THE BULLETIN OF SYMBOLIC LOGIC
Volume 13, Number 1, March 2007

INCOMPLETENESS IN A GENERAL SETTING

JOHN L. BELL

Full proofs of the Gödel incompleteness theorems are highly intricate affairs. Much of the intricacy lies in the details of setting up and checking the properties of a coding system representing the syntax of an object language (typically, that of arithmetic) within that same language. These details are seldom illuminating and tend to obscure the core of the argument. For this reason a number of efforts have been made to present the essentials of the proofs of Gödel's theorems without getting mired in syntactic or computational details. One of the most important of these efforts was made by Löb [8] in connection with his analysis of sentences asserting their own provability. Löb formulated three conditions (now known as the *Hilbert-Bernays-Löb* derivability conditions), on the provability predicate in a formal system which are jointly sufficient to yield the Gödel's second incompleteness theorem for it. A key role in Löb's analysis is played by (a special case of) what later became known as the *diagonalization* or *fixed point* property of formal systems, a property which had already, in essence, been exploited by Gödel in his original proofs of the incompleteness theorems. The fixed point property plays a central role in Lawvere's [7] category-theoretic account of incompleteness phenomena (see also [10]). Incompleteness theorems have also been subjected to intensive investigation within the framework of *modal logic* (see, e.g., [4], [5]). In this formulation the modal operator takes up the role previously played by the provability predicate, and the derivability conditions on the latter are translated into algebraic conditions (the so-called *GL*, i.e., Gödel-Löb, conditions) on the former.

My purpose here is to present a framework for incompleteness phenomena, fully compatible with intuitionistic or constructive principles, in which the idea of a coding system is *retained*, only in a simple, but very general form, a form wholly free of syntactical notions. As codes we shall take the elements of an arbitrary given nonempty set, possibly, but not necessarily, the set of natural numbers. As the things to be encoded we take, not the expressions of a formal language, but rather what we shall call *propositions*, which may be

Received July 2, 2006.

understood as sentences identified under provable equivalence with respect to a theory.[1]

We shall take as our background theory intuitionistic set theory in any of its usual formulations (e.g., that presented in [2]).[2] We assume given two sets: Ω, the set of *propositions*, and C, the set of *codes*: we also assume that both Ω and C contain at least one element. The elements of the exponential[3] Ω^C may be considered propositional functions on C, or formulas with one free variable ranging over C. Constant elements of Ω^C may be identified with propositions. For $\omega \in \Omega$ we write $\tilde{\omega}$ for the map $C \to \Omega$ with constant value ω.

We also assume given:

- A submonoid L of Ω^Ω.[4] The members of L are called *logical maps*.
- A subset K of Ω^C containing each constant function $\tilde{\omega}$. The members of K are called *coded functions*. We suppose that L acts on K, that is, Ω is closed under composition with members of L on the left.
- A *coding map* $k\colon K \to C$. We shall usually write $\ulcorner\varphi\urcorner$ for $k(\varphi)$. By abuse of notation we shall, for $\omega \in \Omega$, write simply $\ulcorner\omega\urcorner$ for $\ulcorner\tilde{\omega}\urcorner$. The elements $\ulcorner\varphi\urcorner, \ulcorner\omega\urcorner$ are called the *codes* of φ and ω respectively.
- A *substitution map* $s\colon C \times C \to C$ satisfying $s(\ulcorner\varphi\urcorner, u) = \ulcorner\varphi u\urcorner$ for $\varphi \in K$, $u \in C$.

We define the associated *diagonal map* $d\colon C \to C$ by

$$d(u) = s(u, u) \text{ for } u \in C.$$

Clearly d satisfies

$$d(\ulcorner\varphi\urcorner) = \ulcorner\varphi(\ulcorner\varphi\urcorner)\urcorner \text{ for } \varphi \in K.$$

We shall assume that K is closed under composition with d on the right: if $\varphi \in K$, then $\varphi \circ d \in K$.

