International Journal of Theoretical Physics, Vol. 33, No. 5, 1994

Predictability in Deterministic Theories

Arne Hole[1]

Received April 7, 1993

A mathematical model for the general notion of a deterministic physical theory is introduced, and incompleteness results analogous to the halting theorem for Turing machines are demonstrated for this model. The discussion is not limited to algorithmic systems.

1. INTRODUCTION

Suppose that we have a deterministic physical theory with an associated set of physical systems to which the theory applies. To what extent is it possible that one of these physical systems governed by the theory can carry out computations within the theory? The aim of this article is to give a partial answer to this question. The results obtained can be considered as analogous to the unsolvability of the halting problem for Turing machines (Davis, 1958) or to the Gödel incompleteness theorem for formal number theory (Shoenfield, 1967). Due to the nonrecursive nature of the discussion, it should maybe most of all be considered as analogous to the problems related to the Tarski truth definition (Tarski, 1956).

It is known that a number of mathematical problems in physics are undecidable within standard axiom systems of mathematics (for instance, ZFC). Da Costa and Doria (1991) use Richardson's theorem (Richardson, 1968) to establish undecidability results about classical mechanics. One can say that results of this type concern incompleteness phenomena in physics inherited from mathematics via the concrete mathematical models actually used in today's physics. In contrast, my approach in this article is general and direct: no known incompleteness results from mathematics are used, instead analogous results are proven (from scratch) for a new mathematical

[1]Department of Mathematics, University of Oslo, Box 1053, Blindern, 0316 Oslo, Norway. E-mail: arneh@math.uio.no. FAX: +47-22 85 43 49.

0020-7748/94/0500-1085$07.00/0

structure intended to model the concept of a deterministic physical theory in general.

I begin in Section 3 with an informal discussion of calculations in classical mechanics. The formal development starts in Section 4.

2. NOTATION

N means $\{1, 2, 3, \ldots\}$. By an *initial segment* of **N** will be meant a set $A \subseteq \mathbf{N}$ such that if $a \in A$ and $b < a$, then $b \in A$ also, for all $a, b \in \mathbf{N}$. Note that both $\emptyset$ and **N** are initial segments of **N**. If f is a mapping, $Dom(f)$ and $Im(f)$ will denote the domain and image of f, respectively. So $f(Dom(f)) = Im(f)$. I put $\mathbf{R}^* = \mathbf{R} \cup \{-\infty, \infty\}$, the usual two-point compactification of **R**. The notation $[0, \infty]$ always refers to the corresponding closed interval in $\mathbf{R}^*$. We take $\mathbf{R}^\infty$ to mean $\mathbf{R} \times \mathbf{R} \times \cdots$, i.e., the countable infinite product of **R** with itself.

3. THE CONCEPT OF CALCULATION

Consider classical mechanics. Assume that we have a system S given by

$$S\colon \quad H(p, q) = \frac{p^2}{2m}$$

$$p(0) = p_0, \qquad q(0) = q_0$$

where p_0 and q_0 are given real numbers. Consider also the system C defined by

$$C\colon \quad H_c(p_c, q_c) = \frac{(p_c)^2}{2m}$$

$$p_c(0) = -p_0, \qquad q_c(0) = q_0$$

Then we have, according to the theory in question, the equivalence

$$\lim_{t \to \infty} q(t) \text{ exists} \Leftrightarrow \lim_{t \to \infty} q_c(t) \text{ exists}$$

In this situation, one can take the point of view that the system C is *calculating* whether or not the limit of $q(t)$ for the system S exists when $t \to \infty$. Alternatively, let C be given by

$$C\colon \quad H_c(p_c, q_c) = \frac{(p_c)^2}{2m}$$

$$p_c(0) = \begin{cases} 1 & \text{if} \quad p(0) = 0 \\ 0 & \text{otherwise} \end{cases} \qquad q_c(0) = 0$$

Then we have

$$\lim_{t\to\infty} q(t) \text{ exists} \Leftrightarrow \lim_{t\to\infty} q_c(t) \text{ does not exist}$$

In this case we also have

$$\lim_{t\to\infty} q(t) \text{ exists} \Rightarrow q_c(t_0) = \frac{t_0}{m}$$

$$\lim_{t\to\infty} q(t) \text{ does not exist} \Rightarrow q_c(t_0) = 0$$

for all $t_0 \geq 0$. Thus in this example (as well as in the first one) it is possible to read out the "answer" from the calculating system already after a finite lapse of time.

Notice that the concept of calculation, defined via arbitrary *equivalences* in the way we are hinting about now, will naturally be somewhat different from the usual algorithmic concept.

The calculating systems C considered so far have only been required to answer *one* yes/no question about the input system S. The simplest situation with two questions involved arises when we consider one yes/no question along with its negation. Still referring to the input system S, we can ask the following two questions:

1. Does $\lim_{t\to\infty} q(t)$ exist? (*Yes* or *No*)
2. Is it true that $\lim_{t\to\infty} q(t)$ does not exist? (*Yes* or *No*)

When we have two questions negating each other like this, we rule out the trivial possibility of S calculating its own behavior by direct simulation. But still we can easily find a system capable of answering both questions, for instance,

$$C: \quad H_c(p_1, p_2; q_1, q_2) = \frac{1}{2m}(p_1^2 + p_2^2)$$

$$p_1(0) = p_0, \qquad q_1(0) = q_2(0) = 0$$

$$p_2(0) = \begin{cases} 0 & \text{if question 1 is to be answered} \\ 1 & \text{if question 2 is to be answered} \end{cases}$$

Here the answer from C can be taken as no if the statement

$$\lim_{t\to\infty} q_1(t) \text{ exists} \oplus \lim_{t\to\infty} q_2(t) \text{ exists}$$

is true, and yes otherwise ($\oplus$ is exclusive-or).

It is clear that if the notion of "calculating system" is to have any interesting meaning, the Hamiltonian of the "system" should not be allowed to vary with the input system (and question) in an arbitrary way.

A simple and plausible solution is to require the Hamiltonian of the calculating system C to be fixed, while the initial conditions of C are allowed to depend on the input. Further, when considering input systems S with many degrees of freedom $(p_1, p_2, \ldots; q_1, q_2, \ldots)$, it is not necessary to let the number of initial values of C coding information about S increase without bounds. Information-theoretically, one input is sufficient, but it is more tidy to use two: one initial value coding the Hamiltonian H_S of the input system along with the question to be answered, and one coding the initial values $S_i(0)$, $i = 1, 2, 3, \ldots$. Here $S_i(t)$ is meant to denote either a specific coordinate or a specific momentum as a function of time; we may, for instance, agree that $S_1(t) = p_1(t)$, $S_2(t) = q_1(t)$, $S_3(t) = p_2(t)$, $S_4(t) = q_2(t)$, and so on. In this sort of scheme we can agree that $S_i(0) = 0$ if the number of degrees of freedom in S is less than $i/2$.

Let us consider an example. A natural way of mapping $\mathbf{R}^\infty$ bijectively into $\mathbf{R}$ is to take the composition of a map $\Lambda: \mathbf{R}^\infty \to [0, 1)^\infty$ with a map $\nu: [0, 1)^\infty \to [0, 1)$ shuffling decimals. For the sake of simplicity (see later), I will use base 2 decimal expansions (without infinite sequences of repeating ones, of course). Let us define $\lambda: \mathbf{R} \to [0, 1)$ by coding minus sign by 01, repeating each digit in integer parts twice, coding decimal point by 01, and keeping decimals. This gives, for example,

$$\lambda(10.0001\ldots) = 0.1100010001\ldots$$

$$\lambda(-1.101\ldots) = 0.011101101\ldots$$

$$\lambda(0.11100\ldots) = 0.000111100\ldots$$

Let $\Lambda: \mathbf{R}^\infty \to [0, 1)^\infty$ be the corresponding map acting componentwise by λ. To define ν, let $p = (p_1, p_2, \ldots) \in [0, 1)^\infty$ be given. Define $\nu(p)$ as the number in $[0, 1)$ constructed as follows:

- Write p_1 on every second decimal.
- Write p_2 on every second of the *remaining* decimals.
- Write p_3 on every second of the *now* remaining decimals.

And so forth. Example: With $p_1 = 0.1100011000111\ldots$, $p_2 = 0.0010111\ldots$, $p_3 = 0.101\ldots$, $p_4 = 0.00\ldots$, and $p_5 = 0.1\ldots$, we get

$$\nu(p) = 0.1011000001101001010111 1011\ldots$$

Let $\chi: \mathbf{R}^\infty \to [0, 1)$ be defined by $\chi = \nu \circ \Lambda$.

