OCR from scan of South African Journal of Psychology, 1973, 3, 23-45

THE PROBLEM OF STIMULUS STRUCTURE IN THE BEHAVIOURAL THEORY OF PERCEPTION*

M.M.TAYLOR
DCIEM, Box 2000, Downsview, Ontario

ABSTRACT
In J. G. Taylor's Behavioural Theory of Perception there is a problem in the description of the stimulus elements which enter into the conditionings which form the basis of perception. If individual receptor responses are chosen as stimulus elements, then the numbers involved are unreasonably large. An attempt is made to resolve this problem through the development of structure by simple association. A model is presented, in which association among receptor outputs leads to the development of a hierarchy of feature detectors, each level of which provides a mathematical transform of its input and conserves the input information in a compressed form. Lateral inhibition is shown to be responsible for the variety of feature detectors needed to retain all the information and to form a more or less complete transform. Lateral inhibition, working within a redundant transform, is shown to permit efficient encoding of individual stimuli into elements that are meaningful in terms of the statistical structure of the environment. These elements may then be used as stimulus elements in the conditioning processes of the Behavioural Theory of Perception.

The publication in 1962 of J. G. Taylor's book *"The Behavioral Basis of Perception"* (Taylor, 1962) was an important event in the history of psychology. Taylor gave a comprehensive account of how a purely behaviouristic set of operations could account for the phenomena of conscious perception. Behaviourism no longer denied thought and imagery; it implied and explained them.

In Taylor's theory, the perceptual world is a consequence of the successful adaptation of an individual to the variety of behavioural requirements imposed by his environment. This adaptation depends on operant conditioning. Reinforced responses are conditioned to the stimuli then present, so that a recurrence of the same stimuli tends to evoke the same response. The connections, called "engrams", between stimuli and responses form the field from which perception grows. Perception is determined simply by the engrams activated by the current stimulation. Most of these engrams will not actually elicit an overt response on any particular occasion, because other stronger engrams interfere. They nevertheless are part of the perceptual field.

From the simple thread of conditioning, Taylor wove a tapestry of many hues, in which was displayed the whole of perceptual experience, from the earliest confusion of uncomprehended light seen by the infant to the connoisseur's appreciation of a work of art. Furthermore, since conscious experience was considered to be a byproduct of adaptation, it required no ghostly "mind" to direct it. Consciousness should be a property of any sufficiently complex and adaptable organism, and the organism should be conscious of those aspects of its environment toward which it must alter its behaviour patterns. It should not be conscious of those aspects of its environment for which its genetically determined behaviour sufficed and in respect of which it had never been forced to alter its behaviour.

It seems strange that such an appealing theory, comprehensive and yet grand in its simplicity, should have had so little impact in the decade following publication of the book. There are many possible reasons for this neglect, important among which is the nagging question of "Would it work?"

One major unresolved problem in the theory is the question of what serves as a stimulus element and what as a response element in the conditioning process. Since the linkage of stimulus elements to response elements is the cornerstone of the theory, ambiguity in this aspect makes prediction from the theory rather difficult. More importantly, though, it leaves the whole mechanism suspect; since the philosophy that the infant starts life as a *tabula rasa* underlies the whole theory, consistency requires that one must regard individual afferent impulses and individual control signals to motor units as the basic elements of sensation and of response. If this assumption is made, however, the sheer numbers of simultaneous stimuli and responses render extremely difficult the problem of how the stimulus and the response which led to a reinforcement can be selected out of the mass. How can the link between them be conditioned without at the same time conditioning myriads of irrelevant stimulus-response links?

Taylor attacks this problem in two ways. One is to define, arbitrarily, stimulus elements as being (for vision) patches of uniform colour rather than the illumination values of individual points of the scene, and

to define movements in terms of the starting and ending points of a motion without consideration of what came between. These definitions greatly reduce the numerical size of the problem. While there are some 10^8 receptors in each retina, there may only be a few thousand patches of colour visible to the eye at any one moment, and a similar though lesser reduction is produced in the number of movements to be considered. Hence the number of potential conditionings is reduced by many orders of magnitude.

Taylor also attacks the problem of numbers in the conditioning process by considering the infant as constructed of a number of non-interactive subsystems. Within each subsystem, stimuli and responses may interact. Reinforcements can affect only stimulus-response linkages within the particular subsystem responsive to that reinforcement. This partitioning again drastically reduces the number of linkages which might be reinforced by a particular conditioning event. It is not clear what determines the bounds of a subsystem, and how a stimulus is determined to belong to any one subsystem.