---

[1] We note that for classical theories propositions in the above sense form a Boolean algebra, but for the intuitionistic theories which we shall have in mind propositions constitute a *Heyting* algebra, that is, a (distributive) lattice $(L, \leq, \wedge, \vee)$ with top and bottom elements $1, 0$ equipped with a binary operation $\Rightarrow$ satisfying $x \wedge y \leq z$ iff $x \leq y \Rightarrow z$. We define the operations $\neg$ and $\Leftrightarrow$ by $\neg x = x \Rightarrow 0$ and $x \Leftrightarrow y = (x \Rightarrow y) \wedge (y \Rightarrow x)$. When the elements of a Heyting algebra are regarded as propositions arising from a theory, the relation $\leq$ represents entailment and the operations $\wedge$, $\vee$, $\Rightarrow$, $\neg$, $\Leftrightarrow$, 0, 1 represent conjunction, disjunction, implication, negation, bi-implication, and refutable and provable propositions, respectively. For a proposition a, the assertion that $a = 1$ expresses the condition that a is *provable*. The *consistency* of the theory is expressed by the assertion that $0 \neq 1$, that is, by the assertion that its corresponding algebra of propositions has at least two elements.

[2] For definiteness we could take our background theory to be Zermelo set theory formulated within intuitionistic first-order logic. For much of the development in the paper, one may take classical set theory as background theory and note that no use of the law of excluded middle is made.

[3] The exponential X^Y of two sets X, Y is the set of functions from Y to X.

[4] That is, a subset of Ω^Ω containing the identity map on Ω and closed under composition.

We shall call a sextuple $\mathcal{A} = (\Omega, C, K, L, k, s)$ subject to the above data a *coding assemblage*. If Ω has at least two elements, $\mathcal{A}$ is said to be *consistent*.

We now assume a fixed coding assemblage to be given.

A self-map f on Ω is *codable* if, for some map $\varphi \in K$, we have $f(\omega) = \varphi(\ulcorner\omega\urcorner)$ for all $\omega \in \Omega$. In this sense f can be represented as a (coded) function of codes. Such a map φ is called a *coding representation* for f. Reciprocally, given $\varphi \in K$, the map $\omega \mapsto \varphi(\ulcorner\omega\urcorner)\colon \Omega \to \Omega$ is called the (necessarily codable) self-map on Ω *induced by* φ.

We can now prove

Proposition 1. *The following conditions are equivalent*:

(i) *every logical map on* Ω *is codable*;
(ii) *the identity map* $1_\Omega\colon \Omega \to \Omega$ *is codable.*

Proof. (i) $\Rightarrow$ (ii) is obvious. For the converse, the assumption gives a map $\tau \in K$ for which $\omega = \tau(\ulcorner\omega\urcorner)$ for all $\omega \in \Omega$. Clearly, for any logical map f, the map $f \circ \tau$ is a coding representation for f. ⊣

A coding representation τ for 1_Ω is called a *Tarski map*: it satisfies the counterpart in our framework of what is called in the literature *Tarski's T-scheme*:

$$\omega = \tau(\ulcorner\omega\urcorner).$$

A Tarski map therefore corresponds to a *truth definition*. From the equation immediately above we see that a Tarski map, or truth definition, is just a *left inverse* for the coding map $\omega \mapsto \ulcorner\omega\urcorner\colon \Omega \to C$.

Our next task is to prove a version of what is usually known as the Diagonalization Lemma [9, Theorem 2.2.1]. This we state as the

Fixed Point Lemma. Any codable self-map on Ω has a fixed point.

