

## Arithmetic With Satisfaction

JAMES CAIN

**Abstract** A language in which we can express arithmetic and which contains its own satisfaction predicate (in the style of Kripke's theory of truth) can be formulated using just two nonlogical primitives: ' (the successor function) and *Sat* (a satisfaction predicate).

Let  $\mathcal{L}$  be a language with vocabulary:

$$, ( ) \exists \neg \vee = ' Sat$$

plus the variables  $x_0, x_1, x_2, \dots$ . A term is a variable followed by zero or more occurrences of '. An atomic formula is any formula of the form  $t_0 = t_1$ ,  $Sat(t_0)$ , or  $Sat(t_0, \dots, t_i)$  (for any finite string of terms  $t_0, \dots, t_i$ ). Nonatomic formulas are defined in the normal way. (Note that, though for simplicity we let  $Sat$  take any number of terms, this is not necessary for our purposes. We could consider just a 5-place predicate,  $Sat(x_0, \dots, x_4)$ . More will be said about this later.)

We will be concerned with partial interpretations of  $\mathcal{L}$  in which the variables range over the natural numbers, ' is interpreted as the successor function, and a disjoint pair of sets  $(S_1, S_2)$  of finite sequences of natural numbers is assigned to  $Sat$ . Let  $\mathcal{L}(S_1, S_2)$  represent such an interpretation of  $\mathcal{L}$ . Let  $s$  be an infinite sequence of natural numbers, and let  $s^*$  be the corresponding assignment of natural numbers to terms (thus  $s^*(x_i) = s(i)$  and  $s^*(t')$  = the successor of  $s^*(t)$ ). Then we say:

$$\mathcal{L}(S_1, S_2) \models Sat(t_0, \dots, t_i)[s]$$

(i.e.,  $\mathcal{L}(S_1, S_2)$  satisfies  $Sat(t_0, \dots, t_i)$  with  $s$ ) iff  $\langle s^*(t_0), \dots, s^*(t_i) \rangle \in S_1$ . On the other hand, we say:

$$\mathcal{L}(S_1, S_2) \not\models Sat(t_0, \dots, t_i)[s]$$

(i.e.,  $\mathcal{L}(S_1, S_2)$  falsifies  $Sat(t_0, \dots, t_i)$  with  $s$ ) iff  $\langle s^*(t_0), \dots, s^*(t_i) \rangle \in S_2$ . And finally,  $\mathcal{L}(S_1, S_2)$  leaves  $Sat(t_0, \dots, t_i)$  undefined with respect to  $s$  if  $\mathcal{L}(S_1, S_2)$  neither

*Received October 20, 1993; revised November 14, 1994*

satisfies nor falsifies  $Sat(t_0, \dots, t_i)$  with  $s$ . We evaluate nonatomic formulas using the Strong Kleene scheme. Thus:

$$\mathcal{L}(S_1, S_2) \models \neg A[s] (\mathcal{L}(S_1, S_2) \models \neg A[s])$$

iff

$$\mathcal{L}(S_1, S_2) \models A[s] (\mathcal{L}(S_1, S_2) \models A[s]).$$

Similarly:

$$\mathcal{L}(S_1, S_2) \models (A \vee B)[s] (\mathcal{L}(S_1, S_2) \models (A \vee B)[s])$$

iff

$$\mathcal{L}(S_1, S_2) \models A[s] \text{ or } \mathcal{L}(S_1, S_2) \models B[s] (\mathcal{L}(S_1, S_2) \models A[s] \text{ and } \mathcal{L}(S_1, S_2) \models B[s]).$$

Finally:

$$\mathcal{L}(S_1, S_2) \models \exists x_i A[s] (\mathcal{L}(S_1, S_2) \models \exists x_i A[s])$$

iff

$$\mathcal{L}(S_1, S_2) \models A[r] (\mathcal{L}(S_1, S_2) \models A[r]) \text{ for some (every) sequence } r \\ \text{such that, for } j \neq i, s(j) = r(j).$$

We let  $(A \wedge B)$  abbreviate  $\neg(\neg A \vee \neg B)$ .

