integers are those sets of integers first order definable over (U,c). Then (V,c), but not its complete diagram, is Δ^0_2. Now axioms ix and x, which are finitely axiomatizable, can be verified in (V,c), using recursive saturation. Hence $Y_3+\varphi$ is consistent.

Lemma 5. *Every arithmetic sentence provable in Y_3 is provable in Y_4.*

Theorem. *Every arithmetic sentence provable in ALPO is provable in PA.*

In fact, we have given a proof of the Theorem within PA. Further considerations show that it is provable in primitive recursive arithmetic.

If induction for sets is added to ALPO, then we obtain a theory whose arithmetic theorems are the same as PA augmented with transfinite induction on all ordinals below the ϵ_0-th critical number (or $R(<\epsilon_0)$ of FRIEDMAN (1977)). If induction for all formulas (or induction for all formulas together with relativized dependent choice for all formulas) is added to ALPO, then we obtain a theory whose arithmetic theorems are the same as PA augmented with transfinite induction on all ordinals below the $\kappa(\epsilon_0)$-th critical number, where $\kappa(\epsilon_0)$ is the ϵ-th critical number (or $R(<\kappa(\epsilon))$).

References

FEFERMAN, S.

[1970] Ordinals and functionals in proof theory, *Proc. Int. Congr. Math., Nice 1970*, **1**, 229–233.

[1975] A language and axioms for explicit mathematics, in: *Lecture Notes in Mathematics* **450** (Springer, Berlin) pp. 87–139.

[1977] Theories of finite type related to mathematical practice, in: *Handbook of Mathematical Logic* (North-Holland, Amsterdam) pp. 913–971.

[1980] Forthcoming book of the Ω-group, in preparation.

FRIEDMAN, H.

[1977] Set theoretic foundations for constructive analysis, *Ann. of Math.*, **105**, 1–28.

TAKEUTI, G.

[1978] A conservative extension of Peano arithmetic, Part II of "Two applications of logic to mathematics," *Publ. Math. Soc. Japan*, **13**.

J. Barwise, H. J. Keisler and K. Kunen, eds., *The Kleene Symposium*

Church's Thesis and Principles for Mechanisms

Robin Gandy
Mathematical Institute, St. Giles, Oxford OX1 3LB, Great Britain

For Stephen Kleene

Beware! Beware!
His flashing eyes, his floating hair!
Weave a circle round him thrice,
And close your eyes in holy dread,
For he on honeydew hath fed,
And drunk the milk of paradise.

Abstract: After a brief review of Church's thesis and Gödel's objection to it, it is argued that Turing's analysis of computation by a human being does not apply directly to mechanical devices. A set-theoretic form of description for discrete deterministic machines is elaborated and four principles (or constraints) are enunciated, which, it is argued, any such machine must satisfy. The most important of these, called "the principle of local causality" rejects the possibility of instantaneous action at a distance. Although the principles are justified by an appeal to the geometry of space-time, the formulation is quite abstract, and can be applied to all kinds of automata and to algebraic systems. It is proved that if a device satisfies the principles then its successive states form a computable sequence. Counter-examples are constructed which show that if the principles be weakened in almost any way, then there will be devices which satisfy the weakened principles and which can calculate *any* number-theoretic function.

1. Introduction

Throughout this paper we shall use "calculable" to refer to some intuitively given notion and "computable" to mean "computable by a Turing machine"; of course many equivalent definitions of "computable" are now available.

Church's Thesis. *What is effectively calculable is computable.*

TURING, by this analysis of the process of calculation in his paper (1936) on computable numbers, gave cogent arguments in support of this thesis. Both Church and Turing had in mind calculation by an abstract human being using some mechanical aids (such as paper and pencil). The word

"abstract" indicates that the argument makes no appeal to the existence of practical limits on time and space. The word "effective" in the thesis serves to emphasize that the process of calculation is deterministic— not dependent on guesswork—and that it must terminate after a finite time.

Gödel has objected, against Turing's arguments, that the human mind may, by its grasp of abstract objects, be able to transcend mechanism. (This objection is stated briefly in footnote** on p. 72 of "The Undecidable" and at greater length on pp. 325–326 of WANG's "From Mathematics to Philosophy"; also in a typescript "Footnote*** to be added at the word 'mathematics' on p. 73, line 3 of: The Undecidable..."). Examples where we appear to use our insight and imaginative grasp to arrive at decisions in advance of any process of mechanical verification are well-known: we recognize that certain formal systems are consistent by imagining models for them; when we have gained a little familiarity with a system of ordinal notions we perceive that it is pre-well-ordered. The question is whether these examples are inspired guesswork or lucky accidents, or whether, as Gödel believed, they are the result of the workings of a non-mechanical intelligence. Gödel's objection can only be properly justified by a theory of intelligence. As he admits, our present understanding of the human mind is far from being penetrating enough for the construction of such a theory. For this purpose the knowledge provided by introspection, the history of ideas, experimental psychology, neurophysiology and artificial intelligence seems meagre indeed. One can only keep an open mind.

What we can say is that Turing outlined a proof of the following:

Theorem T. *What can be calculated by an abstract human being working in a routine way is computable.*

We shall return to Turing's proof and Gödel's objection at the end of this paper. Our chief purpose is to analyze mechanical processes and so to provide arguments for the following:

Thesis M. *What can be calculated by a machine is computable.*[1]

Although some of Turing's arguments can be applied indifferently to men or machines, there are crucial steps in Turing's analysis where he appeals to the fact that the calculation is being carried out by a human being. One such appeal is used to justify the assumption that the calculation proceeds as a sequence of elementary steps. A human being can only write one symbol at a time. But, if we abstract from practical limitations,

[1]I have been interested in the problem of how to justify this thesis for a long time. My earlier attempts were unsatisfactory. I owe to a conversation with Harvey Friedman the renewal of interest which led to the writing of this paper.

we can conceive of a machine which prints an arbitrary number of symbols simultaneously. (I owe to conversations with J. C. Shepherdson a realization that proofs of Thesis M must take parallel working into account. See SHEPHERDSON (1975) for a discussion of some of the problems this raises.) Turing's arguments do not suffice, nor do I think he would have claimed that they suffice, to justify Thesis M. The reader may (like several of those with whom I have talked) feel that, once Turing's ideas have been grasped, Thesis M is so unproblematic as to make arguments for it uninteresting or even unnecessary. This feeling has I think two sources. Firstly, actual machines which calculate fall under a narrow range of stereotypes to each of which Turing's arguments may rather easily be adapted. And the design of the most successful calculating machines—digital computers—was, at least in the early stages of their development, significantly influenced by Turing's ideas. But a slight effort of imagination will suggest devices which differ radically from the practical stereotypes. Conway's construction of a universal machine from his game of life is a good example.[2] There can be no guarantee that a further effort of imagination may not result in a device to which Turing's analysis is inapplicable.

The second source for a lack of interest in Thesis M is the belief that it can *only* function as a definition: if some imagined device can calculate what is not computable it is no machine. But I shall propose criteria for "being a machine" which, on the face of it, differ significantly from the criterion "works in a computable manner". At the very least, then, I hope to explain to the reader who believes that Thesis M is unproblematic the grounds for his belief.

For vividness I have so far used the fairly nebulous term "machine". Before going into details I must be *rather* more precise. Roughly speaking I am using the term with its nineteenth century meaning; the reader may like to imagine some glorious contraption of gleaming brass and polished mahogany, or he may choose to inspect the parts of Babbage's "Analytical Engine" which are preserved in the Science Museum at South Kensington.

(1) In the first place I exclude from consideration devices which are *essentially* analogue machines. In his paper "A notion of mechanistic theory" (1974) KREISEL discusses the ways in which physical theories (including even Newtonian theory) might give rise to non-computable functions. A more extreme possibility than those considered by Kreisel is the following: could some physical theory lead to a linear operator which has an infinite discrete spectrum and which is such that the multiplicity of

[2]An account of "the game of life" will be found in GARDNER (1970 and 1971). J. H. Conway, who invented the game, described how, with an appropriate initial configuration, it could be made to mimic the action of any Turing machine at the Logic Summer School held in Cambridge (U.K.) in 1971.

an eigenvalue is not a computable function of its place in the spectrum? So I shall distinguish between "mechanical devices" and "physical devices" and consider only the former.[3] The only physical presuppositions made about mechanical devices (Cf. Principle IV below) are that there is a lower bound on the linear dimensions of every atomic part of the device and that there is an upper bound (the velocity of light) on the speed of propagation of changes.

(2) Secondly we suppose that the progress of calculation by a mechanical device may be described in discrete terms, so that the devices considered are, in a loose sense, digital computers.

