

NLF-65 3-

A Theoretical Analysis of the Effects of Amygdalectomy and of Organismic Motivation

TERENCE W. BARRETT

Neuropsychology Laboratories, Stanford University, Stanford, California

Communicated by K. M. Colby

ABSTRACT

A theoretical analysis of the effects of amygdalectomy is presented within a general scheme for organismic motivation. By an engineering simulation, a solution for n -person nonzero sum games is outlined; by an automata theory analysis, the effects of amygdalectomy are specifically defined with respect to higher-order learning and the organismic equivalent of "real time" in Turing machines is detailed.

I. A PROPOSED SOLUTION FOR NONZERO-SUM AND N -PERSON GAMES

Preface

Studies of organismic functioning have usually concentrated upon the structural problems and neglected the kinetic. In psychology there exists an area of study known as "motivation" that is at a level of abstraction that leaves no clue as to its derivation from a parent discipline. It is suggested here that the derivation should be from thermodynamics for the following reason. There is a parallel between the empirical studies of cognition and motivation on the one hand, and the theoretical analyses of information theory and thermodynamics on the other. The concern of information theory applied to cognition is with structural analysis and it is advocated here that the principles of thermodynamics be applied to motivation because of their mutual concern with the kinetic. Brillouin [4] has derived information theory from the principles of thermodynamics, reason enough for advocating that the first concern of the brain theorist should be with motivation.

In biology the dynamics of the organism are either taken for granted or the end results are described. In ethology there have been certain hydraulic models, now abandoned. Psychology has fared better. Ashby [1] proposed a homeostat that has inspired others in neuropsychology [11].

An individual's interest within the confines of a scientific enterprise can be left a good deal to personal predilection. Some like to observe, and are motivated by such questions as "What exists?" Others like to explain, and are motivated by such questions as "How does it work?" If, however, the disciplines of mathematics and electrical engineering are applied to biological phenomena, we may arrive at answers to questions beginning with "Why?" It seems to the writer that rarely has the predilection to exercise the question been used in biology.

The analyses to follow will show why a certain region of the mammalian brain exists and why it functions in the way it appears to. The explanations will embed the problem of this region's functioning within the molar one of explaining organismic kinetics. Higher-order structuring in the brain will also be shown to imply a higher-order kinetic system.

Introduction

The conservation of energy principle in thermodynamics would seem analogous to the prescribed rule of zero summing in *n*-person games of the mathematical theory of games. The common interest of these branches of physics and mathematics is the concern with operations for the transferral from one state to another. Grayson [6] has shown how a single-loop unity-feedback system may be a physical structure exhibiting behavior describable by the mathematical models of von Neumann's and Morgenstern's *Theory of Games and Economic Behavior* [15]. Extending this suggestion further, if we confine our attention to electrical activity, it is possible to envisage the brain solely in impedance circuit terms. A simple impedance network with parallel and series impedance and its reduction to feedback circuit form is shown in Fig. 1. This reduction is but an example of the general model of Fig. 2, which has the gain equation

$$Gf = \frac{G}{1 + GH}$$

It is evident that if the input is of step or ramp form and occurring with fixed magnitude at random times, then for the required negative feedback to occur, an optimum value of *G* must be found to minimize the error. Grayson [6] has suggested that, as the input varies in a random manner,

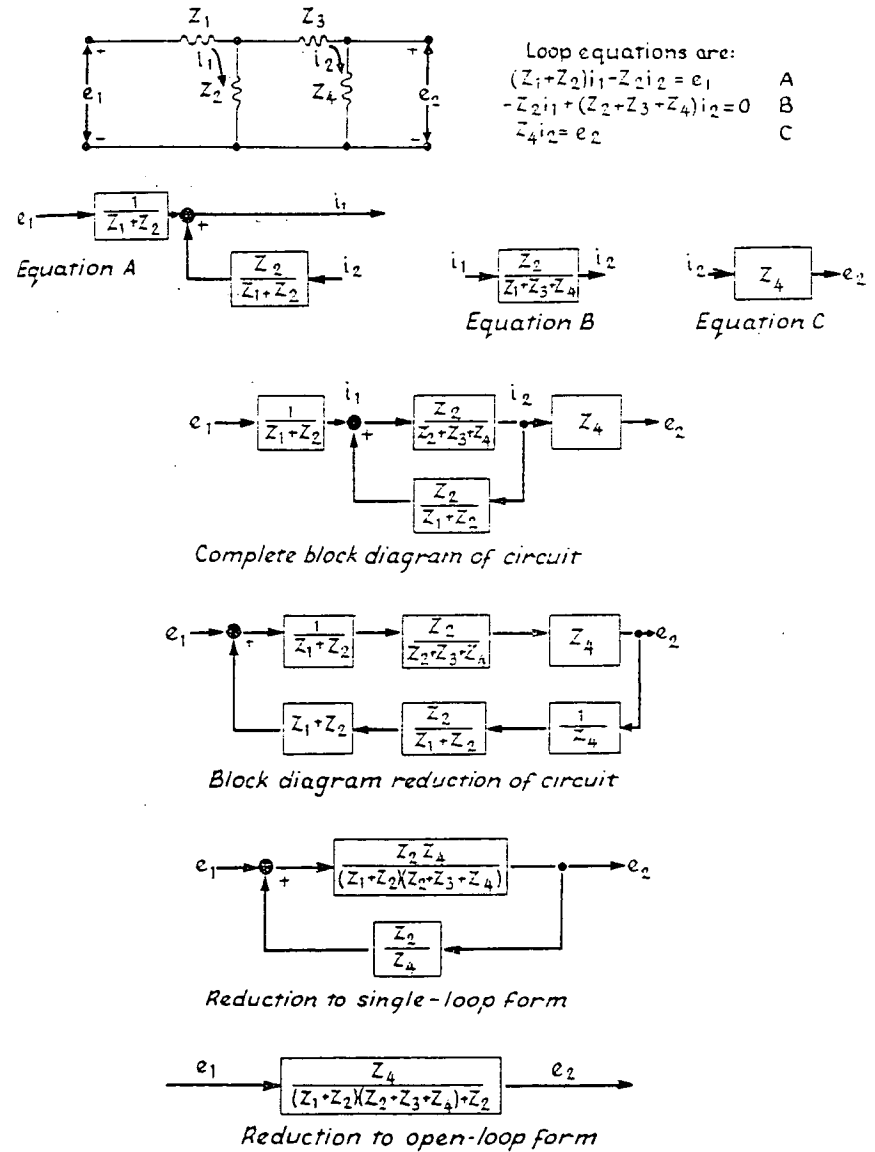


FIG. 1

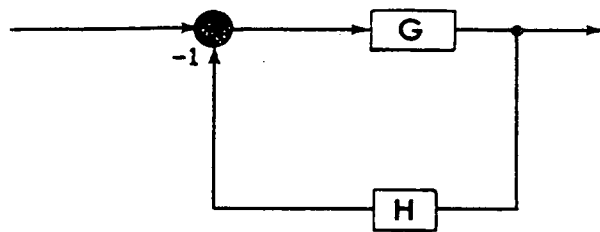


Fig. 2

the more time after the system is in a given state, the less likely it is that it will remain in that state. For this to occur the integral of the product of the error and the weighting function e^{-st} is minimized and stability is achieved thereby. Let $s = \sigma + j\omega$, then

$$F(s) = \int_0^{\infty} e(t)e^{-st} dt. \quad (1)$$

Now the problem of determining an optimum value of G may be treated as a game.