Using χ as input code, we can now put up a quite general (but somewhat informal) scheme for a special type of calculating system C within classical mechanics. First of all, the Hamiltonian H_C and the initial values $C_i(0)$ for $i \geq 3$ shall be fixed for all input systems S. Information about S can be given via the first two initial values of C as follows:

$$C_1(0) = \text{some real number } (H_S \text{ and the question coded here})$$

$$C_2(0) = \chi(S(0))$$

The calculation made by C can take as long as it wishes, but we must assume that the answer lies encoded in the state $C(t) = (C_1(t), C_2(t), \ldots)$ in the limit $t \to \infty$. For instance, C might use the existence/nonexistence of one of its component limits, say $\lim_{t\to\infty} C_3(t)$, as output signal. It could then, for example, work as follows:

$\lim_{t\to\infty} C_3(t)$ exists: The answer is yes

$\lim_{t\to\infty} C_3(t)$ does not exist: The answer is no

Later, when we make the formal definition of a calculating system, we shall be more general about this. We will then only assume that the answer is *algorithmically inferable* from complete information about which limits $\lim_{t\to\infty} C_i(t)$ exist, and the values of those existing (cf. the example above where we used $\oplus$ in defining the output signal).

4. DETERMINISTIC THEORIES

The central properties of a deterministic theory as considered here are the following:

1. Given a point in time t, the state of a system governed by the theory can, from the point of view of the theory, be completely specified by a finite or countably infinite number of information bits. For example, the bits may code the decimal expansions of a collection of real numbers, the coefficients of some sort of series expansion, or a description written in plain English.
2. If the states of two systems are equal at time t, then their states are equal for all times $t' > t$.

These properties form the physical motivation for the formal definition of deterministic theory given below. First we need some language.

Consider the alphabet $\Sigma = \{0, 1, d\}$, and let

$$\Omega^* = \Sigma \times \Sigma \times \Sigma \times \cdots = \Sigma^\infty$$

The elements of Ω^* are infinite words in the alphabet Σ. The ith coordinate of an element $a \in \Omega^*$ will be denoted by a^i, for $i = 1, 2, 3, \ldots$. The element a itself will be written $a = a^1a^2a^3 \ldots$. Example: $a = 1dd0110\ldots$ gives

$a^1 = 1$, $a^2 = d$, $a^3 = d$, and so forth. Let

$$\Omega = \{w \in \Omega^* | w^i \neq d \text{ for all } i \in \mathbf{N}\}$$

The ith coordinate a^i of an element $a \in \Omega$ will sometimes be called the ith *bit* of a. We define a topology on Ω by declaring that a subset $U \subseteq \Omega$ is a neighborhood of $a \in \Omega$ iff there is an $n \in \mathbf{N}$ such that $V_a^n \subseteq U$, where the set V_a^n is defined by

$$V_a^n = \{b \in \Omega | b^i = a^i \text{ for all } i \leq n\}$$

It is trivial to check that this does indeed define a topology on Ω, and that in this topology the sets V_a^n are open for all $a \in \Omega$ and $n \in \mathbf{N}$. We have a canonical surjection $Dec: \Omega \to [0, 1]$ given by

$$[Dec(a)]^i = a^i$$

for $i \in \mathbf{N}$, where $[Dec(a)]^i$ is interpreted as ith decimal in a base 2 decimal expansion for $Dec(a)$. Note that Dec is not injective (infinite sequences of repeating ones may occur in Ω).

Lemma 1. The map $Dec: \Omega \to [0, 1]$ is continuous.

Proof. Let $Dec(a) \in [0, 1]$ and let $\epsilon > 0$ be given. Choose $n \in \mathbf{N}$ such that $2^{-n} < \epsilon$. Then for all $b \in V_a^n$ we have $|Dec(b) - Dec(a)| < \epsilon$. ■

Let $G: [0, \infty] \to \Omega^*$, and let $i \in \mathbf{N}$. If there is an $L \in \Sigma$ and an $a \in [0, \infty)$ such that $G(t)^i = L$ for all $t \in [a, \infty)$, we write $\lim_{t \to \infty} G(t)^i = L$. If there are no such L and a, we say that $\lim_{t \to \infty} G(t)^i$ does not exist.

Definition 1 (∞-maps). A map $S: [0, \infty] \to \Omega^*$ will be called an ∞-map if $S(0) \in \Omega$ and the following conditions are satisfied for all $i \in \mathbf{N}$:

If $\lim_{t \to \infty} S(t)^i = L$, then $S(\infty)^i = L$.
If $\lim_{t \to \infty} S(t)^i$ does not exist, then $S(\infty)^i = d$.

Notice that a given map $S_0: [0, \infty) \to \Omega^*$ with $S(0) \in \Omega$ can always be extended to a unique ∞-map $S: [0, \infty] \to \Omega^*$. The difference between S_0 and S is simply that S contains some extra explicit information about the limits $\lim_{t \to \infty} S_0(t)^i$.

Let *Tur* denote the set of all one-taped, one-sided Turing machines working over the alphabet Σ. By an *input* to M is meant an element $a \in \Omega^*$. It is always understood that the machines $M \in Tur$ start their computations with their head at tape position 1. Given $M \in Tur$ and $w \in \Omega^*$, we define a map $M[w]: [0, \infty] \to \Omega^*$ as follows:

(i) $M[w](t) = a_i$ for all $t \in [i, i + 1)$, where a_i is the tape of M after i operations given input w, for $i = 0, 1, 2, 3, \ldots$. Note that $M[w](0) = a_0 = w$.

(ii) For each $i \in \mathbf{N}$, let $[M[w](\infty)]^i = L$ if $\lim_{t \to \infty}[M[w](t)]^i = L$, and let $[M[w](\infty)]^i = d$ if $\lim_{t \to \infty}[M[w](t)]^i$ does not exist.

The map $M[w]$ is then an ∞-map. It will be called the *development* of M given input w. When I say that $M[w]$ *halts*, I mean that M halts with input w. This is equivalent to

$$(\exists n \in \mathbf{N})(M[w](t) = M[w](t') \text{ for all } t, t' > n)$$

A machine $M \in Tur$ is called *acceptable* if M halts with 0 or 1 at the head position (i.e., has output 0 or 1) for all inputs $w \in \Omega^*$. We define

$$Tur_A = \{M \in Tur \mid M \text{ is acceptable}\}$$

I will use the notation $M(w)$ to denote the output of $M \in Tur_A$ when it gets $w \in \Omega^*$ as input. The output $M(w)$ should not be confused with the development $M[w]$ as defined earlier.

Definition 2 (Deterministic Theories). A *deterministic theory* T is a pair $T = (\Gamma_T, H)$, where

Γ_T is a family of ∞-mappings S: $[0, \infty] \to \Omega^*$
H is an arbitrary mapping with domain of definition Γ_T

such that the following condition (called the determinism axiom) is satisfied for all $S, S' \in \Gamma_T$ and all $t_0 \in [0, \infty]$:

$$\text{If } S(t_0) = S'(t_0) \text{ and } H(S) = H(S'),$$
$$\text{then } S(t) = S'(t) \text{ for all } t \in [t_0, \infty]$$

In this definition, the parameter $t \in [0, \infty]$ can be considered as time. The elements $S \in \Gamma_T$ are called *systems*. The reason for using ∞-maps is that it is convienient to have information about limit behavior coded in an easily accessible manner. In most applications (cf. Examples 1–3 below) the maps $S \in \Gamma_T$ will satisfy $S(t) \in \Omega$ for all $t \in [0, \infty)$. Then $S(\infty)^i = d$ iff $\lim_{t \to \infty} S(t)^i$ does not exist. The symbol d is chosen to rhyme with "does not have a value."

Given a deterministic theory $T = (\Gamma_T, H)$, we denote the image of H by $Ham(T)$, i.e., $H(\Gamma_T) = Ham(T)$. We often write H_S instead of $H(S)$ for $S \in \Gamma_T$. Frequently H_S is called the Hamiltonian of the system S, in analogy with classical mechanics. Note, however, that our use of the word "system" here corresponds (via interpretation) to the *time development* of a system in the usual sense of the word. The map H is included in the definition of deterministic theory only to make the process of representing concrete theories as deterministic theories easier (cf. Examples 1–3 below). The map is not necessary from a theoretical point of view.

By the determinism axiom, the mapping

$$\phi_T\colon \quad \Gamma_T \to Ham(T) \times \Omega$$

defined by $\phi_T(S) = (H_S, S(0))$ is an injection. Its inverse (defined on the image of ϕ_T) will be denoted by ϕ_T^{-1}.

The reason for topologizing Ω is that we sometimes want to think of an element $w \in \Omega$ as representing a point in space, for instance, via the map *Dec*. In this context, the definition below has a natural geometrical meaning.