Proof. Suppose $f\colon \Omega \to \Omega$ is codable with coding representation $\varphi : C \to \Omega$. Let $c = \ulcorner\varphi \circ d\urcorner$ and put $\alpha = \varphi(d(c))$. Then

$$\ulcorner\alpha\urcorner = \ulcorner\varphi(d(c))\urcorner = \ulcorner(\varphi \circ d)(\ulcorner\varphi \circ d\urcorner)\urcorner = d(\ulcorner\varphi \circ d\urcorner) = d(c).$$

Hence

$$f(\alpha) = \varphi(\ulcorner\alpha\urcorner) = \varphi(d(c)) = \alpha.$$

so that α is a fixed point for f. ⊣

It now follows from Proposition 1 and the Fixed Point Lemma that if a coding assemblage has a Tarski map, then every logical map on Ω has a fixed point. This immediately yields

Tarski's Theorem. If the set Ω of propositions in a coding assemblage $\mathcal{A}$ has a logical map with no fixed points, then $\mathcal{A}$ has no Tarski map. ⊣

Tarski's theorem in this formulation applies in particular when Ω is a Heyting algebra with at least two elements, and the negation map $\neg$ on Ω—which of course has no fixed points—is included among the logical maps. This in turn enables us to recapture the usual formulation of Tarski's theorem on the undefinability of truth. For the algebra of propositions of a consistent theory T has at least two elements, so, provided the language of T meets the modest requirements for generating a coding assemblage (along the lines of example 1 immediately below), it follows that the associated coding assemblage has no Tarski map. This means that the language for T contains no truth definition for T; in a word, truth for T is *undefinable* in T.

EXAMPLES.

1. PEANO ARITHMETIC. In this case the ingredients of the coding assemblage $\mathcal{P}$—the *Peano assemblage*—are as follows: Ω is the set of sentences of the language $\mathcal{L}$ of first-order intuitionistic arithmetic $\mathbf{P}$ identified modulo provable equivalence from $\mathbf{P}$, C is the set N of natural numbers, K is the set of maps of the form $\varphi\colon n \mapsto \psi(\mathbf{n})$ where $\psi(x)$ is a formula of $\mathcal{L}$ with at most one free variable and $\mathbf{n}$ is the term of $\mathcal{L}$ representing n. The coding map k is given by $k(\varphi) = \#\psi$ where # is any standard Gödel numbering of the formulas of $\mathcal{L}$. The map $s\colon N \times N \to N$ is a recursive substitution function on Gödel numbers.[5] Finally L consists of the maps $\sigma \mapsto \sigma$, $\sigma \mapsto \neg\sigma$, $\sigma \mapsto \neg\neg\sigma$.

Assuming that $\mathbf{P}$ is consistent, it follows, as observed above, that $\mathcal{P}$ has no Tarski map, and so truth in $\mathcal{P}$ is undefinable in $\mathcal{P}$.

For any consistent recursively axiomatized theory $\mathbf{T}$ extending $\mathbf{P}$, a corresponding assemblage $\mathcal{T}$ can be constructed along similar lines to $\mathcal{P}$, and a similar conclusion drawn concerning the undefinability of truth there.

2. INTUITIONISTIC SET THEORY. Just as in classical set theory the power set $\mathrm{P}A$ of any set A is a Boolean algebra under the usual set-theoretic operations, so in intuitionistic set theory the power set is, under the same operations, a Heyting algebra. In particular, writing 1 for the one-element set $\{0\}$, $\mathrm{P}1$ is a Heyting algebra which we shall denote by Π. If σ is a sentence of set theory, we write $\{0 \mid \sigma\}$ for the element $\{x\colon x = 0 \wedge \sigma\}$ of Π. From the axiom of extensionality it follows that $\{0 \mid \sigma\} = \{0 \mid \sigma'\}$ iff $\sigma \leftrightarrow \sigma'$. Thus the elements of Π correspond naturally to what we have termed *propositions*, in this case, to sentences identified under provable equivalence from the axioms of intuitionistic set theory. Under this correspondence each element $\omega \in \Pi$ is correlated with the proposition $0 \in \omega$, and each proposition σ with the element $\{0 \mid \sigma\}$ of Π.

Π also plays the role of a *subset classifier*. That is, for each set A, subsets of A are correlated bijectively with maps $A \to \Pi$: each subset $X \subseteq A$ is correlated with the map $x \mapsto \{0 \mid x \in X\}\colon A \to \Pi$, and each map $f\colon A \to \Pi$ with the subset $f^{-1}(1)$ of A. The top element 1 (bottom element $\varnothing$) of Π is identified with the true (false) proposition(s). In this way Π^A is seen to be naturally isomorphic to $\mathrm{P}A$.