For the moment let us consider a two-person zero-sum game, a zero-sum game because wealth is neither created nor destroyed. Figure 3 is an example of a possible payoff matrix (columns: $i = 1$ to 4 ; rows: $j = 1$ to 4), where we have arbitrarily considered a 4×4 matrix. The argument runs as follows. Let nature be considered as attempting to maximize the error and the input is either step or ramp. For stability, the

		A			
		1	2	3	4
B	1	10	3	5	1
	2	8	9	14	4
	3	7	1	3	8
	4	6	4	8	9

Fig. 3

payoff to be minimized is described by Eq. (1), which is the definition of the Laplace transform, if s is real and e is the base of the natural logarithm. This being the case, then less oscillation will occur if the payoff minimized by nature's opponent is

$$\text{pyf} = C \cdot \frac{1 + GH}{1 + GH + G} \quad (2)$$

where C is a constant determined by the height of the step input or the slope of the ramp input. If there is an element in the payoff matrix that is simultaneously the maximum of its column (i) and the minimum of its row (j), the game is completely determined; such an element is called a "saddle point." The saddle point of this game is seen to be when G is at ∞ . Whatever strategy nature uses, the payoff will always be at a minimum. Such would theoretically be the case of the sleeping organism (i.e., with $G = \infty$, environmental perturbations are at a minimum, i.e., input is inhibited). Waking is explained later. Inhibition thus consists of sending a brain parameter to its maximum value, thus imparting constancies to a fully joined system so that autonomous systems are created.

Now, for reasons that will become apparent later, I believe that two nonzero-sum type of subgames within a zero-sum continuous game are called for to explain the functioning of those parts of the mammalian brain known as the limbic systems [10]. I envisage the game to be conducted in the following manner. If the brain is a player opposing its internal environment and its external environment, then if both games simultaneously played are of the nonzero-sum type, the total loss could be recuperated if a loss in one game constituted a win in the other. Thus two n -person nonzero-sum games played simultaneously can, at a level of abstraction, be reduced to a two-person zero-sum game. Such a dual game would exist in some form of equilibrium.

Unfortunately, no general method of solution for n -person or for nonzero-sum games has yet been proposed. However, Fig. 4 outlines the system I intended and a solution will be followed through.

My reasoning is as follows. With hunger or thirst, G_1 is manipulated to permit nature to win; that is, G_1 is less than ∞ . (The reason feeding should be considered a win for nature is: for input to occur, perturbations must be permitted to the organism's stability. Intake of food from the external environment must be a win for nature with G_1 less than ∞ ; G_1 returns to ∞ after intake of a learned amount of food. After ingestion leading to a return to normal of blood sugar level, G_2 goes to ∞ by delayed

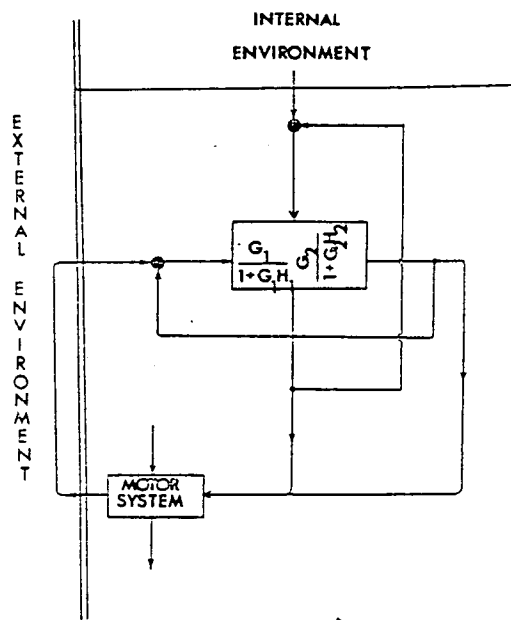


FIG. 4

feedback coupling. With blood sugar level falling, G_2 is less than ∞ , permitting input from the *internal* environment. This produces a win for the internal environment, producing appetitive behavior and setting conditions for consummatory responses: G_1 goes to a value less than ∞ . And the process repeats itself.) With satiation, G_2 is manipulated to permit the organism to win against the internal environment (i.e., G_2 goes to ∞). It is postulated that nature brings about satiation, and the internal environment, hunger. We will also provisionally state the axiom: G_1 and G_2 are assumed synchronous, that is, if $G_1 = \infty$, then $G_2 = \infty$. And if $G_1 < \infty$, then $G_2 < \infty$.

This postulate is provisional because it is made unworkable in this form in the light of the contention that both nonzero-sum games when taken together form a zero-sum game. This being the case, if the external environment is permitted to win, the internal environment cannot at the same time also be permitted to win, for in a zero-sum game wealth is neither created nor destroyed. Thus the provisional postulate would entail that the external environment, in being permitted to win, also permits the brain to win against the internal environment, for in this fashion wealth is neither created nor destroyed. However, this is but saying that G_2 is

maximized to ∞ . But by the provisional postulate, if G_2 is maximized, then G_1 is also. And if G_1 is maximized, then the external environment cannot win, which is paradoxical, as the external environment must lose even while it is winning. A similar argument can be made using the internal environment as first player.

To avoid this deadlock, delayed feedback is introduced and G_1 and G_2 may be modified out-of-phase; that is, if the external environment is the first player, then G_1 leads G_2 . This permits the external environment to win in order that it may lose later.

In the light of these conditions my argument is thus a formal reason *why* biological systems require delayed feedback in the hunger and thirst systems. The system depicted here has similar characteristics to two coupled harmonic oscillators in thermodynamic equilibrium, and the requirement is necessitated by this arrangement.

The application of this model to aspects of functioning of the mammalian brain is quite evident. Lesions of the amygdaloid complex, a structure of the limbic systems, have resulted in an animal that will become obese from overeating if food is available to visual inspection; if, however, food is out of sight but can be obtained by simple learnable methods, then the amygdalotomized animal will eat less than normal [2]. It is as if incentives have a protracted value for such a preparation and are unmodifiable by the fluctuations of the internal environment; while perception of the incentive will always elicit appetitive behavior, appetitive behavior cannot be self-initiated for an incentive that is not in the visual field.

A formal definition of the amygdalotomized organism would be: *An amygdalotomized organism is exemplified by the flow diagram of Fig. 4 but such that although G_2 varies as in the case of a normal organism, G_1 is static.*

It is noteworthy that MacKay [8] has written acutely on the logical indeterminacy of a free choice. His argument has an intuitive appeal that is compatible with the network of Fig. 3. It is my contention that the amygdalotomized animal is operating with only one feedback loop fully functional and as such, cannot be said to be "free."

