A SUGGESTED MODEL FOR INFORMATION REPRESENTATION IN A COMPUTER THAT PERCEIVES, LEARNS, AND REASONS*

Peter H. Greene
Committee on Mathematical Biology
The University of Chicago
Chicago, Illinois

Summary. This paper investigates how computers might represent enough of the structure of the percepts and concepts they handle so that they may sensibly be said to deal with the meaning of these things, rather than just to sort and recombine mere labels for the operator's percepts and concepts. One of the main requirements is that each element of information contain partial representations of many other elements and schemata for their interconnection. Some of these requirements may be met if it proves feasible to represent information in the form of vectors (such as modes of oscillation of a complicated network or resonator) which may be resolved into components in various coordinate systems. These systems represent various points of view from which the information may be regarded, and some of the information in each system is elicited by a probabilistic mechanism for use by a conventional computer.

This paper forms part of a program[1] of investigating mechanisms whereby computers can represent enough of the structure of the percepts and concepts they handle so that they, rather than the human operator, can be said to deal with the meaning of these things. The earlier paper stated some of the main goals; this paper proposes a model which may achieve some of these goals. We saw[1] that present-day computers deal primarily with external relations among concepts which are given in a form that does not represent their inner structure. Thus the concepts may be sorted and combined, but their meaning resides in the mind of the operator. We saw that an important task which precedes comparison and abstraction is the formation of impressions into representations adequate to sustain abstraction. Moreover, this primary challenge of forming impressions into logical elements is inseparably connected with the formation of a rule which gives the ordering and interdependence of the logical elements. Each element must contain partial representations of many other elements and schemata for their interconnection. We explained that if we regard the concepts of the computer as copies of a definite world of facts, we are doing the computer's job of carving out significant units. The particular units constructed by the computer would depend upon the generative principles of connection we have mentioned, and we were led to expect clusters of related thought precursors formed around individual generative principles built into the computer. We spoke of the need for analyses of meaningful units richer than their description in terms of conjunctions of atomic parts or their negations. Finally, we discussed problems involved in the important task of becoming acquainted with causal relations and the potentialities of things and actions—what would happen in situations not actually existing. The bibliography cited important studies of psychological behavior exemplifying all the preceding considerations. These studies, together with the philosophical studies there cited and with comments on the relevance of these ideas to computers, are reviewed in a previous paper.[2]

This paper proposes a new model for the representation of information in a computer which, if it proves feasible to realize, would lead to all the features of behavior that we have called for in the preceding summary, including certain structural properties of behavior resembling many of the psychological properties of perception, learning, and higher mental functions. A problem which will remain to be solved before such a model can become useful is to specify the precise way in which the properties will appear and get those situations to be just the ones we are interested in for practical reasons. Some of the motivation for part of the model was presented in an earlier paper by the author.[3] The present paper explains the model itself in a more systematic way and introduces features with numerous consequences not then known. Table 1 summarizes the various parts of the argument and indicates their interdependence.

Initially the model was intended to simulate important features of Gestalt perception, with the understanding that a more adequate theory could not arbitrarily separate perception and thinking. It was therefore very gratifying to discover that upon the introduction of a missing part of the model required for the understanding of Gestalt perception, the model without further extension predicted a type of behavior resembling theoretically inferred mechanisms of learning, thought processes, and certain integrated action patterns performed by animals.

In a crude pattern recognizer capable of recognizing a class of patterns such as the front door lock of an apartment building, the discriminating features, or "perceptual units," are rigidly built into the objects to be discriminated. But we know that more sophisticated recognizers cannot be provided with such ready-made units, and it is our job to understand how the task preceding discrimination, namely the

*This research was supported by the Office of Naval Research under Contract No. Nonr 2121(17) NR 049-148. Reproduction in whole or in part is permitted for any purpose of the United States Government.

formation of stable units to be recognized and discriminated, may be accomplished. A machine (or person) lacking this process would combine all sorts of stimuli into meaningless groups. The model presented here is intended for that part of a machine which produces the stable, meaningful units, and is thus intended not to replace but to supplement other types of perceiving or reasoning devices.

Now let us state the kind of mechanisms which will be used, although only the subsequent discussion will reveal what connection they have with the problems that we have outlined. Information is to be coded as patterns of oscillation of complicated resonators or networks of oscillators. In particular, these patterns may be resolved into superpositions of normal modes of various kinds, and these normal modes serve as the symbols for certain stable percepts or concepts—e.g., for "good" Gestalten. The same pattern may be resolved into superpositions of any one of a number of complete sets of normal modes. Such transformations are carried out reversibly, and the modes of any such set will be thought of as representing possible outcomes of the elicitation of information regarding one particular aspect of the total pattern. The above processes are intended to simulate human processes which are normally never at the level of awareness. Elicitation of this information for responses and simulated perception at the level of awareness will require the following mechanisms (for reasons which cannot be guessed before the explanations which follow): First, the total pattern must be split into one of the possible sets of normal modes. Second, each such mode must be multiplied and averaged with the same shot noise, and the mode thus yielding the largest result selected to produce the percept or response and then returned to the population of total patterns. Subject to slight elaboration later, these are all the mechanisms we shall need: coding as modes of oscillation, together with a mechanism for elicitation involving random selection and recombination. For concreteness, I like to think of microwave modes in some sort of complicated resonator, but I do not know whether such a realization might be feasible. Even though no realization is known, it still seems worthwhile to derive unexpected psychological properties of mechanical processes.

The argument will be presented by indicating the new perceptual features introduced as the properties of the model are incorporated, one by one. At each stage of the argument appropriate behavioral desiderata, chosen from the topics of the first paragraph, will be shown to motivate the incorporation of the next mathematical property of the model. Thus the specific mathematical model just outlined is derived from such topics, which will form the bulk of this paper, and must not be regarded merely as philosophical background. Once these arguments are understood, and not before, the simulated psychological aspects of the behavior of the model may be pointed out rather simply.

Since this paper will propose a new means of information representation in computers, it is important first to have for a guide an intuitive idea of the kind of thing which is being attempted. We wish to build a machine that performs the necessary steps prior to discrimination and abstraction, and we are particularly interested in the versatility displayed in the processes of pattern formation and stabilization. Perception automatically adjusts itself to variations of the stimuli in a multitude of integrated ways, and it continually leads to the formation of new meaningful perceptual elements. Thus we seek a model which offers the possibility for such flexibility. Heuristically speaking, we shall not seek numerous models in each of which some class of response patterns is "built in," but rather a model which may be conceived of intuitively as a sort of arena in which complicated patterns can spring up, interact, and evolve "by themselves." We cannot hope to specify perceptual processes in all their complexity, but must allow them to arise in an evolutionary process of elaboration in which what is given consists of certain primordial elements, certain laws of transformation, and the opportunity to carry out these transformations. This is in accord with many mechanisms which are believed to operate in human perception and thought.

In the previous presentation of motivation for part of the model,[3] reasons were explained for wishing to represent "good" Gestalten by normal modes which had mathematical properties similar to those of eigenstates and transformation amplitudes in quantum mechanics (although the physical realization has nothing whatever to do with quantum mechanics). This is because these mathematical structures lead to theorems specifying behavior depending upon relational aspects of complex configurations and resembling perceptual phenomena. The first part of this paper will add further psychological features to the model and will be self-contained with no essential reference to quantum mechanics.

In accord with the program of studying the initial stages of perception which provide stable perceptual units for discrimination, we assume the existence of a device that can recognize perceptual units which we have transformed into sufficiently stable and standard form. Thus we start with the existing level of technology of pattern recognition. What might we add by developing the idea of coding as modes of oscillation of complicated and changeable networks or resonators?