Both Taylor's solutions to the problem of complexity in the conditioning process seem to imply an innate structure of a kind at variance with his assumption that the infant is born with no prior knowledge of the world. If to the newborn infant the world is composed importantly of patches of light and shade, then why should not the assumption be carried further to the point that the infant's world is composed of objects, that the infant can discriminate perspective transformations, and so forth? The important principle of the book is that the meaningful relationships in the world can be derived from conditioning. But if the primary meaningful relationship of Gestalt continuity can be built in by assumption in the form of visual patches, then a major problem is not resolved but ignored. How do the patches become stimulus elements?

The same sort of quarrel can be taken with respect to the subsystem approach. While no doubt it is valid to regard different aspects of the infant's stimulation and responses as independent of other aspects, to a first approximation, it seems akin to an *a priori* solution of the problem tackled by Taylor. If particular stimulus elements and particular response elements are to be referred only to a particular sub-system, and reinforcement applied only to those elements in that sub-system, the segregation of the world involved in perception has been largely brought about by fiat. Apart from the fact that the subsystems are nowhere very well categorized as to type, the *a priori* connection of reinforcement types with particular stimulus elements or groups is not clear. Certainly visual stimulation is not restricted to the use of any one subsystem, and neither is the movement of an arm. But somehow when these are linked by a reinforcement, only a particular subsystem is involved.

In the present paper, I shall attempt to show that various ideas current in the perceptual literature can resolve the problem of numbers with less appeal to the *a priori.* We require that the infant be supplied with a hierarchy of analytic structures — in some contexts akin to what have been called "feature detectors", in others to "linear transforms", and yet again to "template matching devices" — but we do not require that these analytic structures be initially attuned to any particular type of analysis. We therefore do not impute any *a priori* knowledge of the structure of the world to the infant. Under certain simple (and over-simplified) assumptions, the analytic structures will change in such a way that they analyse the qualities of the statistical structure of the stimuli to which they are exposed. They do not involve conditioning which depends on reinforcement, but work entirely on simple association. They result in great simplification of the input in terms of meaningful elements which correspond to statistical regularities of the world. The outputs of the analytic structures are taken to be the stimulus elements in the Behavioural Theory. Responses can be produced in a similar way by elaboration, according to statistical rules, of simple commands. Both stimuli and responses, in this scheme, are "meaningful" in terms of structures in the environment. They therefore fit comfortably into the scheme of independent subsystems for conditioning. They are relatively very few in number — far fewer than the visible patches used by Taylor — and hence the problem of numbers becomes manageable.

Template matching, feature detectors, and linear transforms

Elementary discussions of pattern recognition (e.g. Lindsay & Norman, 1972; Neisser, 1967) often begin with consideration of a template matching scheme. Suppose that one wants to recognize every occurrence of the letter A. The template matching scheme provides a template shaped like the target letter A. The test

letter is imaged onto the template, and if it is an A, light from the letter hits the template and not the surround, while light from the background hits the surround and not the template. If the letter is much brighter or dimmer than its background, the average illumination on the template will differ greatly from that on the surround, and a detection of an A will be recorded. Other templates are looking for B, C and so forth, and the one with the best output is selected as correct. This scheme is usually rejected for practical pattern recognition devices on the grounds that it would need a new template for each position of the A within the image field, for each different size ofA, for each orientation of the A, and for each distortion involved with different type faces or varieties of handwriting. Indeed, there are an enormous number of possible varieties of A, and the scheme as stated would be incredibly unwieldy.

In place of the one-stage pattern recognition scheme given by the template-matching process, a two-stage operation might be proposed in which little templates or some other unspecified operations provide information about properties of the target letter. For example, A has a line sloping up to the right, a line sloping up to the left, a horizontal bar connecting two lines, an acute angle at the top, and so forth. These "features" are considered as a list and matched against lists of features to be expected of A, B, and so forth. The best list match is selected as the target letter. We are not concerned here with the success of the matching operation, nor yet with the use of feature lists in pattern recognition. The item of interest here is that the template matching scheme has been made less unwieldy by the use of parts of the target instead of consideration of the whole target at once. The hierarchy is simpler, because many letters share common features. For example, A,B,D,E,F,H,I,K,L,M,N,P,R,U,V,W, all share the feature that their leftmost part is a vertical or near vertical line. Hence a device that detected a near-vertical line with nothing to its left would be useful in the discrimination of this set of letters from the remainder. A device which detected a closed area in the letter would provide a feature possessed by A,B,D,O,P,Q,R, and not by the others. If a letter had both features, it could only be one of A,B,D,PorR.