Say that sequence  $s$  *extends*  $\langle n_0, \dots, n_i \rangle$  provided, for  $j \leq i$ ,  $s(j) = n_j$ . We say that  $\mathcal{L}(S_1, S_2) \models A[\langle n_0, \dots, n_i \rangle]$  iff, for every  $s$  extending  $\langle n_0, \dots, n_i \rangle$ ,  $\mathcal{L}(S_1, S_2) \models A[s]$ .  $\mathcal{L}(S_1, S_2) \models A[\langle n_0, \dots, n_i \rangle]$  iff, for every  $s$  extending  $\langle n_0, \dots, n_i \rangle$ ,  $\mathcal{L}(S_1, S_2) \models A[s]$ .

We will be interested in those interpretations of  $\mathcal{L}$  in which  $Sat$  can be understood as expressing a satisfaction predicate for the language. Assume that we have a Gödel numbering of the formulas of  $\mathcal{L}$  by the natural numbers (we place no further restrictions on the Gödel numbering—it can even be nonrecursive). We say that  $\Phi(S_1, S_2) = (S_3, S_4)$ , where  $S_3 = \{\langle n_0, \dots, n_i \rangle \mid n_0 \text{ is the Gödel number of a formula } A \text{ such that } A \text{ contains at most } x_0, \dots, x_{i-1} \text{ as free variables and } \mathcal{L}(S_1, S_2) \models A[\langle n_1, \dots, n_i \rangle]\}$ , and  $S_4 = \{\langle n_0, \dots, n_i \rangle \mid \text{either } n_0 \text{ is not the Gödel number of a formula which contains at most } x_0, \dots, x_{i-1} \text{ free, or } n_0 \text{ is the Gödel of such a formula, } A, \text{ and } \mathcal{L}(S_1, S_2) \models A[\langle n_1, \dots, n_i \rangle]\}$ . It should be clear that if  $S_1$  and  $S_2$  are disjoint then  $\Phi(S_1, S_2)$  will be a disjoint pair. We say that  $Sat$  *expresses a satisfaction predicate for*  $\mathcal{L}(S_1, S_2)$  iff  $\Phi(S_1, S_2) = (S_1, S_2)$ , in which case we say that  $(S_1, S_2)$  is a *fixed point* of  $\Phi$  and  $\mathcal{L}(S_1, S_2)$  is a *fixed point language*.

Say that  $(S_1, S_2) \leq (S_3, S_4)$  iff  $S_1 \subseteq S_3$  and  $S_2 \subseteq S_4$ . Clearly  $\Phi$  is monotonic (in the sense that if  $(S_1, S_2) \leq (S_3, S_4)$  then  $\Phi(S_1, S_2) \leq \Phi(S_3, S_4)$ ), and  $(\Lambda, \Lambda) \leq \Phi(\Lambda, \Lambda)$  (where  $\Lambda$  is the empty set). It follows that  $\Phi$  has fixed points, including a smallest fixed point.<sup>1</sup>

We need to define the notion of definability in a partially interpreted language. We say that an  $i$ -place relation,  $R$ , is *weakly defined* in  $\mathcal{L}(S_1, S_2)$  by a formula  $A$  provided that  $A$  contains at most  $x_0, \dots, x_{i-1}$  free and  $\mathcal{L}(S_1, S_2) \models A[s]$  for exactly those  $s$  which extend elements of  $R$ .  $R$  is *strongly defined* by  $A$  iff it is weakly defined by  $A$  and  $N^i - R$  is weakly defined by  $\neg A$ .  $R$  is weakly (strongly) definable iff it is weakly

(strongly) definable by some formula. A function is said to be strongly definable iff its graph is. (To handle definability of a set,  $S$ , of numbers, we treat  $S$  as a set of 1-tuples and let  $\langle n \rangle = n$ .)