First of all, we have the obvious fact that our perceptual units will be very complicated. They might be sorted and combined in the same ways as conventional symbols, but in addition they possess elaborate internal structure which may be used as a vehicle for the expression of relationships with other symbols and an agent or object of transformations induced by interactions with other symbols. Next, we have the equally obvious fact that complicated modes of oscillation "spring up by themselves" in the sense of our intuitive requirement, specified perhaps only by some frequencies or simple symmetries of the impressed energy. These patterns are more complicated than any we could build directly, and this phenomenon of a complicated pattern springing up from a much simpler one might allow for richness of perception and thought without making unreasonable demands upon the complexity of the coding and memory. Evidence for the storage of ideas in the form of thought precursors is cited in a previous bibliography[1] and reviewed,[2] and such a method might provide a logical extension of the procedure of Shaw, Newell, Simon, and Ellis[4] in constructing their Information Processing Language. In this language the data are not inert and structureless but are provided in the form of data programs, and the data are obtained by executing these programs. A list of data may be specified by a list of processes that determine the data. These authors explain that their approach leads to a computer that contains at any given moment a large number of parallel active programs frozen in the midst of operation and waiting until called upon to produce the next operation or piece of data. Development of our model might provide feasible ways of "freezing" extremely complicated programs.

Two topics which immediately come to mind in connection with modes of oscillation are the superposition of patterns into complicated structures and the resolution of complicated patterns into linear combinations of basic oscillations. Although there is no more information in a linear superposition than in the components, we shall deliberately be working with a language in which, as befits perception, some information will be in the focus of attention while other information will temporarily be relegated to the background. Otherwise, only chaos would be perceived. We shall see later how our symbols will contain the latter kind of information in a latent form and how this information can be brought into the awareness of the recognizing part of the machine by a process which consigns other information to latent status. Thus the available information in superposition may be different from the available information in the components.

The preceding paragraph suggests that we are interested in observing the same pattern from different points of view in order to bring out different pieces of information. It seems natural to investigate the simplest possibility consistent with our concern for the superposition and resolution of modes, namely that the various points of view are represented by the components of the complicated symbol in various coordinate systems. That is to say, we resolve our symbol into combinations of various kinds of modes. What new perceptual features would this introduce? Since the components of the symbol in various coordinate representations are related by linear transformations, we shall have, once the observation procedure is explained, a schema for the connection and interdependence of observations called for in the first paragraph. This rule of connection will not be as rigid as it may seem at first because flexibility and variety will be introduced by the "latent" information in each representation which shows up only in other representations, so that from the point of view of the recognizing system, one representation does not tell all about the others. Second, the process of utilization of information will itself introduce probabilities and possibilities for recombination and variation of symbols, and this process can lead to the evolution of more complex forms.

In accord with the first paragraph, we see that any perceptual element in one coordinate system will contain partial representations of many other elements in that system or other systems. Moreover, linear transformations between representations do not coordinate particular parts of patterns in one representation with particular parts in another but are of a holistic nature, in accord with the first paragraph and with elementary facts of Gestalt perception.

Let us introduce more structure by considering the mathematical properties of the coordinate systems. The natural kind of coordinate system is provided by normal modes of oscillation, or more precisely, eigenfunctions of linear operators which characterize the system, that is, functions which are merely multiplied by a constant when the operators are applied to them. Later on we shall explain how these normal modes will provide precisely defined patterns which we shall identify with "good" Gestalten. What new perceptual characteristics are thus introduced? First of all, we have the general feature that stable perceptual units would have to appear in certain discrete, reproducible, integrated forms, which may change discontinuously from one to another with nothing in between. This is such a characteristic feature of human perception that we often tend to forget those instances in which we do see indefinite and merging forms. Evidence that stable percepts rest upon a basis of merging, streaming, scintillating, and reduplicating forms is adduced by Schilder.[5] This is the kind of thing our model tries to do; that is, to base stable perceptual units upon wave-like functions which permit superposition, decomposition, and transformation.

Specifying the operators of our system determines these normal modes which will constitute the stable perceptual units. Thus the system actively "carves out" units, in the terminology of the first paragraph, by means of operators built into the system or attained through a process of learning. When we discuss the process of elicitation and utilization of these perceptual elements, we shall see the interrelations among clusters of elements belonging to the same operators, and we shall derive behavior related to the clusters of thought precursors formed around individual generative principles mentioned in the first paragraph. We shall also derive a certain interchangeability in the early stages of perception among the modes belonging to a single operator.

Our last observation at this stage pertains to the linearity of the system. Characteristics of the system will be represented by operators, and perceptual units by combinations of their eigenfunctions. As a consequence, we have the fact that although the operators may be related by complicated nonlinear relations, the perceptual units will be linearly related. This results from the general mathematical fact that the eigenfunctions of any of the operators we shall consider can be linearly superposed to produce any function we shall require.

Our observations so far have been very general. Nonetheless, they are essential both to the derivation of the model and to the comprehension of the purpose of the model once derived. Our method so far has been to assume that a useful model can be constructed and to examine what the various parts will do. Only when we know this can we know how to put them together. Our conclusions so far, supplemented by the equally general comments upon selective awareness and group properties of perceptual transformations, will, as indicated in Table 1, lead to conclusions with a great deal of content.

Our next stage will be to see what the operators and eigenfunctions mean. We shall frequently be interested in the results of a number of different types of observation of the perceptual unit forming system by the recognizing system. Let us suppose that an observation of some type, which we shall label A, might reveal any one of the states a_1, a_2, etc., an observation of type B one of the states b_1, b_2, etc., and similarly for observation C, and so on. In describing events at a perceptual or mental level of organization, we customarily use a form of language in which we are able to formulate laws governing the influence of past events upon the present without considering the details of trace structures which exist at all intermediate times. We therefore seek a formalism which connects behavior at a single earlier observation with that at a single later observation, rather than providing a continuous description of observations. In particular, we shall often be interested in the conditional probabilities $P(a_i|a_j)$, $P(a_i|b_j)$, etc. that an observation of A will reveal the state a_i, given that a previous observation has revealed a particular state. In situations where an A observation must yield some a state, we have the equality $\sum_i P(a_i|b_j) = 1$ for an arbitrary state b_j.

We characteristically do not see all aspects of an object simultaneously but typically find one aspect at the focus of attention, while other aspects remain indefinite until some effort is made to perceive them clearly. If the perceptual experience depends in some statistical fashion upon a set of "observations," i.e., elicitations of coded information by the recognizing part of the machine, we might expect that observations of the first type all yield the same result, while results for the second type of observation will be less coherent. If we interpret our P's above to apply to two observations in this observed set, we might then have a situation in which $P(a_k|a_i) = \delta_{ki}$ (i.e., 1 if k = i and 0 otherwise), while $P(b_j|a_i)$ might take on various values. This statement is intended to cover a number of situations. For instance, we may suppose that two observations are performed in immediate succession, the second being applied to the information elicited as the result of the first. Then the requirement states that if an observation of type A applied to the original information has revealed the information a_i, then further observations of type A applied to the elicited information a_i can only continue to reveal a_i, while an observation of a different type B applied to a_i might reveal any of the b_j. According to a second interpretation, elicited information functions slightly differently. Suppose that a particular perception or action of type A is not correlated with a single element a_i, but rather with a particular distribution of weights among all the elements of the set $\{a_1, a_2, \ldots\}$. A natural way of assigning weights is to perform a number of independent A observations upon the same information, thus obtaining a population of a's, and to let the weight of a_i be the proportion of a_i's in that population. Then the probability $P(a_k|a_i)$ may be interpreted as the probability that a randomly chosen member of the population is a_k, given that a previously randomly chosen member was a_i. The condition $P(a_k|a_i) = \delta_{ki}$ then states that all the members of the population are the same, so that the percept or act has only one nonvanishing component, and thus the condition allows for the existence of such "elementary" acts. A weaker requirement might allow for the existence of several, but not all components. This requirement, general as it may be in isolation, will in connection with our other general requirements shortly turn out to impose very specific requirements on the mathematical structure of the model.