Definition 3. A deterministic theory $T = (\Gamma_T, H)$ is called *closed* if for each fixed $S \in \Gamma_T$ the set

$$\{a \in \Omega | (\exists S' \in \Gamma_T)(H(S') = H(S) \wedge S'(0) = a)\}$$

is a closed subset of Ω.

Example 1 (Classical mechanics). In the spirit of Section 3, we can define a *classical theory* as a pair $\Theta = (Clas, h)$, where *Clas* is a nonempty set consisting of mappings $s\colon [0, \infty) \to \mathbf{R}^\infty$ and $h\colon s \mapsto ham(s)$ is a map defined on *Clas*. (To avoid ambiguities in notation, we now denote the elements of *Clas* by small Latin letters instead of capital ones.) We can represent Θ as a deterministic theory as follows. For each $s \in Clas$, define an associated map $S\colon [0, \infty] \to \Omega^*$ by extending $t \mapsto Dec^{-1}(\chi(s(t)))$ to an ∞-map, where $s(t)$ is represented with the base 2 decimal expansion where no infinite sequence of repeating ones occurs. Define $T = (\Gamma_T, H)$ by putting $\Gamma_T = \{S | s \in Clas\}$ and $H_S = ham(s)$ for all $S \in \Gamma_T$. We then have a canonical bijection $\xi\colon \Gamma_T \to Clas$ given by $S \mapsto s$.

Example 2 (Turing machines). Let D be the set of all Turing machine developments starting in Ω, i.e.,

$$D = \{M[w] | M \in Tur \text{ and } w \in \Omega\}$$

Let $\Gamma_T \subseteq D$, and let $H\colon \Gamma_T \to Tur$ be such that for all $M, M' \in Tur$ the following is satisfied: If $H(M[w]) = M'$, then $M'[w] = M[w]$. Then $T = (\Gamma_T, H)$ is a closed deterministic theory. Note that Turing machines play the role of Hamiltonians here.

Example 3 (Billiard balls). Suppose we are given an infinite array of cells

$$c_1, c_2, c_3, \ldots$$

each of which can contain a billiard ball or be empty. Then the state of the array at a given instant of time t can be considered as a point $w(t) \in \Omega$, where for each $i \in \mathbf{N}$:

$$w(t)^i = \begin{cases} 1 & \text{if there is a ball in cell } i \text{ at time } t \\ 0 & \text{otherwise.} \end{cases}$$

Suppose also that we are given a set X of mechanical devices (not necessarily algorithmic) that can move the billiard balls in the array around in such a way that the state $w(t)$ of the array at all times $t \geq 0$ is uniquely determined by the state $w(0)$ at time $t = 0$ and the system $h \in X$ operating on the array. Assuming discrete time, we can represent this situation as a deterministic theory exactly as we did with Turing machines. For each $h \in X$, let $h[w]: [0, \infty] \to \Omega^*$ be the ∞-map corresponding to the situation where h starts with $w \in \Omega$ as input. Let $D = \{h[w] | h \in X \text{ and } w \in \Omega\}$, let $\Gamma_T \subseteq D$, and let $H: \Gamma_T \to X$ be such that if $H(h[w]) = h'$, then $h'[w] = h[w]$. Then $T = (\Gamma_T, H)$ is a closed deterministic theory.

In connection with Example 3, it should be mentioned that the billiard ball computer is a well-known physical model of algorithmic computation. Consult Fredkin and Toffoli (1982). In these models, the computations themselves are done by billiard balls.

5. CALCULATING SYSTEMS

Given $T = (\Gamma_T, H)$, we want to investigate the ability of a system $S \in \Gamma_T$ to answer questions about other systems in Γ_T. To do this, first of all we must decide what should be meant by a question. We restrict ourselves to the following format: Given $M \in Tur_A$, $S \in \Gamma_T$, and $t \in [0, \infty]$, what is the output $M(S(t))$? Technically, it is advantageous to let the question be specified by t and M alone. We are then led to the definition below.

Definition 4 (The question set). Let $Q = Tur_A \times [0, \infty]$. The set Q is called the *question set* associated with deterministic theories. An element $q \in Q$ is often written $q = (M_q, t_q)$.

As an example, let $M \in Tur_A$ be such that $M(w) = 1$ iff $w^i = 0$ for some $i \leq 10$. Then $q = (M, 17) \in Q$ represents the question "Are any of the first ten coordinates zero at time $t = 17$?" To ask this question about a system $S \in \Gamma_T$ is to ask for the value of $M(S(17))$.

Since for all $i \in \mathbf{N}$ and $x \in \Sigma$ there is an $M \in Tur_A$ such that $\forall w \in \Omega^*(M(w) = 1 \Leftrightarrow w^i = x)$, all bits of a given point $S(t) \in \Omega^*$ can be extracted via machines in Tur_A. Thus all information about an arbitrary system S in a deterministic theory can be extracted via questions in Q. We are now ready to state our formal definition of a calculating system.

Definition 5 (Calculating systems). A *calculating system* C for a deterministic theory $T = (\Gamma_T, H)$ is a 4-tuple

$$C = (H_C, f_C, i_C^0, Out_C)$$

where

$H_C \in Ham(T)$

$f_C\colon \quad \Omega \times \Omega \to \Omega$

$i_C^0\colon \quad Ham(T) \times Q \to \Omega$

$Out_C \in Tur_A$

such that for all $(S, q) \in \Gamma_T \times Q$ there is a system $C[S, q] \in \Gamma_T$ satisfying

(i) $H(C[S, q]) = H_C$
(ii) $C[S, q](0) = i_C(S, q)$

where $i_C\colon \Gamma_T \times Q \to \Omega$ is defined by $i_C(S, q) = f_C(i_C^0(H_S, q), S(0))$. The map i_C will be called the *input map* (or the *input code*) of C.

Definition 6. Let C be a calculating system for a theory $T = (\Gamma_T, H)$, and let $(S, q) \in \Gamma_T \times Q$, with $q = (M_q, t_q)$. If

$$Out_C(C[S, q](\infty)) = M_q(S(t_q))$$

then (S, q) is called *solvable* for C. The set of all $(S, q) \in \Gamma_T \times Q$ which are solvable for C is called the *domain* of C, and is denoted by D_C.

According to this definition, a calculating system C is characterized by a fixed Hamiltonian H_C, an output signal Out_C, and an input mapping i_C. The situation where C is calculating the answer to the question $q \in Q$ about $S \in \Gamma_T$ is represented by the system $C[S, q] \in \Gamma_T$. The answer to the question, namely $M_q(S(t_q))$, is given by $Out_C(C[S, q](\infty))$. My reason for splitting the input map i_C into two parts f_C and i_C^0 is that it will be convenient to have the dependence of $C[S, q](0)$ on $S(0)$ under extra control. Assuming that $C[S, q](0)$ is allowed to depend on H_S, q, and $S(0)$ only, this splitting represents no loss of generality. To be precise, given a map $\xi\colon Ham(T) \times Q \times \{S(0) | S \in \Gamma_T\} \to \Omega$, if we choose i_C^0 injective (which we assume can always be done) and put $f_C(w, z) = \xi((i_C^0)^{-1}(w), z)$, then

$$f_C(i_C^0(H_S, q), S(0)) = \xi_C(H_S, q, S(0))$$

Most of the results in this article will concern calculating systems having input codes of a special type called *explicit* codes. I will now proceed to define this.

A strictly increasing map $\eta: A \to \mathbf{N}$ will be called an *initial sequence* if $A = Dom(\eta) \subseteq \mathbf{N}$ is an initial segment. If η is an initial sequence, there exists a unique initial sequence $\bar{\eta}: B \to \mathbf{N}$ (the complement of η) defined on an initial segment $B \subseteq \mathbf{N}$ such that

$$Im(\eta) \cup Im(\bar{\eta}) = \mathbf{N}$$

$$Im(\eta) \cap Im(\bar{\eta}) = \emptyset$$

Here we adopt the convention that if $Im(\eta) = \mathbf{N}$, then $\bar{\eta}$ is empty, i.e., $Dom(\bar{\eta}) = Im(\bar{\eta}) = \emptyset$. Conversely, if $Im(\eta) = \emptyset$, then $Dom(\bar{\eta}) = \mathbf{N}$ and $\bar{\eta}$ is the identity. Now let η be an initial sequence. Given $a, b \in \Omega$, define $\sigma_\eta(a, b) \in \Omega$ as follows:

$$[\sigma_\eta(a, b)]^{\bar{\eta}(i)} = a^i \qquad \text{for all} \quad i \in Dom(\bar{\eta})$$

$$[\sigma_\eta(a, b)]^{\eta(i)} = b^i \qquad \text{for all} \quad i \in Dom(\eta)$$

Here $[\sigma_\eta(a, b)]^k$ means kth coordinate of $\sigma_\eta(a, b)$, in accordance with our conventions. I will sometimes write $\sigma_\eta(a, b)^k$ instead of $[\sigma_\eta(a, b)]^k$.