(3) Lastly we suppose that the device is deterministic; that is, the subsequent behaviour of the device is uniquely determined once a complete description of its initial state is given.

After these clarifications we can summarize our argument for a more definite version of Thesis M in the following way.

Thesis P. *A discrete deterministic mechanical device satisfies principles I–IV below.*

Theorem. *What can be calculated by a device satisfying principles I–IV is computable.*

The principles were arrived at by considering, schematically, examples that might at least in principle be realized by mechanical or electrical means. But (not too surprisingly in view of the generality aimed at) there are many abstract structures for which they also hold. Hence the principles and the theorem may be of interest even to those who are not concerned with Thesis M. Some examples are given in the concluding discussion.

Note on terminology. In the above statement we used "discrete deterministic mechanical device" to emphasize the somewhat restricted significance we are giving to the term "machine". Now that the point has been made we shall, for brevity, revert to the word "machine"; for the sake of variety, and for their flavor, we shall also sometimes use the words "device" and "mechanism" (for an object, not for a tenet).

2. The form of description

Since we are considering discretely acting machines, we may without loss of generality suppose that the action of a machine is described by

[3] POUR-EL (1974) investigates the computing power of a particular class of analog machines. But, in principle, any physical phenomenon might be used for analog computation.

describing the sequence $S_0, S_1, \ldots$, of its states; the input or initial arguments are encoded in S_0. It is of little importance whether we designate certain states as "halt" states encoding the output, or whether we consider, as Turing did, an infinite sequence which enumerates a (possibly empty) set whose members are encoded by certain of the S_i.

Our use of the term "discrete" presupposes that each state of the machine can be adequately described in finite terms. In order that we can apply any insights which we may have about mechanisms we want this description to reflect the actual, concrete, structure of the device in a given state. On the other hand, we want the form of description to be sufficiently abstract to apply uniformly to mechanical, electrical or merely notional devices. We have chosen to use hereditarily finite sets; other forms of description might be equally acceptable. We suppose that labels are chosen for the various parts of the machine—e.g., for the teeth of cog wheels, for a transistor and its electrodes, for the beads and wires of an abacus. Labels may also be used for positions in space (e.g., for squares of the tape of a Turing machine) and for physical attributes (e.g., the color of a bead, the state of a transistor, the symbol on a square). Starting from a potentially infinite set L of labels we form sets, sets of sets and so on; but we do not use the empty set in this construction, since it is an abstract object while our descriptions by sets may be of completely concrete structures. (We do, however, allow the empty structure.)

2.1 Definition.

$$\mathbf{P}_F(X) =_{\mathrm{Df}} \{Y : Y \subseteq X \wedge Y \neq \varnothing \wedge Y \text{ is finite}\}.$$

$$\mathrm{HF}_0 =_{\mathrm{Df}} \varnothing,$$

$$\mathrm{HF}_{n+1} =_{\mathrm{Df}} \mathbf{P}_F(L \cup \mathrm{HF}_n).$$

$$\mathrm{HF} =_{\mathrm{Df}} \bigcup \{\mathrm{HF}_n : n \in \omega\} \cup \{\varnothing\}.$$

The variables a, b, c, d (perhaps decorated with subscripts etc.) will always range over L, while s, t, u, v, w, x, y, z range over HF or designated subsets of HF. Note that $\varnothing \notin \mathrm{HF}_n$, $a \notin \mathrm{HF}$, and $\mathrm{HF}_n \subseteq \mathrm{HF}_{n+1}$.

2.2 It is convenient at this point to introduce various definitions and notations which will be used later. Since $\in$ is a well-founded relation on HF, we can use $\in$-recursion as a method of definition.

(1) $\quad \mathrm{Sup}\, x =_{\mathrm{Df}} \{a : a \in x \vee \exists y \in x.\, a \in \mathrm{Sup}\, y\}.$

$\mathrm{Sup}\, x$, the *support* of x is the set of labels which occur in the construction of x.

(2) $\quad |x| =_{\mathrm{Df}} \mathrm{Card}(\mathrm{Sup}\, x).$

$|x|$, the *size* of x is the number of labels occuring in x. It is to be distinguished from $\bar{\bar{x}}(=\mathrm{Card}(x))$

Let $A \subseteq L$; we define the *restriction* $x\restriction A$ of x to A by

(3) $$x\restriction A =_{\mathrm{Df}} (x \cap A) \cup \{y\restriction A : y \in x \wedge \mathrm{Sup}\, y \cap A \neq \emptyset\}.$$

It is the structure whose construction follows that of x omitting all labels not in A. The condition $\mathrm{Sup} y \cap A \neq \emptyset$ is necessary to ensure that $x\restriction A \in \mathrm{HF}$. If $\mathrm{Sup} x \cap A = \emptyset$, then $X\restriction A = \emptyset$.

The *transitive closure* of x, $\mathrm{TC}(x)$ is defined by

(4) $$\mathrm{TC}(x) =_{\mathrm{Df}} \bigcup \{\mathrm{TC}(y) : y \in x\} \cup (x \cap L).$$

2.3 As our terminology suggests, labels, like co-ordinates, are necessary for the description of concrete devices but do not by themselves have direct physical reference. A particular state of a machine corresponds not to a particular $x \in \mathrm{HF}$, but rather to the $\in$-isomorphism-type (which we shall call the *stereotype*) of x. All the information about that state which is relevant to the operation of the machine must be encoded in any structure x which is used to describe it. Relations and functions over HF which have concrete reference must be invariant under isomorphisms.

However, it is natural to suppose that if a label refers to a particular element of a mechanism in the state described by x, then it will refer to the same element in the next state; most things preserve their identity as time passes. In general the next state will also contain some new elements; for example, when a Turing machine moves left from the leftmost square so far used a new square must be created. A transition function F which determines the description Fx of the next state must specify new labels for the new elements, but no physical significance attaches to that specification. We now incorporate these ideas in a series of definitions.

(1) The variable π ranges over permutations of L. The effect of such a permutation on a structure x is defined in the obvious way:

$$a^{\pi} = \pi(a); \qquad x^{\pi} = \{y^{\pi} : y \in x\} \cup \{a^{\pi} : a \in x\}.$$

(2) Two structures x, y are *isomorphic over a set A of labels*, written "$x \simeq_A y$" just in case there is a permutation π which is the identity on A (i.e., $\forall a \in A.\pi(a) = a$) and which carries x into y (i.e., $x^{\pi} = y$). They are *isomorphic* ("$x \simeq y$") if they are isomorphic over the empty set. We shall write "$x \simeq_z y$" for "$x \simeq_{\mathrm{Sup} z} y$". Note that if $x \simeq_A y$, then $x\restriction A = y\restriction A$.

(3) A property P of structures and the corresponding class P' $(= \{x : P(x)\})$ are *structural* iff they are closed under isomorphism; i.e., if

$$P(x) \wedge x \simeq y. \rightarrow P(y).$$

(4) $X \subseteq \mathrm{HF}$ is a *stereotype* iff

$$\exists x. X = \{y : y \simeq x\}.$$

(5) A function $F : \mathrm{HF} \rightarrow \mathrm{HF}$ is *structural* iff for all π

$$(Fx)^{\pi} \simeq_{x^{\pi}} Fx^{\pi}.$$

This is stronger than the condition $x \simeq y \rightarrow Fx \simeq Fy$, but not as strong as requiring invariance (i.e., $(Fx)^{\pi} = Fx^{\pi}$). It expresses precisely the requirement that F describes the transition between physical states with some persistent elements.

2.4. *We can now state:*

Principle I ***The form of description***

Any machine M can be described by giving a structural set $S_M \subseteq \mathrm{HF}$ of state-descriptions together with a structural function $F : S_M \rightarrow S_M$. If $x_0 \in S_M$ describes an initial state, then $Fx_0, F(Fx_0), \ldots$, describe the subsequent states of M.

2.5. Examples. (1) The state of a Turing machine can be described by a structure of the form

$$\langle x, y, z, a_i, c_j \rangle$$

where x is the graph of the relation "a' is the label for the square standing immediately to the right of the square with label a", y is the graph of the relation "b is the label for the symbol printed on the square with label a", z codes the programme for the machine, a_i is the label for the scanned square and c_j is the label for the current instruction. Notice that the persistence of symbols (even those which do not occur on the tape at a given instant) is guaranteed by the occurrence of their labels in the (fixed) structure z.

(2) In describing an abacus one cannot treat the beads on a wire as an unstructured set, since, if one did, the label for a bead to be removed could not be structurally determined and the transition function would not be structural.