The next stage in the development of the model is to try to combine this statistical formalism with the system of linear transformations which we have previously discussed. In order to simulate important features of human perception, we wish to find ways of relating the conditional probabilities between pairs of sets so that transformations may be compounded, or so that they may be decomposed into transformations involving intermediate sets, as for example by expressing $P(c|a)$ in terms of $P(c|b)$ and $P(b|a)$. These transformations must preserve the sums of probabilities, so that they add up to one as before. Our first thought is to let the previously discussed linear transformation scheme be identical with the well-known matrix multiplication of probabilities, $P(c|a) = \sum_j P(c|b_j)P(b_j|a)$.

However, if we now refer back to the discussion of selective awareness in the preceding paragraph, we see that we shall in general be unable to obtain the situation there described, in which $P(a_k|a_i) = 0$ if $k \neq i$. The reason is that if the same matrix combination is to hold for all kinds of observations, we find by replacing c by a_k and a by a_i in the multiplication rule that $P(a_k|a_i)$ will be the sum of several positive terms, and this contradicts the requirement that it equal zero. That is to say, we can get from a_i to a_k via some of the b's, contrary to our desire, unless we can define a transformation according to which many of these paths cancel each other.

Landé,[6,7,8] in works which gave me my first notions of what to do about my conviction that something useful might come of analogies between quantum mechanics and perception, points out that there is just one known system of transformations which will do. If we restrict ourselves to the case of symmetrical probabilities, $P(b|a) = P(a|b)$, (a restriction for which certain heuristic justifications may be offered), then the transformations may be defined by complex valued matrices

$\|\Psi_{kj}\| = \|\Psi(a_k|b_j)\|$, such that the squared modulus $|\Psi_{kj}|^2$ of the element Ψ_{kj} equals the corresponding probability $P(a_k|b_j)$. Then it can immediately be seen that the requirements that the probabilities add up to one and that they satisfy the condition for selective awareness are equivalent to the condition that the matrices just defined be unitary matrices (the complex analogue of orthogonal matrices). We shall call the Ψ's by the name which analogous quantities have in physics, probability amplitudes, to distinguish them from the probabilities P.

Of course, we are not denying the validity of the well-known law of matrix multiplication of probabilities in favor of the matrix multiplication of probability amplitudes. The two rules apply to different situations, the former determining, as it must, the probability that c is elicited given that b and a have been elicited, and the latter determining the probability that c is elicited if only a has been elicited, expressing this probability in terms of two conditional probabilities that refer to observations not actually performed. This difference of course requires that we introduce a suitable precisely defined way in which the elicitation of b modifies the information.

Now we may connect this discussion of probability amplitudes with the previous groundwork concerning transformations between coordinate systems by remembering that the unitary matrices are precisely the matrices which transform between coordinate systems defined by sets of orthogonal and normalized vectors. And now we may turn the argument around and say: given a coding of information as complex oscillations, with components in various coordinate systems, we wish to define a process which will elicit the information corresponding to a particular component with a probability proportional to the squared modulus of the amplitude of that component. As emphasized in Table 1, we are led to this result only by combining the above desiderata at the behavioral level, and with this result we begin to collect the promised fruit of all the discussion that has gone before. Once we define a suitable mechanism for elicitation of information, we shall be able to use the discussions of qualitative properties of mathematical structures to show how the model will simulate certain important structural characteristics of thought. Then, when we add a suitable mechanism for the change of the Ψ in time (or with respect to some other parameter), we shall be able to derive some typical properties of Gestalt perception.

We shall shortly define a mechanism for the elicitation of information with a probability equal to the squared modulus of its amplitude. Accepting for the moment the assertion that this can be done, we may immediately derive operator and eigenfunction equations and thereby specify the eigenfunctions which will be the coordinate vectors. We may expect the observable properties of the oscillatory states to depend upon numerical functions which take different values in different states, because the input of the recognizing part of the machine can always be expressed in such a form. These values will be precisely defined without any statistical spread (we shall say "sharp") in the set of states for which some perceptual aspect is definite. In such a set of observations of type A, let us denote the precisely defined numerical value in state a_k by the same symbol a_k. Then for the mean value of a over all states corresponding to observation A, we clearly have $\bar{a} = \sum_k a_k P(a_k|s) = \sum_k \Psi^*(a_k|s) a_k \Psi(a_k|s)$, by definition of the relation between Ψ and P, where s denotes the state of the system in an arbitrary coordinate system. What will be the mean value of a in terms of the B coordinate system? If in the expression for $\bar{a}$ we perform the linear transformation to the B representation, we find that $\bar{a} = \sum_m \Psi^*(b_m|s)\mathbf{A}\Psi(b_m\ s)$, where $\mathbf{A}$ is the linear operator represented in the B system by the matrix having elements $A_{mn} = \sum_k \Psi^*(a_k|b_m)a_k\Psi(a_k|b_n)$. According to this formalism, the operator $\mathbf{A}$ is represented in the A system by the diagonal matrix $A_{kk'} = \delta_{kk'}a_k$. This standard kind of computation shows the origin of the operators mentioned in our discussion of stable configurations.

Now all one has to do to obtain eigenfunction equations defining the coordinate systems is to multiply both sides of the defining equation for A_{mn} by $\Psi(b_n|a_i)$ and sum over n, making use of the unitary nature of the Ψ's. The result is the eigenfunction equation $\mathbf{A}\Psi(b_m|a_i) = a_i \cdot (b_m|a_i)$. Given the matrix representing $\mathbf{A}$, one may solve this equation to obtain the various a_j and corresponding $\Psi(b_m|a_i)$. This equation states that the transformation amplitude matrix from any set of states to the set in which a particular type of observation is sharp must be an eigenfunction of the operator corresponding to that observation in the former set of states, and the numerical value a_i identifying the state must be the associated eigenvalue. We began by merely requiring a_i to be some number associated with the state. The present conclusion follows entirely from the requirement that probabilities be the squared magnitudes of the Ψ's, and we were led to this requirement by considering general behavioral desiderata.

Examination of the mathematical meaning of the Ψ's reveals that the eigenfunction $\Psi(b_m|a_i)$ is simply the set of coordinate values representing an a vector in the B system, so that we may alternatively consider basic coordinate vectors to be eigenvectors of operators, or the transformation amplitudes to be eigenfunctions of operators. It is easy to show[6,7,8] that our assumptions have implied that the operators must be Hermitian, and their eigenvectors will thus be orthogonal and provide suitable coordinate systems.

In a previous paper[3] I discussed a mechanism for the successive transformations of a set of inputs to a network, which produced a group of transformations depending upon a parameter τ. This parameter was the analogue of time in quantum mechanics, although in the case of the network it denoted spatial distance across a network from an input mosaic to an output mosaic. It was briefly noted that if we had a network containing a large number of alternate paths for the signals, each path being specified by distributed phase lags, we would obtain transformations of coordinate systems described by equations analogous to the Schrödinger equation, namely $\partial\Psi/\partial\tau = i\mathbf{H}\Psi$, where $\mathbf{H}$ is a Hermitian operator defined in terms of the phase lags. Such a formalism provides one interpretation for the earlier statement that normal modes will be stable, for the τ dependence of eigenfunctions of $\mathbf{H}$ will be entirely exponential with an imaginary exponent and will cancel out of the observation probabilities when we take the squared magnitudes in the elicitation process. We must ultimately be more specific about the detailed nature of $\mathbf{H}$ and other operators which determine the behavior of the system, because we will not be satisfied merely with quasi-perceptual behavior in unspecified situations. However, we shall be able to draw a large number of conclusions which follow from the formalism alone and which will thus be true independently of the precise nature of the operators. These sometimes enable us to work backwards from behavioral desiderata and thereby eliminate whole classes of possible operators. However, since this paper will be confined to conclusions which follow from the general formalism alone, we shall not discuss the problems of further specification of operators. We shall explicitly refer to the transformation equation of this paragraph only in connection with certain so-called "Gestalt constancies."