With the normal animal, it would seem that the feedback loop from the internal environment always, as it were, has the last word. In this deterministic sense, the animal may be said to be free. With amygdalotomy and the "freezing" of variable G_1 , the external environment has the first and last word. The animal is completely determined from without

If E_k^* is the input dictionary or all the sets of words formed from E_k , where a word or a tape means any finite sequence of letters, then the domain of the direct transition function M may be extended to $S \times E_k^*$ by requiring

$$(\forall_s) M(s, \Lambda) = s,$$

$$(\forall_s)(\forall_x) (\forall_{E_k^*}) (\forall \sigma) M(s, x\sigma) = M(M(s, x), \sigma).$$

The response function of a machine $S = (S, M, a)$ to any word $x \in E_k^*$ is defined by

$$\text{rps}(x) = M(a, x).$$

Now, R is a right congruence on E_k^* iff R is an equivalence relation on E_k^* that satisfies the substitution property

$$(\forall x, y, z)_{E_k^*} xRy \Rightarrow xzRyz \quad [\text{where } (\forall x)_X \text{ signifies every } x \text{ in } X]$$

Now, the equiresponse relation of a sequential machine S is a relation $\perp(S)$ on E_k^* defined as $(\forall x, y)_{E_k^*} x \perp y(S)$ iff $\text{rps}(x) = \text{rps}(y)$. Then $\perp(S)$ is a right congruence on E_k^* . Then, if R is a right congruence on E_k^* , the quotient sequential machine modulo R is defined as $\perp(R) = (T, N, b)$ where

$$T = (R[x] \mid x \in E_k^*),$$

$$N(R[x], \sigma_i) = R(x\sigma_i), \quad i = 0, \dots, k - 1,$$

$$b = R(\Lambda).$$

It is stated ([7], page 285) that "whenever we define an algebraic structure whose domain is a set of equivalence classes, we must show that the operations defined on the equivalence classes are independent of the choice of representatives of the equivalence classes. To shorten the terminology, we shall say that the operations must be well-defined."

Paralleling this definition of a well-defined operation with respect to classes, I also define a well-defined operation with respect to relations, which would read, in an analogous fashion:

Whenever we define an algebraic structure whose domain is a set of equivalence relations, we must show that the operations defined on the equivalence relations are independent of the choice of the representatives of the equivalence relations. To shorten the terminology, we will say that the operations must be well-defined.

From this I draw the following conclusion. Both normal and amygdalotomized animals show equiresponse relations ($\perp(S)$) because both are able to substitute responses to various outcomes. Only normal animals, however, exhibit responses that are well defined with respect to both

and any degree of choice is obfuscated. The "last word" of the internal environment of a normal animal makes a free choice logically indeterminate. The exclusive disjunctive to such an animal as the amygdalotomized, creates a dilemma as great as the two bales of hay to Buridan's ass and for which it has no machinery to cope. This has been shown experimentally with amygdalotomized rhesus monkeys [2].

Sections 2 and 3 of this paper are concerned not with motivation per se, but with higher-order regulation of organismic functioning. As regulation implies its own kinetics, the next two sections follow from the first in that the strategies used to gain the payoffs referred to in this section are regulations subservient to these goals. Each section is self-contained but it is taken as axiomatic that the aim or "meaning" of organismic functioning is motivational.

2. AN ANALYSIS OF HIGHER-ORDER LEARNING

Burks, in his introduction to a collection of papers by John von Neumann [14], concludes that this mathematician considered that automata theory should start with mathematical logic and move toward analysis, probability theory, and thermodynamics. However, because of the analogy to the precedence of thermodynamics over information theory already mentioned, it will be the guiding principle for explaining the working of the brain, here, to move in exactly the reverse direction. It seems likely that once the organismic motivational principles are solved, the remaining problems of brain functioning present engineering difficulties of a less complex kind.

In what follows we will attempt a description of the functioning of the amygdalotomized animal in precisely defined terms. If this description is accurate, a programming of the amygdalotomized beast's behavior is possible.

Using the notation of Harrison [7], we let E_k denote the input alphabet. Then, "a sequential machine without output over E_k is a triple $S = (S, M, a)$ where S is a non-empty set called the set of internal states, $a \in S$ is the initial state, and $M: S \times E_k \rightarrow S$ is called the direct transition function" [7], page 273). As an example, let $S = (S, M, a)$ where $S = (a_0, \dots, a_3)$, then M is defined by its graph as

$s \in S$	a_0	a_1	a_2	a_3	a_0	a_1	a_2	a_3
$M(s, a_0)$	a_1	a_2	a_3	a_0	a_1	a_2	a_3	a_0
$M(s, a_1)$	a_2	a_3	a_0	a_1	a_2	a_3	a_0	a_1
$M(s, a_2)$	a_3	a_0	a_1	a_2	a_3	a_0	a_1	a_2
$M(s, a_3)$	a_0	a_1	a_2	a_3	a_0	a_1	a_2	a_3

classes and relations; amygdalectomized animals do not exhibit responses well defined with respect to relations. I make this statement because, from the data I have gathered on learning sets, it can be shown that amygdalectomized animals do not form such a set. Therefore, the operations defined on the equivalence relations are not independent of the choice of representatives of the equivalence relations. The operations of these subjects, therefore, are not well defined in this sense.

It may be said, then, that whereas normal animals are quotient-sequential machines with respect to classes and relations, the amygdalectomized animal is a quotient-sequential machine only with respect to classes.

The operations involved in the solution of a learning set task can be said to be well defined as the normal animals displayed an ability to solve the task. The failure of the amygdalectomized animal to learn relational equivalences may be said to indicate the nonexistence of a homomorphism between the normal animal defined as a quotient-sequential machine with respect to classes and relations and the amygdalectomized animal defined as a quotient-sequential machine with respect to classes only, where homomorphism is defined ([7], page 21) as "a mapping α from an algebraic system D into (onto) another system E is said to be a *homomorphism* of D into (onto) E if α preserves the algebraic structure of D ." The reader is cautioned that algebraic structure is not specified. Obviously, if the algebraic structure is defined in terms of classes alone, then the normal and amygdalectomized animals are homomorphisms.

Following Harrison's suggestion that "intuitively, the behavior of a machine will be the set of tapes recognized by the machine" ([7], page 296), then, if an *automaton* or *sequential machine* over E_k with *output* is a quadruple (S, M, a, F) where S is a nonempty set of internal states; M is the transition function; $M: S \times E_k \rightarrow S$, $a \in S$, is the initial state; and $F \subseteq S$ is the set of states giving a one output, then let R be an equivalence relation on E_k^* and let β be a subset of E_k^* . Then R is said to *refine* β iff xRy implies that $x \in \beta$ iff $y \in \beta$. Symbolically,

$$xRy \Rightarrow (x \in \beta \Leftrightarrow y \in \beta).$$

Now, if β were treated as an equivalence relation of rank 2, then this definition would be simply the definition of the equivalence relation R refining the equivalence relation β . Let β be a subset of E_k^* ; β is the behavior of a finite sequential machine over E_k iff there exists a right-congruence relation R of finite rank that refines β . It is interesting that

amygdalectomized animals fabricate "templates of control" (higher-order schematic image for action; in automata theory terms, β) but they are unable to utilize the economic information involved in such a fabrication. This is quite clearly seen in a study by Schwartzbaum and Pribram [13]. It would appear, therefore, that the equivalence relation can be formed by amygdalectomized beasts. What appears to be missing is the right-congruence relation operation that refines β . It therefore seems that if β is any subset of E_k^* , then the right congruence induced by β and defined as