While detection of the feature-properties of letters would require fewer templates overall than would template matching of the entire letters, still the number is very large, and if templates are to be used to detect the features the matching errors due to the rest of the letter must be ignored somehow. If nothing else, parts of the letter not potentially involved in the feature being detected must be masked off so that they do not indicate mismatch. This leads to one of the most difficult problems associated with template matching devices, that of segregating the part of the scene being tested against the template from the rest of the scene. How do we know whether a mismatch is real or is due to irrelevant elements like the right-hand-side of a letter whose left side is being tested for a near-vertical straight line?

The obvious solution to the segregation problem is always to examine only a very small area of the scene. That way, other parts of the figure cannot interfere. But if this is done, features such as "enclosed area" and "straight edge near vertical at the left side of figure" are not detected because they refer to large parts of the letter. These properties must themselves be composed from lists of little features. The latter, for example, could be from a list of colinear little lines, none of which shows a leftward projection. There must be fewer useful features of this microscopic kind than of the kind we were previously considering. A partial list might include lines, T-junctions, angles, and crosses at various orientations. Most features involved in letters could be composed from these elements. An A, for example, could be described as (starting from the bottom left) "upward line, right-stem T, more upward line, angle down-turn to the right, downward line, left-stem T, more down line, two T stems joined". The first three or four items on the list specify that the left side of the letter forms a more or less vertical straight line.

The features described here are still used like the big templates with which we started. At any one location, either one feature or another is present. If there is a T-junction, there is not a line. This is an inefficient way of handling the information. Suppose that the A were carelessly handwritten, so that the cross-bar just failed to meet the left leg of the A. Strictly speaking, no T-junction is present on the left leg, and the description list for the A is very different from the prototype list. But the detector template for the T-junction will be quite well matched despite the failure of the cross-bar quite to reach the line. There will be a strong output from the T-junction detector, and a less-then-perfect output from the line detector because of the unwanted proximity of the cross-bar. The whole situation can be better described by reporting both line and T-junction as being reasonably well matched. Indeed, the line template will give a reasonably large output for any T-junction, although not as large as for a clear line. Hence we might suppose that instead of

a list of properties possessed by the pattern, a better system might be to defer decision as to the possession of the properties and to report instead the degree to which each feature is characteristic of the location in question. The location where the cross-bar of the careless A nearly meets the left leg might then be characterised by "strong line, strong right stem T, weak cross, no left-stem T". The character list of the A would then be a set of numbers representing the relative strengths of the different features to be expected from the various locations, and the best match to this prototype profile would be given by a prototype A.

The profile of matches between an input pattern and the members of a set of feature templates is exactly the same as what is known mathematically as a transform. The Fourier transform is probably the best known example of a mathematical transform, and it is constructed in exactly this way. An input pattern is compared to a variety of templates in the form of sinusoids, and the degree to which it matches each template sinusoid is taken as the amount of that component in the input. Formally, the match to a single template is determined by forming a weighted sum of the input elements. The template is the list of weights assigned to the various input elements. If the input pattern matches one particular template perfectly, in a Fourier transform, it will not match any of the others at all. Each of these other templates will give a match output of exactly zero. This is a property of "orthogonal" transforms; the pattern that matches any one individual template exactly will give a zero output from any of the others. It is not a necessary property of transforms in general. Orthogonal transforms have many convenient properties for mathematical analysis, and have been studied at great length. A reasonable introduction may be found in any lextbook on Linear Algebra (e.g. Finkbeiner, 1966), and more com- prehensive consideration of the Fourier transform is given by Bracewell (1965). As we shall see, however, non-orthogonal transforms have great value in the analysis of patterns when the best form of the analysis is not known *a priori,* as is the case in pre-perceptual analysis of incoming stimulation.

The reader accustomed to Fourier transforms and the rest of the apparatus of linear algebra, will perhaps not recognize the approach taken here. We have come to the notion of a transform by successively subdividing and generalizing the notion of a template matching device, and in the process have arrived at a transform that is only a partial version of the transforms taught in mathematical environments. The transforms we have generated consist of degrees of match to templates useful in the recognition of letters. They do not necessarily convey all the information in the stimulus pattern, although a sufficient set of templates can form a transform which does. We must now discuss the way in which information may be concentrated and can be transmitted through the process of transformation.