The last mechanism we shall have to introduce is the one which selects normal modes with a probability proportional to their amplitudes. The problem is as follows: Suppose the oscillatory pattern representing, for instance, all possible ways of looking at a complicated stimulus is a function ψ of a number of variables, which we shall not specify in detail. For illustrative purposes we may consider the excitation at some point to be a function of time, $\psi(t)$. We have supposed all along that ψ may be resolved into superpositions of various kinds of normal modes, or eigenfunctions. Thus, alternatively, $\psi(t) = \sum_n a_n\phi_n(t)$ or $\psi(t) = \sum_n b_n\xi_n(t)$, etc., depending upon the point of view from which the total configuration is being observed. Suppose the point of view is such that ψ is split into the ϕ_n. The coefficients a_n will be complex numbers. We want to specify a mechanism which will pick out one of the normal modes and discard the rest, in such a way that the probability of picking the kth mode ϕ_k is equal to $|a_k|^2$, or actually to $|a_k|^2/\sum_n|a_n|^2$ in case ψ is not normalized. How does one pick something with a probability proportional to the square of a complex number? Wiener[9,10] has considered this question in connection with the identical situation in quantum mechanics, and his conclusion is that he can think of only one non-quantum mechanical process which could do that. Therefore I decided to examine the consequences of incorporating such a process in the present model, and it was at this point that the model turned out automatically to possess many interesting structural properties of thought, together with the ability to learn by being "rewarded" in a simple way.

This selection process is quite simple and will now be described. Wiener gives two variants, of which only one will be considered here. Let $Y(t)$ be the integrated complex-valued output of a shot noise generator. Then dY, if it existed, would be the instantaneous output. Actually, this is a physically unrealizable idealization of a Gaussian random process. What one actually measures is the increment ΔY for a small increment of time. But the intuitive notation can be given a precise sense. It turns out that the integrals $\int\phi_n(t)dY(t)$ are all random variables with Gaussian distributions, and in case the ϕ_n are all normalized the distributions will all be the same. Moreover, if the ϕ_n are orthogonal to each other, that is, $\int\phi_n^*(t)\phi_m(t)dt = 0$ for $n \neq m$, then the Gaussian distributions will be mutually independent. Our general considerations led us to consider Hermitian operators, which have orthogonal eigenfunctions, so that this condition is satisfied. Now let us calculate the time averages of the normal components of ψ multiplied by the shot noise: $A_n = \int a_n\phi_n(t)dY(t)$. This is accomplished by a physical device which finds the statistical correlation of two inputs. Finally, let these numbers A_n control a gate which lets through the mode having the largest A_n. Since the A's are random variables, this process picks the modes statistically, and the probability of picking the kth mode turns out, as shown by Wiener, to be just what we want, namely $|a_k|^2/\sum|a_n|^2$. We shall not go into detailed consideration of the suitability and implications of different ways of comparing the A's,[10] except to say that if comparisons are made in pairs, the order in which the pairs are picked makes little difference, and that it is also possible to compare pairs of superpositions of several modes and then split these up further. This latter point will have a useful interpretation. A possible way of performing the process is indicated in Figure 1. This general type of process turns out to have numerous interesting consequences, which we are finally ready to understand, all the groundwork having been prepared.

Restrictions on the length of this paper preclude all but the briefest discussion of these consequences. Further information, especially in connection with an attempt to understand the representation of information in the brain, will be published elsewhere. However, making use of the background of information which we have developed about the properties of the mathematical structures, we may derive most of the interesting features in a few sentences. Here is a list of them.

We begin with the ways in which information coded as modes of oscillation is modified in the process of elicitation by the recognizing part of the machine. The information exists in a population of "wave functions" ψ, which are split up in various ways into superpositions of components ϕ_n or ξ_n, etc. The process of elicitation operates upon one of these resolutions and selects one or more components, ignoring the rest. The selected components will then be identified or produce some action or be subjected to further transformations, and they may or may not subsequently be returned to the population of wave functions, depending upon the purposes to be served. Observing the same information from a different point of view is interpreted as splitting the wave function into a different set of components and then selecting some of these components. The various components in any particular resolution correspond to various ways the pattern could appear when looked at from the corresponding point of view, or to various activities in a particular mode of behavior.

If there is just one ψ (instead of a population of them), then an observation from one point of view might elicit any one of the ϕ's, but subsequent observations from the same point of view will continue to elicit the same component. This is one way in which an ambiguous "percept" becomes clarified in the process of bringing it to the "awareness" of the recognizer. Another, and possibly more significant, way will be explained when we come to the effect of the shot noise on the latent information.

If the information selected by the elicitation process is returned to the population of wave functions, an interesting phenomenon results. There will be a repeated process of splitting the wave function into components of some type, discarding some of the components, returning the remaining randomly selected components to the population, splitting them into other types of components, and so on. This process brings about the reappearance of components which have previously been discarded. For instance, if only the ϕ_1 component is retained, and it is split into a superposition of several ξ components, and only one of the ξ's is retained, this ξ will in general when resolved in the ϕ system contain many of the ϕ's, and thus contain ϕ's which had previously been discarded. This is a perfectly elementary fact in any vector space, which has two immediate consequences for us. The first is that if an input excites one mode, then the result of the resolution and selection process is to excite other modes in the same coordinate system. In other words, the model exhibits a form of <u>association</u>. It is a standard problem in probabilities to calculate the association strength as a function of such things as the number of iterations of the resolving-selecting process, and the problem is being investigated. The second consequence will be explained in connection with learning.

The probability of finding a particular value for the numerical function which identifies the possible results of some type of observation is determined by the squared magnitude of the corresponding amplitude and is thus independent of any factor of the form $e^{j\alpha}$. This means that an arbitrarily large amount of information may be included in the phases of complex exponential factors of the amplitudes of modes belonging to some type of observation without making any difference in the results of that observation. However, this information will affect the results of a different kind of observation. For example, suppose that $\psi = a_1 e^{j\alpha_1}\phi_1 + a_2 e^{j\alpha_2}\phi_2$, and that $\phi_i = \Psi(1|i)\xi_1 + \Psi(2|i)\xi_2$. Then in an observation which reveals ϕ's, the probability of finding ϕ_1 and its associated numerical value will be $|a_1|^2$ independent of α_1 and α_2. However, in the ξ system $\psi = [\Psi(1|1)a_1 e^{j\alpha_1} + \Psi(1|2)a_2 e^{j\alpha_2}]\xi_1 + [\Psi(2|1)a_1 e^{j\alpha_1} + \Psi(2|2)a_2 e^{j\alpha_2}]\xi_2$, so the probability of observing ξ_1 will be $|\Psi(1|1)|^2 \cdot |a_1|^2 + |\Psi(1|2)|^2 |a_2|^2 + \Psi^*(1|1)\Psi(1|2)a_1^* a_2 \cdot$

$e^{j(\alpha_2-\alpha_1)} + \Psi(1|1)\Psi^*(1|2)a_1 a_2^* e^{j(\alpha_1-\alpha_2)}$, which depends very much upon α_1 and α_2. This is what was meant by the statement that in any representation there would be latent information which did not show up in that representation but only in others. We first introduced the idea of selective awareness by requiring that the states in the focus of attention be clearly distinguishable, while the others are uncertain. The present consequence shows that each point of view will reveal information which will not be revealed by other types of observation. We may look at this fact in two ways. One way is to say that a perceiver must at any time disregard many things in the stimulus pattern in order to have any intelligible perception. The converse way is to say that knowledge of one aspect or even several will not include all the information needed to understand other aspects and their interrelations. The additional information required is the rule for the ordering and interconnection of impressions[1] mentioned in the first paragraph. This and several other interpretations of the meaning of the mathematical formalism are well known in the analogous situations in quantum mechanics, and credit has previously been given[3] to the sources of some of these formulations.