Example 4. Let $\eta: \mathbf{N} \to \mathbf{N}$ be given by $\eta(j) = 2j$ for all $j \in \mathbf{N}$. Then $\bar{\eta}(j) = 2j - 1$ for all $j \in \mathbf{N}$, and $Dom(\bar{\eta}) = \mathbf{N}$. We have $\sigma_\eta(a, b) = a^1 b^1 a^2 b^2 a^3 b^3 \ldots$.

Definition 7 (Explicit input code). Let $T = (\Gamma_T, H)$ be a theory, and let $C = (H_C, f_C, i_C^0, Out_C)$ be a calculating system for T. The input map i_C of C is called *explicit* if there is an initial sequence $\eta: A \to \mathbf{N}$ such that $f_C = \sigma_\eta$, i.e., such that

$$i_C(S, q) = \sigma_\eta(i_C^0(H_S, q), S(0))$$

for all $(S, q) \in \Gamma_T \times Q$.

A basic thing to notice about this definition is that the class of explicit input codes is sufficiently rich to contain codes whose maps $f_C = \sigma_\eta$ conserve all information about $i_C^0(H_S, q)$ and $S(0)$. For instance, the map σ_η considered in Example 4 is conservative in this sense.

Definition 8 (Universal systems). A calculating system $C = (H_C, f_C, i_C^0, Out_C)$ for a theory $T = (\Gamma_T, H)$ is called a *universal system for* T if

$$D_C = \Gamma_T \times Q$$

In other words, a universal system for a theory T is a calculating system which can answer any question $q \in Q$ about any system $S \in \Gamma_T$ correctly. The main result of this article can (in its weakest form) now be stated in a mathematically precise form as follows: There exists no closed deterministic theory which has a universal system with an explicit input code. The next two sections of the paper are purely mathematical, and their aim is to prove this theorem.

6. FRACTAL POINTS

Consider an initial sequence $\eta: A \to \mathbf{N}$, and let $\sigma_\eta: \Omega \times \Omega \to \Omega$ be the map defined above. For given $a, z \in \Omega$ we put $\xi_0 = z$ and

$$\xi_{i+1} = \sigma_\eta(a, \xi_i)$$

for $i = 0, 1, 2, 3, \ldots$. Then we define

$$Frac_\eta^z(a) = \lim_{i \to \infty} \xi_i$$

where lim refers to the Ω-topology.

Lemma 2. Let $\eta: A \to \mathbf{N}$ be an initial sequence. Then for fixed $z \in \Omega$ the map $Frac_\eta^z: \Omega \to \Omega$ is well defined, and

$$\sigma_\eta(a, Frac_\eta^z(a)) = Frac_\eta^z(a)$$

for all $a \in \Omega$. Moreover, if $\eta(1) > 1$, then $Frac_\eta^{z_1} = Frac_\eta^{z_2}$ for all $z_1, z_2 \in \Omega$ [i.e., $Frac_\eta^z(a)$ is independent of z]. If the complement of $Im(\eta)$ in $\mathbf{N}$ is infinite, then the map $Frac_\eta^z: \Omega \to \Omega$ is injective.

Proof. First assume $Dom(\bar{\eta}) = \emptyset$, where $\bar{\eta}$ is the complement of η. Then $\eta(i) = i$ for all $i \in \mathbf{N}$, so $\sigma_\eta(a, b) = b$ for all $a, b \in \Omega$. It follows trivially that $Frac_\eta^z(a) = z$ for all $a, z \in \Omega$. We also have $\eta(1) = 1$ and $Im(\eta) = \mathbf{N}$.

Assume now $Dom(\bar{\eta}) \neq \emptyset$. Given $a = a^1 a^2 a^3 \cdots \in \Omega$, construct $b = b^1 b^2 b^3 \cdots \in \Omega$ in three stages as follows.

(i) For all $i < \bar{\eta}(1)$, put $b^i = z^i$.
(ii) Put $b^{\bar{\eta}(i)} = a^i$ for all $i \in Dom(\bar{\eta})$.
(iii) Put $b^{\eta(i)} = b^i$ for all $i \in Dom(\eta)$ (starting from $i = 1$, recursively).

Note that $\eta(i) = i \Leftrightarrow i < \bar{\eta}(1)$, in which case (iii) simply states $b^i = b^i$. I claim that the sequence ξ_i defined by a converges to b. To see why, we prove by induction that ξ_i has the first i coordinates in common with b.

If $\bar{\eta}(1) > 1$, then $b^1 = z^1$ and $(\xi_1)^1 = \sigma_\eta(a, z)^1 = \sigma_\eta(a, z)^{\eta(1)} = z^1$. If $\bar{\eta}(1) = 1$, then $b^1 = b^{\bar{\eta}(1)} = a^1$ by (ii), while $(\xi_1)^1 = \sigma_\eta(a, z)^1 = \sigma_\eta(a, z)^{\bar{\eta}(1)} = a^1$ also. So the result holds for $i = 1$.

Assume it holds for $i = k$, $k \geq 1$. We must prove that $(\xi_{k+1})^j = b^j$ for all $j \in \{0, \ldots, k+1\}$.

Case 1. $j < \bar{\eta}(1)$. Then $\eta(j) = j$. So

$$(\xi_{k+1})^j = \sigma_\eta(a, \xi_k)^j = \sigma_\eta(a, \xi_k)^{\eta(j)} = (\xi_k)^j$$

By repetition, $(\xi_{k+1})^j = (\xi_0)^j = z^j$. Also, $b^j = z^j$ by (i).

Case 2. $j = \bar{\eta}(1)$. Here we get

$$(\xi_{k+1})^j = \sigma_\eta(a, \xi_k)^j = \sigma_\eta(a, \xi_k)^{\bar{\eta}(1)}$$
$$= a^1 = b^{\bar{\eta}(1)} = b^j$$

Case 3. $j > \bar{\eta}(1)$. Assume first $j \in Im(\bar{\eta})$, say $j = \bar{\eta}(m)$. Then

$$(\xi_{k+1})^j = \sigma_\eta(a, \xi_k)^j = a^m = b^{\bar{\eta}(m)} = b^j$$

Assume next $j \in Im(\eta)$, with $j = \eta(m)$. We now have $m < j$, so $m \leq k$. Then

$$(\xi_{k+1})^j = \sigma_\eta(a, \xi_k)^j = (\xi_k)^m$$
$$= b^m = b^{\eta(m)} = b^j$$

where I used the induction hypothesis at the third equality and (iii) at the fourth. Thus the induction is complete. So $Frac^z_\eta(a) = b$. Thus the map $Frac^z_\eta : \Omega \to \Omega$ is well defined. If $a, c \in \Omega$, $i \in Dom(\bar{\eta})$, and $a^i \neq c^i$, then $[Frac^z_\eta(a)]^{\bar{\eta}(i)} = a^i \neq c^i = [Frac^z_\eta(c)]^{\bar{\eta}(i)}$. The injectivity of $Frac^z_\eta$ when the complement of $Im(\eta)$ in $\mathbf{N}$ is infinite follows directly from this, since in that case $Dom(\bar{\eta}) = \mathbf{N}$. If $\eta(1) > 1$, then $\bar{\eta}(1) = 1$, so the independence of b on z in this case is trivial from the explicit construction of b given above. Finally, assume that $i \in Im(\bar{\eta})$, where $i = \bar{\eta}(m)$. Then, with $b = Frac^z_\eta(a)$ as before, we get

$$\sigma_\eta(a, b)^i = a^m = b^{\bar{\eta}(m)} = b^i$$

If $i \in Im(\eta)$, where $i = \eta(m)$, then

$$\sigma_\eta(a, b)^i = b^m = b^{\eta(m)} = b^i$$

So we have $\sigma_\eta(a, b) = b$. ∎

Let $\Omega_0 = \{a \in \Omega | (\forall i \in \mathbf{N})(\exists j \in \mathbf{N})(j > i \wedge a^j \neq 1)\}$, i.e., let Ω_0 be the set of elements in Ω which have no infinite sequence of repeating ones. For some applications the following result is useful:

Lemma 3. Let η be an initial sequence. If $z, a \in \Omega_0$, then $Frac^z_\eta(a) \in \Omega_0$.