(3) An example which played an important part in our development of the theory is "The game of life" (GARDNER, 1970, 1971) or, more generally, the crystalline automata of VON NEUMANN (1966). Such an automaton consists of a (finite portion of a) rectangular lattice of cells, each of which may be in one of a fixed number of states (or, equivalently, may bear one of a fixed list of symbols). The next state of each cell is determined, by a fixed table, from its own state and that of its immediate neighbors. Initially only finitely many cells are in non-quiescent states, so that the state can

always be adequately specified by giving the states of the finitely many cells which have so far been brought into play. This example is important both because it involves parallel action (the symbols on all cells in play may change simultaneously) and because it raises problems about identifications between new labels (see 4.8).

Of course, one can also consider three-dimensional crystalline automata. In principle the state of any concrete device can be adequately described by specifying, to a certain degree of approximation, the relevant physical parameters (chemical composition, pressure, current flow and so on) of sufficiently small regions of space. (By using the words "concrete", "mechanical" and the like, we intend that these regions will always be very much larger than the size of an atom). This would not be a good way of describing most devices, but we shall appeal to the possibility of so doing when we come to justify Principle IV.

(4) For a given machine one can always arrange that each label in a fixed finite list is structurally distinguished from all others; for example in $x' = \langle x, a, b, c \rangle$ the labels a, b, c are distinguished. We shall use such *distinguished labels* in our examples without explicitly giving the (additional) structure which distinguishes them.

3. Conditions on S

The remaining three principles place certain restrictions on the set S of state-descriptions and on the transition function F. We shall show that if any of the principles be significantly weakened in (almost) any way then every function becomes calculable. More precisely, let α be an arbitrary predicate of natural numbers; we shall exhibit a machine (S, F) which satisfies the weakened condition and all the other conditions and which calculates α in some obvious sense. We put this picturesquely by saying that the machine displays free will. (Actually it is the class of machines satisfying the weakened conditions which displays free will.)

Each of the principles involves certain finiteness or boundedness conditions. We say that a quantity is *bounded* if there is a number k which may depend on the machine considered, but which does not depend on the state for which the quantity is being evaluated, such that in all states the value of the quantity is less than k.

Principle II ***The Principle of Limitation of Hierarchy***

The set-theoretic rank of the states is bounded. I.e.

$$\exists k . S \subseteq \mathrm{HF}_k.$$

It may well be natural or convenient to describe a machine or a method of storing information in hierarchical terms. The multiplier in a computer, or a series of parallel processors may be subordinate to the control unit; some kinds of data are treated as lists of lists. It is natural when describing such a hierarchical structure by a member of HF to think of a high level part as having its subordinates as members. The principle asserts that for a given machine the maximum height of its hierarchical structure must be bounded.

In describing real or imagined devices one tends to keep the overall rank low. I think that this corresponds to a real limitation of the human intellect. Anyone who has worked with the theory of types knows that above about type 4 one can only work formally, not conceptually; whereas there is no such difficulty in handling functions of large numbers of arguments. One gets over the conceptual block by thinking of a function of high type as operating on *names* for its arguments; in just the same way, a bibliography of bibliographies lists their titles in preference to transcribing their contents. But because of the block it is hard to think of devices which would make good use of objects of high rank.

Of course it is always possible to give a first-order description of HF: one treats the sets as objects rather than as structures. But it does not follow, as Counterexample 3.1 shows, that Principle II is vacuous. The first-order description requires the introduction of extra labels, and this, in conjunction with the other principles reduces the computational power.

3.1. Counterexample. Let $S = \{\iota^n(a) : a \in L,\ 1 \leqslant n\}$ (where $\iota(x) = \{x\}$), and let F be defined by

$$F(\iota^n(a)) = \begin{cases} \iota a & \text{if } \alpha(n), \\ \iota^2 a & \text{otherwise.} \end{cases}$$

It is trivial to verify that (S, F) satisfies the remaining principles; plainly this machine displays free will.

3.2. The next principle says, roughly, that any device can be assembled from parts of bounded size, and that these parts can be so labelled that there is a unique way of putting them together. Model construction-kits aim, not always successfully, to satisfy this principle. However, unlike construction-kits, we shall consider parts which overlap. Thus the tape of a Turing machine can be uniquely reassembled from the collection of *all* pairs of consecutive squares with their symbols; two such pairs are glued together with overlap if they both contain the same label for some square.

The proper formulation of this idea in terms of HF requires care. The simplest way (which I at one time thought was the correct way) would be to take as the parts of a structure x the restrictions $x \restriction A$ of x to subsets of

the support of x of bounded size. However, these parts are insufficiently structured to insure unique assembly.

3.3. Counterexample. Let

$$x=\{\langle a,b_1\rangle,\langle a,b_2\rangle,\ldots,\langle a,b_n\rangle\}.$$

Let $\langle a\rangle=\langle a,b_i\rangle\restriction\{a\}$; for any reasonable definition of ordered pair this will be independent of b_i. Let $y=x\cup\{\langle a\rangle\}$. It is easily seen that if $A\subseteq\mathrm{Sup}\,x$ and $\bar{\bar{A}}<n$, then $x\restriction A=y\restriction A$. Thus x cannot be uniquely assembled from parts of the form $x\restriction A$ with A of bounded size.

What has gone wrong here is that, since

$$\langle a,b_i\rangle\restriction\{a,b_j\}=\{\langle a\rangle\}\quad\text{for } i\neq j,$$

$x\restriction\{a,b_j\}$ contains not only the intended $\langle a,b_j\rangle$ but also the "floating" part $\langle a\rangle$. To put matters right we need to know the restrictions of x to *structured* parts (such as $\langle a,b_j\rangle$) not merely its restrictions to lists of labels.

3.4. Definition. Let $P\subseteq L\cup\mathrm{HF}$.

(i) The set $\mathrm{Part}(x,P)$ of parts of x from the list P is defined by

$$\mathrm{Part}(x,P)=\begin{cases}\{x\} & \text{if } x\in P,\\ \bigcup\{\mathrm{Part}(y,P):y\in x\}\cup(x\cap P\cap L) & \text{otherwise.}\end{cases}$$

(ii) The *restriction $x\restriction P$ of x to the list of parts P* is defined by

$$x\restriction P=\begin{cases}x & \text{if } x\in P,\\ \{y\restriction P:y\in x\wedge\mathrm{Part}(y,P)\neq\varnothing\}\cup(x\cap P\cap L) & \text{otherwise.}\end{cases}$$

(iii) The list P *covers* x iff $x\restriction P=x$; if in addition $P\subseteq\mathrm{TC}(\{x\})$, then P is a set of parts *for* x.

3.5. Remarks. (1) If $P\subseteq L$, then $\mathrm{Part}(x,P)=\mathrm{Sup}\,x\cap P$.

(2) P covers x iff every $\in$-chain $a\in x_n\in\cdots\in x_1\in x$ contains a member of P.

(3) If $P\subseteq Q$, then $x\restriction P=(x\restriction Q)\restriction P$. But this equation is not a consequence of $P\subseteq\mathrm{TC}(Q)$.

(4) $x\restriction P\cup Q$ is not, in general, determined by $x\restriction P$ and $x\restriction Q$ (see 3.7).

(5) The sets $\{x\},x,\mathrm{Sup}\,x$ are all sets of parts for x.

(6) The word "part" is used ambiguously to denote the stereotype, a particular set z belonging to this stereotype, and the located part $x\restriction\{z\}$. This is illustrated by: "I need a new sparking plug since the ones in the garage are all duds and something is wrong with the plug in the first cylinder".

3.6. Let P be a set of parts for x of bounded size; it is not to be expected that x will be uniquely determined by all the $x\restriction\{z\}(z\in P)$. But many structures can be uniquely reassembled from lists of bounded size of parts of bounded size (see Theorem 3.8). We shall refer to such lists (located or unlocated) as *sub-assemblies*.

Definition. Let $Q\subseteq\mathbf{P}_F(\mathrm{TC}(x))$. The structure x can be *uniquely reassembled* from the set Q of sub-assemblies iff x is the unique object y satisfying

(i) $y\in\mathrm{HF}$;
(ii) $\bigcup Q$ covers y;
(iii) $\forall s\in Q.\ x\restriction s=y\restriction s$.

Principle III ***The Principle of Unique Reassembly***

There is a bound q and for each $x\in S$ a set $Q\subseteq\mathbf{P}_F(\mathrm{TC}(x))$ from which x can be uniquely reassembled such that $|s|<q$ for each $s\in Q$.

Remarks. (1) If S satisfies II, then q determines a bound for the number of members of each $s\in Q$ as well as for its size. If II is not satisfied, then the cardinality of s might be unbounded.