The next three consequences concern Gestalt perception. If a clear perception or an integrated action depends upon the coherent behavior elicited from a population of wave functions ψ, then we might have a situation in which the clarity depends upon all the ψ's being the same. This is called a <u>pure case</u> of an assembly in quantum mechanics; otherwise the assembly is a <u>mixture</u>. In connection with the preceding consequence, we may note that an observation from one point of view will turn a pure case into a misture with the same response probabilities <u>in that point of view</u>, but will disrupt the coherence of responses from other points of view. It is a simple mathematical fact that a composite system may be pure while its subsystems are not. Moreover, an observation of one subsystem may produce a change in the pureness of another subsystem. This fact has been discussed[3] in relation to the well-known facts of Gestalt perception that (a) a stable perceptual unit, or "good Gestalt," may become unstable and hard to see if embedded in a larger pattern, and (b) what was originally a single Gestalt may be broken up in various ways by the addition of new lines in the visual stimulus pattern, so that the new Gestalten will not coincide completely with any of the old ones. Illustrations were given in that paper. This is of great importance in a perceptual machine, because we might wish to scan a pattern to find a meaningful pattern hidden it it, while at the same time we do not wish every chance grouping of lines in a stimulus to be perceived as a meaningful pattern.

The next Gestalt property rests upon the previously considered assumption that the transformation given by an equation like $\partial\Psi/\partial\tau = jH\Psi$, the solution of which is some function $\Psi = \sum_n a_n\Phi_n$. In this case the following situation is well known in quantum mechanics. Suppose that conditions are changed in such a way that H becomes a function of a parameter μ. Then the solution of the equation will likewise become a function of μ which we may write $\Psi(\mu) = \sum_n a_n(\mu)\Phi_n$, where $a_n(0) = a_n$. The new response probabilities will be $|a_n(\mu)|^2$. Now one might expect the change in a_n to be roughly of the same order as the change in μ. It turns out instead that the change in a_n is of the same order as the rate of change of μ, that is, $d\mu/d\tau$. Thus, provided that μ changes very gradually, it can become quite large without inducing a significant change in the response probabilities. Examples were given[3] of very important analogous phenomena in perception, the Gestalt constancies. In order to perceive objects as meaningful units, one must ignore certain inhomogeneities, and it is well known, for example, that large spatial variations of luminance are not perceived, provided that they are gradual; but they become prominent when any discontinuity is introduced. An illustration has been explained in detail.[3]

The third Gestalt property likewise requires a wave equation. There is some reason to believe that while the constancy phenomena would occur with a wide class of wave equations, these equations must be first order in τ, as written above. In this case the time dependence of a wave function which is an eigenfunction of H will be exponential with an imaginary exponent, and thus will cancel out of the response probability. Probabilities arising from a superposition of such steady state solutions will, however, be time dependent, so that one source of change is derived from the interference of such steady state solutions for different states associated with H. It is possible to make the interpretation that certain perceptual changes, inexplicable when only one point of view is considered, come from the information, not available to awareness in that point of view, about the range of possibilities of other aspects. Moreover, one can calculate the frequencies at which the perceptual transitions will occur. Because of the determinate form and possible discreteness of the normal modes, a change might be as abrupt as in the perception of the well-known picture of a staircase which can look as though it is being viewed either from above or from below. In general, the resolutions of ψ into ϕ's and into ξ's given in connection with latent information may be used to predict a possible interference effect of a second stimulus introduced into the field. The situation is analogous to the interference effects observed when an electron has the opportunity to go through two holes.

Next we shall consider a general feature of the model which resembles the mode of organization of certain highly integrated action patterns in animal behavior. Reference to Figure 1 reveals that the information is elicited by a process which has two stages: first, splitting the wave function into normal modes, and second, selecting a mode by a mechanism utilizing shot noise. The first stage requires either very specific filters which separate the modes, or else (as shown) precise copies of the modes which can be used to obtain the various components in the same way one finds terms in a Fourier expansion. The second stage requires that the noise level be high enough so that the results of the integrations may exceed any thresholds which exist in the component which compares them. There is evidence from the study of instinctive behavior in animals (previously cited[1] and reviewed[2]) that complicated action patterns are broken down into a sequence of acts, each one released under the two conditions that a very specific stimulus be present and that a drive level be high enough. People are trying to build machines that do remarkable things, and it seems wise to see how those things actually are accomplished in nature. Even if the detailed mechanisms of instinctive behavior are not the same as those of the present model, it may be of interest that they can be described by the same flow charts at some level of abstraction.

Next we shall examine some ways in which the behavior of the model resembles structural characteristics of thinking. We have mentioned the importance of the fact that the wave function is split into many modes and some are selected randomly. The random selection, aided by systematic biases, of which we shall later give one example in connection with learning, can lead to the evolution of complex forms. It seems to be typical of thought and action processes that many alternatives are proposed out of which some are selected. In perception we have, for instance, the previously mentioned evidence that stable percepts emerge from a conglomeration of wavelike images. In instinctive behavior there is evidence that the animal behaves in a searching or random way until the two previously mentioned conditions release the block to action and allow selection of a particular action pattern. In thought pathology there is some evidence that a large variety of symbolically related ideas can emerge as the final thought product, which would be suppressed precursors of normal thought processes. In the development of thought in an infant, according to hypotheses of Freud, various rudimentary images become associated with drives. In the absence of the thing which will satisfy the drive and reduce its level, these images become activated in some way which raises them to hallucinatory vividness (like the images in dreams). If the activation level is below threshold, the necessary concentration may be achieved by utilizing energy from another image, or by fusing several images related to the same drive so as to pool their activation energy. Any such image may stand for the drive in this form of thought, termed the primary process by Freud. Ideas may stand for their opposites, because at this stage there is no mechanism to distinguish between evoking an image to affirm it or to deny it. Thus ideas are associated by their relevance to the same drive, not because of formal logical relationships. Moreover, mutually contradictory ideas can exist side by side, for there is no mechanism for comparing them and rejecting one or both. We can find familiar examples of such thinking in the substitutions and fusions of images in dreams and in slips of the tongue, but it is the aim of the arguments[1] cited in the first paragraph of the present paper to make clear the fact that any organized and independent thought processes, however logical their outcome, must employ such a process. While such thinking is indispensable, it must be supplemented by a secondary process which enables the infant to get along in the world. Under the influence of learning by checking his ideas with the real environment, he must voluntarily delay the drive induced discharge of images so they are in accord with reality, and he must make connections, not between ideas which are related merely to the same wish, but between ideas which are related by virtue of having the same relation to actual experience. The absence of this unlimited fusion of ideas results in less pooling of the activation energy, and the representations become less vivid, taking the form of thoughts instead of images of a hallucinatory nature. These thoughts are now manipulated logically so that they may lead to actions producing gratifying changes in the environment rather than merely the hallucinations of gratification. In logic itself it appears,[1] that not only on the lowest levels of the synthesis of meaning but at the highest propositional levels, too, reasonable statements may be produced only by a process of rejection of a multitude of lawfully produced but unreasonable statements.

If these hypotheses are true, thought, while remarkable, does not appear inconceivable, and it seems worthwhile to see whether such processes may be built into machines. An example of a machine program which has some features of this generation of clusters of ideas around individual goals followed by selection of some of them by means of an evaluative process is described by Newell, Shaw, and Simon[11] in a discussion of chess playing machines. Their program contains a subprogram built around a set of about a dozen goals, each corresponding to some feature of the chess situation. Each goal has associated with it a move generator and an analysis and evaluation procedure. The move generator associated with a goal proposes alternative moves associated with that goal, regardless of suitability, on the basis of any connection with that goal. The evaluation and analysis procedures determine the value of the move from the point of view of that goal alone. The analysis procedure is concerned only with the acceptability of a move once it has been generated by the move generator. The consequences of the moves generated by all the goals are explored by evaluating possible game continuations generated by a different generator from the one which proposes moves. An executive routine makes the final choice of an acceptable move from the fifty or so proposed moves.