Proof. By the explicit construction of $Frac^z_\eta(a)$ given in the proof of Lemma 2, it follows that if neither a nor z contains any infinite sequence of repeating ones, then $Frac^z_\eta(a)$ will not contain such a sequence either. ∎

Example 5. Let $\eta(i) = 2i$ for all $i \in \mathbf{N}$, and let

$$a = 0110011101101\ldots, \qquad z = 1101010111001\ldots$$

Then $\bar{\eta}(i) = 2i - 1$ for $i \in \mathbf{N}$, and

$$\xi_1 = \sigma_\eta(a, z) = 01111001001110110111100011\ldots$$

$$\xi_2 = \sigma_\eta(a, \xi_1) = 00111101011010110010110111\ldots$$

$$\xi_3 = \sigma_\eta(a, \xi_2) = 00101101011110110011110011\ldots$$

etc. It turns out that

$$Frac_\eta^z(a) = 00101100011110100011110111\ldots$$

Notice that $Frac_\eta^z(a)$ contains smaller and smaller copies of itself. Reading every second bit, one gets the point $Frac_\eta^z(a)$ itself, and reading every *fourth* bit one also gets the same point. And so on.

7. THE BASIC RESULT

We are now ready to prove our first nonexistence result concerning universal systems. The proof is typical diagonal argument.

Theorem 1. There exists no closed deterministic theory which has a universal system with explicit input code.

Proof. Assume that $T = (\Gamma_T, H)$ and $C = (H_C, \sigma_\eta, i_C^0, Out_C)$ is a counterexample. Since $Out_C \in Tur_A$, there exists a Turing machine $M \in Tur_A$ such that

$$M(w) = \begin{cases} 1 & \text{if} \quad Out_C(w) = 0 \\ 0 & \text{if} \quad Out_C(w) = 1 \end{cases}$$

(we can take M as the negation of Out_C). Let $q = (M, \infty)$; then $q \in Q$. Since T has a calculating system, it follows that it is nonempty. Choose $R \in \Gamma_T$, and let

$$S^1 = C[R, q]$$

(Here 1 is a usual index, it does *not* indicate first coordinate.) Then consider the sequence $\{S^i\}_{i=1}^\infty \subseteq \Gamma_T$ defined by

$$S^{i+1} = C[S^i, q]$$

where $S^1 \in \Gamma_T$ is defined above. Let $i \in \mathbf{N}$ be given. By definition of $C[S^i, q]$, we have $S^{i+1}(0) = i_C(S^i, q)$. Then, since

$$H(S^i) = H_C \tag{1}$$

it follows by Definition 7 that

$$\begin{aligned} S^{i+1}(0) &= i_C(S^i, q) \\ &= \sigma_\eta(i_C^0(H(S^i), q), S^i(0)) \\ &= \sigma_\eta(i_C^0(H_C, q), S^i(0)) \end{aligned}$$

From Lemma 2 it now follows that the sequence $\{S^i(0)\}_{i=1}^\infty$ converges in the Ω-topology to the point $Frac_\eta^z(a)$, where $a = i_C^0(H_C, q)$ and $z = S^1(0)$. Since T is closed, it follows from (1) that there is a system $S \in \Gamma_T$ such that

$$H_S = H_C$$

$$S(0) = Frac_\eta^z(a)$$

with a and z defined above. By injectivity of ϕ_T, the system S is uniquely determined by these conditions. Now

$$\begin{aligned} i_C(S, q) &= \sigma_\eta(i_C^0(H_S, q), S(0)) \\ &= \sigma_\eta(i_C^0(H_C, q), S(0)) \\ &= \sigma_\eta(a, Frac_\eta^z(a)) = Frac_\eta^z(a) = S(0) \end{aligned}$$

where I used Lemma 2 at the fourth equality. Then we get

$$\begin{aligned} C[S, q] &= \phi_T^{-1}(H_C, i_C(S, q)) \\ &= \phi_T^{-1}(H_S, S(0)) = S \end{aligned}$$

This gives

$$\begin{aligned} M(S(\infty)) = 0 &\Leftrightarrow Out_C(C[S, q](\infty)) = 0 \\ &\Leftrightarrow M(C[S, q](\infty)) = 1 \\ &\Leftrightarrow M(S(\infty)) = 1 \end{aligned}$$

The first equivalence here is because C is universal, the second is by definition of M, and the third uses the result derived above. ■

8. AN EXAMPLE FROM CLASSICAL MECHANICS

Referring to Example 1 in Section 4, let $T = (\Gamma_T, H)$ be the representation of a classical theory $\Theta = (Clas, h)$. We can make the notion of "calculating system" for *Clas* as considered in Section 3 precise by referring to calculating systems for T. A calculating system for T is said to be *classical* if the following two conditions are satisfied:

(i) For all $(S, q) \in \Gamma_T \times Q$,

$$\xi(C[S, q])_2(0) = \chi(\xi(S)(0))$$

(ii) For all $i \neq 2$ and all $q \in Q$, we have

$$\xi(C[S, q])_i(0) = \xi(C[S', q])_i(0)$$

for all $S, S' \in \Gamma_T$ such that $H(S) = H(S')$.

Further, we define a *classical universal system* for T to be a calculating system for T which is both classical and universal.

Theorem 2. Let $T = (\Gamma_T, H)$ be a representation of $\Theta = (Clas, h)$. If for each fixed $s \in Clas$ the set

$$\{a \in \mathbf{R} | (\exists s' \in Clas)(h(s') = h(s) \wedge s'_2(0) = a \wedge s'_j(0) = s_j(0) \text{ for all } j \neq 2)\}$$

is a closed subset of $\mathbf{R}$, then T has no classical universal system.

Proof. Let $C = (H_C, f_C, i_C^0, Out_C)$ be a classical universal system for T Let

$$\Gamma_T^+ = \{S \in \Gamma_T | \xi(S)_2(0) \in [0, 1)\}$$

and let $T^+ = (\Gamma_T^+, H^+)$, where H^+ is the restriction of H to Γ_T^+. Then since $\xi(C[S, q])_2(0) \in [0, 1)$ for all $(S, q) \in \Gamma_T \times Q$, the tuple $C^+ = (H_C, f_C, i_C^{0+}, Out_C)$ is a universal system for T^+, where i_C^{0+} is the restriction of i_C^0 to $Ham(T^+) \times Q$. Let i_C^+ be the input map of C^+. Using the definition of χ and remembering that $\chi(s(0)) \in [0, 1)$ for all $s \in Clas$, it is easily seen that i_C^+ is explicit, with associated sequence

$$\eta(j) = 4j + 14$$

for $j \in \mathbf{N}$. {The coordinates occupied by the a_2 bits in $\chi(a)$ are given by $\beta(j) = 4j - 2$, and when $a_2 \in [0, 1)$ the first four of these bits code 0001.} The remainder of the proof is analogous to the proof of Theorem 1. In the sequence $\{S^i\}_{i=1}^{\infty}$ constructed there, we have

$$[S^i(0)]^{\bar{\eta}(k)} = [S^j(0)]^{\bar{\eta}(k)} \qquad \text{for all} \quad i, j, k \in \mathbf{N}$$

and it follows that $\xi(S^i)_n(0) = \xi(S^j)_n(0)$ for all $i, j \in \mathbf{N}$ and all $n \neq 2$. In other words, only the second component moves. Further, by Lemma 2 the sequence $\{S^i(0)\}_{i=1}^{\infty}$ converges in the Ω-topology to a point $p \in \Omega$, and since *Dec* is continuous (Lemma 1), the sequence $\{\xi(S^i)_2(0)\}_{i=1}^{\infty}$ converges in $\mathbf{R}$. By the closure condition stated in the theorem combined with Lemma 3 it then follows that $p \in \Gamma_T^+$. ■

9. FINITELY DESCRIBABLE SYSTEMS

So far we have demanded from our universal systems that they are able to treat any system S in a given theory. To be reasonable, however, we should only require that a universal system can handle cases where the input system S can be described (from the point of view of the deterministic theory in question) by a *finite* amount of information. The finiteness condition will be implemented via the notion of *interpreting systems*, which is to be introduced below.

A point $a \in \Omega^*$ is called *recursive* if there is a Turing machine $M \in Tur$ such that $M[000\ldots](\infty) = a$, i.e., such that when starting from input $w = 000\ldots$, M writes a on the tape. A number $t \in [0, \infty)$ is called recursive if $Dec^{-1}(\lambda(t))$ is recursive, where λ is the map defined in Section 3. A number $t \in [0, \infty]$ is called recursive if $t = \infty$ or t is recursive considered as an element of $[0, \infty)$. A question $q = (M_q, t_q) \in Q$ is called recursive if t_q is recursive. Finally, let $A \subseteq \mathbf{N}$ be an initial segment. A map $\eta: A \to \mathbf{N}$ is called recursive if it is a total, recursive map from A to $\mathbf{N}$, in the usual sense.

For each $n \in \mathbf{N}$, define $Cut\{n\}: \Omega \to \Omega$ by

$$[Cut\{n\}(w)]^i = \begin{cases} w^i & \text{for} \quad i \leq n \\ 0 & \text{for} \quad i > n \end{cases}$$

Example: $Cut\{3\}(101001\ldots) = 101000\ldots$. It is convenient to define $Cut\{\infty\}(w) = w$ for all $w \in \Omega$.