Transform as description

The usual orthogonal transform has two important properties. Firstly, it is complete, in the sense that an "inverse" transform can be found that will recreate the original input exactly. Secondly, it has exactly as many elements as did the original input. It must be stressed here that we are dealing always with a finite number of individual input elements, and so do not consider continuous transforms. The "dimensionality" of the input is the same as that of the output of the transform. This means that it takes as many templates to make a transform as there are independent items in the input. A good example of this condition is seen in the various additions and subtractions involved in an analysis of variance, which can be considered as a transform. The total degrees of freedom allotted to the various effects and interactions (components of the transform) is always equal to the number of data points entering into the analysis. The analysis of variance can actually be used quite effectively as an example of a transform, because the sums of squares which are commonly used to indicate significance levels are actually estimates of the amount of the input (data) involved in each component (effect) of the transformation (analysis). The underlying mathematics is identical.

The analysis of variance example can be pushed further. Often the statement is made in a report that "85% of the variance is accounted for by the main effect". This means that if the main effect variable were used to predict the data, the remaining error in prediction would give a variance only 15% as large as the predicted variance. The prediction would be rather good. Often it turns out that only a few effects and interactions suffice to describe the data from an experiment as well as it can be described, because all the other effects and interactions contribute no more than the underlying noise or measurement error. Hence, all we know

about the patterns underlying the data can be stated by giving the predictions from these few effects and interactions, and ignoring the rest. In terms of the transform approach, this is the same as saying that all we know about an input pattern can be retrieved from only a few components of the transform, because all the other components are so small that they probably derive from noise or measurement error and not from anything that matters.

The object of performing pattern analysis by means of transforms is to reduce, to as small a number as possible, the number of transform components that must be considered without inducing any more error in the reconstruction from the inverse transform than is inevitably introduced by the noise processes.
The point is important enough to be worth restating. A transform can be sure of conveying all the information in a particular input pattern only if it has as many components as there are elements in the input pattern. Well chosen transforms will have a few large components where the match of the pattern to the template is good and many small components whose sizes are no greater than would be expected from noise or measurement error in the input. The original input pattern can be restored from the complete transform by inverse transformation. If the small components are ignored, the error in the restoration will be no greater than the difference due to noise between two successive samples of the same underlying input pattern. The best transform for analysing a particular class of stimulus pattern is that which on average minimizes the number of large components that must be considered in the reconstruction.

Transforms have another aspect worthy of consideration. The components of a transform each represent some aspect of the stimulus that is independent of the aspects represented by the other components. As an example, consider a trivial case in which two successive measurements of an object yield values P and Q. It is not helpful to present these two measurements when we want to know how big the object is. Better we should report (P + Q)/2, the average of the two measures, and if some indication of the reliability of the measurement is wanted, we can report (P - Q)/2 as well. These two reports constitute a kind of transform imposed on the data. One component tells about the object, and one about the measurement. Probably only the information about the object is interesting, and the other component can be dropped without loss regardless of how large it is. We could not reconstitute the original data without it, but we can get all the information possible about the object itself from the first component alone. This is a typical effect of a transform. The information about some aspects of interest are contained in one or a few components. The other components carry information that may be vital to an accurate reconstruction of the original input, but refer to aspects of the input that are of no interest at all to the later analysis processes because of the reasons for which that analysis is needed.

Referring back to the Behavioural Theory of Perception, for a moment, the distinction between two reasons for dropping components from a transform can be regarded as non-behavioural and behavioural. Components may be dropped regardless of their behavioural implications cither because they are uninformative or because they inform about matters that are not behaviourally relevant. The latter case will occur when the relevant components are linked by conditioning to elements of behaviour while the irrelevant components have not been so linked. They may be available for conditioning, but have not been found to differentially affect behaviour. Hence they do not appear in the perceptual field according to Taylor's theory.

In the course of this discussion, "pattern recognition" has been quietly de-emphasised. At this point it can be discarded almost completely. The function of the analytic structure was originally to provide information efficiently in such a form that a decision might be made about what is present in the world. Taylor's theory does not require that such a decision ever be made. It requires only that a decision be possible about what to do in respect of what is present. Sometimes this may entail recognition of objects, but sometimes it may involve only the glimpse of something moving rapidly towards you, which means "duck". Reaction to properties of objects rather than categorization of objects is the function of perception according to the behavioural theory. Properties of the world can be obtained more readily through appropriate transformation than directly from patterns of stimulation.

Elaborative transforms for output

We have reduced the dimensionality of the stimulus input by means of transforms. Later we shall consider ways In which it is further reduced and the extent to which it may be reduced overall. Now a brief digression will indicate how a similar approach can reduce the dimensionality of the output required to control motion, and thus can reduce the size of the response field that must be conditioned according to the theory.