[5] See, e.g., [3, Example 7.4.5].

Now let us attempt to build a coding assemblage *using* Π *as the set of propositions*. Here it is natural to take C, the set of codes, to be any set containing at least one element, and to take $L = \Pi^{\Pi}$ and $K = \Pi^{C}$. Using the observation immediately above, we may then identify K with PC. Take the coding map k to be an arbitrary map P$C \to C$; for $X \in$ PC, write $\ulcorner X \urcorner$ for $k(X)$.

For $\omega \in \Pi$, the constant map $\tilde{\omega}\colon C \to \Pi$ is correlated with the element $\omega^* = \{x \in C : 0 \in \omega\}$ of PC; it will be convenient to write $\ulcorner \omega \urcorner$ for $\ulcorner \omega^* \urcorner$. The map $\omega \mapsto \ulcorner \omega \urcorner\colon \Pi \to C$ is the *coding map* on Π.

The quintuple $\mathcal{Q} = (\Pi, C, \mathrm{P}C, \Pi^{\Pi}, k)$ comprises five of the six ingredients required to specify a coding assemblage. Can $\mathcal{Q}$ itself be enlarged to a coding assemblage? As we shall see, this cannot be done when the coding map on Π satisfies the modest requirement of being injective.[6]

For under the assumption that $\ulcorner \bullet \urcorner\colon \Pi \to C$ is injective, it is easy to check that it has a left inverse τ given by

$$\tau(x) = \bigcup \{X \subseteq 1 : \ulcorner X \urcorner = x\}.$$

So if $\mathcal{Q}$ could be enlarged to a coding assemblage, Tarski's theorem would be violated, since Π evidently has at least two distinct elements.

In fact, if $\ulcorner \bullet \urcorner$ is injective, an argument akin to that for Russell's paradox shows directly that $\mathcal{Q}$ cannot even be augmented by a diagonal map. For any such diagonal map d would then have to satisfy

$$d(\ulcorner X \urcorner) = \ulcorner \{0 \mid \ulcorner X \urcorner \in X\} \urcorner \tag{$*$}$$

for $X \in$ PC. Now define

$$U = \{x \in C : d(c) = \ulcorner \varnothing \urcorner\}.$$

Then, using $(*)$ and the injectivity of $\ulcorner \bullet \urcorner$,

$$\begin{aligned} \ulcorner U \urcorner \in U &\leftrightarrow d(U) = \ulcorner \varnothing \urcorner \\ &\leftrightarrow \ulcorner \{0 \mid \ulcorner U \urcorner \in U\} \urcorner = \ulcorner \varnothing \urcorner \\ &\leftrightarrow \{0 \mid \ulcorner U \urcorner \in U\} = \varnothing \\ &\leftrightarrow \ulcorner U \urcorner \notin U \end{aligned}$$

and we have a contradiction.

We now turn to Gödel's theorems.

We have seen that every codable self-map g on Ω has a fixed point. Let us call an element $\alpha \in \Omega$ for which $g^{-1}(\alpha) = \{\alpha\}$ a *strong* fixed point for g.

[6] The modesty of this requirement is more easily seen when the background theory is classical (i.e., the law of excluded middle holds). For then $\Omega = \{\varnothing, 1\}$ and injectivity of $\ulcorner \bullet \urcorner$ boils down simply to $\ulcorner \varnothing \urcorner \neq \ulcorner 1 \urcorner$: that is, the true and the false receive different codes.

