Notre Dame Journal of Formal Logic
Volume XIX, Number 3, July 1978
NDJFAM

# TRUTH IN CONSTRUCTIVE METAMATHEMATICS

JOHN STAPLES

1 *Introduction* Metamathematics may be divided into two parts: that which relies on the use of logic, and that which does not. Most metamathematics which uses logic uses classical (two-valued) logic: we shall call it *classical* metamathematics. It is of course possible to consider metamathematics which uses logic, but which intends for the logic a constructive interpretation; we shall call that *constructive* metamathematics. There is a need for greater interest in constructive metamathematics, for example, to deal with theories which are in some way not standard, but which are claimed to admit a constructive interpretation. This paper is (apart from some asides) intended as a contribution to constructive metamathematics.

2 *Tarski's notion of truth*

2.1 In classical metamathematics the word "truth" has been given a technical meaning by Tarski, apparently without causing too much confusion with whatever is one's primary, intuitive notion of the meaning of the word. Tarski's notion of truth is not however confined to classical metamathematics. Its role is to describe the meaning of the logical concepts in a theory, in terms of the logical notions underlying the metatheory; it can do that in constructive metamathematics just as well as in classical metamathematics. When one considers the fundamental role of Tarski's notion in classical metamathematics, it is of some interest to know whether it can have any comparable role in constructive metamathematics.

Consider Tarski's definition of truth. It defines the truth of compound formulas and open formulas of a first order theory in terms of the truth of the closed atomic formulas, as follows. The definition is recursive, and it is assumed that the variables of the theory are intended to range over a fixed set $S$ of objects; also that each relation symbol is intended to refer to a specific relation on that set of objects. We can also assume that the theory has a constant corresponding to each object in $S$ (if it does not, extend the theory by adding such constants, define truth for the extended theory and take truth for the original theory to be the restriction of that

*Received April 7, 1976*

notion to the original theory). For the sake of brevity I shall ignore the distinction between objects of $S$ and their names.

(i) an open formula $A(x_1, \ldots, x_n)$ with just the free variables displayed, is true just when for all objects $c_1, \ldots, c_n$ of $S$, $A(c_1, \ldots, c_n)$ is true,
(ii) a closed formula $A \;\&\; B$ is true just when $A$ is true and $B$ is true,
(iii) a closed formula $A \vee B$ is true just when either $A$ is true or $B$ is true,
(iv) a closed formula $\neg A$ is true just when $A$ is not true,
(v) a closed formula $A \supset B$ is true just when $B$ is true if $A$ is true,
(vi) a closed formula $(\exists x)A(x)$ is true just when for some object $c$ of $S$, $A(c)$ is true,
(vii) a closed formula $(\forall x)A(x)$ is true just when $A(c)$ is true for all objects $c$ in $S$.

It is this notion of truth which we shall discuss below. We are not intending to confuse it with one's primary, intuitive notion of truth, but we are interested in how well the two correspond.

**2.2** Consider first Kleene's notions of realizability ([2], p. 95). Truth as defined above fails to correspond to "is realizable", for theories where both are defined, because (v) and (vii) are not like the corresponding clauses in the definitions of realizability. For a closed formula $A \supset B$ to be realizable, one requires a uniform (recursive) procedure for obtaining objects which realize $B$ from objects which realize $A$. Similarly, for $(\forall x)A(x)$ to be realizable, one requires a uniform (recursive) procedure for obtaining objects which realize $A(c)$ from any object which codes $c$.

Gödel's method for interpreting constructive theories also does not relate to the above definition of truth, though in this case the differences are more fundamental. On the other hand, when one looks at the description of truth given by Bishop ([1], pp. 7-8) one finds that it corresponds quite closely with the above definition of truth. Taking Bishop's "prove" to mean "show to be true (in the intuitive sense)", we consider his comments on the interpretation of implication and of universal quantification. They are, in paraphrase:

To show $P \supset Q$ is true we must show that $Q$ is true whenever $P$ is true.

To show $(\forall x)A(x)$ true we must show $A(c)$ is true, for every $c$ which can be substituted for $x$.

Moreover, Bishop says explicitly ([1], p. 8) that universal statements have the same meaning in constructive mathematics as in classical mathematics, reinforcing the idea that explicit uniformities such as are required in the realizability notion do not form part of Bishop's notion of truth.