Movements may be considered as sequences of combinations of motor unit operations. Individual motor units do not perform meaningful movements. Only combinations of them performed in proper sequence can do that. Even the simple flexing of a joint requires the reciprocal relaxation of one muscle and the tension of its antagonist. Such simple component combinations join to provide more and more complex sequences of behaviour controlled by simple commands and not by control of the individual motor units (e.g. Easton, 1972). These sequences obviously involve feedback and are thus on the face of it more complicated than the transforms we have been considering. But the feedback may be considered merely to control the proper execution of the commanded element of motion at any particular point in the sequence, and thus it can be ignored in this consideration of the structure of the motion command system. The elaboration of the command can be considered as the inverse of the kind of transform which analyses the stimulus input. For output, a few input dimensions control an apparently higher dimensional output because each control dimension evokes a whole pattern of output. The combinations of these patterns produce the compJex overt movements.

An example of the artificial elaboration of dimensionality in control sequences is provided by the artificial speaking machine often used in speech simulation studies (e.g. Schafer, 1972). The machine consists of a set of four or five filters whose gains, bandwidths (sometimes), and centre frequencies can be controlled from a computer, plus two alternate sound sources, one a pulse train whose frequency can be controlled, the other a white noise. The total control sequence provided to this device from the computer produces satisfactory speech at an information rate of no more than 600 bits per second (e.g. Schafer, 1972), but the resulting waveform has a bandwidth of more than 5 kHz and an apparent information rate of more than 5×10^4 bits/sec. This high bit rate is, of course, illusory, since the waveform can be transformed without loss back to the control sequence. Even the 600 bits per second of the control sequence is high, since some aspects of the control are statistically related to other aspects because of mechanical restrictions on allowable speech sounds and on rates with which speech sounds may vary. Another situation with illusory high bit rates on output is in the rapid movements of a skilled pianist's fingers. These movements could presumably be analysed in terms of interrelated musical phrases each of which can be executed from simple low-rate commands.

Development of transforms

The discussion thus far has been based on the usefulness of appropriate transforms for providing information concentrated in ways adapted to the needs of the organism. We cannot, however, assume that the infant is provided with all the proper transforms at birth, since this would be committing it to a prior knowledge of the world into which it was to be born, in violation of Taylor's basic assumption. Even if the infant were provided with some of the proper transforms, the question would still arise of how these transforms could have evolved in the growth of species. The problem would not be solved, it would merely have been pushed back a stage.

In this section, I address the question of how useful transforms can develop from the inchoate mass of possibilities available to the newborn infant. Whether they are hardwired by genetic mechanisms or are learned from experience, the proper transforms must have somehow been acquired over the course of time. Probably some are hardwired; the early stages of processing probably are more or less independent of the particular environment, depending on the structure of illumination as light and dark patches, on the structure of sound as harmonic relationships, and so forth. Similarly, the later processing stages probably develop in accord with the particular environment of the individual; whether his world is urban linear or rural patchy, for example. Turnbull (1961) reported that pygmies of the equatorial African rain forest did not possess size-distance constancy over ranges greater than a few tens of meters. Blakemore and Cooper (1970) found that kittens raised in an environment with no vertical lines do not develop cortical line-

detecting cells responsive to vertical lines, and those raised in an environment devoid of horizontal lines were likewise devoid of horizontal line detectors. These kittens later acted as if they could not see sticks oriented in the direction for which they had no detectors.

In the following, no distinction is made between evolved structure and individually adapted structure. Evolution presumably succeeds because there is an advantage to individuals born with structures initially adapted to the environment, if the environment is stable enough not to require differential adaptation across individuals. Hence the discussion on adaptation can proceed as if there were no evolved structure; any correct initial structure merely speeds the adaptive process.

The analytic structure with which we shall deal consists of conceptual units called, for convenience, "cells". No serious attempt is made to match the detailed properties of these cells with those of real neurones, although similarities of properties may be noted. Our "cells" may involve several real neurones, or it is conceivable that the functions of several cells might be incorporated in a single neurone. The model is a conceptual one, intended to show one way in which an initially unstructured analysis network can develop without external feedback into a computer of useful transforms.

In the later part of the discussion of the model, lateral inhibition is shown to have a crucial role in the development of good transforms and in the efficient coding of individual input patterns. Indeed, it is not too much to say that lateral inhibition makes the whole scheme possible.