(2) Observe that if x can be uniquely reassembled from Q, then $\bigcup Q$ covers x.

3.7. A good idea of what S must be like if it is to satisfy III can be gained from the following

Example. Let $S_n=\{x:x\subseteq\mathbf{P}_F(L)\wedge\bar{\bar{x}}=n\}$. Then $n<q$ is a necessary and sufficient condition for S_n to satisfy III with the bound q.

Proof. Since the members of $x\in S_n$ may be arbitrarily large, the only sub-assemblies we need consider are subsets of L of size q. Suppose $q\leqslant n$. Let $x=\{\{a,b_i\}:\ 1\leqslant i\leqslant n\}$, where $a,b_1,\cdots,b_n$ are distinct labels. Then for any $s\subseteq L$ with $\bar{\bar{s}}\leqslant q$,

$$x\restriction s=(x\cup\{\{a\}\})\restriction s,$$

and unique reassembly fails. To show the sufficiency of the condition, suppose that $x\in S_n$, $x\neq y$, $v\in y-x$ and $x=\{u_1,\ldots,u_n\}$.[1] For $1\leqslant i\leqslant n$ pick $a_i\in u_i/v$ (the symmetric difference of u_i and v) and let $b\in v$. Set

[1] See 'Notes added in proof' on p. 147.

$s=\{a_1,\ldots,a_n,b\}$. Then $v\restriction s\neq\varnothing$ and for each $1\leqslant i\leqslant n$, $u_i\restriction s\neq v\restriction s$. Thus

$$x\restriction s\neq y\restriction s$$

and the result follows by contraposition.

The proof of sufficiency may readily be generalized to other situations. One can replace L by some other list of parts, and one can consider structures which have sets like S_n in their transitive closures. In particular one readily proves:

3.8. Theorem. *Let σ be a finite similarity type of a first-order structure. Let S be the set of sets which code, in some natural way, first order-structures whose similarity type is σ and whose domain is a finite subset of L. Then S satisfies* III.

Thus almost any kind of automaton can be described in such a way that III is satisfied.

3.9. Counterexample. Let $A_n=\{a_1,\ldots,a_n\}$ be a set of n distinct labels, and e, 0 be distinguished labels. Let

$$y^{n,r}=\left\{v: v\subseteq A_n\wedge\bar{\bar{v}}=n-r\right\}\quad\text{for } 0\leqslant r<n;$$

$$z_\alpha^{n,r}=\begin{cases} y^{n,r} & \text{if } r=0, \text{ or if } r\neq 0 \text{ and } \alpha(n),\\ y^{n,r}\cup y^{n,r-1} & \text{if } 0<r<n \text{ and } \neg\alpha(n);\end{cases}$$

$$x_\alpha^{n,r}=\begin{cases} \{\langle a_1,a_2\rangle,\langle a_2,a_3\rangle,\ldots,\langle a_{n-r},e\rangle\}\cup z_\alpha^{n,r} & \text{if } 0\leqslant r<n,\\ \{e\} & \text{if } r\geqslant n \text{ and } \alpha(n),\\ \{0\} & \text{if } r\geqslant n \text{ and } \neg\alpha(n).\end{cases}$$

Let $S_\alpha=\bigcup\{\{y: y\simeq x_\alpha^{n,p}\}: n,p\in\omega\}$, and let F_α be defined by $F_\alpha x_\alpha^{n,p}=x_\alpha^{n,p+1}$. It is obvious that this is a machine displaying free will and which satisfies I and II. We shall verify (see Example 4.4) that it satisfies IV. III fails because, for example, if $s\subseteq A_n$ and $\bar{s}<n$, then

$$y^{n,1}\restriction s=(y^{n,1}\cup y^{n,0})\restriction s.$$

Further, if we weaken III by substituting "$y\in S$" for "$y\in\mathrm{HF}$" in the definition of unique reassembly, then S_α satisfies the weakened principle. For if $y\in S_\alpha$ and $\mathrm{Sup}\, y=\mathrm{Sup}\, x_\alpha^{n,p}$ and for each $1\leqslant i\leqslant n$

$$y\restriction\{e,a_i\}=x_\alpha^{n,p}\restriction\{e,a_i\},$$

then $y=x_\alpha^{n,p}$. One might say that the weakened principle is satisfied

because S_α is freely determined. It is therefore worth remarking that the principles do not require that S be computable. For example

$$S=\left\{y: y\subseteq L\wedge\alpha(\bar{\bar{y}})\right\}$$

satisfies I-III for any α.

4. Local causation

We now come to the most important of our principles. In Turing's analysis the requirement that the action depend only on a bounded portion of the record was based on a human limitation. We replace this by a physical limitation which we call the *principle of local causation*. Its justification lies in the finite velocity of propagation of effects and signals: contemporary physics rejects the possibility of instantaneous action at a distance.

Principle IV (Preliminary version). *The next state, Fx, of a machine can be reassembled from its restrictions to overlapping "regions" s and these restrictions are locally caused. That is, for each region s of Fx there is a "causal neighborhood" $t\subseteq\mathrm{TC}(x)$ of bounded size such that $Fx\restriction s$ depends only on $x\restriction t$.*

4.1. Complications in formulating this principle precisely arise from three sources.

(1) Since we wish our analysis to apply to abstract as well as to concrete devices, we do not wish to distinguish spatial structure from other structure.

(2) We have to be able to determine the causal neighborhoods of x without knowing in advance what are the regions s of Fx.

(3) If Fx has new labels these cannot be determined by $x\restriction t$ (see 2.3).

We cannot require that for *every* $s\subseteq\mathrm{TC}(Fx)$ of bounded size there be a causal neighborhood. For example, if $s=\{a\}$ where a is a particular cell-state of a crystalline automaton then whether or not $Fx\restriction\{a\}$ is empty depends globally, not locally, on x. Because of this and because of (2) above it turns out to be easiest first to decide what are causal neighborhoods.

4.2. In his analysis Turing writes "The new observed squares must be immediately recognizable by the computer". For us the natural analogue of this requirement is that the causal neighborhoods be structurally determined. Thus there must be a list T of stereotypes of bounded size; t is a

causal neighborhood of x if $t \subseteq \mathrm{TC}(x)$ and $t \in \bigcup T$. It is, however, very convenient to introduce a slight complication. We permit some of the stereotypes in T to be parts of others, and then we take as causal neighborhoods only those $t \subseteq \mathrm{TC}(x)$ which are included in no larger stereotype. For example for a Turing machine we need not specifically indicate which is the leftmost square of the tape; if a causal neighborhood contains $\langle a_i, a_{i+1}\rangle$, then a_i cannot be the leftmost square and we do not get another causal neighborhood by willfully omitting $\langle a_i, a_{i+1}\rangle$. For further examples see Example 4.4 below.

4.2.1. Definition. (i) We say that t' *subsumes* t, and write $t \preccurlyeq t'$, iff $t \subseteq \mathrm{TC}(t')$.

(ii) The set $\mathrm{CN}_1(x, T)$ of causal neighborhoods of x determined by the set T of stereotypes is defined by

$$\mathrm{CN}_1(x,T) =_{\mathrm{Df}} \{t : t \subseteq \mathrm{TC}(x) \wedge t \in \bigcup T \wedge \forall t' \subseteq \mathrm{TC}(x).\ t' \in \bigcup T \wedge t \preccurlyeq t' \rightarrow t \preccurlyeq t'\}.$$

Since T is fixed for a given device we shall usually omit references to it.

4.3. We first consider the case $\mathrm{Sup}\, Fx \subseteq \mathrm{Sup}\, x$, that is, no new labels are introduced. The effect of the located causal neighborhood $x \restriction t$ will be given by $G(x \restriction t)$ for some structural function G; this effect, which we call a *determined region*, is $Fx \restriction s$ for some $s \subseteq \mathrm{TC}(Fx)$. Because we are supposing $\mathrm{Sup}\, Fx \subseteq \mathrm{Sup}\, x$ we must require

(1) $\quad \mathrm{Sup}\, G(x \restriction t) \subseteq \mathrm{Sup}\, t \quad$ for all $t \in \mathrm{CN}_1(x)$;

that is a causal neighborhood must include the region it affects. Typically $Fx \restriction s$ will describe the state of some bounded region V space, and $x \restriction t$ will describe the state at the previous instant of the region consisting of all points (or cells) within a distance ct of V, where c is the velocity of light and t is the time between instants. At this point it is convenient to introduce a bit of notation:

$$u \subseteq^* y \leftrightarrow_{\mathrm{Df}} \exists s \subseteq \mathrm{TC}(y).\ u = y \restriction s.$$

(If one considers y as a tree of its $\in$-chains, then $u \subseteq^* y$ implies that u is a subtree with the same vertex as y). Now let

$$D = \{v : v \subseteq^* Fx \wedge \exists t \in \mathrm{CN}_1(x).\ v = G(x \restriction t)\};$$

this is the set of determined regions. It might happen that not every causal neighborhood determined a region in D. But if that were allowed, we could construct a machine displaying free will by deciding arbitrarily which of two causal neighborhoods was to take effect (see Counterexample 4.9.1).