The present approach is to try to get things like this to happen by themselves without having to specify precise programs in advance, in order to allow the spontaneous elaboration of symbolic complexity. Let us examine the way information is elicited in the present model. There are two stages in this process. In the first stage the wave function ψ, from which all points of view may potentially be extracted, and which may even represent a large number of ideas in some integrated portion of the machine's "mind," is split into components in one or another coordinate system. Two possible resolutions of the wave function are represented by the hexagons and squares in Figure 1B. As we have proved mathematically, each of the components contains some latent information, represented by the shading, which is not available to the "awareness" of the recognizing part of the machine, but which becomes transformed in other resolutions into information which can be elicited. This latent information is thus necessary to the reversible transformations between different points of view. At this stage of merging and splitting all information is present only in the form of potentialities. Suppose that ψ is resolved into ϕ components and the shot noise generator is turned on. One of the ϕ's will then be elicited, provided that the noise level is high enough so that the threshold of the comparing component is exceeded. The expansion coefficients give the probabilities for elicitation of the various ϕ's, but there is no way of knowing precisely which one will be elicited in a given case. Thus the noise level acts in a way similar to the drive level which elicits images. Now suppose that the noise begins at a very low level, too low for the elicitation of one of the ϕ's. The threshold may be exceeded, however, by the result of the combination of the noise with a superposition of two or more of the ϕ's, since this result will be $\int(a_k\phi_k + a_l\phi_l + \ldots)dY$ instead of merely $\int a_k\phi_k dY$, and the corresponding elicitation probability will be proportional to $|a_k|^2 + |a_l|^2 + \ldots$ instead of merely $|a_k|^2$. Thus superpositions of images will be elicited, in analogy to the condensations of dream images. In general, at this stage there is no way to refer to a mode except by exciting it, and the association between modes comes from belonging to the same resolution of the wave function, i.e., in our interpretation, belonging to the same point of view.

Now let us examine the second stage, that is, the transformations which the symbols may undergo after having been elicited by the random process. Let us suppose that the elicited output does not come directly from the resolving apparatus (dotted line in Figure 1A) but comes instead via the multiplication component which receives its input from the resolving apparatus (solid line).

Let $\Delta Y(t)$ be the instantaneous output of the noise generator, which would have no well-defined meaning if the noise were ideal shot noise, but in practice will be $Y(t + \Delta t) - Y(t)$ over some short interval of time Δt. Then $\Delta Y(t)$ will be a random variable with a Gaussian distribution which is independent of the distribution of $\Delta Y(t')$ if $|t - t'| > \Delta t$. Under these assumptions, the output will no longer be $a_k \phi_k(t)$, but will be $a_k \phi_k(t) \Delta Y(t)$. Let us suppose that ϕ_1 and ϕ_2 have been elicited with their respective probabilities, so that we now have a new population of wave functions $\phi_1(t) \Delta Y(t - t_1)$ and $\phi_2(t) \Delta Y(t - t_2)$. The ΔY's will be at different times because the modes in the population are separated out at different times. The two modes will be present in the relative proportions $|a_1|^2 : |a_2|^2$. This population, as we shall see in a moment, may be described by the wave function $a_1 \phi_1(t) \Delta Y(t - t_1) + a_2 \phi_2(t) \Delta Y(t - t_2)$, which gives the right probabilities of elicitation, provided that the comparing element averages a population of inputs, because the average of $|\Delta Y(t)|^2$ is always the same constant for all t. Now let us return to our previous discussion of the meaning of the latent information, which represents potentialities in other modes of observation. In the expressions for Ψ in the ϕ and ξ systems, let us replace the exponentials $e^{j\alpha 1}$ and $e^{j\alpha 2}$ by $\Delta Y(t + t_1)$ and $\Delta Y(t + t_2)$, respectively. Averaging over the population and making use of the independence of $\Delta Y(t + t_1)$ and $\Delta Y(t + t_2)$, we find that only the first two terms of the expression in the ξ system remain, the interference terms having disappeared. The elicitation probability for ξ_1 will thus be $|\Psi(1|1)|^2 |a_1|^2 + |\Psi(1|2)|^2 |a_2|^2$, which is the same probability we would have predicted by considering the actual population of wave functions $\phi_1 \Delta Y(t + t_1)$ and $\phi_2 \Delta Y(t + t_2)$ in the relative proportion $|a_1|^2 : |a_2|^2$ rather than the combined wave function $\phi_1 \Delta Y_1 + \phi_2 \Delta Y_2$. This justifies our use of the combined wave function. However, before the wave function has interacted with the eliciting apparatus, $\phi_1 + \phi_2$ will not give the result as a population of ϕ_1's and ϕ_2's because of the interaction terms.

It may also be seen that the probability which has just been given for ξ_1 is precisely that which would have been obtained by using the matrix combination of probabilities instead of the matrix multiplication of probability amplitudes. The preceding argument has thus fulfilled the promise made in the beginning of this paper that an elicitation mechanism would be designed which would make the group properties of transformations between coordinate systems consistent with the necessarily valid rule of composition of response probabilities. This argument is, of course, familiar in quantum mechanics, the only difference being that in quantum mechanics the random multipliers are slightly different (α is a random variable in the factor $e^{j\alpha}$), and that we are using an ordinary mechanical device to do the multiplication. What is new about the argument is that the resulting model is being proposed for a concrete device which has many features of thinking, so we shall now examine the behavioral implications of the above formal mathematical property.

Prior to the interaction with the eliciting device, the modes of any resolution contain all the information needed for their recombination and subsequent resolution in another coordinate system. The original wave function could have been subjected to any kind of observation and thus could have exhibited any one of a number of properties. These properties may be mutually incompatible, so that no individual elicited wave function could exhibit more than one, although they might all be found in a population of responses. An individual wave function elicited by one of the kinds of observation may no longer contain the information which tells about the other kinds of observation, so that awareness of one aspect may preclude simultaneous awareness of other aspects. But all these mutually incompatible properties may simultaneously exist in the form of incompletely developed potentialities in the original wave function. This is another respect in which the first stage of the present model resembles primary process thinking.

We may look at this another way. The original wave function was potentially describable by a wide variety of predicates applicable to the ideas represented by the elicited modes. But interaction with the eliciting apparatus has revealed some potentialities and at the same time has in a precise mathematical way destroyed the information needed to put the modes together again and resolve them in a new way which can reveal other potentialities. Thus the wave function after such interaction can reveal a smaller set of potentialities. It may not yet have been recognized by the recognizing part of the machine, so that from the standpoint of the "awareness" of the machine it may still be unknown, but it will now be an unknown ϕ or else an unknown ξ, rather than an unknown something which could reveal ϕ aspects or ξ aspects, or any number of other potential aspects. Even if the properties coded as ϕ_1, ϕ_2, etc. exhaust all logical possibilities with respect to that aspect of the total pattern, so that in one sense saying that it is a ϕ_1 or ϕ_2, etc. is a tautology, this information nonetheless has important behavioral consequences for the machine even before it is recognized by the machine. We saw mathematically that this will make a difference in any subsequent elicitation of information from another point of view. In addition, knowing which aspect is relevant has important behavioral consequences because it entails using the appropriate rule of ordering and interdependence[1,2] mentioned in the first paragraph of this paper. This is a necessary stage of concept formation which precedes abstraction, and it is important because it tells the machine, in a sense, along what dimension to think. We may summarize this paragraph by saying that the elicitation process, even before its results are "noticed," has partially crystallized the information by limiting the set of potential predicates. If, in our example, ϕ_1 and ϕ_2 each represented a superposition of many modes, this crystallization would have been only partial, for while the ϕ_1 and ϕ_2 have been multiplied by independent ΔY's so that the interaction terms between them have been destroyed, all the modes within ϕ_1 or ϕ_2 have been multiplied by the same ΔY, which makes no difference within the subspace of their superpositions (since $|\Delta Y|^2$ = constant). Thus within these subspaces potentialities exist which may be destroyed by further specification. All the above corresponds both to theoretical interpretations of underlying thought mechanisms and to common sense observations about gradual articulation of concepts which are not clear to begin with.