Consider an array of receptors responsive to patterns of stimulation from the environment. Each of these cells can fire independently of each other, and the rate of firing increases with the illumination. For the time being, saturation and adaptation effects are ignored, although they may enhance the operation of the analysis structure. The "logical" independence of the receptor cells does not imply that their firing is independent when the array is exposed to a real scene. In real scenes, neighbouring points are more likely to be similarly illuminated than are distant points, and hence neighbouring receptors are more likely to be firing at similar rates than are distant cells. Hence a cell that fires is more likely to have a neighbour that fires in the same short interval than is a cell that has not fired for a while.

From the point of view of information transmission, the dependence between two neighbouring cells suggests that it would be less effective to report the firing rate of each than to report the average firing rate of both and perhaps the difference in their firing rates. By analogy with our example of the two measurements of an object, a transform whose components were essentially the sum and the difference of the two rates would be useful.

Real patterns involve more than two receptors, but the principle remains that if a group tend to have correlated outputs, then the report is informationally simpler if the group output is reported along with deviations from that output. Mathematically, the transform which most effectively .performs this function of concentrating information onto a few components is called the Karhunen-Loeve transform. (Watanabe, 1967; 1969; Atal, 1972) This transform is defined for a particular statistical ensemble of patterns, and differs from ensemble to ensemble. It is even conceivable that there might be an ensemble for which the Karhunen-Loeve transform was a Fourier transform, or that there might be one whose optimum transform was a simple identity. The latter would be the case if all inputs were statistically uncorrelated; no transform could concentrate the information better than simple report of each element of the input in order from the one with the largest expected value to the one with the smallest expected value.

The Karhunen-Loève transform for a particular statistical ensemble of input patterns is determined from the autocorrelation function that relates each input element to each other. In terms of our cell model, we would want an analysis structure that would respond to associations among receptor firings, associations between firings being the closest approximation to autocorrelation among continuous-valued functions.

The analytic structure model

The output of each receptor cell is connected, initially at random, to a number of analytic cells, and each of

these analytic cells has input from many receptor cells. Associated with each connection is a weight which might be positive or negative, again chosen at random initially (genetically determined structure would replace the random initial conditions with a more ordered set of weights). If a receptor cell fires, its weight is added to the weights of other recently firing receptors connected to a particular analytic cell, and if the sum of these weights becomes strong enough, the analytic cell fires. Clearly, if receptors with strong positive weights fire and no receptors with strong negative weights fire, the analytic cell will fire readily. Conversely, if receptors with negative weights but not those with positive weights fire, the analytic cell will be inhibited. We assume also that the analytic cell has some spontaneous firing rate, so that increased inhibition is represented by a reduction in its firing rate.

The network as so far described has two layers, a receptor layer and an analytic layer which is connected randomly to the receptor layer. This system will perform a transform on the input, but it is a random transform and is most unlikely to be relevant to the world of input patterns. It must be provided with a means of change, related to the associations among receptor firings. We want analytic cells to come to recognise common patterns in the world and to report receptor firings in terms of those patterns. We do not want to impose any "correct" patterns as do most trainable networks of this general structure (e.g. Rosenblatt, 1962).

To create the required changes in the system, we impose a growth rule. This rule is fairly arbitrary, and is chosen largely because it seems to work, rather than because it agrees with any neurological phenomena. The rule is that if the analytic cell fires, the weights attached to input connections from receptor cells that have recently fired are increased, and those on the other input connections are decreased. This makes the analytic cell more likely to fire in response to the same pattern if it is again presented, and less likely to fire in response to the complement of that pattern, since the receptors sensitive to the complement will tend to have more inhibitory connections to the analytic cell than previously.

Let us follow what happens to the network in two extreme situations. In the first situation, one pattern is always presented, with no variation. In the second, patterns are presented at random, and there is no correlation between any pair of receptors. In the first situation, some analytic cells will initially be preferentially biased to fire in response to the single pattern, and others will tend to be inhibited by it. AIl analytic cells will fire at some time or other, so that on balance all will tend to become more responsive to the pattern. Of course, not all brightly litreceptors will have fired just before any particular analytic cell fires, so that the progress of each analytic cell towards becoming a template for the presented pattern may be somewhat erratic. Overall, though, the brightly lit receptors are more likely to have fired recently than are the dimly lit ones, so that the weights are more likely to increase for bright regions of the pattern and more likely to decrease for dim regions. Eventually, all analytic cells will be templates for the single pattern, and will respond to other patterns only insofar as they match the prototype.