Since the treatment of the other logical notions also corresponds well with the constructive interpretation of Tarski's notion of truth, we have the perhaps unexpected situation that Bishop's notion of truth can be described in much the same way as Tarski's definition. The explanation of course lies in the constructive interpretation which we are giving to the logical notions which are used in Tarski's definition. As Tarski's definition seems a reasonable model of Bishop's intuitive notion, we shall study it further.

**3** *Properties of Tarski's notion of truth* In this section "true" is to be interpreted according to Tarski's definition, except where stated otherwise.

**3.1** *The axioms of intuitionist predicate calculus are true* To see that, one simply checks in turn each of the axiom schemes. The result is essentially well-known, so we do not argue it in detail.

**3.2** We do not of course expect to extend **3.1** to classical predicate calculus, but we do now enquire whether there are logical statements additional to those of the intuitionist predicate calculus which are true.

**3.3** To begin with we define (after Gentzen, and Gödel) for each formula $A$ of the theory under consideration, a formula $G(A)$ as follows:

(i) For atomic formulas, $G(A)$ is $\neg\neg A$,
(ii) $G(A \,\&\, B)$ is $G(A) \,\&\, G(B)$
(iii) $G(A \vee B)$ is $\neg(\neg G(A) \,\&\, \neg G(B))$
(iv) $G(\neg A)$ is $\neg G(A)$
(v) $G(A \supset B)$ is $G(A) \supset G(B)$
(vi) $G((\forall x)A(x))$ is $(\forall x)G(A(x))$
(vii) $G((\exists x)A(x))$ is $\neg(\forall x)\neg G(A(x))$.

The following observation is essentially also due to Gentzen, and Gödel.

**3.4** *For every formula $A$, $\neg\neg G(A) \supset G(A)$ is provable in the intuitionist predicate calculus*

The proof is straightforward, by induction following the definition of $G(A)$, and is therefore omitted. From **3.1** we therefore have:

**3.5** *For every formula $A$, $\neg\neg G(A) \supset G(A)$ is true*

We can conclude that

**3.6** *Classical derivability can be given a constructive interpretation*

By that one simply means that if a formula $B$ can be derived using classical logic from formulas $A_1, \ldots, A_n$, then the following assertion is true: $G(B)$ can be derived from $G(A_1), \ldots, G(A_n)$ using intuitionist logic. For, observe that

$$\frac{A,\ A \supset B}{B} \text{ translates to } \frac{G(A),\ G(A) \supset G(B)}{G(B)},$$

and generalization,

$$\frac{A(x)}{(\forall x)A(x)} \text{ translates to } \frac{G(A(x))}{(\forall x)G(A(x))};$$

both the rules of inference obtained are obviously valid. More particularly,

**3.7** *In a theory such that $G(A)$ is true for every axiom $A$, $G(B)$ is true for every $B$ which is derivable from the axioms by classical logic*

An instance of a theory as supposed in **3.7**, is intuitionist first order arithmetic, as was shown by both Gentzen and Gödel. Other instances are known.

## 4 *A possible further property of Tarski's notion of truth*

**4.1** Whether the axiom scheme which we consider in this section is true according to Tarski's definition, depends on the intuitive notion of truth underlying the metamathematics. For example it is trivially true if we interpret the metamathematics classically. Our interest, however, is in whether it has any plausibility in constructive metamathematics. We continue to use "true" to refer to Tarski's definition, and we abbreviate "not true" to "false". Consider an arbitrary formula $A(x)$ with one free variable, and the formulas $A(c)$ which are obtained from it by substituting an arbitrary constant $c$ for $x$. The statement we wish to examine the plausibility of, is the assertion that the first of the following two possibilities excludes the second:

(i) not all $A(c)$ are true,
(ii) no $A(c)$ is false.

Now if we had required, as for example would be the case in a realizability interpretation, that "all $A(c)$ are true" is a justified assertion only when one has some explicit uniform means for justifying each $A(c)$, then the assertion (i) would not exclude (ii); the lack of such a uniform means would justify (i), even if no individual instance of $A(c)$ were false. In fact however, we require for "all $A(c)$ are true" to be justified, only that each individual instance $A(c)$ be true. If it is not the case that each individual instance $A(c)$ is true, it may well be considered plausible that (ii) is excluded.