By a *counting machine* will be meant a machine $M \in Tur$ such that for all $w \in \Omega$ the following condition is satisfied:

- If $M[w]$ halts, then the output tape of M (i.e., the tape of M at the halting moment) is of the form $111\ldots110a^1a^2a^3\ldots$.

If M is a counting machine and $M[w]$ halts, we define $Num(M[w]) \in \mathbf{N}$ as the number of elements in the initial group of ones on the output tape. Example: If $M[w]$ halts with the output tape $11101d01\ldots$, then $Num(M[w]) = 3$.

Definition 9 (Interpreting systems). An *interpreting system* is a pair $P = (M_1, M_2)$, where $M_1 \in Tur$ is a counting machine, and where $M_2 \in Tur$ is such that $M_2[w](\infty) \in \Omega$ for all $w \in \Omega$.

The set of all interpreting systems will be denoted by Π. Given an interpreting system $P = (M_1, M_2)$, define the map $int_P: \Omega \to \Omega^*$ by

$$int_P(w) = \begin{cases} M_2[Cut\{Num(M_1[w])\}(w)](\infty) & \text{if} \quad M_1[w] \text{ halts} \\ ddd\ldots & \text{otherwise} \end{cases}$$

We now state the definition of our new "weak" version of a calculating system. Roughly speaking, these systems are only required to handle recursive inputs.

Definition 10 (Weak calculating systems). A *weak calculating system C* for a deterministic theory $T=(\Gamma_T, H)$ is a 4-tuple $C=(h_C, f_C, i_C^0, Out_C)$, where

$$h_C: \Pi \times \Pi \to Ham(T)$$

$$f_C: \Omega \times \Omega \to \Omega$$

$$i_C^0: Ham(T) \times Q \to \Omega$$

$$Out_C \in Tur_A$$

such that for all $(S, q) \in \Gamma_T \times Q$ and all interpreting systems P, P' there is a system $C[S, q, P, P'] \in \Gamma_T$ satisfying

(i) $H(C[S, q, P, P']) = h_C(P, P')$
(ii) $C[S, q, P, P'](0) = i_C(S, q)$

where the map $i_C: \Gamma_T \times Q \to \Omega$ (called the input map of C) is defined by $i_C(S, q) = f_C(i_C^0(H_S, q), S(0))$.

Definition 11. Let C be a weak calculating system for $T=(\Gamma_T, H)$, and let $(S, q) \in \Gamma_T \times Q$. If for all $P, P' \in \Pi$ such that $int_P(i_C(S, q)) = i_C^0(H_S, q)$ and $int_{P'}(i_C(S, q)) = S(0)$, we have

$$Out_C(C[S, q, P, P'](\infty)) = M_q(S(t_q))$$

then (S, q) is called *solvable* for C. The set of all solvable $(S, q) \in \Gamma_T \times Q$ is called the *domain* of C, and is denoted by D_C.

Given an ordinary calculating system $Z=(H_Z, f_Z, i_Z^0, Out_Z)$, we can construct a weak calculating system $C=(h_C, f_C, i_C^0, Out_C)$ by putting $h_C(P, P') = H_Z$ for all $P, P' \in \Pi$ (constant map), $f_C = f_Z$, $i_C^0 = i_Z^0$, and $Out_C = Out_Z$. Then $D_Z \subseteq D_C$. In general, we can interpret the map h_C of a weak calculating system as an assignment of different variants of the calculating system, corresponding to using different "preprocessing units." Each preprocessing unit corresponds to a certain pair of interpreting systems, such that $h_C(P, P')$ denotes the variant where the preprocessing unit representing (P, P') is used. Working with a given pair $P=(M_0, M_0')$ and $P'=(M_1, M_1')$, we can extract complete information about $int_P(i_C(S, q))$ and $int_{P'}(i_C(S, q))$ [which are interpreted as $i_C^0(H_S, q)$ and $S(0)$, respectively] from $i_C(S, q)$ by performing a finite number of algorthmic steps, as follows:

1. Find $Num(M_0[i_C(S, q)]) = n$.
2. Read the first n coordinates of $i_C(S, q)$, so that $w_0 = Cut\{n\}(i_C(S, q))$ is exactly known.
3. Conclude that $int_P(i_C(S, q)) = M'_0[w_0]$, where M'_0 is known.
4. Find $Num(M_1[i_C(S, q)]) = m$.
5. Read the first m coordinates of $i_C(S, q)$, so that $w_1 = Cut\{m\}(i_C(S, q))$ is exactly known.
6. Conclude that $int_{P'}(i_C(S, q)) = M'_1[w_1]$, where M'_1 is known.

We picture that when the above process is completed, the calculating system C guesses that $int_p(i_C(S, q)) = i^0_C(H_S, q)$, $int_{P'}(i_C(S, q)) = S(0)$, and bases its subsequent calculations on these assumptions. What we demand of C, then, is that *if these assumptions happen to be correct*, C shall always answer the question q about S correctly. For inputs such that step 1 or 4 does not halt, we imagine that C never gets past the preprocessing stage. No requirements are made on C in these cases.

Definition 12. A weak calculating system $C = (h_C, f_C, i^0_C, Out_C)$ with $f_C = \sigma_\eta$ is called *finitely describable via its own code* if there is a $P_C \in \Pi$ such that for all P, $P' \in \Pi$ and all recursive $q \in Q$ we have

$$int_{P_C}(Cut\{n\}(i^0_C(h_C(P, P'), q))) = i^0_C(h_C(P, P'), q)$$

where n is the cardinality of $Dom(\bar{\eta})$ [if $Dom(\bar{\eta})$ is infinite, then $n = \infty$].

Note that $Cut\{n\}(i^0_C(h_C(P, P'), q))$ is the only information about $i^0_C(h_C(P, P'), q)$ available in $i_C(S, q)$. The definition states that $i^0_C(h_C(P, P'), q)$ should be algorithmically reconstructable from this information alone. The question remains whether i^0_C codes $(h_C(P, P'), q)$ faithfully, i.e., whether complete information about $h_C(P, P')$ and q is contained in $i^0_C(h_C(P, P'), q)$. Definition 12 leaves this open. The important point is that the class of input codes we consider is sufficiently rich for such a faithful coding to be *possible*. This is certainly the case with the class of explicit input codes, since there $i^0_C(h_C(P, P'), q)$ is an *arbitrary* coding of $h_C(P, P')$ and q into a point in Ω. And if $Dom(\bar{\eta})$ is infinite, then all information about $i^0_C(h_C(P, P'), q)$ is contained in $i_C(S, q)$.

Notice also that a weak calculating system which is both finitely describable via its own code and faithfully described by i^0_C still can have a potentially infinite data storage capacity. Interpreting in terms of classical mechanics, it can, for instance, have an infinite number of degrees of freedom, with all the "storage" degrees of freedom being assigned a common, trivial initial condition at the start of each calculation.

Definition 13 (Weak universal systems). A *weak universal system* for $T = (\Gamma_T, H)$ is a weak calculating system C for T such that $D_C = \Gamma_T \times Q$.

Theorem 3 (Main result). Let $T = (\Gamma_T, H)$ be a theory, and let $\eta: A \to \mathbf{N}$ be a recursive initial sequence. Assume that C is a weak calculating system for T such that $f_C = \sigma_\eta$ and such that C is finitely describable via its own input code. Then:

(i) If T is closed, C is not universal.
(ii) If $Im(\eta)$ is finite or empty, C is not universal.