$$xR_\beta y \Leftrightarrow (\forall z) \in E_k^* (xz \in \beta \Leftrightarrow yz \in \beta)$$

also defines a learning set. Consider this also: I have shown [2] that, over 15 reversals on a reversal learning task, both normal and amygdalectomized rhesus monkeys show three peaks in the curves displaying the number of trials per reversal. These peaks are not there by chance but at the over-the-horizon level of significance (0.06007⁶). I do not think that this is a "magical number" at all; but suppose, thinking in the machine language of programming, that there is a commutator shift involved in the formation of a learning set or the development of an equivalence relation R refining β . Then let there be a count or ISZ (increment a negative number and skip the next instruction if zero) command in the program. Suppose the counter is set in two's complement arithmetic, say, at 7775. After three shifts in both amygdalectomized and normal animals, the next instruction is skipped and the following instruction may be JMS higher-order learning (where JMS is "jump to subroutine"). Essentially this would be the development of an equivalence relation of refinement. Now I am proposing that my data show that although both the normal and amygdalectomized animals possess the ISZ commands in their "programs" (from the peaks in the learning curves), only the normals have the JMS command or possess the higher-order learning subroutine. This is because, whereas the normals improve their performance, the amygdalectomized animals plod along at a lower level of efficiency.

Turning attention to recursive functions, we have that a subset R of X^* is (1) recursive if its characteristic function

$$R^{(n)} = \begin{cases} 0 & \text{if } n \notin R, \\ 1 & \text{if } n \in R, \end{cases}$$

is a recursive function; (2) recursively enumerable if it is the range of some recursive function f (i.e., $R = \{f(x) \mid x \in X\}$, or is the empty set \emptyset).

Now a set S is recursive iff both S and \bar{S} are recursively enumerable. It will be recalled that behavior was defined as the set of tapes recognized by a machine. Finally, there exists a recursively enumerable set that is not recursive, for: a recursively enumerable set U can be exhibited for which \bar{U} is not recursively enumerable. Thus in this manner, define the set U by the equivalence

$$n \in U \Leftrightarrow n \in S_n$$

where (S_1, S_2, \dots) is the effective enumeration of the recursively enumerable sets; U is recursively enumerable. For if f is defined by

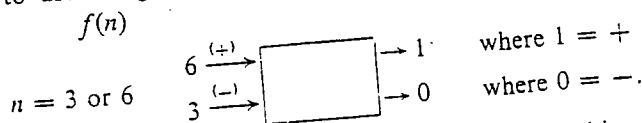
$$f(n) = \begin{cases} y & \text{if } n = T(x, y) \text{ and } y \text{ is the } x\text{th element of } S \text{ and where } \\ & T \text{ enumerates the pairs of integers,} \\ \text{undefined} & \text{if not,} \end{cases}$$

then $U = \{f(n) \mid n \in N\}$ and so is recursively enumerable. \bar{U} can be shown to be not recursively enumerable. For were \bar{U} recursively enumerable then $\bar{U} = S_{n_0}$ for some n_0 . But then

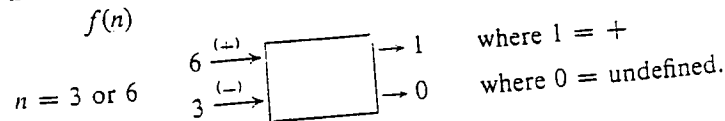
$$n_0 \in U \Leftrightarrow n_0 \in S_{n_0} \Leftrightarrow n_0 \in \bar{U},$$

which is a contradiction.

Turning now to amygdallectomized and normal machines, it is apparent that for a right congruence relation upon β to occur in the case of a reversal problem on two stimuli, say, a 3 and a 6, both stimuli must be categorized by the machine. That is to say, when one stimuli is not negative, when it formerly was positive, there must be a comparator that registers "regret," to use the game-theoretic term. Such a machine would be as follows.



Both input functions are defined and such a machine produces a subset R of X^* that is both recursive and recursively enumerable, as $f(n)$ is defined in all cases. Now, it appears to me that the amygdallectomized machine learns each reversal as a new problem. This means it has no notion of "regret" being produced by its machinery, and would look thus:



It will be seen that although such a machine would produce a subset R of X^* that is recursively enumerable, it would not be recursive, because $f(n)$ is not defined for all cases. (The design of the experiments was as follows. Two were of the sequential variety, whereby the animal had to perform in the present in the light of his previous activities. A learning set task depended on this subject's remembering prior performances over a 4-month period of reversals of reinforcement using the same stimuli.)

In support of the notion that amygdallectomized animals do not process the data irrelevant to the solution of a problem, the data of Douglas, Barrett, and Pribram [5] are presented. In this study, discrimination problems were given normal rhesus monkeys and animals with amygdala and hippocampal lesions. In the first instance two stimuli were presented and one was rewarded; then three stimuli were presented and one was rewarded; finally, five stimuli were presented and again only one was rewarded. The problems differed, therefore, in the variable of the number of irrelevant stimuli.

There are certain mathematical theories of learning that predict learning rates. Blehert [3] has shown how such a description can deepen data analysis. The equation used to describe the learning curves obtained was

$$\Pr(X_n) = \begin{cases} \frac{1}{X} & \text{for } n \leq N', \\ 1 - \frac{1}{X}(1 - \theta)^{n-N'} & \text{for } n > N'. \end{cases}$$

where $\Pr(X_n)$ is the probability of a correct response on trial n and X is the number of stimuli in the problem; N' is the number of the trial learning commenced and n is the trial number. The parameter θ was calculated from the data by the formula

$$\theta = \frac{1}{X} \left[\sum_{n=N'}^{N''} \Pr(Y_n) \right]^{-1}$$

where $\Pr(Y_n)$ is the probability of an incorrect response on trial n and N'' is the number of the trial when criterion was reached.

Figure 5 displays the various θ values for the three groups. It can be seen that whereas normal animals decrease their learning rate (decreased θ values) when the number of irrelevant cues are increased, amygdallectomized animals increase their learning rate (increased θ value). This seems to indicate that not only are amygdallectomized animals increasing their

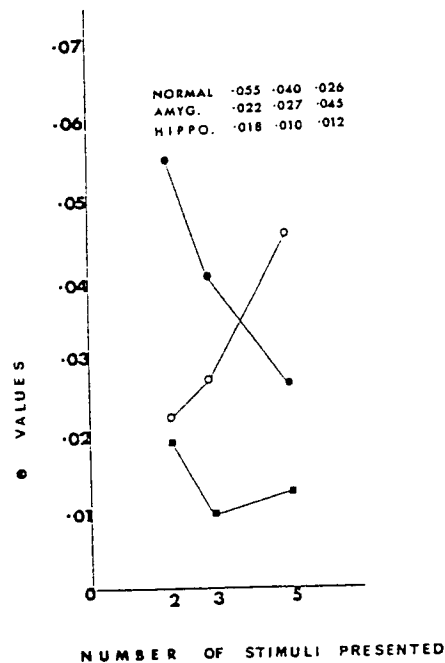


FIG. 5

efficiency of responding to the rewarded cue, but they are also treating each task as equivalent in difficulty. We can only draw the conclusion, therefore, that they are not processing the irrelevant data (relevance being that which is reinforced by food reward).