**Theorem 1** *Every relation definable in the first order language of arithmetic (with vocabulary:  $+ \times ' 0 =$ ) is strongly definable in any fixed point language  $\mathcal{L}(S_1, S_2)$ .*

*Proof:* Suppose that  $\mathcal{L}(S_1, S_2)$  is a fixed point language. Since  $\mathcal{L}(S_1, S_2)$  contains  $=$  and  $'$ , it will suffice to show that the relations  $x = 0$ ,  $x + y = z$ , and  $x \times y = z$  are strongly definable.  $x = 0$  is of course definable by  $\neg\exists y(x = y')$ . We show that addition is definable as follows.

Consider the formula:

$$(x_1 = 0 \wedge x_2 = x_0) \vee \exists x_4 \exists x_5 (x_1 = x'_4 \wedge \text{Sat}(x_3, x_0, x_4, x_5, x_3) \wedge x_2 = x'_5).$$

Suppose the Gödel number of this formula is  $m$ . Let  $\text{Sum}(x_0, x_1, x_2)$  be the formula  $\text{Sat}(m, x_0, x_1, x_2, m)$  (which in turn abbreviates the formula  $\exists x_6 (\neg\exists x_7 x_6 = x'_7 \wedge \text{Sat}(x_6^{(m)}, x_0, x_1, x_2, x_6^{(m)}))$ ).

$\text{Sum}(x_0, x_1, x_2)$  strongly defines the addition function. We prove this by induction. Suppose that we are given  $n_0$ . We first need to show that for each  $n_1$  and  $n_2$ ,  $n_0 + n_1 = n_2$  iff  $\mathcal{L}(S_1, S_2) \models \text{Sum}(x_0, x_1, x_2)[\langle n_0, n_1, n_2 \rangle]$ . Suppose  $n_1 = 0$ .  $\mathcal{L}(S_1, S_2) \models \text{Sum}(x_0, x_1, x_2)[\langle n_0, 0, n_2 \rangle]$  iff  $\langle m, n_0, 0, n_2, m \rangle \in S_1$ , which holds, since  $\mathcal{L}(S_1, S_2)$  is a fixed point, iff

$$\mathcal{L}(S_1, S_2) \models (x_1 = 0 \wedge x_2 = x_0) \vee \exists x_4 \exists x_5 (x_1 = x'_4 \wedge \text{Sat}(x_3, x_0, x_4, x_5, x_3) \wedge x_2 = x'_5)[\langle n_0, 0, n_2, m \rangle],$$

which in turn holds iff

$$\mathcal{L}(S_1, S_2) \models (x_1 = 0 \wedge x_2 = x_0)[\langle n_0, 0, n_2, m \rangle],$$

which holds iff  $n_0 + 0 = n_2$ . Suppose  $n_1 = k + 1$ .  $\mathcal{L}(S_1, S_2) \models \text{Sum}(x_0, x_1, x_2)[\langle n_0, k + 1, n_2 \rangle]$  iff  $\langle m, n_0, k + 1, n_2, m \rangle \in S_1$ , which holds, since  $\mathcal{L}(S_1, S_2)$  is a fixed point, iff

$$\mathcal{L}(S_1, S_2) \models (x_1 = 0 \wedge x_2 = x_0) \vee \exists x_4 \exists x_5 (x_1 = x'_4 \wedge \text{Sat}(x_3, x_0, x_4, x_5, x_3) \wedge x_2 = x'_5)[\langle n_0, k + 1, n_2, m \rangle],$$

which holds iff

$$\mathcal{L}(S_1, S_2) \models \exists x_4 \exists x_5 (x_1 = x'_4 \wedge \text{Sat}(x_3, x_0, x_4, x_5, x_3) \wedge x_2 = x'_5)[\langle n_0, k + 1, n_2, m \rangle],$$

which holds iff

$$\mathcal{L}(S_1, S_2) \models \exists x_4 \exists x_5 (x_1 = x'_4 \wedge \text{Sum}(x_0, x_4, x_5) \wedge x_2 = x'_5)[\langle n_0, k + 1, n_2 \rangle],$$

which, by the induction hypothesis, holds iff  $n_0 + (k + 1) = n_2$ .