In the second extreme case, with purely random patterns being presented, it is easy to see that while the individual analytic cells might change the pattern to which they were most responsive, no structure will emerge from the network. The transform that it performs will be as random as ever, while individual components drift from one preferred pattern to another. There may be a tendency for the network to prefer patterns like the few most recently presented ones, but this does not imply the existence of a coherent structure.

Real stimulus patterns derived from objects in the world fall between these two extremes. No one pattern is continuously repeated (except for short periods in studies such as those on figural after-effects) and never is the system exposed to random stimulation (except for short periods in studies such as those on sensory deprivation). Aspects of patterns are, however, frequently repeated. Just as we earlier broke down the features of letters into aspects such as lines, T-junctions, and crosses, so real-world patterns repeatedly display elementary features such as bright or dark spots, bright-dark edges which bound patches of more or less constant colour, and lines and bands which may be considered as composed of opposing pairs of edges or as strings of spots. Simple association would suggest that the most common relation between two nearby receptors is that they are similarly illuminated, and hence that the most common form for the analytic cells should come to be a simple brightness detector. Some cells should be responsive to edges, if they start with such a configuration, but basically the growth of the systems should converge toward the most common

pattern feature. Such an analysis network would be pretty useless. In order to save the analysis network from this fate of collapse into a system which detects only a single feature, we need one more component commonly reported in neural systems — lateral inhibition.

Lateral inhibition

Two distinct types of inter-cellular connection have been called "lateral inhibition". These have been labelled "recurrent" and "non-recurrent". Non-recurrent lateral inhibition is the kind considered at length by von Bekesy (e.g. 1967) and which is supposed to be responsible for effects like Mach bands (e.g. Milner, 1970; von Bekesy, 1967; Ratliff, 1965). It also can maintain good resolution of detail over several similar analytic levels (Harth & Pertile, 1972). In the simple model we are considering, non-recurrent inhibition occurs when a receptor cell has an inhibitory connection with an analytic cell. As we shall see, the associative growth of the analytic network, when combined with the other sort of lateral inhibition, is likely to lead to inhibitory connections being arranged around the borders of excitatory regions; such arrangements are what have led to the non-recurrent inhibition being termed "lateral", when "forward inhibition" would have been a less confusing name.

The "lateral inhibition" that is important to the growth of the analytic network is sometimes termed "recurrent", because it connects cells of the same level in the analytic network, and so can in principle participate in feedback circuits. It should more properly be called just "lateral" inhibition, and in what follows that terminology will be used without the addition of"recurrent".

Lateral inhibition, when viewed in the framework of the growth of an analytic network, seems to have two rather similar functions. Be- cause it tends to prevent too many analytic cells from firing in response to any particular pattern, it tends to diversify the features being detected. Cells that would respond to a common pattern, but only weakly, will be suppressed by cells that respond more strongly to that pattern, but can respond to and move toward some other aspect of the input to which their neighbours are not suited. Lateral inhibition creates a sort of pressure for diversity like that which in biological evolution ensures that all viable niches are occupied by some species or other. Cells compete for patterns in the same way as species compete for food.

The second function of lateral inhibition in the transform is to simplify and make orthogonal the representation of a single stimulus pattern. Without lateral inhibition, all cells responsive to a particular aspect of a pattern would fire, but the inhibition suppresses all but the few strongest, thus permitting more precise description of the aspect being reported. This is very much akin to the "sharpening" of detail commonly quoted (e.g. von Bekesy, 1967; Ratliff, 1965) as a function of lateral inhibition.

Let us now consider the place of lateral inhibition in the growth of the network. A rule is required, and as with the original growth rules it is somewhat arbitrary. It seems to work, but does not necessarily con- form to neurological function. The rule is that if an input to an analytic cell fires shortly after the cell itself fires, or simultaneously with the firing of the cell, then that input becomes more inhibitory. Note that this refers equally to the connections from receptors, although that feature is not used in the subsequent analysis.

If two analytic cells respond preferentially to the same stimulus pattern, then they will tend to fire at similar times after that pattern is presented. Moreover, they will tend to fire rapidly in response to the pattern. Hence each will tend to fire often when the other fires or while the other is recovering from firing. Hence, according to the rule, each will tend to inhibit the other. Of course, if A fires after B, then B fired before A, which would tend to make the B connection into A become more excitatory. This competition between excitatory and inhibitory effects will be resolved one way or the other, in favour either of inhibition or of facilitation. Which function wins will be determined by the period defined as simultaneous, because if they fire "simultaneously", thc cells become mutually inhibitory. If the "simultaneous" period is long enough compared with the period over which excitatory impulses summate to induce cell firing, then the mutually inhibitory tendency between analytic cells with similar responses will predominate.