Our final observation about the second stage of information elicitation concerns logic. In distinction to the mode of representation of information in the first stage, which permits the same wave function to contain mutually incompatible potentialities, and which permits merging and resplitting, the information is now in a form in which only one of the potentialities has been realized. If the information in this form is utilized by the recognizing part of the machine, it will be displayed in a form which follows the usual rules of probability, as we have seen, and which may be manipulated by the usual logical procedures employed in computers. In fact, it is well known in quantum mechanics that the mathematical formalism used to describe the analogous features of observation—namely the projections of wave functions on particular coordinate axes—can be partially described in terms of the Boolean algebra of propositions.[12] Thus we may summarize this discussion by saying that information is changed in the process of elicitation into a form more amenable to logic, but at the expense of richness of interconnections. Thus the machine may use this modified form to explore the logical consequences of some concept, but in order to do something more creative it may have to return to the original form and start again.

So far we have seen how the model behaves in ways resembling some aspects of perception and thought. Now I shall very briefly indicate one of the ways in which it can learn. We shall assume that in some part of the machine information is elicited from some pool of wave functions, used for some purpose, and then returned to the pool of wave functions. This process is shown in Figure 2, which in addition summarizes some of the preceding discussion. Suppose that the contents of the pool consists of N copies of the wave function Ψ, and that at the moment they are split by the ϕ filters, so that $\Psi = \sum_i a_i \phi_i$, where the ϕ's are symbols for certain responses. Suppose that the shot noise generator is operating so that one of the ϕ's will be elicited, and let us suppose that we are interested in teaching the machine to perform one of the responses, say ϕ_1, rather than any of the other ϕ's. For the first elicitation the probability of obtaining ϕ_1 is $|a_1|^2$. In case this occurs, the operator (or some other part of the machine) "rewards" the machine by turning off the noise generator. Meanwhile, the ϕ_1 has returned to the pool, which now contains N - 1 of the ϕ's and one ϕ_1. The probability of eliciting ϕ_1 has now become $|a_1|^2(N-1)/N + 1/N$, which is larger than before. The wave functions will be recombined and resolved again in a number of ways, but since the noise generator has been turned off, no information will travel around the cycle, so that no information will be discarded, and there will be no further change in the probability of finding ϕ_1. Now suppose instead that the first trial resulted in the response represented by one of the other ϕ's, say ϕ_2. This time the operator leaves the noise generator on for a while. The ϕ_2 returns to the pool, so that now there are N - 1 of the ϕ's and one ϕ_2. These wave functions become resolved in other coordinate systems, just as in the preceding case, but this time, since the noise generator is operating, information will go around the cycle and some of the components will be lost. Now let us split the resulting wave function into ϕ components for the next trial, and let us see what the new response probabilities will be. Referring back to the discussion of association, we observe that in the process of splitting and recombination the discarded component ϕ_1 will have returned to some extent. Thus the probability for ϕ_2 on the next trial will have increased somewhat, but not as much as that of ϕ_1, had the latter occurred on the first trial. The probability of any one of the ϕ's, say $\phi_{j'}$, will become $(1/N)\{1 + (N-1)\cdot \sum_{n,m,\ldots,k,j} |\Psi(j'|n)|^2 |\Psi(n|m)|^2 \ldots |\Psi(k|j)|^2 |a_j|^2\}$, in which the subscripts n, m,...k stand for the various states in the intermediate coordinate systems. The probabilities are not so easy to compute on subsequent trials and depend upon details of how wave functions in the pool become accessible to elicitation, so that further discussion will be postponed until a time when it can be done in a thorough and systematic way. However, it appears that some reinforcement learning is possible, reinforcement acting directly only upon the noise generator and not directly upon the wave functions themselves. The general idea is that the desired function remains, while the others are chopped up. The fact that the probability of elicitation of one of the ϕ's depends upon both the amplitude of the ϕ and the level of the noise, with reinforcement acting only upon the noise, resembles parts of certain theories of learning in animals, but this topic too will be reserved for more systematic discussion when that is possible.

This paper will conclude with a few brief remarks of a more general nature. First of all, the mathematical language employed is well suited to describe both the fluidity of the initial stages of perception and thought, in which ideas have the potentiality of being examined on the basis of a large number of cross-cutting systems, and the final stages in which definite and stable relations may be perceived. The language is also suited to describe a system in which structures are not always defined in advance. As in the illustrative example of microwave modes, a source of excitation which is only moderately complex can interact with a structure which is only moderately complex to produce modes which are tremendously complex, and which could not feasibly be constructed in any other way. In addition, these structures themselves are given the opportunity to become still more highly developed by an evolutionary process of selection and recombination. In this sense the model resembles the desired arena in which patterns spring up and develop by themselves. In such a system the information is stored in the form of incompletely defined potentialities, which are realized only in interaction with the eliciting and recognizing parts of the system. The percepts, for instance, are in the combination of stages, not in either alone; for in the first stage by itself they contain all points of view at once and none alone, while in the second stage by itself they are in the form of the conventional computer representation of information, which is no more perceptual and subjective than the image on a television screen. The full computer might have to contain a conventional component for reception and useful transformations of the input, followed by the present model, followed by a more or less conventional component to discriminate among the stabilized images or ideas and produce an output. Of course if we were trying to design a machine which could have a knack for skilled actions, the same problems of fluidity of transformations would recur, as we may be convinced by noting that our signature comes out the same even if we hold the pen in the mouth or between the toes.

Since the information exists only in the combined input, wave function system, and recognizer, one might wonder where the boundaries are between the observed system, the observing mechanism, and the observer. Von Neumann has shown[12] in the precisely similar mathematical situation in quantum mechanics, that even though the transformations are different in the three subsystems, the combined result is independent of where the boundaries are drawn. Thus although in this system, as well as in the workings of the brain, it is impossible to draw sharp boundaries, perhaps that is not necessary for understanding the processes involved.

Finally, we have seen that this model makes a distinction between items of information which can merge and then exert interacting effects in new coordinate systems, and items which may be sorted and combined but remain separate and have independent effects in any new coordinate systems. It is conceivable that the evolutionary process of selection and recombination could lead to the elaboration of a complicated wave function which contained a large body of information which had become integrated in a way which is necessary for complex behavior. Whether or not such a situation could occur depends upon how clever we are in designing the proper rules of evolution, but the point to be made here is that the mathematical language we are using is (as is known in quantum mechanics) a good language in which to describe such a situation should it occur, or for the purpose of trying to make it occur.

