \sigma_1.
\end{aligned}$$

Similarly,

$$\sigma_2^+ \leqslant \Pr(\lceil \sigma_1 \wedge \sigma_2 \rceil) \rightarrow \sigma_2,$$

so that

$$egin{aligned} (\sigma_{1\ \wedge}\,\sigma_{2}) &\leqslant \sigma_{1\ \wedge}^{+}\,\sigma_{2}^{+} \ &\leqslant \operatorname{Prv}(\, \lceil \sigma_{1\ \wedge}\,\sigma_{2}\, 
ceil) {
ightarrow} (\sigma_{1\ \wedge}\,\sigma_{2}) \ &= (\sigma_{1\ \wedge}\,\sigma_{2})^{+}\,. \end{aligned}$$

The required result now follows from Theorem 5.3.

Theorems 5.3 and 5.6 show that, for certain suitable theories, the set  $\nabla \sigma = \{\tau : \tau^+ = \sigma^+\}$  is a filter-like block below  $\sigma^+$  in B(T).

6. Final remarks. It is clear that suitable theories could be defined more abstractly along the lines of Smullyan's representation systems. The results of this paper could then be proved in this abstract setting. The standard interpretation of these abstract suitable theories are the first order theories of this paper; however, there quite probably will be other interpretations. For instance, theories in high order languages, free variable systems, intuitionistic systems and modal systems are possible interpretations.

It would be interesting to see such a program carried out. We point out that here the only connectives used explicitly are  $\rightarrow$  and ] (no quantifiers are used), compactness is not used, and one or two tautologies are used.

In what sense is diagonalization needed for Gödel's theorem (G)? We can also ask the same question about the other properties. (P856)

One way of answering these questions is to find general principles which are weaker than  $(\Delta)$  but stronger than (some of) the other properties.

For each of several properties (X) we have seen that  $(FX) \Rightarrow (X)$ . In what sense is this a general phenomenon? Which properties can have formal versions? (**P857**)

It is interesting to compare the proofs of  $(FG) \Rightarrow (G)$  and  $(K) \Rightarrow (G)$  (see Theorem 4.2). Each proof goes *via* a contradiction, however, we have

- $(FG) + not(G) \Rightarrow explicit contradiction,$ 
  - $(K) + not(G) \Rightarrow too many sentences are provable.$

In general, the two contradictions are not equivalent but are equivalent for first order theories.

Another approach to suitable theories T could be via its Lindenbaum algebra B(T).

In what sense is B(T) different from other Boolean algebras? Are there any reasonable conditions on B(T) which are equivalent to  $(\Delta)$  or any of the other properties? (**P858**)

The two operations  $(\cdot)^+$  and  $(\cdot)^-$ , as well as several other obvious operations, deserve further study.

## REFERENCES

- [1] S. Feferman, Arithmetization of metamathematics in a general setting, Fundamenta Mathematicae 49 (1960), p. 35-92.
- [2] G. Kreisel, On weak completeness of intuitionistic predicate logic, The Journal of Symbolic Logic 27 (1962), p. 139-158.
- [3] M. H. Löb, Solution of a problem of Leon Henkin, ibidem 20 (1955), p. 115-118.
- [4] A. Macintyre, On the forcing companion of peano arithmetic, Cambridge Summer School and Logic Colloquium (1971), abstract.
- [5] Yu. V. Matijasevič, Enumerable sets are diophantine, Soviet Mathematics 11 (1970), p. 354-358.

Reçu par la Rédaction le 8. 5. 1972