Hence we require

(2) $\quad \forall t \in \mathrm{CN}_1(x).\ \exists v \in D.\ v = G(x \restriction t).$

This can be expressed picturesquely as "every cause has an effect".

Finally we require that Fx can be reassembled (perhaps uniquely) from D: every region of Fx must have a cause. D consists of *located* subassemblies, and s is not uniquely determined by $Fx \restriction s$. We might have required that G determines s as well as $Fx \restriction s$, but that is not necessary. Instead we modify Definitions 3.4(iii) and 3.6.

4.3.1. Definition. (i) A set C of located subassemblies *covers* x if

$$\exists Q.\ \bigcup Q \text{ covers } x \wedge C = \{x \restriction s : s \in Q\}.$$

(ii) Let $C \subseteq \{v : v \subseteq^* x\}$; then x can be *uniquely reassembled* from C iff

$$\forall y.\ [C \text{ covers } y \wedge \forall v \in C.\ v \subseteq^* y] \rightarrow y = x.$$

Then our last requirement for the case $\mathrm{Sup}\, Fx \subseteq \mathrm{Sup}\, x$ is

(3) $\quad D$ covers Fx, and if III is satisfied, then Fx can be uniquely reassembled from D.

4.4. Example. We show that the transition function of Counterexample 3.9 satisfies 4.3(1)–(3). First we give the stereotypes for the causal neighborhoods

$$t_1^i = \{a_i'\}, \quad t_2^i = \{\langle a_i, a_{i+1}\rangle\}, \quad t_3^i = \{\langle a_i, a_{i+1}\rangle, \langle a_{i+1}, e\rangle\},$$
$$t_4^i = \{\langle a_i, e\rangle\}, \quad t_5^{i,j} = \{\langle a_i, e\rangle, \{a_i'\}, \{a_i', a_j'\}\}, \quad t_6 = \{0\}, \quad t_7 = \{e\},$$

where $a_i' = \{0, a_i\}$ and in the definition of $y^{n,r} A_n$ is to be replaced by $\{a_i' : 1 \leqslant i \leqslant n\}$. But $t_2^i \preccurlyeq t_3^i, t_4^i \preccurlyeq t_5^{i,j}, t_6^i \preccurlyeq t_1^i, t_7 \preccurlyeq t_3^i, t_7 \preccurlyeq t_4^i, t_7 \preccurlyeq t_5^{i,j}, t_1^i \preccurlyeq t_5^{i,j}, t_1^j \preccurlyeq t_5^{i,j}, t_4^{i+1} \preccurlyeq t_3^i, t_6 \preccurlyeq t_5^{i,j}$. From these we compute:

$$\mathrm{CN}_1(x_\alpha^{n,r}) = \{t_1^i : 1 \leqslant i \leqslant n\} \cup \{t_2^i : 1 \leqslant i < n-r-1\} \cup \{t_3^{n-r-1}\} \quad \text{if } r < n-1;$$

$$\mathrm{CN}_1(x_\alpha^{n,n-1}) = \begin{cases} \{t_4^1\} & \text{if } \alpha(n), \\ \{t_5^{1,j} : 1 \leqslant j \leqslant n\} & \text{otherwise;} \end{cases}$$

$$\mathrm{CN}_1(x_\alpha^{n,r}) = \begin{cases} \{t_6^1\} & \text{if } \alpha(n) \text{ and } r \geqslant n, \\ \{t_7^1\} & \text{if } \neg\alpha(n) \text{ and } r \geqslant n. \end{cases}$$

Then define $G(x \restriction t_3^i) = \{\langle a_i, e\rangle\}$, $G(x \restriction t_4^i) = \{0\}$, $G(x \restriction t_5^{i,j}) = \{e\}$ and $G(x \restriction t_k^i) = x \restriction t_k^i$ in all other cases. It is straightforward to verify that conditions

(1)–(3) of 4.3 are satisfied. Note that in the case of t_5 a number of different causal neighborhoods all have the same effect.

4.5. We now turn to the case $\mathrm{Sup}\,Fx \not\subseteq \mathrm{Sup}\,x$. The ground was prepared in 2.3. We cannot require $Fx\restriction s = G(x\restriction t)$ but only $Fx\restriction s \simeq_x G(x\restriction t)$; and even this is too strong since the new labels ($\notin \mathrm{Sup}\,t$) in $G(x\restriction t)$ might by accident belong to $\mathrm{Sup}\,x$. And we want the new labels in $G(x\restriction t)$ to correspond to new labels in Fx; that is we require $\mathrm{Sup}\,s \cap \mathrm{Sup}\,x \subseteq \mathrm{Sup}\,t$. This condition includes 4.3(1) as a special case. We are thus led to the following definition of the determined regions of Fx (cf. the definition of D in 4.3).

4.5.1. Definition. The set $\mathrm{DR}_1(Fx,x)$ of determined regions of Fx is given by

$$\mathrm{DR}_1(Fx,x) =_{\mathrm{Df}} \{v : v \subseteq^* Fx \wedge \exists t \in \mathrm{CN}_1(x).\ v \simeq_t G_1(x\restriction t) \wedge \mathrm{Sup}\,v \cap \mathrm{Sup}\,x \subseteq \mathrm{Sup}\,t\}.$$

Corresponding to 4.3(2) we give the following formulation of "every cause has an effect".

4.5.2. $\forall t \in \mathrm{CN}_1(x).\ \exists v \in \mathrm{DR}_1(Fx,x).\ v \simeq_t G_1(x\restriction t)$. Notice that, in view of the last clause of the definition of DR_1, 4.5.2 implies the apparently stronger version obtained by replacing "$\simeq_t$" by "$\simeq_x$" provided the function G_1 satisfies $\mathrm{Sup}\,G_1(x\restriction t) \cap \mathrm{Sup}\,x \subseteq \mathrm{Sup}\,t$. In the future we shall suppose, without explicit mention, that G_1 satisfies this proviso.

4.6. Example. *A simple meitotic machine.* For any x let $\pi[x]$ be a permutation of L satisfying $\mathrm{Sup}\,x \cap \mathrm{Sup}\,x^{\pi[x]} = \varnothing$. Let Dx, the duplicate of x be defined by

$$Dx = x \cup x^{\pi[x]}.$$

Fix some y_1; set

$$S = \{x : \exists n.\ x \simeq D^n(\{y_1\})\}, \qquad T = \{t : t \simeq \{y_1\}\},$$

$$Gu = Du.$$

Then if we take $F = D$, F satisfies 4.5.2. For if $x \in S$, then $x = \{y_1, y_2, \ldots, y_{2^n}\}$ where each y_i is a distinct isomorph of y, and $\mathrm{CN}_1(x,T) = \{\{y_i\} : y_i \in x\}$, and $x\restriction\{y_i\} = \{y_i\}$ and $G(x\restriction\{y_i\}) = \{y_i, y'\}$ where y' is a distinct isomorph of y_i.

A comparison of this extremely simple self-reproducing device with von Neumann's complex 29-state crystalline self-reproducing automaton shows both the power and the limitation of our abstract approach. Unlike, say, the action of a Turing machine, the copying process is done in one stroke.

But the mechanism gives no guidance for the construction of a concrete self-reproductive machine situated in space.

4.7. The example also illustrates an inadequacy of our formulation thus far. If we took, for example, F to be D^2, then the same S, T, G as above could still be used; 4.5.2 would still be satisfied, and $\mathrm{DR}_1(Fx,x)$ would still provide an assembly for x. The definition of DR_1 does not put any bounds on the number of distinct regions which arise from a given causal neighborhood. This lack of determination allows free will to be displayed (see Counterexample 4.9.2). To prevent this we require, roughly speaking, that every cause have a unique effect. More precisely:

4.7.1.

$$\forall v \in \mathrm{DR}_1(Fx,x).\ \forall s \subseteq \mathrm{TC}(Fx).\ Fx\restriction s \simeq_x v \to Fx\restriction s = v.$$

This requirement could be met, for example, by a meitotic machine if instead of building the daughters of x from new labels, we built them from, say, pairs $\langle a, x\rangle$ so that each daughter cell showed, so to speak, its ancestry within itself. For concrete devices the requirement is met naturally. New labels refer to new regions of space or to new states and these have a distinctive geometrical or structural connection to parts of the causal neighborhood which gave rise to them. All that is required is that this connection be encoded in the description Fx.