It will be noted that I am not claiming that the amygdallectomized machine is Epicurean and only zeros in on positive reinforcement. What I am saying is that initially, to the normal animal the unrewarded stimulus is neutral; then, after being rewarded and then unrewarded, it has a history of past reinforcements that the normal animal "regrets." It is not, therefore, a negative cue, and if it were, the amygdallectomized animal would define the function $f(n)$ fully. The unrewarded cue can only be "defined," therefore, within the context of past reinforcements. This so-called context I see as none other than a sequential ordering of experiences. Let it suffice to remark here that regret only exists in a temporal sequence. The normal animal, therefore, exhibits behavior or produces a subset R of X^* that is both recursive and recursively enumerable; an amygdallectomized animal, on the other hand, exhibits behavior

or produces a subset R of X^* that is recursively enumerable but not recursive.

3. AN ANALYSIS OF THE NEED FOR INCREASED CHANNEL CAPACITY IN THE BIOLOGICAL MACHINE WITH THE FORMATION OF HIGHER CENTERS OF CONTROL

Rabin [12] introduced the concept of real-time computation by a Turing machine. In doing so, he was able to establish the relative strengths of one-tape versus two-tape machines by proofs of impossibility in real time of actual computations. We will follow his arguments through and draw analogies with respect to the increased channel capacity of the biological machine in the ascent of the phylogenetic tree and remark upon the concomitant development of abstract learning. It is felt that this concomitance is no accident, but the result of the need of a control head of a higher order, so that the work of the n control heads may be supervised. The "programming" of this control head of a higher order is considered to be essentially learning set formation.

First, Rabin stipulates the limitation that if the problem (or input data), consists of n symbols, then the computation will be performed in n basic steps, one step per input symbol. If the result of a computation on every input sequence is always 1 or 0, then the machine can be viewed, as I have already outlined, as defining a set, namely, the set of those input sequences that yield 1. Rabin shows that there exists a recognition problem that can be done in real time using two tapes but cannot be done in real time on a single tape.

Now, a multitape Turing machine over the input alphabet Σ is a finite automaton M having a finite set S of states and a working alphabet $W = (\alpha_1, \dots, \alpha_n)$. One state, call it s_0 , is distinguished as the initial state of M . A subset $F \subseteq S$ is stipulated as the set of designated final states. The machine has k two-way infinite linear work tapes t_1, \dots, t_k , which are divided into squares. There is a reading printing head, which at any given time scans one square of each of the work tapes. M is capable of receiving inputs $\sigma \in \Sigma$. The working alphabet is assumed to contain a blank symbol and at least one other symbol so that $2 \leq n$.

The function

$$M(\sigma, s, \alpha_{i_1}, \dots, \alpha_{i_k}) = (s', X_1, \dots, X_k, \alpha_{j_1}, \dots, \alpha_{j_k})$$

specifies the operation of the machine, where $\sigma \in \Sigma$, $s, s' \in S$, $\alpha_i, \alpha_j \in W$, $X_i \in (0, 1, -1)$. This function is referred to as the *machine-table* of the

k -tape, Turing machine M . Rabin's interpretation is that if the input is σ and M is in state s and is reading α_r on the tape t_r , $1 \leq r \leq k$, then M will go into state s' , print α_r on the scanned square on t_r , and move each tape t_r one square left, or one square right, or not at all, according as to whether X_r equals 1, -1 , or 0. This action of M is called an *atomic move*. The set of all finite sequences on the alphabet Σ will be denoted by Σ^* .

Further defining conditions are as follows. A sequence $x = \sigma_1 \cdots \sigma_p \in \Sigma^*$ is said to be *accepted* by M if, when started in s_0 and with blank work tapes, M will go under the input sequence x through its atomic moves and end in a state F (i.e., the state of M at the p th time unit is designated).

The set of all sequences accepted by M is called the set *defined* by M and is denoted by $T(M)$.

A set $T \subseteq \Sigma^*$ is called *real-time definable* (recognizable) if there exists a multitape Turing machine M such that $T = T(M)$. In particular, T is called *k -tape real-time definable* if for some M with k work tapes, $T = T(M)$.

The set T_2 is defined: Let $\Sigma = (a, b, 0, 1, \alpha, \beta)$. Words on (a, b) will be called *ab words* and the set of *ab words* will be denoted by A . Words on $(0, 1)$ will be called *01 words* and the set of *01 words* will be denoted by Z .

If $x = \sigma_1 \sigma_2 \cdots \sigma_{n-1} \sigma_n$, then, by definition, $x^* = \sigma_n \sigma_{n-1} \cdots \sigma_2 \sigma_1$. Let $T_2 = (uvxu^* \mid u \in A, v \in Z) \vee (uv\beta v^* \mid u \in A, v \in Z)$; then,

LEMMA 1. *The set T_2 is real-time definable by a two-tape machine.*

The proof of this lemma is by description of the mode of operation of a two-tape machine M for which $T_2 = T(M)$: as the *ab word* u is coming in, M will print it on its second tape. According to whether the input following uv is α or β , M will start tracing back its first or second tape. M will end in a designated state if and only if the sequence w of inputs following α (or β) coincides with the sequence being traced backward on the first (second) tape. ■

Rabin's Theorem 1 is as follows. The set T_2 is not real-time definable by a one-tape machine. Consequently, two-tape real-time computation can do more than one-tape real-time computation. To prove this, it is assumed by way of contradiction that the one-tape machine M does define T_2 in real time. Let the number of states of M be m and the number of letters in its working alphabet be n . Then, if M has input w , the *work space* $t(w)$ of M on w is the sequence of tape squares covered by the motion of M while

having the input sequence w . If x is a sequence of squares on the tape or a sequence of symbols, then $l(x)$ will denote the *length* (i.e., the number of elements) of x . Let x be an input sequence; by the *coding* of x is meant the sequence of symbols in the squares of the work space $l(x)$, the state of M , and its position on the tape, at the end of the input x .

LEMMA 2. *There exists a numerical constant $c > 0$ such that for every $u \in A$ and every integer $i > 0$ there exists a $v \in Z$ such that $l(v) = i$ and $ci \leq l(t(uv))$.*

Proof. There are 2^i sequences $v \in Z$ such that $l(v) = i$. Since the input uv may be followed by β , if $v_1 \neq v$, then uv_1 and uv must be coded differently. Otherwise, $uv\beta v^*$ and $uv_1\beta v^*$ will both be accepted by M .