Observing that the power set of Ω is a Heyting algebra, it is clear that α is a strong fixed point for g iff

$$\Omega \subseteq (g^{-1}(\alpha) \Leftrightarrow \{\alpha\}).^{7} \qquad (*)$$

A coding representation π for g is then easily seen to satisfy

$$\alpha = \omega \leftrightarrow \pi(\ulcorner \alpha \urcorner) = \omega. \qquad (**)$$

We next prove another version of the Diagonalization Lemma, namely, the

DISTINGUISHED ELEMENT LEMMA. Suppose that Ω has a codable self-map g with a strong fixed point α. Then, for any logical map f on Ω there is $\beta \in \Omega$ such that

$$f^{-1}(\alpha) \cap \{\beta\} = \{\alpha\} \cap \{\beta\}$$

or equivalently,

$$f(\beta) = \alpha \leftrightarrow \beta = \alpha,$$

or equivalently (cf. $(*)$),

$$\{\beta\} \subseteq (f^{-1}(\alpha) \Leftrightarrow \{\alpha\}).$$

PROOF. Let π be a coding representation for g. Given $f : \Omega \to \Omega$, define $c = \ulcorner f \circ \pi \circ d \urcorner$, $\beta = \pi(d(c))$. Then

$$\begin{aligned} \ulcorner f(\beta) \urcorner &= \ulcorner f(\pi(d(c))) \urcorner = \ulcorner (f \circ \pi \circ d)(\ulcorner f \circ \pi \circ d \urcorner) \urcorner \\ &= d(\ulcorner f \circ \pi \circ d \urcorner) = d(c). \end{aligned}$$

Hence

$$\begin{aligned} f(\beta) = \alpha &\leftrightarrow g(f(\beta)) = \alpha \leftrightarrow \pi(\ulcorner f(\beta) \urcorner) = \alpha \\ &\leftrightarrow \pi(d(c)) = \alpha \leftrightarrow \beta = \alpha. \end{aligned} \qquad \dashv$$

Now we can formulate Gödel's First Incompleteness Theorem in the present setting. Here we require Ω to have a distinguished element $\top$: we think of $\top$ as the representing the *provable* propositions: thus, for $\alpha \in \Omega$, to assert that $\alpha = \top$ is to assert that α is (*actually*) *provable*. We suppose given a map $\pi \in K$ which we shall term a *provability* map, in the sense that, for each $\omega \in \Omega$, the element $\pi(\ulcorner \omega \urcorner)$ of Ω shall be construed as the proposition *expressing* the assertion "ω is *provable*". Now let g be the (necessarily codable) self-map on Ω induced by π. We shall call g a *Gödel map* if it has $\top$ as a strong fixed point. If g is a Gödel map, then the provability map π satisfies

$$\omega = \top \leftrightarrow \pi(\ulcorner \omega \urcorner) = \top.$$

This may be construed as asserting that a proposition ω is provable iff the proposition $\pi(\ulcorner \omega \urcorner)$ expressing the provability of ω is itself provable. Notice that the self-map on Ω induced by a Tarski map is a Gödel map.

[7]Recall that $\Leftrightarrow$ is the bi-implication operator in Heyting algebras.

Now call a coding assemblage *Gödelian* if it has a Gödel map and there is an element $\perp \in \Omega$ different from $\top$ together with a logical map v on Ω which interchanges the two.

We think of v as the *negation* operation on propositions and $\perp$ as representing the *refutable* propositions.

We can now prove

GÖDEL'S FIRST INCOMPLETENESS THEOREM. The set of propositions of any Gödelian coding assemblage has at least 3 elements.

PROOF. Given a Gödelian coding assemblage there is, by the Distinguished Element Theorem, an element $\beta \in \Omega$ for which $v(\beta) = \top \leftrightarrow \beta = \top$. In that case $\beta \neq \perp$ and $\beta \neq \top$, so that Ω has the three distinct elements $\perp, \top, \beta$. $\dashv$

An element $\beta \in \Omega$ differing from $\perp$ and $\top$ evidently represents an *undecidable* proposition, so the theorem just proved may be taken to assert that *any Gödelian coding assemblage contains undecidable propositions*.