We shall now examine the consequences of assuming that (i) excludes (ii). It should be made clear however that a proof that some logical principle (such as this one) is true according to Tarski's definition, will involve the assumption that the same principle is true in the intuitive sense; and if some discussion such as the above convinces one that (i) excludes (ii) is true in the intuitive sense, then there is no difficulty in completing the proof that (i) excludes (ii) is true in the sense of Tarski.

**4.2** First we examine the symbolic expression of the assertion that (i) excludes (ii). A direct translation would be $\neg(\forall x)A(x) \supset \neg\neg(\exists x)\neg A(x)$, which is equivalent in the intuitionist predicate calculus to $\neg(\forall x)A(x) \supset \neg(\forall x)\neg\neg A(x)$, and hence to

**4.3** $(\forall x)\neg\neg A(x) \supset \neg\neg(\forall x)A(x)$,

which is well known as an assertion which is not provable in the intuitionist predicate calculus.

Though our discussion so far has involved formulas $A(x)$ with only the free variable displayed, it is clear from our definition of truth that the truth of **4.3** for such formulas implies its truth for all formulas. Thus in examining the consequences of the assumption that (i) implies (ii), we shall allow the use of **4.3** for formulas $A(x)$ with free variables other than the one displayed. The result which we particularly wish to note is:

**4.4** *For all formulas $A$, the formulas $A \supset G(A)$ are derivable from the axioms of the intuitionist predicate calculus, together with the scheme* **4.3**

To see that, one proves all instances of the equivalent scheme

$$\neg A \equiv \neg G(A),$$

by induction following the definition of $G(A)$. The proof is straightforward, and the only place where **4.3** is used is at the stage where one wishes to prove

$$\neg(\forall x)A(x) \supset \neg G((\forall x)A(x)),$$

which is just

$$\neg(\forall x)A(x) \supset \neg(\forall x)G(A(x)).$$

The proof of that is as follows. Suppose $\neg(\forall x)A(x)$; if $(\forall x)G(A(x))$ then $(\forall x)\neg\neg A(x)$, since $\neg A(y)$ implies, by inductive hypothesis, $G(\neg A(y))$, i.e., $\neg G(A(y))$, contradiction. Thus by **4.3**, $\neg\neg(\forall x)A(x)$, contrary to our initial hypothesis, so we conclude as required that $\neg(\forall x)G(A(x))$. Since the other steps are straightforward and rely only on the axioms of the intuitionist predicate calculus, we omit them.

**5** *Properties of the scheme $A \supset G(A)$* The author has called this scheme the subclassical axiom scheme, because any theory which includes it is consistent if and only if the corresponding classical theory (i.e., the theory obtained by adjoining the law of excluded third) is consistent. To see that, we first note that:

**5.1** *The corresponding classical theory can be interpreted within the original theory*

To see that, we simply take the interpretation of a classical formula $A$ to be $G(A)$. The subclassical axiom scheme tells us that for each axiom $A$ of the original theory, $G(A)$ is a theorem of the original theory, so the stated result follows from **3.7**. Now, to reach the desired conclusion, we show that $\neg A$ is provable if and only if it is provable in the corresponding classical theory. One part is trivial, since the corresponding classical theory is an extension of the original theory; suppose then that $\neg A$ is classically provable. From **5.1**, $G(\neg A)$, i.e., $\neg G(A)$, is provable, hence from the axiom scheme, $\neg A$ is provable.

**6** *Other interpretations for the subclassical axiom scheme* The discussion of **4.1** should not be taken as implying that theories including the subclassical axiom scheme cannot be subjected to constructive interpretations other than truth according to Tarski. Provided that the corresponding classical theory is consistent, it is reasonable to suppose that there will be other interpretations; for some examples of realizability-type interpretations see [2], p. 253, [3].

## REFERENCES

[1] Bishop, E., *Foundations of constructive analysis.* McGraw-Hill (1967).

[2] Staples, Jr., "Combinator realizability of constructive finite type analysis," in Mathias, A. R. D., and H. Rogers (eds.), *Cambridge Summer School in Mathematical Logic*, Springer-Verlag, Lecture Notes in Mathematics, vol. 337 (1973), pp. 253-273.

[3] Staples, J., "Combinator realizability of a constructive Morse set theory," *The Journal of Symbolic Logic,* vol. 39 (1974), pp. 226-234.

*Queensland Institute of Technology*
*Brisbane, Queensland, Australia*