4.8. We now come to the last, and most complex, condition. In general new determined regions may overlap. For example, a crystalline automaton may introduce unboundedly many new cells at a given instant. The geometric relations (in particular the neighborhood relation) between these new cells must be causally determined; since the determined regions will be of bounded size, there must be overlaps between them. It is obvious (see Counterexample 4.9.3) that such overlaps must be determined, or else free will will be displayed. Further this determination must be local. It would, I suppose, be possible to specify exactly which new labels of the regions corresponding, say, to $G_1(x\restriction t_1), G_1(x\restriction t_2)$ should be identified. But it seems easier to determine, to within isomorphism over x, $Fx\restriction s$ for a region s which includes $s_1, \ldots, s_r$ where $Fx\restriction s_1, \ldots, Fx\restriction s_r$ are overlapping determined regions. $Fx\restriction s$ is to be determined by an appropriate causal neighborhood. Thus there must be a list T_2 of stereotypes from which the set $\mathrm{CN}_2(x) = \mathrm{CN}(x, T_2)$ is got by Definition 4.2.1, and a new structural function G_2 such that for suitable r the following condition is satisfied.

4.8.1$_r$.

$$\forall V \subseteq \mathrm{DR}_1(Fx,x).\ \left[\overline{\overline{V}} < r \wedge \bigcap \{\mathrm{Sup}\,v : v \in V\} \not\subseteq \mathrm{Sup}\,x\right] \to \exists v \in \mathrm{DR}_2(Fx,x).\ \forall v_1 \in V.\ v_1 \subseteq^* \bar{\bar{v}}.$$

Remarks. (1) Here DR_2 is got from $CN_2(x)$ and G_2 as in 4.5.1., i.e.

$$DR_2(Fx,x) =_{Df} \{v: v \subseteq^* Fx \wedge \exists t \in CN_2(x).\ v \simeq_t G_2(x \restriction t) \wedge \mathrm{Sup}\, v \cap \mathrm{Sup}\, t \subseteq \mathrm{Sup}\, t\}.$$

(2) Counterexample 4.9.4 shows that it will not suffice to take $r=2$. In Lemma 5.1 we shall show that an appropriate value for r is one greater than the maximum number of new labels in any determined region.

(3) By substituting $\{v_1\}$ for V in the condition we see that T_2 and G_2 must in effect include T_1 and G_1. However, since we need to determine the overlaps between members of DR_1, but not between those of DR_2, and since moreover we do not require the uniqueness condition 4.7.1 for DR_2, we still need to single out T_1 and G_1.

(4) A weaker form of $4.8.1_r$ could be got by requiring V to range only over subsets of a subset of DR_1 from which Fx could be (uniquely) reassembled. In the absence of III this is too weak. I do not know if the weaker form would suffice to avoid free will in the case of unique reassembly.

Exactly as in 4.5 we need to insist that every cause has an effect; we simply take 4.5.2 with subscript 1 replaced by 2.

The problems of identifying labels in new regions correspond to real problems in real life. Embryologists investigate what causes sheets of tissue to join up, and how migrating groups of cells recognize that they have reached their destination; the problem is to discover local causes (e.g., a gradient of concentration of some substance) for these phenomena. For crystalline automata 4.8.1 is not automatically satisfied. Consider, for example, two adjacent perpendicular blocks of cells already in play forming, as it were, two sides of a rectangular courtyard. Suppose the extremities of the blocks start to "grow" new cells so as to make the other two sides of the courtyard. When these sides meet identifications will have to be made, but these cannot be locally determined. Free will could be displayed; if one chose not to make the identification, the automaton would be growing on a Riemann surface rather than on the plane. One way of ensuring local determination would be to supplement the automaton with a computer which kept track of the coordinates of all cells in play; at each step the computer would output instructions (part of the bounded $t \in CN_2$) about identifications to be made. This technique was used by Eupalinus in the sixth century B.C. so as to ensure that the excavations of a tunnel from both sides of Mount Castro on Samos should meet in the middle (see VAN DER WAERDEN (1954)).

Another way around the difficulty is to arrange that the cells in play of a crystalline automaton always form a convex set. Suppose two cells Q, R in play both require the bringing in to play of a new cell at P. The distances

PQ and PR, and hence also QR are bounded. And since the cells already in play form a convex set, the description x will include a neighborhood of bounded size which determines the spatial relations between Q and R, and hence can be used to prescribe the necessary identification at P. Applying this idea in three dimensions we see that 4.8.1 can be justified for concrete machines (see 2.5(3)).

4.9. We sum up all the requirements we have made:

Principle IV ***The Principle of Local Causality***

There are sets T_1, T_2 of stereotypes of bounded size and structural functions G_1, G_2 such that the conditions below are satisfied. $CN_k = CN(x, T_k)$ *and* $DR_k = DR_k(Fx, x)$ *are given, for* $k = 1, 2$, *by definitions* 4.2.1 *and* 4.5.1.

(1) $\forall t \in CN_k.\ \exists v \in DR_k.\ v \simeq_t G_k(x \restriction t)$ for $k = 1, 2$.

(2) $\forall v \in DR_1.\ \forall v' \subseteq^* Fx.\ v' \simeq_x v \rightarrow v' = v$.

For each r

$(3)_r$ $\forall V \subseteq DR_1.\ \left(\overline{\overline{V}} \leqslant r \wedge \cap \{\mathrm{Sup}\, v: v \in V\} \not\subseteq \mathrm{Sup}\, x\right) \rightarrow \exists v \in DR_2.\ \forall v_1 \in V.\ v_1 \subseteq^* v$.

(4) DR_1 covers Fx, and if principle III is satisfied, then Fx can be uniquely reassembled from DR_1.

Counterexamples. We hope that by now the interested reader will be able to flesh out our rather schematic treatment. In particular we shall not specify S nor T_1 and T_2, but rather give x, Fx, $CN_1(x)$ etc. explicitly; it should be obvious that they correspond to structural properties and functions. We use n^* to stand for some structure which encodes the number n, and $0, 1, 2$ for distinguished labels.

4.9.1. (Necessity of IV(1) for $k=1$).
Let $x = \{n^*, 0, 1\}$,

$$Fx = \begin{cases} \{0\} & \text{if } \alpha(n), \\ \{1\} & \text{otherwise.} \end{cases}$$

Take $CN_1(x) = \{\{0\}, \{1\}\}$ and G_1 the identity function. Evidently II, III and IV(2), (4) are satisfied.

4.9.2. (Necessity of IV(2)).
Let $x_0=\{n^*,0,1\}$. If $\alpha(n)$, let

$$Fx_0=\{0,1,\{a\}\}\ (=y, \text{ say}) \quad \text{and} \quad F(Fx_0)=\{0\};$$

while, if $\neg\alpha(n)$,

$$Fx_0=\{0,1,\{a\},\{b\}\}\ (=z, \text{ say}) \quad \text{and} \quad F(Fx_0)=\{1\}.$$

Let $t_1=\{0,1\}, t_2=y, t_3=z$; so $t_1\leqslant t_2\leqslant t_3$. Let $G(\{0,1\})=\{0,1,c\}, G(y)=\{0\}, G(z)=\{1\}$. Then II, III, IV(1), (4) are satisfied. The conditions on CN_2 and G_2 are not required as there is no overlapping. The second application of F, and t_2, t_3 were used only to reduce the two non-isomorphic values of Fx_0 to some standard output. In the remaining counterexamples we omit this step.

4.9.3. (Necessity of IV(1) for $k=2$).
Let $x_0=\{n^*,0,1\}$; let

$$Fx_0=\begin{cases}\{\{0,a\},\{1,a\}\}\ (=y \text{ say}) & \text{if } \alpha(n),\\ \{\{0,a\},\{1,b\}\}\ (=z \text{ say}) & \text{if } \neg\alpha(n).\end{cases}$$

Let $CN_1(x_0)=\{\{0\},\{1\}\}$, $G_1(\{0\})=\{\{0,c\}\}$, $G_1(\{1\})=\{\{1,d\}\}$.
Let $CN_2(x_0)=\{\{0\}\}$ and $G_2(\{0\})=\{\{0,c\},\{1,c\}\}$.
The overlap causal neighborhood takes effect only when $\alpha(n)$ is true. II, III, IV(1) (for $k=1$), (2), (3) and (4) are satisfied.