All this applies in particular to the Peano coding assemblage $\mathcal{P}$. Let *Prov* be a provability predicate for **P**. Then, by standard arguments,[8] we have, for any arithmetical sentence ψ,

$$\vdash_{\mathbf{P}} \psi \leftrightarrow \vdash_{\mathbf{P}} Prov(\#\psi). \qquad (\P)$$

Now let $g : \Omega \to \Omega$ be the map given by $g(\sigma) = Prov(\#\sigma)$. It then follows from (¶) that $g(\sigma) = \top \leftrightarrow \sigma = \top$; in other words, g is a Gödel map for $\mathcal{P}$. (Notice that *Prov* is essentially the coding representation of g.) Assuming that **P** is consistent, $\mathcal{P}$ is then Gödelian, and accordingly contains undecidable propositions.

Finally let us set about formulating Gödel's Second Incompleteness Theorem in the present setting. To do this we need to introduce the concept of a Hilbert-Bernays-Löb, or HBL- map. Let us assume that $\mathcal{A}$ is a coding assemblage in which Ω is a Heyting algebra. An *HBL-map* in $\mathcal{A}$ is a map $\chi \in K$ satisfying the following conditions:[9]

(i) $\chi(\ulcorner 1 \urcorner) = 1$,
(ii) $\chi(\ulcorner \omega \Rightarrow \omega' \urcorner) \leq (\chi(\ulcorner \omega \urcorner) \Rightarrow \chi(\ulcorner \omega' \urcorner))$,
(iii) $\chi(\ulcorner \omega \urcorner) \leq \chi(\ulcorner \chi(\ulcorner \omega \urcorner) \urcorner)$.

If χ is an HBL-map, the self-map $\Box$ on Ω induced by χ satisfies the conditions:

(a) $\Box 1 = 1$,
(b) $\Box(\omega \Rightarrow \omega') \leq (\Box\omega \Rightarrow \Box\omega')$,
(c) $\Box\omega \leq \Box\Box\omega$.

---

[8] See, e.g., [6, Ch. 16].

[9] These conditions correspond to the Derivability Conditions D1–D3 presented on p. 827 of [6].

The map $\Box$ is called the *HBL-operator* on Ω *associated* with χ; it may be considered a modal operator satisfying the K4 axioms.[10] It follows immediately from (a) and (b) that $\Box$ is order-preserving.

If we think of χ as a provability map, then $\Box$ may be thought of as a provability operator on propositions: for each proposition ω, $\Box\omega$ is the proposition expressing "ω is provable" and $\neg\Box\omega$ is the proposition expressing "ω is unprovable". If for brevity we agree for the moment to suppress the term "expressing", then (a) above asserts: *if ω is a provable proposition, then so is the proposition* "ω is *provable*"; (b) asserts: *the proposition* "ω *implies* ω' *is provable*" *implies the proposition* ' "ω *is provable*" *implies* "ω' *is provable*" '; and (c) asserts: *the proposition* "ω *is provable*" *implies the proposition* " 'ω *is provable*' *is provable*". Note also that, since 0 represents the refutable propositions, $\neg\Box 0$ may be taken as the proposition expressing the *internal consistency*—and $\Box 0$ the *internal inconsistency*—of the coding assemblage with respect to the given HBL-map χ.

Now let us call a coding assemblage $\mathcal{A}$ *suitable*[11] if (i) Ω is a Heyting algebra, (ii) $\mathcal{A}$ has an HBL-map, and (iii) for each $\alpha \in \Omega$ and $\varphi \in K$, the function $\omega \mapsto (\varphi(\omega) \Rightarrow \alpha)\colon C \to \Omega$ is in K.

We can now prove a version of

LÖB'S THEOREM[12]. Let $\mathcal{A}$ be an suitable coding assemblage with HBL-map χ and associated HBL-operator $\Box$. Then, for any $\alpha \in \Omega$

(i) $\Box(\Box\alpha \Rightarrow \alpha) \leq \Box\alpha$.

(ii) $\Box\alpha \leq \alpha \to \alpha = 1$.