Proof. Assume that $T = (\Gamma_T, H)$ and $C = (h_C, \sigma_\eta, i_C^0, Out_C)$ is a counterexample. Let $M = \neg Out_C$ (the negation of Out_C); then $q_0 = (M, \infty) \in Q$. Since C is finitely describable via its own code and η is recursive, there is a $P_0 \in \Pi$ such that

$$int_{P_0}(\sigma_\eta(i_C^0(h_C(P, P'), q_0), w)) = i_C^0(h_C(P, P'), q_0)$$

for all $P, P' \in \Pi$. Further, there is a $P_1 \in \Pi$ such that

$$int_{P_1}(\sigma_\eta(i_C^0(h_C(P, P'), q_0), w)) = Frac_\eta^w(i_C^0(h_C(P, P'), q_0)))$$

Let $H_C = h_C(P_0, P_1)$. Pick $R \in \Gamma_T$, and construct the sequence $\{S^i\}_{i=1}^\infty \subseteq \Gamma_T$ by putting $S^1 = C[R, q_0, P_0, P_1]$ and

$$S^{i+1} = C[S^i, q_0, P_0, P_1]$$

for $i \in \mathbf{N}$. Then since $H(S^i) = h_C(P_0, P_1) = H_C$, we get

$$S^{i+1}(0) = \sigma_\eta(i_C^0(H(S^i), q_0), S^i(0)) = \sigma_\eta(i_C^0(H_C, q_0), S^i(0))$$

for $i \in \mathbf{N}$. As in the proof of Theorem 1, we obtain a system $S \in \Gamma_T$ with $H_S = H_C$ and $S(0) = Frac_\eta^z(a)$, where $a = i_C^0(H_C, q_0)$ and $z = S^1(0)$. Note that if $Im(\eta)$ is finite, the sequence $\{S^i\}_{i=1}^\infty$ is eventually constant. So the closure assumption on T is not necessary to ensure $S \in \Gamma_T$ in the case where $Im(\eta)$ is finite. We now get

$$\begin{aligned} i_C(S, q_0) &= \sigma_\eta(i_C^0(H_S, q_0), S(0)) \\ &= \sigma_\eta(a, Frac_\eta^z(a)) = Frac_\eta^z(a) = S(0) \end{aligned}$$

by Lemma 2, and so

$$C[S, q_0, P_0, P_1] = \phi_T^{-1}(H_S, i_C(S, q_0)) = S$$

But

$$\begin{aligned} int_{P_0}(i_C(S, q_0)) &= int_{P_0}(\sigma_\eta(i_C^0(H_S, q_0), S(0)) \\ &= i_C^0(H_S, q_0) \end{aligned} \tag{2}$$

Further,

$$\begin{aligned} int_{P_1}(i_C(S, q_0)) &= int_{P_1}(\sigma_\eta(i_C^0(H_S, q_0), S(0)) \\ &= Frac_\eta^{S(0)}(i_C^0(H_S, q_0)) \\ &= Frac_\eta^{S(0)}(a) \end{aligned}$$

Assume first $Dom(\bar{\eta}) \neq \emptyset$. Then from the explicit construction in the proof of Lemma 2 we get

$$[S(0)]^i = [Frac_\eta^z(a)]^i = z^i$$

for all $i < \bar{\eta}(1)$, so $Frac_\eta^{S(0)}(a) = Frac_\eta^z(a) = S(0)$. If $Dom(\bar{\eta}) = \emptyset$, then trivially $Frac_\eta^{S(0)}(a) = S(0)$ also. So in any case,

$$int_{P_1}(i_C(S, q_0)) = S(0) \tag{3}$$

Because of (2) and (3), we now get

$$\begin{aligned} M(S(\infty)) = 0 &\Leftrightarrow Out_C(C[S, q_0, P_0, P_1](\infty)) = 0 \\ &\Leftrightarrow M(C[S, q_0, P_0, P_1](\infty)) = 1 \\ &\Leftrightarrow M(S(\infty)) = 1 \end{aligned}$$

as in the proof of Theorem 1. ∎

Theorem 3 was proved by giving the calculating system C an input problem (S, q) which was directly related to its own ability to solve problems, thus creating self-reference. Interpreting physically, it is important to understand that seen from the inside of C, nothing special needs to be noted about the problematic input $i_C(S, q)$. It contains a finite amount of information which can be read off in C within a finite amount of time and serve as the basis for a calculation. The problem may be very complex, but the complexity is finite. Further, if i_C^0 conserves all information about $(h_C(P, P'), q)$ and (for instance) both $Dom(\eta)$ and $Dom(\bar{\eta})$ are infinite, then $i_C(S, q)$ describes the (S, q) problem *correctly*. And we know that (i) C is free to use *any* method in its calculations, (ii) C may have an infinite data storage capacity, and (iii) there is no time limit on the calculation. Thus there is, in this sense, a limit on the possibilities of making predictions within a deterministic theory. It is sometimes said (informally) that when the state of a system governed by a deterministic physical theory (for instance, classical mechanics) is known *exactly*, one can "in principle" answer any question about the future state of the system. Theorem 3 shows that in a certain sense, this is *not* the case.

10. SELF-REFERENCE IN CALCULATING SYSTEMS

In the proof of Theorem 3, we started from a given theory T together with a calculating system C for T, and proved (using some assumptions on C and T) the existence of two interpreting systems P_0, P_1 and a system $S \in \Gamma_T$ such that

$$\begin{aligned} &C[S, q_0, P_0, P_1] = S \\ &int_{P_0}(i_C(S, q_0)) = i_C^0(H_S, q_0) \qquad (4) \\ &int_{P_1}(i_C(S, q_0)) = S(0) \end{aligned}$$

where q_0 is the negation of the output signal Out_C. To investigate these matters more closely, I will now give (somewhat different) explicit constructions of such S, P_0, and P_1 in two special cases covered by the theorem.

First Case. Let $T = (\Gamma_T, H)$ be a deterministic theory, and assume that $C = (h_C, \sigma_\eta, i_C^0, Out_C)$ is a weak calculating system for T, where $\eta(j) = 2j$ for $j \in \mathbf{N}$. Then $\bar{\eta}(j) = 2j - 1$ for $j \in \mathbf{N}$, and both $Im(\eta)$ and $Im(\bar{\eta})$ are infinite. The sequence η is recursive. Let $M_C \in Tur$ be a counting machine such that $M_C[w]$ halts iff w is on the form

$$a^1a^1a^2a^2 \cdots a^m a^m 01 b^1 b^2 b^3 \ldots \qquad (5)$$

for some $m \in \mathbf{N}$, in which case $Num(M_C[w]) = 2m + 2$. Let $M'_C \in Tur$ be the identity, i.e., $M_C[w](\infty) = w$ for all $w \in \Omega^*$. Then $P_C = (M_C, M'_C)$ is an interpreting system.

Observation 1. For all $w \in \Omega$, $int_{P_C}(w) = w$ iff w is of the form

$$a^1a^1a^2a^2 \cdots a^m a^m 01000 \ldots \qquad (6)$$

Proof. Left to the reader. ■

Suppose $i_C^0(h_C(P, P'), q)$ is on the form (6) for all $P, P' \in \Pi$ and all recursive $q \in Q$. Then C is finitely describable via its own code, P_C being the witness system. Let $M_0 \in Tur$ be a counting machine such that for all $w \in \Omega$ the following condition is satisfied:

- If the point $c \in \Omega$ defined by $c^i = w^{\bar{\eta}(i)}$ for $i \in \mathbf{N}$ is of the form (5) for some $m \in \mathbf{N}$, then $M_0[w]$ halts with $Num(M_0[w]) = \bar{\eta}(2m + 2)$. If c is not of the form (5), then $M_0[w]$ does not halt.

Let $M'_0 \in Tur$ be such that for all $w \in \Omega$ we have $M'_0[w](\infty) = a$, where $a^i = w^{\bar{\eta}(i)}$for all $i \in \mathbf{N}$. Then $P_0 = (M_0, M'_0)$ is an interpreting system.

Observation 2. If $w \in \Omega$ is of the form (6) and $z \in \Omega$, then

$$int_{P_0}(\sigma_\eta(w, z)) = w$$

Proof. Left to the reader. ■

Let $M_1 \in Tur$ be a counting machine such that for all $w \in \Omega$ the following condition is satisfied:

- If the point $c \in \Omega$ defined by $c^i = w^{\eta(\bar{\eta}(i))}$ for $i \in \mathbf{N}$ is of the form (5) for some $m \in \mathbf{N}$, then $M_1[w]$ halts with $Num(M_1[w]) = \eta(\bar{\eta}(2m+2))$. If c is not of the form (5), then $M_1[w]$ does not halt.

Clearly such a machine exists. Next we use the following:

Observation 3. Let η be as before. Then there is an $M \in Tur$ such that for all $z, w \in \Omega$ we have

$$M[\sigma_\eta(z, w)](\infty) = Frac_\eta^w(a)$$

where $a^i = w^{\bar{\eta}(i)}$ for all $i \in \mathbf{N}$.

Proof. Left to the reader. ■

Let $M_1' \in Tur$ be a machine that satisfies the condition of Observation 3. Then $P_1 = (M_1, M_1')$ is an interpreting system. Note that the way P_1 is constructed now, it has the property that $int_{P_1}(\sigma_\eta(y, w))$ depends only on w. I will refer to this as the *one-eye* property of P_1.