4.9.4. (Necessity of IV(3)$_r$ with $r>2$).
Let $x_0=\{n^*,0,1,2\}$, $p_0=\{a_1,a_2\}$, $p_1=\{a_0,a_2\}$,

$$p_2=\begin{cases}\{a_0,a_1\} & \text{if } \alpha(n),\\ \{a_2,a_3\} & \text{otherwise.}\end{cases}$$

Let

$$Fx_0=\{0,1,2,p_0,\varphi_1,\iota^2p_2\}.$$

(Fx_0 describes a triangle with vertices a_0,a_1,a_2 if $\alpha(n)$ and a star with center a_2 otherwise).

Take $CN_1(x_0)=\{\{0\},\{1\},\{2\}\}$ and $CN_2(x_0)=\{\{0,1\},\{1,2\},\{2,0\}\}$. Let

$$G_1(\{k\})=\{k,\iota^k\{b,c\}\} \quad (k<3),$$

and

$$G_2(\{j,k\})=\{j,k,\iota^j\{b,c\},\iota^k\{c,d\}\} \quad (j\neq k, j,k<3).$$

One can compute that

$$DR_1=\{\{k,\iota^kp_k\}:k<3\},$$
$$DR_2=\{\{j,k,\iota^jp_j,\iota^kp_k\}:j,k<3, j\neq k\}.$$

All the principles are satisfied if in IV(3) r is given the value 2.

5. The Main Theorem

First we prove a key lemma which is a consequence of IV alone.

5.1. The Key Lemma. *Let S,T_1,T_2,G_1,G_2 be given and let G_1 satisfy*

$$\forall u.\ \mathrm{Card}(\mathrm{Sup}\,G_1u-\mathrm{Sup}\,u)\leqslant r.$$

Suppose that F,F' satisfy IV (1), (2), $(3)_{r+1}$, (4); *then*

$$DR_1(Fx,x)\simeq_x DR_1(F'x,x).$$

We first prove several subsidiary lemmas. Throughout the proofs, the placing of a prime on any expression indicates the result of replacing Fx by $F'x$ and priming all other introduced constants. In particular $DR_1'=DR_1(F'x,x)$.

5.2. Lemma. *Let $DR_1=\{v_i: i<m\}$; then $DR_1'=\{v_i': i<m\}$ where v_i' is the unique $v'\subseteq^*F'x$ which is isomorphic over x to v_i.*

Proof. By the definition of DR_1, given v_i there is some $t_i\in CN_1$ such that $v_i\simeq_x G_1(x\restriction t_i)$. By IV(1) ($k=1$) and (2) there is a unique $v'\subseteq^*F'x$ such that $v'\in DR_1'$ and $v'\simeq_x v_i$. This argument is symmetrical between DR_1 and DR_1' so the result follows.

5.3. Lemma. *Let r and m be as in Lemmas* 5.1 *and* 5.2. *If $K\subseteq\{0,1,\ldots,m-1\}$, $\overline{\overline{K}}\leqslant r+1$ and*

$$\bigcap\{\mathrm{Sup}\,v_i: i\in K\}\not\subseteq\mathrm{Sup}\,x,$$

then there is a permutation π, which is the identity on x, such that $v_i'=v_i^\pi$ for $i\in K$.

Proof. By IV(3) there is a $v\in DR_2$ with $v_i\subseteq^*v$ ($i\in K$). By IV(1) (with $k=2$) there is a $v'\simeq_x v$. Let $v'=v^\pi$. Then $v_i^\pi\subseteq^*v'$, since $\subseteq^*$ is a structural relation. But then, by Lemma 5.2, $v_i^\pi=v_i'$.

We define $\mu(a)=\{i: a\in \mathrm{Sup}\, v_i\}$ and call it the *signature* of a. We write A for $\mathrm{Sup}\, Fx - \mathrm{Sup}\, x$.

5.4. Lemma. *For any* $a_0\in A$, $b_0\in A'$,
(i) $\mathrm{Card}\{a\in A: \mu(a_0)\subseteq\mu(a)\}\leqslant \mathrm{Card}\{b\in A': \mu(a_0)\subseteq\mu'(b)\}$;
(ii) $\mathrm{Card}\{b\in A': \mu'(b_0)\subseteq\mu'(b)\}\leqslant \mathrm{Card}\{a\in A: \mu'(b_0)\subseteq\mu(a)\}$.

Proof. The slight tiresomeness of this lemma and its proof arises from the fact that DR_2 does not tell us which v_i's do *not* overlap. By symmetry it is sufficient to prove (i).

Since, by IV(4), DR_1 covers $Fx, \mu(a_0)\neq\emptyset$; for simplicity, suppose $a_0\in v_0$ so that $0\in\mu(a_0)$. For each $b\in \mathrm{Sup}\, v_0'\cap A'$ such that $\mu(a_0)\not\subseteq\mu'(b)$ let i_b be the least number such that $i_b\in\mu(a_0)-\mu'(b)$, and let $I=\{i_b: b\in \mathrm{Sup}\, v_0'\cap A'\}$. Now the condition on r implies that $\mathrm{Card}(\mathrm{Sup}\, v_0'\cap A')\leqslant r$ and so $\bar{I}\leqslant r$. Hence by Lemma 5.3, (taking $K=I\cup\{0\}$ and observing that $a_0\in\mathrm{Sup}\, v_i$ for $i\in K$) there is a π, the identity on $\mathrm{Sup}\, x$, with $v_i^\pi=v_i'$ for $i\in K$.

We claim that if $a\in\mathrm{Sup}\, v_0\cap A$ and $\mu(a_0)\subseteq\mu(a)$, then $\mu(a_0)\subseteq\mu'(a^\pi)$. For if not, since $a^\pi\in\mathrm{Sup}\, v_0^\pi=\mathrm{Sup}\, v_0'$, there is an $i\in K$ such that $i\notin\mu'(a^\pi)$ and $i\in\mu(a_0)$. Thus $a^\pi\notin\mathrm{Sup}\, v_i'=\mathrm{Sup}\, v_i^\pi$, and so $a\notin\mathrm{Sup}\, v_i$ which contradicts $i\in\mu(a_0)\subseteq\mu(a)$. Thus

$$\{a\in A: \mu(a_0)\subseteq\mu(a)\}^\pi\subseteq\{b\in A': \mu(a_0)\subseteq\mu'(b)\}.$$

5.5. Lemma. *For all* $K\subseteq\{0,1,\ldots,m-1\}$

$$\mathrm{Card}\{a\in A: \mu(a)=K\}=\mathrm{Card}\{b\in A': \mu'(b)=K\}.$$

Proof. Let ν_K, ν_K' stand for the LHS and RHS of the equation. Then

$$\nu_K=\mathrm{Card}\{a\in A: K\subseteq\mu(a)\}-\sum\{\nu_{K'}: K\subset K'\}.$$

The result follows readily from Lemma 5.4 by downward induction on $\bar{\bar{K}}$.

Proof of 5.1. μ partitions A into disjoint sets $\{a\in A: \mu(A)=K\}$ for those K with $\nu_K\neq 0$. By Lemma 5.5 there is an exactly corresponding partition of A'. Hence we can define a permutation π which is the identity on x, such that

$$\mu'(a^\pi)=\mu(a)\quad\text{for all } a\in A.$$

But then $a\in\mathrm{Sup}\, v_i\leftrightarrow a^\pi\in\mathrm{Sup}\, v_i'$, for all $i<m$. So $v_i'=v_i^\pi$ and hence $\mathrm{DR}_1^\pi=\mathrm{DR}_1'$.

5.6. Corollary. *If* S *satisfies* III, *then the conclusion of Lemma* 5.1 *can be strengthened to*: $Fx\simeq_x F'x$.

5.7. The Main Theorem. *If* S *and* F *satisfy* I–IV, *then, to within isomorphism,* F *is computable.*

Proof. There are only a finite number of stereotypes of a given rank and given size. Hence the lists T_1 and T_2 may be taken as finite, and there are also only finitely many stereotypes for $\{x\restriction t: x\in S, t\in T_1\cup T_2\}$. Thus, to within isomorphism, the domains of G_1 and G_2 are finite and these functions can be given by enumeration. This also gives a bound r to $\mathrm{Card}\,(\mathrm{Sup}\, Gu-\mathrm{Sup}\, u)$. By III and Corollary 5.6 the stereotype of Fx (over x) is uniquely determined by the conditions of IV (with $(3)_{r+1}$). Hence it can be computed by a search procedure; for all the conditions of IV involve quantification only over finite lists.