If $l(t(uv)) \leq k$ for all $v \in Z$, $l(v) = i$. Then there are at most $n^k \cdot k \cdot m$ different codings of the inputs uv . Hence $2^i \leq n^k \cdot k \cdot m$. If i is large, this forces k to be large so that we may assume that $km \leq n^k$ (it is assumed that $2 \leq n$). Thus $2^i \leq n^{2k}$ and hence

$$\frac{1}{2} \frac{\ln 2}{\ln n} i \leq k.$$

It may be taken, then, that $c_1 = \frac{1}{2}(\ln 2 / \ln n)$. Rabin comments that this c_1 will do for all i larger than some i_0 ; for a suitable smaller c the lemma will hold for all i . ■

LEMMA 3. *There exists an integer $d > 0$ (depending only on M) such that for every $u \in A$ and every integer $i \geq l(u)$ there exists a sequence $v \in Z$, $l(v) = i$, such that (a) $ci \leq l(t(uv))$, and (b) no more than one fifth of the squares of $t(uv)$ are covered by M more than d times.*

This is proven by choosing a sequence $v \in Z$, $l(v) = i$, for which (a) holds. Let d_1 be a number such that more than one fifth of the squares of $t(uv)$ are covered by M more than d_1 times. Then the total number of moves of M exceeds $d_1 \frac{1}{5} [l(t(uv))] \geq \frac{1}{5} d_1 ci$. But since M operates in real time, the number of moves of M by the input uv is exactly $l(u) + l(v) \leq 2i$. Thus, $\frac{1}{5} d_1 ci \leq 2i$ and $d_1 \leq 10/c$. The number $d = (10/c + 1)$ satisfies (b).

Now, Rabin's proof of the main theorem rests on the idea that in working in certain input sequences, the machine M develops bottleneck squares on its work tape through which information cannot flow in sufficient quantity. A bottleneck square is defined as follows. Let $u \in A$,

$v \in Z$. A square B on $t(uv)$ is called a bottleneck square of $t(uv)$ if (1) under input uw the machine passes through B no more than d times; (2) B lies outside the work space $t(u)$; (3) the length of the section of $t(uv)$ determined by B that does not contain $t(u)$ exceeds $l(u) + 1$. From this follows Lemma 4.

LEMMA 4. For every $u \in A$ there exists a $v \in Z$ such that the tape $t(uv)$ has a bottleneck square.

This is proved as follows. Let i be an integer such that $5l(u) + 5 < ci$ and also $l(u) \leq i$. By Lemma 3 there exists a sequence $v \in Z$ such that $ci = l(t(uv))$ and fewer than one fifth of the squares of $t(uv)$ are covered more than d times.

$$l(t(u)) \leq l(u) + 1 < \left(\frac{ci}{5}\right) \leq \frac{l(t(uv))}{5}$$

Dividing $t(uv)$ into five equal parts, we see that on either the left or the right end of $t(uv)$ there is an interval of length $\frac{2}{5}l(t(uv))$ that does not contain any squares of $t(u)$. In this interval the one fifth of $t(uv)$ is considered that does not run to the end. As fewer than one-fifth squares of $t(uv)$ are covered more than d times by M , there is a square B in this one fifth of $t(uv)$ that is covered at most d times, as there are at least $l(t(uv))/5 \geq (ci/5) > l(u) + 1$ squares between B and the end of $t(uv)$. Thus B is a bottleneck square. ■

The main theorem is proved as follows. Let $u \in A$ and $v \in Z$ be such that $t(uv)$ has a bottleneck square B . Assume that B is to the right of $t(u)$ (although this may not be so). As the input uw is coming in, there is a first time that M enters the right-most square E of $t(uv)$. Let $w \in Z$ be the initial section of v such that uw is the sequence leading to the first visit of M at E . Thus $t(uv)$ and $t(uw)$ have the same right-hand end square E and B is also a bottleneck square of $t(uw)$.

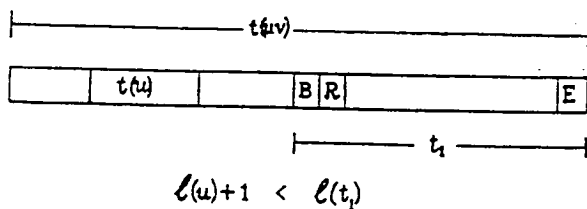


FIG. 6. From [12], page 207, by permission.

Mathematical Biosciences 4 (1969), 153-178

The square immediately to the right of B is denoted by R . A passage of M through B means either a move of M from B to R or a move from R to B . The state of M during a passage is the state M has when it reaches R in the first case, and the state M has when it reaches B in the second case. Under the input uw the machine M will first cover the tape $t(u)$ and then, under the w portion of the input, move to the square E . Let p_1, p_2, \dots, p_r be the consecutive passages through B ($r = 1$ is not excluded). The passage p_1 is a move from B to R , p_2 is a move from R to B , and so on. Let the state of M during the passage p_i be s_i , $1 \leq i \leq r$. The scheme of the bottleneck square B is the $(r + 1)$ -tuple (e, s_1, \dots, s_r) where e is 1 if B is to the right of $t(u)$ and e is -1 if B is to the left of $t(u)$, and s_1, \dots, s_r are as above.

Now, the number of r passages through B is at most d . Thus, there are at most N ,

$$N = 2 \cdot m + 2 \cdot m^2 + \dots + 2 \cdot m^d,$$

different schemes of bottleneck squares, where m is the number of states of M .

Let g be a number such that $N < 2^g$. For each $u \in A$, $l(u) = g$, let $v \in Z$ be a 01 sequence such that $t(uv)$ has a bottleneck square B_u and let w denote the section of v leading to the first visit of M to the end E_u of $t(uv)$. There must be two different sequences $u_1, u_2 \in A$, $l(u_1) = l(u_2) = g$, such that the bottleneck squares B_{u_1} and B_{u_2} have the same scheme, say $(1, s_1, \dots, s_r)$. Note that $e = 1$, which means that B_{u_i} is to the right of $t(u_i)$, $i = 1, 2$.

Let

$$u_1 w_1 = u_1 \varepsilon_1 \cdots \varepsilon_{n_1} \cdots \varepsilon_{n_2} \cdots \varepsilon_{n_r} \cdots \varepsilon_{n_{r+1}},$$

$$u_2 w_2 = u_2 \delta_1 \cdots \delta_{m_1} \cdots \delta_{m_2} \cdots \delta_{m_r} \cdots \delta_{m_{r+1}},$$

where $\varepsilon, \delta \in (0, 1)$, ε_{n_1} is the input where M visits B_{u_1} during the first passage, ε_{n_2} is the input when M visits B_{u_1} during the second passage, and so on up to ε_{n_r} ; similarly for $\delta_{m_1}, \delta_{m_2}, \dots$, in the second sequence $u_2 w_2$. After receiving the input $\varepsilon_{n_{r+1}}$ ($\delta_{m_{r+1}}$), M visits for the first time the right-hand end square E_{u_1} (E_{u_2}).

The main point of Rabin's argument is as follows. In the sequence $u_1 w_1$ replace, for each odd $1 \leq i \leq r - 2$, the segment $\varepsilon_{n_{i+1}} \cdots \varepsilon_{n_{(i+1)-1}}$ by the sequence $\delta_{m_{i+1}} \cdots \delta_{m_{(i+1)-1}}$. Furthermore, $\varepsilon_{n_{r+1}} \cdots \varepsilon_{n_{(r+1)-1}}$ is replaced by $\delta_{m_{r+1}} \cdots \delta_{m_{(r+1)-1}}$. The resulting sequence is called $u_1 w'_1$. It is noted that all the changes were made in the w_1 portion of $u_1 w_1$. Now $u_1 w_1$ and $u_2 w_2$ have the same schemes of states in the passages of M

Mathematical Biosciences 4 (1969), 153-178

through B_{u_1} and B_{u_2} , respectively, and our changes in u_1w_1 were made only in the inputs between visits to B_{u_1} , while M was on the right of B_{u_1} , or after the last visit to B_{u_1} . We can see by finite induction over $1 \leq i \leq r+1$ that u_1w_1' again has the same scheme $(1, s_1, s_2, \dots, s_r)$ and that at each input δ_{n_j} , j odd and $2 \leq j \leq r+1$, the portion of the tape right of B_{u_1} is identical with the portion of the tape $t(u_2w_2)$ right of B_{u_2} at input δ_{m_j} and the states of M at the corresponding inputs are the same.