PROOF. Given $\alpha \in \Omega$, the map $\omega \mapsto (\chi(\ulcorner\omega\urcorner) \Rightarrow \alpha)\colon \Omega \to \Omega$ is clearly codable and so by the Fixed Point Lemma has a fixed point β. That is,

$$\beta = (\Box\beta \Rightarrow \alpha). \tag{$*$}$$

A fortiori $\beta \leq (\Box\beta \Rightarrow \alpha)$, whence

$$\Box\beta \leq \Box(\Box\beta \Rightarrow \alpha) \leq (\Box\Box\beta \Rightarrow \Box\alpha).$$

Hence

$$\Box\beta = \Box\beta \wedge \Box\Box\beta \leq \Box\alpha. \tag{$**$}$$

It follows that $(\Box\alpha \Rightarrow \alpha) \leq (\Box\beta \Rightarrow \alpha)$, whence

$$\begin{aligned} \Box(\Box\alpha \Rightarrow \alpha) &= \Box(\Box\beta \Rightarrow \alpha) \\ &= \Box\beta \quad (\text{by } (*)) \\ &\leq \Box\alpha \quad (\text{by } (**)). \end{aligned}$$

This gives (i).

[10] See [4, p. 5].

[11] I.e., suitable for proving Gödel's second incompleteness theorem: see below.

[12] Theorem 4.1.1 of [6].

For (ii), we assume $\Box\alpha \leq \alpha$, so that $(\Box\alpha \Rightarrow \alpha) = 1$. It now follows from (i) that

$$1 = \Box 1 = \Box(\Box\alpha \Rightarrow \alpha) \leq \Box\alpha.$$

Therefore $\Box\alpha = 1$, and since $\Box\alpha \leq \alpha$, we conclude that $\alpha = 1$. ⊣

REMARKS. (1) From (i) of Löb's Theorem we see that $\Box$ satisfies the so-called *GL* (Gödel-Löb) *axiom*[13] for a normal modal logic, i.e., the scheme

$$\Box(\Box A \to A) \to \Box A.$$

(2) If $\Box$ is the identity map, then we deduce from (ii) of Löb's Theorem that Ω has just one element. This fact is also a consequence of Tarski's Theorem: if $\Box$ is the identity, then χ is a Tarski map, which can only exist when Ω has only one element.

(3) Taking $\alpha = 0$ in (i), we get $\Box\neg\Box 0 = \Box(\Box 0 \Rightarrow 0) \leq \Box 0$. Since $\Box$ is order-preserving, we also have $\Box 0 \leq \Box\neg\Box 0$. Hence (a) $\Box 0 = \Box\neg\Box 0$; whence (b) $\neg\Box 0 = \neg\Box\neg\Box 0$. (a) may be taken to assert that *the proposition expressing internal inconsistency is* (*equivalent to*) *the proposition expressing the provability of the proposition expressing internal consistency*; and (b) *that the proposition expressing internal consistency is* (*equivalent to*) *the proposition expressing the unprovability of the proposition expressing internal consistency*.

(4) Let us call the fixed points of the map $\omega \mapsto \neg\Box\omega$ *Gödel propositions*. Gödel propositions may be considered as expressing their own unprovability. Statement (b) above tells us that $\neg\Box 0$, the internal consistency proposition, is also a Gödel proposition. In fact it follows from Löb's theorem that $\neg\Box 0$ *is the only Gödel proposition*. For let α be a Gödel proposition, i.e., $\alpha = \neg\Box\alpha$. Now $\Box 0 \leq \Box\alpha$, whence $\alpha = \neg\Box\alpha \leq \neg\Box 0$. It follows that $\Box\alpha \leq \Box\neg\Box 0 = \Box(\Box 0 \Rightarrow 0) \leq \Box 0$ by Löb's theorem. Since also $\Box 0 \leq \Box\alpha$, it follows that $\Box 0 = \Box\alpha$, whence $\alpha = \neg\Box\alpha = \neg\Box 0$.