We next need to show that

$$\begin{aligned} \mathcal{L}(S_1, S_2) \models & \neg\text{Sum}(x_0, x_1, x_2)[\langle n_0, n_1, n_2 \rangle] \text{ iff } n_0 + n_1 \neq n_2 \\ \text{i.e., } \mathcal{L}(S_1, S_2) \models & \text{Sum}(x_0, x_1, x_2)[\langle n_0, n_1, n_2 \rangle] \text{ iff } n_0 + n_1 = n_2. \end{aligned}$$

The proof is again by induction, only now we replace  $\models$  with  $\models$ ,  $\in S_1$  with  $\in S_2$ , and  $= n_2$  with  $\neq n_2$ . The case is similar for  $\times$ . Take the formula:

$$(x_1 = 0 \wedge x_2 = 0) \vee \exists x_4 \exists x_5 (x_1 = x_4' \wedge Sat(x_3, x_0, x_4, x_5, x_3) \wedge Sum(x_0, x_5, x_2)).$$

Suppose that this formula has Gödel number  $k$ .  $Sat(k, x_0, x_1, x_2, k)$  defines  $\times$  in any fixed point. The proof is parallel to the case for addition.

**Remark 2** Note that the satisfaction predicate is used in the above proof only in the form  $Sat(t_0, t_1, t_2, t_3, t_4)$ . We could have let  $\mathcal{L}$  contain just a 5-place predicate  $Sat(x_0, x_1, x_2, x_3, x_4)$  in addition to  $'$ . Then, given the theorem, the language will contain adequate resources to code finite sequences and talk about its own syntax. One will then be able to define a more general notion of satisfaction as a relation between a Gödel number for a formula and a code for a finite sequence. The approach taken in the paper is simpler and less artificial.

**Remark 3** The proof also works if we use the van Fraassen supervaluation scheme instead of the Strong Kleene scheme. On the other hand, the proof will not go through if the Weak Kleene scheme is used. This is so because any formula of the form  $\dots \exists x (\dots Sat(x, \dots)) \dots$  will be paradoxical (i.e., neither satisfied nor falsified in any fixed point by any sequence) since for some instances  $Sat(x, \dots)$  is undefined (e.g., instances in which the value of  $x$  is a paradoxical sentence).

**Remark 4** Of course the strength of the fixed point languages go well beyond that of arithmetic, since they contain their own satisfaction predicates. So, for example, in the minimal fixed point the  $\Pi_1^1$  relations are weakly defined and the hyperarithmetical relations are strongly defined.<sup>2</sup>

## NOTES

1. Of course there will be no fixed point in which  $Sat$  is totally defined. The formula  $\neg Sat(x_0, x_0)$  (cf., “ $x_0$  is heterological”) will be neither satisfied nor falsified by its own Gödel number in any fixed point. On the other hand,  $Sat(x_0, x_0)$  (cf., “ $x_0$  is autological”) will sometimes be satisfied by its own Gödel number, sometimes falsified by it, and sometimes neither satisfied nor falsified by it.
2. The basic trick involved in the proof of the theorem (the construction of appropriate self-referential formulas without the use of a substitution function) came to me while contemplating remarks of Kripke on diagonalization and the recursion theorem. It has been brought to my attention that Visser [2], pp. 666–667 also uses this trick in his proof of the “Prediagonal Lemma for SAT,” though he does so while considering a language in which it is already given that a pairing function is available and  $Sat$  expresses a two-place relation between a Gödel number for a formula and a code for a finite sequence.

## REFERENCES

- [1] Kripke, S., “Outline of a Theory of Truth,” *Journal of Philosophy*, vol. 72 (1975), pp. 690–716. [Zbl 0952.03513](#)

- [2] Visser, A., "Semantics and the Liar Paradox," pp. 617–706 in *Handbook of Philosophical Logic*, vol. 4, edited by D. Gabbay and F. Guentner, Reidel, Dordrecht, 1983.  
[Zbl 0875.03030](#) 1

*Philosophy Department  
University of Louisville  
Louisville, KY 40292*