The work spaces $t(u_1w_1')$ and $t(u_2w_2)$ have squares B_{u_1} and B_{u_2} , respectively, with the following properties. The work space $t(u_i)$ is completely to the left of B_{u_i} , $i = 1, 2$. The portions of $t(u_1w_1')$ and $t(u_2w_2)$ beyond B_{u_1} and B_{u_2} are strictly longer than $l(u_1) = l(u_2) = g$. By the previous paragraph, at the end of the inputs u_1w_1' and u_2w_2 , M is at the end of squares E_1 and E_2 of the respective work spaces and the portions of tape from B_{u_1} to E_1 and B_{u_2} to E_2 as well as the states are identical. Assume now that both u_1w_1' and u_2w_2 are followed by the input αu_1^* . Since $u_1 \neq u_2$

$$u_1w_1'\alpha u_1^* \in T_2, \quad u_2w_2\alpha u_1^* \notin T_2.$$

But $l(\alpha u_1^*) = g + 1$ is less than the distance from E_i to B_{u_i} , $i = 1, 2$. Since M operates in real time and makes one move per input, it will stay, throughout the input portion αu_1^* , to the right of B_{u_i} . Thus M will start in both cases in the same state and will move through identically printed portions of tape. It will therefore be in the same state at the end of $u_1w_1'\alpha u_1^*$ and $u_2w_2\alpha u_1^*$ and hence cannot accept one and reject the other; a contradiction.

4. ANALYSIS AND COMMENTS

There appear to be the following crucial steps to the proof.

1. The introduction of the integer d whose value is limited by the concept of real time (i.e., $l(u) + l(v) \leq 2i$), so that no more than one fifth of the squares of $t(uv)$ are covered by M more than d times. By the concept of real time, therefore, Turing machines have a channel capacity.

2. The introduction of g , which is a number such that the number N of different schemes of bottleneck squares $< 2^g$ and now $l(u)$ must equal g . This I interpret as follows. The capacity of the channel is now delimited so that B really is a bottleneck square in the input sense, whereas before it was defined in an output sense (cf. Lemma 3). Therefore, the existence of a bottleneck square does not warrant the organism-machine's development of more than one tape; it is the delimitation of g (i.e., the lemma), talking

now in biological language, that should state that if the organism's input capacity is less than its behavioral repertoire (i.e., the portions of $t(u_1w_1')$ and $t(u_2w_2)$ beyond B_{u_1} and B_{u_2} are strictly longer than $l(u_1) = l(u_2) = g$), so that it is, as has been shown, delimited in two senses, then that organism must develop two tapes if it is to recognize efficiently. It is noteworthy that the work space is longer than the input length in the critical case. The work space in the Turing machine had been altered by a previous input uv . Now, in the biological machine, there is a further delimitation that only tapes of uniform length are accepted as input, and the case used in Rabin's proof could hardly occur biologically. (The real time of a Turing machine is equivalent to the restriction of a channel capacity, or bounded input, of an organism. Although Rabin's proof only works because of the assumption that input length is unbounded, this does not preclude drawing parallels with the biological machine, which has a bounded input but unbounded command set. A Turing machine is thus in the reverse predicament to that of a biological machine. The parallels drawn here are more in the nature of mirror images.) The organism is not frustrated by the restrictions on its channel capacity; it is, rather, in a state of not knowing. Therefore, the input $l(uv) > l(u)$ or $l(v)$ should be conceived as not really an input but a metaphysical goal that the all-knowing outsider knows the organism could achieve but cannot now with its limited channel capacity, and limited sampling or hold capacity.

3. The work space is longer than input lengths, forcing a decision when a critical area is not reachable after a certain set of instructions have been executed. Thus the inability to reach a decision, or rather, a correct one of exclusion, is due to the unreachability of the criteria. We have already commented upon the unrealistic assumption that organisms, as well as Turing machines, are able to engorge any length of their "life space," given enough real time to do so. However, the proof still stands for biological machines provided that the tape $l(uv)$ is taken as containing information that could be obtained by the organism if not only did it have two tapes but also were its sampling capacity not limited. Rabin's Turing machine stands with the criteria for recognition within its long-term store, as it were, and is unable, because of real-time limitations, to bring input and criteria together. The biological machine, on the other hand, must fabricate its criteria from its finite sampling, and for such a machine ignorance is bliss, for if the criteria are in its long-term store, then any input will reach it. Only the all-knowing experimenter can take a look at the possible modifications that tape $l(uv)$ could make; and this is the crux

of experimental child psychology, in which the adult experimenter may take a look at what would be possible if the child had a bigger channel capacity. Thus, Rabin's Turing machine is in the awful predicament of actually "knowing" reality (i.e., the modifications on its tape by input $l(uv)$) but is prevented by real time from making a decision on this information when input of a certain form arrives. The biological machine, however, never knows reality and if its decisions are "off," then it is unaware of it.

(These remarks should be considered in the light of the consideration that the real time of a Turing machine is equivalent to the restriction of a channel capacity (bounded input) of an organism; that is, these are complementary bounds: time for the Turing machine within which to execute instructions (i.e., a bounded set of commands), and input limitations for the organism (i.e., a bounded set of addresses upon which to execute the commands). Rabin's theorem only has relevance to biology when translated in this manner. Parallel processing in the organism at a lower level of hierarchical control is brought about by channel-capacity limitations at a higher level (i.e., main program level) at which there is only serial processing. Giving a machine two tapes is not the same as giving it two heads, for an enlarged command language is required for a higher-order monitoring head to be used and this is forbidden by real-time limitations.)

We return now to the problem raised earlier that really to simulate the organism-machine, Rabin's Turing machine must be further delimited so that its tape sampling is of equal length. If this limitation is accepted, then it follows that giving a machine two tapes is not the same as giving it two instruction heads, and thus problems solvable by one solution are not necessarily solvable by the other. ("Know" for the Turing machine is in the sense that it has unbounded input; on the other hand, our neuronal events do not "know" more because we process everything carried by them within the limits of selective attention. The Turing machine, however, because of real-time limitations does not ingest everything on its tapes.) That is, by the limitation of finite sampling of tape length, or (using the real-time equation) of finite time, it will not solve anything to give the machine two tapes to solve Rabin's recognition problem, in this case $u_i > w_i$, for it might be stipulated that both u_1 and u_2 have the same scheme for length w_1 or w_2 . Then Rabin's proof may be reread as showing that a Turing machine with two tapes and finite sampling space cannot categorize certain inputs correctly after being modified by certain previous inputs.