From Löb's Theorem one finally derives:

GÖDEL'S SECOND INCOMPLETENESS THEOREM. Given a suitable consistent coding assemblage with an HBL map χ. Then, for any $\alpha \in \Omega$, $\chi(\ulcorner\alpha\urcorner) \neq 0$; or equivalently $\neg\chi(\ulcorner\alpha\urcorner) \neq 1$. In particular, $\chi(\ulcorner o\urcorner) \neq 0$; or equivalently $\neg\chi(\ulcorner 0\urcorner) \neq 1$.

PROOF. If $\chi(\ulcorner\alpha\urcorner) = 0$, then $\Box 0 \leq \Box\alpha = 0$; hence by Löb's theorem $0 = 1$, and it follows that Ω has just one element. ⊣

Thinking of χ as a provability map, $\neg\chi(\ulcorner\alpha\urcorner)$ is, as we know, the proposition (expressing) "α *is unprovable*", so that asserts that "α *is unprovable*" is provable. In that case the second incompleteness theorem may be taken to assert that in any suitable consistent coding assemblage, *there is no proposition whose unprovability is provable*. In particular, the unprovability of the proposition 0 is also unprovable, in other words, the *proposition expressing internal consistency is unprovable*. This applies in particular to the Peano

[13] See [5, p. 5].

assemblage, whose suitability follows from the fact that the provability predicate *Prov* in **P** satisfies the conditions on an HBL-map.[14] The same applies to the coding assemblage associated with any consistent recursively axiomatizable extension of **P**.

On the other hand, recall that $\Box 0$, i.e., $\chi(\ulcorner 0 \urcorner)$, expresses what we have termed *internal inconsistency*. In that case $\chi(\ulcorner 0 \urcorner) = 1$ asserts that internal inconsistency is provable, or, in a word, that the coding assemblage is *internally inconsistent*. If we take the HBL map χ to be $\tilde{1}$, then we are, in effect, asserting the provability of every proposition of the form "α is provable". In particular "0 is provable" is provable, in other words, we have internal inconsistency. But this is perfectly compatible with consistency, that is, the condition that Ω has two elements. This just shows that internal consistency need have little to do with consistency, or, more generally, that provability maps need have little to do with actual provability.[15]

In conclusion, we point out that while in stating and proving these results we have used ordinary set-theoretic language, they can be formulated in toposes (see, e.g., [1]) or more general categories (cf. the discussion in [10]).

REFERENCES

[1] JOHN L. BELL, ***Toposes and local set theories***, Clarendon Press, Oxford, 1988.

[2] ———, ***Set theory: Boolean-valued models and independence proofs***, Clarendon Press, Oxford, 2005.

[3] JOHN L. BELL and M. MACHOVER, ***A course in mathematical logic***, North-Holland, 1977.

[4] G. BOOLOS, *Gödel's second incompleteness theorem explained in words of one syllable*, ***Mind***, vol. 103 (1994), no. 409, pp. 1–3.

[5] ———, ***The logic of provability***, Cambridge University Press, 1995.

[6] G. BOOLOS and R. JEFFREY, ***Computability and logic***, Cambridge University Press, 1974.

[7] F. W. LAWVERE, *Diagonal arguments and cartesian closed categories. Category theory, homology theory and their applications, II*, ***Battelle Institute Conference, Seattle, Washington, 1968***, Springer, Berlin, 1969, pp. 134–145.

[8] M. H. LÖB, *Solution of a problem of Leon Henkin*, ***The Journal of Symbolic Logic***, vol. 20 (1955), pp. 115–118.

[9] C. SMORYNSKI, *The incompleteness theorems*, ***Handbook of mathematical logic*** (J. Barwise, editor), North-Holland, 1977, pp. 821–866.

[10] N. YANOFSKY, *A universal approach to self-referential paradoxes, incompleteness and fixed points*, this BULLETIN, vol. 9 (2003), no. 3, pp. 362–386.

DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF WESTERN ONTARIO, CANADA
E-mail: jbell@uwo.ca

[14] See, e.g., [6, Ch. 16].

[15] An observation also made in [6].