6. Discussion

(1) It is perhaps worth emphasizing how unrestrictive the principles are. Unlike most automata and algorithms which have been proposed, our treatment does not depend on singling out any set of "elementary" operations. The concept of an algorithm introduced by KOLMOGOROV and USPENSKY (1953) shares this feature; but at each step only a bounded portion of the whole state is changed. The "elementary" operations of most procedures can be carried out in a single step by a device satisfying I–IV. For example it is not too hard to construct a machine which will carry out all the outermost reductions in a formula of the λ-calculus in a single step. One exception (pointed out to me by William Boone) is a Markov algorithm. The process of deciding whether a particular substitution is applicable to a given word is essentially global.

(2) I am sorry that Principle IV does not apply to machines obeying Newtonian mechanics. In these there may be rigid rods of arbitrary lengths and messengers travelling with arbitrary large velocities, so that the distance they can travel in a single step is unbounded. I tried to construct a Newtonian device which should calculate some non-computable function without displaying free will, but was quite unsuccessful. Perhaps some elliptic equation—e.g. Laplace's equation—would permit the construction of such a device.

Problem. Find an alternative to IV which would be satisfied by Newtonian machines, but which would not allow free will to be displayed.

(3) I think it fair to say that the main theorem provides a better proof of Turing's Theorem T (see the Introduction) than any given so far. Turing's own analysis makes clear that calculation by a human being will satisfy I;

II and III can always be satisfied by using a suitable form of description for the record and the program of the calculation; and Turing's arguments *almost* forces one to accept IV (local causation) without any further investigation of particulars of the record and its description.

(4) Since abandoning any of II–IV allows free will, Gödel's objection might be met by abandoning any of them. And GÖDEL himself (1958) showed how the use of functionals of unbounded type could be used to transcend finitistic reasoning (though not computability). But I think it plain that a theory of non-mechanical human intelligence would in fact need only to conflict with IV. The non-mechanical intelligence would, so to speak, see the state x as a *Gestalt*, and by abstract thought make global determinations which could not be got at by local methods.

(5) Despite the liberality advertised in (1) above there is a limit to what a machine can do in a single step. The number of stereotypes in HF_n of size not greater than q can be computed from n and q by a function in Grzegorczek's class $\mathcal{E}_3$ (in which the number of iterations of exponentiation allowed is bounded). Hence, as in the proof of the main theorem, one can compute in $\mathcal{E}_3$ a bound for the number of possible stereotypes of values of Fx from the size of x. In particular, if every number can be uniquely coded by some member of S, then any numerical function which can be calculated in a single step must be bounded by a function in $\mathcal{E}_3$.

Problem. Find a weakening of II which does not permit free will to be displayed.

(6) Because our analysis is not tied to any particular choice of elementary operations it serves to emphasize the familiar view that recursive procedures can be characterized without reference to any kind of recursion or inductive process other than iteration. What is essential is the total exclusion of the infinite. POST (1936) entitled his paper "Finite combinatory processes" and TURING wrote (in (1936)) "The computable numbers may be briefly described as the real numbers whose expressions as a decimal are calculable by finite means". Such characterizations may exorcise intimations of the supernatural from theorems (such as those of BOONE and HIGMAN (1974) and HIGMAN (1961)) which define recursive or recursively enumerable structures in a purely algebraic way. In passing we may note an easily proved, trivial, generalization of Higman's theorem.

Theorem. *Let L, L' be two disjoint sets of labels. Let $X \subseteq HF_n(L)$ be Σ_1 over $HF(L)$. Then we can find a machine (S, F) with $S \subseteq HF(L \cup L')$ satisfying* II-IV *and an $x_0 \in S$ such that*

$$X = \{x: \exists m\, x = F^m x_0 \wedge \mathrm{Sup}\, x \subseteq L\}.$$

(7) The heavy use made of restrictions, and the complications involved in fitting them together (for unique reassembly and the proof of Lemma 5.1) suggest that a treatment using concepts analogous to those of sheaf theory or topos theory might be worth developing. However, it seems to me that the concepts from category theory which would be necessary would be too abstract to allow one to use them (as we have used the more concrete notions of set theory) as a justification for the main thesis of this paper.

Notes added in proof

To 3.7: In the proof of sufficiency we must also consider the case $y \subset x$. Suppose $v = u_1$ and $v \notin y$. Let $s = \{a_2, \ldots, a_n, b\}$ where $a_i \in v/u_i$ for $2 \leqslant i \leqslant n$. Then $x \restriction s \neq y \restriction s$.

To 4.2.1: Observe that if $x \subseteq TC(t) - t$ and $t' = t \cup x$, then $t \preccurlyeq t'$ and $t' \preccurlyeq t$.

To 4.3.1: It is not clear whether the condition given for unique reassembly from located sub-assemblies is equivalent to that given in 3.6 for unique reassembly from (unlocated) sub-assemblies. Our counter-examples do in fact satisfy the principle in either sense. But I would now favour adopting the suggestion made in the last paragraph of 4.3.

To 4.8: The statement that if unboundedly many new labels are added, then there *must* be overlaps is false. It is sufficient for our purpose that there *may* be overlaps.

References

BOONE, W. W. and G. HIGMAN
[1974] An algebraic characterization of groups with soluble word problem, *J. Austral. Math. Soc.*, **18**, 41–53.

DAVIS, M. (Editor)
[1965] *The Undecidable* (New York)

GARDNER, M.
[1971] Mathematical games, *Sci. Am.*, **223**(4), 120–123; **224**(2), 112–116.

GÖDEL, K.
[1958] Über eine bisher noch nicht benützte Erweiterung des finiten Standpunkts, *Dialectica*, **12**, 280–287.

HIGMAN, G.
[1961] Subgroups of finitely presented groups, *Proc. Roy. Soc. Ser. A*, **262**, 455–475.

KOLMOGOROV, A. N. and USPENSKI, V. A.
[1953] On the definition of an algorithm, *Uspehi Math. Nauk.*, **8**, 125–176; *Am. Math. Soc. Translations*, **29** (1963), 217–245.

KREISEL, G.
[1974] A notion of mechanistic theory, *Synthese*, **29**, 11–26.
POST, E. L.
[1936] Finite combinatory processes. Formulation 1, *J. Symbolic Logic*, **1**, 103–105; reprinted in Davis (1965).
POUR-EL, M. B.
[1974] Abstract computability and its relation to the general purposes analog computer, *Trans. Am. Math. Soc.*, **199**, 1–28.
SHEPHERDSON, J. C.
[1975] Computation over abstract structures: serial and parallel procedures and Friedman's effective definitional schemes, in: *Logic Colloquium '73*, edited by H. E. Rose and J. C. Shepherdson (Amsterdam), pp. 445–513.
TURING, A. M.
[1936–7] On computable numbers, with an application to the Entscheidungs problem, *Proc. London Math. Soc. Ser. 2*, **42**, 230–265; reprinted in Davis (1965).
VAN DER WAERDEN, B. L.
[1954] *Science Awakening* (Groningen).
VON NEUMANN, J.
[1966] *The Theory of Self-Reproducing Automata*, edited by A. W. Burks (Urbana and London).
WANG, H.
[1974] *From Mathematics to Philosophy* (London).

J. Barwise, H. J. Keisler and K. Kunen, eds., *The Kleene Symposium*

Recent Advances in the Theory of Higher Level Projective Sets*

Alexander S. Kechris
California Institute of Technology, Pasadena, CA 91125, U.S.A.,
and University of Paris VII, Paris, France

Dedicated to Professor S. C. Kleene on the occasion of his 70th birthday

Abstract: An outline is given of certain aspects of the current theory of higher level projective sets, especially concentrating on the structure theory of Π^1_3 sets, from determinacy hypotheses. Some of the key open problems in this area are also discussed.

1. Introduction

Many decades of work in descriptive set theory, from the beginning of this century until the early sixties, resulted in an extensive theory of the projective sets of the first two levels of the projective hierarchy and its effective analog, the analytical hierarchy of Kleene. As it is by now well-known, this work was done in two almost consecutive and originally independent stages. The first one, which today we call *classical descriptive set theory*, lasted until the late 1930's. The second one, which we call today *effective descriptive set theory*, originated, independently of the classical work, in Kleene's pioneering researches in recursion and definability theory during an almost 20 year period starting in the mid-1930's, but it was not before the work of Mostowski and mainly Addison in the early 1950's, that it was recognized as providing an effective refinement and strengthening, and as a consequence an extension, of the classical theory.

Despite these considerable achievements in comprehending the nature of sets of the first two levels of the projective hierarchy, those of level higher than two remained totally inaccessible. Why this was the case was explained by the work in the metamathematics of set theory, originated by Gödel and Cohen. As it turned out, most of the important questions that

*Preparation of this paper was partially supported by NSF Grant MCS76-17254 A01. The author is an A. P. Sloan Foundation Fellow.