REFERENCES

1. P. H. Greene, "An Approach to Computers That Perceive, Learn, and Reason," Proc. Western Joint Computer Conference; 1959. pp.181-186.
2. P. H. Greene, "The Fundamental Units of Perception: An Outline," Air Force Office of Scientific Research Technical Report, to appear in 1960. Revision of unpublished outline of series of lectures given in the Dept. of Psychology, Northwestern Univ., Spring, 1959.
3. P. H. Greene, "Networks for Pattern Perception," Proc. National Electronics Conference, Vol. 15; 1959.
4. J. C. Shaw, A. Newell, H. A. Simon, and T. O. Ellis, "A Command Structure for Complex Information Processing," Proc. Western Joint Computer Conference; 1958, and Rand Corp. P-1277.
5. P. Schilder, "Mind: Perception and Thought in Their Constructive Aspects," Columbia University Press, New York, N.Y.; 1942.
6. A. Landé, "Quantum Mechanics," Sir Isaac Pitman & Sons, Ltd., London, Eng.; 1951.
7. A. Landé, "Foundations of Quantum Theory: A Study in Continuity and Symmetry," Yale University Press, New Haven, Conn.; 1955.
8. A. Landé, "Quantum Mechanics, From Duality to Unity," Amer. Scientist, 47:341; 1959.
9. N. Wiener, "Nonlinear Problems in Random Theory," The Technology Press of The Massachusetts Institute of Technology and John Wiley & Sons, Inc., New York, N.Y.; 1958.
10. N. Wiener and A. Siegel, "A New Form for the Statistical Postulate of Quantum Mechanics," Phys. Rev., 91:1551-1560; 1953.
11. A. Newell, J. C. Shaw, and H. A. Simon, "Chess Playing Programs and the Problem of Complexity," IBM Journal of Res. and Devt., 2:320-335; 1958.
12. J. von Neumann, "Mathematical Foundations of Quantum Mechanics," translation by R. T. Beyer, Princeton Univ. Press, Princeton, N.J.; 1951.

Table 1. Interdependence of the Arguments

<u>Symbols as modes of oscillation</u>
Much internal structure.
"Spring up," so easy storage.

<u>Superposition and resolution</u>

<u>Components in various coordinate systems represent various points of view</u>.
Gives rule of interconnection and dependence.
Holistic nature of transformations.

<u>Normal modes provide coordinate systems</u>.
Will be identified with stable percepts.
Reproducible, integrated, often discrete.
Units "carved out" by system.
Relation between two systems always linear.

<u>Elicitation of information: Split into one of the coordinate sets of eigenfunctions and correlate with shot noise</u>.
Destroys the latent information.
Representations now amenable to logic at expense of richness of interconnections.
Allows learning by reinforcement which acts by stopping the shot noise source.
Simulates some properties of thoughts (primary process) and integrated action patterns (instinctive acts).
Random selection and recombination can lead to elaboration of more complicated symbol structures.
Existence of many possible systems of categorizations cutting across one another, so that what is definite in one is incompletely defined in others.
One cannot and need not draw a definite line between environment and observer.

<u>Observations of various types</u>
Conditional probabilities relating pairs of observations provide a useful description.

<u>Selective awareness</u>
$P(a_k|a_i) = \delta_{ki}$
$P(a_k|b_j)$ has various values.

<u>Transformations should form a group</u>.

<u>The P's cannot be the entities (Ψ) related by the group of transformations, but must be their squared moduli</u>.
The Ψ's will be the transformation coefficients between coordinate systems specified by different sets of normal modes.
Each set of coordinates will contain latent information in the form of phases, which will be elicited only in other coordinate systems and thus represents potentialities.

<u>Operator and eigenfunction equations</u>.
Sharply defined numerical values elicited must be eigenvalues.

<u>Kinds of system which give wave equation for change in Ψ's with time (or other parameter)</u>.
E.g., multiple paths, specified by phase lags.

<u>Abilities for Gestalt perception</u>.
Pure states or normal modes analogous to "good" configurations.
Whether part of a pattern is seen as a Gestalt will often depend upon other parts of the pattern. Problem of seeing parts in a larger pattern.
Gestalt constancies, such as brightness constancy.

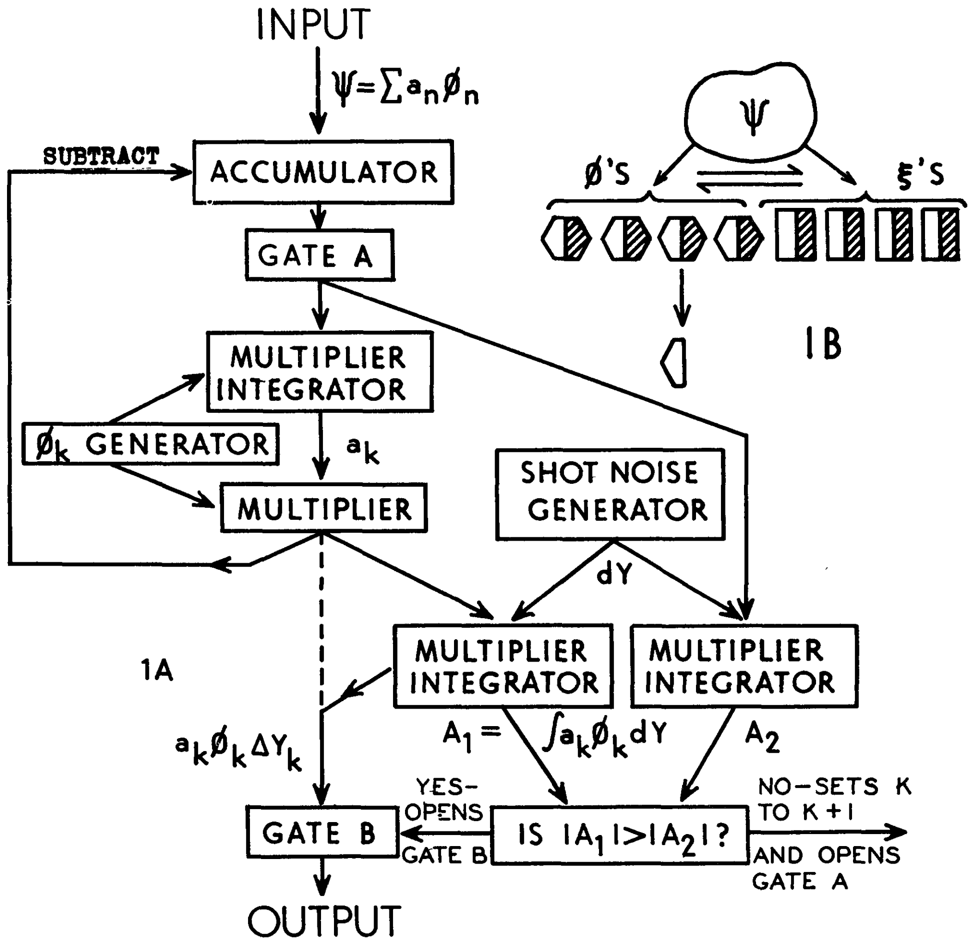

Figure 1 Flow Chart of the Model

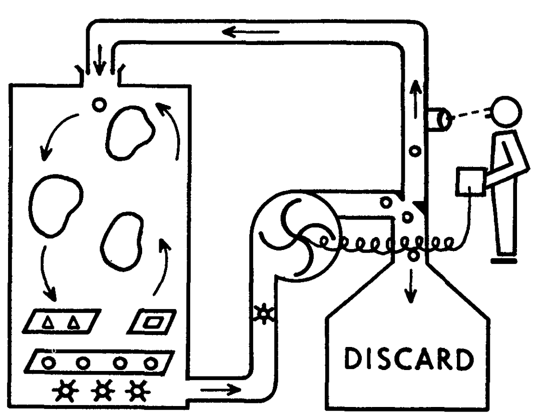

Figure 2
Pictorial representation of the functions of the various parts of the model, illustrating reinforcement learning.

A population of wave functions in the pool (depicted by the large objects circulating in the tank at the left) are split into components of some type (the small circular objects) by the appropriate separating mechanism (the grate with round holes), although they could have been separated into different sorts of components (by the other grates). The components contain information which enables them to combine again (the fringe on the circular objects). The noise which acts as the eliciting mechanism is depicted by the pump at a height representing the threshold. There is only room for one component at a time to pass into the pipe leading back to the information pool, so that all but the first component to reach that pipe will drop down and be discarded. In passing through the pump the latent information allowing recombination has been destroyed (the circles are now smooth). The man represents the operator of the machine or another part of the machine which can observe the action represented by the component which has been elicited. To reinforce that response the man turns off the pump; if another component has come through instead, he merely leaves the pump on for a while. He acts only on the pump; the elicited component returns to the pool in either case.