If nature wanted the biological machine to be more in touch with reality, that is, more able to accept tapes of longer length such as $l(uv)$, then, if she already had laid down the restriction of finite sampling (for other good engineering reasons, no doubt), she must have increased the channel capacity of the biological machine by employing more than one instruction head, and each instruction head receives its own finitely sampled tape or portion of tape. This may seem like saying that the biological machine resembles a commonwealth, or that it may perform two or more things at once, but for a further restriction:

In biological machines (I now think of these machines as Rabin-Turing machines with the additional restriction of finite tape-length sampling), it is the case that sometimes a decision such as "greater than" or "less than" can only be made when the final states of two operations are compared simultaneously with each other and not, let it be emphasized, with some previously existing set specifications in the machine. In this case two instruction heads would be required, but a third would also be needed to clear the accumulator of the subtraction or addition into its own memory space. This third head would contain instructions of a kind similar to those of a main program, while the other heads would contain instructions similar to those in subroutines. The transferability of the results of the third head's operations is obvious. It would enable operations to be performed upon operations (i.e., operations would be monitored). In such a case, relations (and not merely classes), would be formed in that the attainment of E_1 by submachines does not entail the additional operation of a comparison with E_2 (attained by other submachines) concerning the relations "greater than," "less than," and "equals." To be sure, a third instruction head would not be required if all that were needed were classifying behavior. But if relational "judgments" are also to be made, then a third head is required.

If this third head operates upon instructions or is a function of a function, then its operation might be felt to define the notion of "abstract" or higher-order learned performance, or even the learning set. A function of a function would, therefore, concern itself with the more important matters of measuring, weighing, and (the offspring of these operations) relations. It has been shown by Piaget [9] that the concept of number is a fusion of classification and iteration. In the young child, these activities appear separate. With the appearance of a third monitoring instruction head with an executive program for the subroutines, there could be a fusion. And this takes place when the growing child increases its channel

capacity (by myelination?). So it is that the existence of bottleneck squares plus the restriction of limited sampling forces the phylogenetically growing Turing machine to increase its channel capacity by the provision of multiple instruction heads—if it wants to know reality at all. If it wants to have relations as well as classes, for they are very useful, then it must have a higher-order instruction head to monitor the subheads. This instruction head is, I repeat, an instruction head; and the nature of the abstract is that it is a *command language*. I make this point because there are some who will still try to understand the abstract in a classificatory (and not relational) manner. Such people usually are attracted to Plato's philosophy, with its emphasis on "forms." A form, I suggest, is the outcome of an abortive attempt to understand a function of a function in a classificatory manner. If the computer engineer were suddenly afflicted with Platonism, he would believe that the existence of an executive program warranted the belief in a ghostly memory bank somewhere, that is, the existence of a form.

Thus it is that I believe that the solution to a learning set or reversal problem such as "go for the odd one" or "first this then that then this . . ." implies the fabrication of an executive program to monitor subroutines. The "odd one" implies a relational judgment such as a "skip if zero" command might perform; and the reversal problem implies an application of an operator that multiplies the economic value by -1 on every reversal when the criterion for reversal has been recognized or categorized. Notice the low-level, even if necessary, activity of classification.

The intention has been, therefore, to compare and contrast (a) Turing-Rabin machines with (b) biological machines. Turing machines are hampered in their computational ability by real-time limitations in their command mode; biological machines are hampered in their computational ability by channel capacity limitations; that is, their accumulators are only designed to handle a *limited word length*. Rabin offers no way of removing real-time limitations and it would appear that limitations on word length are obviated by hierarchial control of executive programs over subprograms. Rabin's proof is obviously irrelevant to the solution. It is relevant to the limitations because they reflect the limitations of the biological machine in mirror image, that is, a machine with unlimited input but limited command mode (real time) has computational limitations for which no solution is given; a machine with limited input (channel capacity) but an unlimited command mode has computational limitations for which a solution is hierarchial control.

There is only a similarity in reverse and I have attempted to draw the analogy.

SUMMARY

The work of Rabin was reviewed in detail showing that in real time there are some classificatory problems unsolvable by a one-tape Turing machine that are solvable by a Turing machine with two tapes.

By introducing the limitation of finite tape sampling, it was indicated that the two-tape machine is not reducible to the two-head machine with one tape. Also, the biological machine, because of this restriction plus that of real time, is an n -head one-tape machine.

Turning away from the old problem of classification to the new one of drawing relations (judgments), it was felt that a monitoring head of instructions is needed. If this is adopted, it becomes apparent that the activity of this instruction head is of the nature of the "abstract." The abstract is therefore an offspring of this relational activity, as an executive program can consist of jump commands alone and it would be wrong to suggest a classificatory function to this monitoring activity.

It is of the nature of the learning set and of reversal learning to be abstract, which leads to the conclusion: the amygdalectomized machine is incapable of abstract learning. It thus suffers from ablation of the appraisal function necessary to the operation of its executive programs. Relational thought is gone; egocentricism obtains. Only subroutines are available.

REFERENCES

- 1 W. R. Ashby, *Design for a brain* (2nd ed.). Wiley, New York, 1960.
- 2 T. W. Barrett, *The function of the amygdaloid complex*. Unpublished Ph.D. dissertation, Stanford University, Stanford, California, 1967.
- 3 S. Blehert, Pattern discrimination learning with rhesus monkeys, *Psychol. Rept.* 19, (1966) 311-324.
- 4 L. Brillouin, *Science and information theory* (2nd ed.). Academic Press, New York, 1962.
- 5 R. J. Douglas, T. W. Barrett, and K. H. Pribram, Limbic lesions and error reduction; to appear in *J. Comp. Physiol. Psychol.*
- 6 L. P. Grayson, Analysis of competitive situations, in *Adaptive control systems* (E. Mishkin and L. Braun, Jr., eds.), pp. 452-478. McGraw-Hill, New York, 1961.
- 7 M. A. Harrison, *Introduction to switching and automata theory*. McGraw-Hill, New York, 1965.

- 8 D. M. MacKay, The logical indeterminacy of a free choice, *Mind* 69(1960) 31-40.
- 9 J. Piaget, *The child's conception of number*. Routledge and Kegan Paul, London, 1952.
- 10 K. H. Pribram, Limbic system, in *Electrical stimulation of the brain* (D. Sheer, ed.), pp. 311-320. University of Texas Press, Austin, 1961.
- 11 K. H. Pribram, Reinforcement revisited: A structural view, in *Nebraska symposium on motivation* (M. R. Jones, ed.), pp. 113-159. University of Nebraska Press, Lincoln, Nebraska, 1963.
- 12 M. O. Rabin, Real time computation, *Israel J. Math.* 1(1963), 203-211.
- 13 J. S. Schwartzbaum and K. H. Pribram, The effect of amygdectomy in monkeys on transposition along a brightness continuum, *J. Comp. Physiol. Psychol.* 53(1960), 396-399.
- 14 J. von Neumann, *Theory of self-reproducing automata* (A. W. Burks, ed.). McGraw-Hill, New York, 1966.
- 15 J. von Neumann and O. Morgenstern, *Theory of games and economic behavior*. McGraw-Hill, New York, 1953.