Observation 4. Let $y, w \in \Omega$. If the point b given by $b^i = w^{\bar{\eta}(i)}$ is of the form (6), then

$$int_{P_1}(\sigma_\eta(y, w)) = Frac_\eta^w(b)$$

Proof. Left to the reader. ■

Define $H_C = h_C(P_0, P_1)$, and construct $S \in \Gamma_T$ as in the proof of Theorem 3. We have $i_C(S, q_0) = S(0)$, $H_S = H_C$, and $C[S, q_0, P_0, P_1] = S$, where q_0 is the negation of the output signal Out_C of C. Since C is finitely describable via P_C, the point $i_C^0(H_S, q_0)$ is of the form (6). So

$$\begin{aligned} int_{P_0}(i_C(S, q)) &= int_{P_0}(\sigma_\eta(i_C^0(H_S, q_0), S(0))) \\ &= i_C^0(H_S, q_0) \end{aligned}$$

by Observation 2. Let $a = i_C^0(H_C, q_0)$ and $z = S^1(0)$. By Lemma 2,

$$\begin{aligned} [\sigma_\eta(a, Frac_\eta^z(a))]^{\eta(\bar{\eta}(i))} &= [Frac_\eta^z(a)]^{\bar{\eta}(i)} \\ &= [\sigma_\eta(a, Frac_\eta^z(a))]^{\bar{\eta}(i)} = a^i \end{aligned}$$

so if $w = i_C(S, q)$, then the point c defined by $c^i = w^{\eta(\bar{\eta}(i))}$ is given by $c = i_C^0(H_C, q_0)$. Thus again since C is finitely describable via P_C, the point c is of the form (6). Using Observation 4,

$$\begin{aligned} int_{P_1}(i_C(S, q)) &= int_{P_1}(\sigma_\eta(i_C^0(H_S, q_0), S(0))) \\ &= int_{P_1}(\sigma_\eta(a, Frac_\eta^z(a))) \\ &= Frac_\eta^z(b) \end{aligned}$$

where we have

$$b^i = [Frac_\eta^z(a)]^{\bar{\eta}(i)} = a^i$$

i.e., $b = a$. So $int_{P_1}(i_C(S, q_0)) = S(0)$. In conclusion, we have constructed two interpreting systems $P_0, P_1 \in \Pi$ and a system $S \in \Gamma_T$ such that the three conditions in (4) hold.

Second Case. Let $T = (\Gamma_T, H)$ be a theory (not necessarily closed), and let $C = (h_C, \sigma_\eta, i_C^0, Out_C)$ be a weak calculating system for T, where we now assume $\eta = \emptyset$, i.e., $\bar{\eta}(j) = j$ for all $j \in \mathbf{N}$. Let $P_C = (M_C, M'_C)$ be as in the first case, and suppose C is finitely describable via P_C, i.e.,

$$int_{P_C}(i_C^0(h_C(P, P'), q)) = i_C^0(h_C(P, P'), q)$$

for all $P, P' \in \Pi$ and all recursive $q \in Q$. Notice that since $Im(\bar{\eta})$ is infinite, no *Cut* operation appears here. Let $P_0 = P_1 = P_C$, and let $H_C = h_C(P_0, P_1)$. Pick $R \in \Gamma_T$, and let $S^1 = C[R, q_0, P_0, P_1]$. Define $S = C[S^1, q_0, P_0, P_1]$. Then $H_S = H_C$, and

$$\begin{aligned} S(0) &= i_C(S^1, q_0) \\ &= \sigma_\eta(i_C^0(H(S^1), q_0), S^1(0)) = i_C^0(H_C, q_0) \end{aligned}$$

Also,

$$\begin{aligned} i_C(S, q_0) &= \sigma_\eta(i_C^0(H_S, q_0), S(0)) \\ &= i_C^0(H_C, q_0) = S(0) \end{aligned}$$

So $C[S, q_0, P_0, P_1] = S$, as before. We have

$$\begin{aligned} int_{P_0}(i_C(S, q_0)) &= int_{P_C}(i_C^0(H_C, q_0)) \\ &= i_C^0(H_C, q_0) = i_C^0(H_S, q_0) \end{aligned}$$

and, in the same manner,

$$int_{P_1}(i_C(S, q_0)) = i_C^0(H_C, q_0) = S(0).$$

Once again, we have constructed S, P_0, and P_1 such that (4) holds.

Discussion of the Two Cases. Case 2 is analogous to the construction employed in the standard proof of the Halting Theorem for Turing machines. It involves using interpreting systems P, P' such that the given initial value $i_C(S, q)$ is interpreted *both* as a coded description of H_S and q, *and* as giving $S(0)$ explicitly. Thus, thinking physically, the assumptions are:

- The system $C[S, q]$ can read off (high-level) coded information about H_S and q from its initial value $C[S, q](0)$.
- The system $C[S, q]$ can read off the individual bits of its initial value $C[S, q](0)$.

In case 1, the situation is different. Here high-level information about H_S and q is coded into a *part* of $C[S, q](0)$, and the initial value $S(0)$ is encoded in the complementary part. So the assumptions are:

- The system $C[S, q]$ can read off coded information about H_S and q from a subset of its initial value bits. We may picture this as a situation where high-level information about H_S and q is coded into a certain physical part of the calculating system. To take an extreme example, the information could be given as written text in a book contained in the system. However, by the one-eye property of P_1, the individual bits of the initial value corresponding to this part of the calculating system (the low-level description) need not be readable within the system.
- There exists an infinite subset of bits in $C[S, q](0)$ which can be read off individually within the system.

Thus, in a sense, the implicit assumptions about introspective abilities of the calculating system are weaker in case 1 than in case 2.

11. CONCLUDING REMARKS

It should be emphasized that although our proofs of Theorems 1 and 3 employ a self-reference diagonal trick, the unpredictability effects described by the theorems need not in general be tied to self-reference of any type. Information-theoretically, it seems quite natural that it is impossible for a calculating system to treat input systems of complexity comparable to (or greater than) its own. However, one should not be fooled by the intuition here. Remember that, for instance, a universal Turing machine *is* in a way able to treat itself. Chaitin (1982) discusses diagonal arguments versus information-theoretic approaches in the case of Gödel's theorem.

Anyway, if there is a system S which is unpredictable for C, then from an interpretational point of view we must conclude that physical systems which are *copies* of S are also C-unpredictable. It also seems, for instance, reasonable that *smaller* copies of S are C-unpredictable. This particular line of thought, I believe, is interesting, since it is conceivable that such smaller copies of S might be contained in $C[S, q, P, P'] = S$ itself. If we imagine a calculating system as an isolated "world," then in this self-similar case the world will contain small, unpredictable fractal subsystems. Further, in classical mechanics it is known that when a system is scaled down conformally, its time development tends to speed up. With effects of this type, it is even possible that the infinity of time involved in the self-referential problems we employed in our proofs can be converted into an infinity in scale, so that the behavior of the small subsystems is unpredictable in finite time as well. To describe this formally, we could add to our deterministic theories some extra structure providing analogs of the scaling (and geometrical union) properties found in classical mechanics. The reason why I find it interesting to speculate slightly at this point is that unpredictability effects of this type build a theoretical bridge from determinism to indeterminism. One obtains a picture where an underlying deterministic theory breaks down on the subjective level and appears *in principle* indeterministic in a given "world."

I would like to finish by stating three problems concerning generalization of the results in this paper.

1. The self-reference problem. Let C be a weak calculating system for T. Find necessary and/or sufficient conditions on C and T for the existence of $S \in \Gamma_T$ and $P, P' \in \Pi$ such that

$$C[S, q, P, P'] = S$$
$$int_P(i_C(S, q)) = i_C^0(H_S, q)$$
$$int_{P'}(i_C(S, q)) = S(0)$$

where $q = (\neg Out_C, \infty)$.

2. The negative universality problem. Find conditions α and β (as weak as possible) such that no deterministic theory T satisfying α has a weak universal system C satisfying β.

3. The positive universality problem. Find a condition α (as weak as possible) such that all deterministic theories T satisfying α have a weak universal system which is finitely describable via its own code.

Concerning problem 3, it can be remarked that it is not very difficult to construct an example of a nonclosed theory T together with a weak universal system C for T such that C has explicit input code and is finitely describable via its own code. The proof is omitted. Concerning problems 1 and 2, I have shown that the conditions on f_C assumed in Theorems 1 and 3 can be somewhat relaxed. For instance, the monotonicity assumption on the sequence η defining f_C can, under quite mild assumptions, be removed. The precise results are omitted here.

REFERENCES

Chaitin, G. J. (1982). *International Journal of Theoretical Physics*, **21**, 941–954.

Da Costa, N. C. A., and Doria, F. A. (1991). *International Journal of Theoretical Physics*, **30**, 1041–1073.

Davis, M. (1958). *Computability and Unsolvability*, McGraw-Hill, New York.

Fredkin, E., and Toffoli, T. (1982). *International Journal of Theoretical Physics*, **21**, 219–253.

Richardson, D. (1968). *Journal of Symbolic Logic*, **33**, 514–520.

Shoenfield, J. R. (1967). *Mathematical Logic*, Addison-Wesley, Reading, Massachusetts.

Tarski, A. (1956). *Logic, Semantics, Metamathematics*, Oxford University Press, Oxford.