How does the rule affect cells which respond to statistically independent aspects of stimulation? At first glance, it might seem that they will also become mutually inhibitory, because there will be the same set of relative probabilities that a cell firing comes simultaneously, before, or after a particular other cell firing as is the case for cells reporting the same aspect of the stimulus. This, however, is not true, because the actual firings of the related cells should be correlated in time, whereas the firings of the unrelated cells should not. The related cells should be correlated impulse by impulse, because the only reasonably likely way that they can respond to the same aspect of the input is to be connected by a similar pattern of weights to the same receptors. Hence each is responsive to the same moment-by-moment fluctuations in the firings of their receptor inputs; each will tend to fire following a burst of firings from a group of receptors with excitatory inputs to them both; each will therefore tend to fire at about the same moment. Hence the related pair of analytic cells will tend to fire "simultaneously" more often than will the unrelated pair. If we assume that inhibitory and excitatory growth effects are roughly balanced if the analytic cells respond to statistically uncorrelated aspects of the stimulation, then the inhibitory effects will dominate the connections among related analytic cells.

If two analytic cells have weighting patterns that are similar, but not identical, then a stimulus that matches one more precisely than the other will tend to fire that one more strongly and probably with a shorter latency (fewer of its excitatory inputs will be required to make it fire). If it fires first, then its already inhibitory connection with the other will tend to make the other not fire, or at least to delay further its firing, which is equivalent to a momentary reduction in its firing rate. Hence the cell which matches most closely the input pattern will more strongly tend to shift its responsiveness toward that pattern, thus tending to separate the response patterns of the two analytic cells. Once they have separated to the point of statistical independence, there should be no further interaction between them (on average).

Lateral inhibition cannot force all cells to adopt mutually independent patterns. Often several cells will be responsive enough to the input pattern to fire regardless of their mutual inhibition, and all will then tend toward the pattern that caused them to fire. They will become more related rather than less so, but also will become more mutually inhibitory. The effect should be a sharpening of discrimination among cells in the neighbourhood of common pattern features. Features diverging slightly from common or prominent aspects of the world should be better discriminated than are features in less well populated regions of the feature space. Such inhibitory clustering of detectors may well explain the common anchoring effects such as the increased discrimination as a function of physical separation for closely separated points. The effect is to permit the component most nearly describing the total stimulus to dominate; lateral inhibition would then produce a configuration of more or less orthogonal components, permitting an optimal transform for the particular stimulus.

This last point may perhaps be better understood by reference to a concrete example. Consider a particular stimulus which consists of an auditory pulse train having a sharp impulse every 10 msec. Over the period of, say, 1 sec, there would be 100 pulses, and the stimulus might well be described in terms of the amplitude and time of each pulse. A more efficient description would be in terms of the pitch of the pulse train, 100 Hz, and a term indicating the "quality" of the buzz. In fact, what is heard is a 100 Hz buzz, and not individual pulses. An alternate, but less efficient description of the stimulus is in terms of its Fourier analysis. There is a component at every multiple of 100 Hz; in the region below 10,000 Hz, there are 100 components. One does not hear any of these components.

Now let us alter the stimulus by deleting one of the Fourier components. Clearly, it would not be efficient to describe the stimulus in terms of the components that remain, since there would still be 99. Rather, it makes sense to describe the signal as the original buzz with an added sinusoidal component whose phase is opposite to that of the deleted component. Indeed, this is what is heard (e.g. Duifhuis, 1970). The predominant component of the percept is the buzz, as before; but added to the buzz is a tone corresponding to the sinusoid omitted. There is no energy at that frequency in the stimulus, but the heard sinusoid could be added to the heard buzz to make the actual stimulus. Sinusoids added out of phase cancel one another. The two heard components — buzz and sinusoid —are orthogonal to one another, and both belong to the redundant transform which could have grown through the processes described.

Lateral inhibition can account for the development of on-centre off-surround detector units of the kind that

give rise to non-recurrent lateral inhibition. If the connective fields of two analytic cells overlay, then these cells will respond to some of the same receptors. The most prominent dimension in the display is the average brightness of the region; but this average brightness will be much the same for neighbouring regions, except for occasions when a boundary crosses between the centres of the two overlapping regions. Hence the analytic cell outputs will not ordinarily be orthogonal. But lateral inhibition will tend to prevent one cell from firing when the other does, so that either cell will fire preferentially when its neighbour does not. Under these conditions, part of the outer region of the receptive field of the cell will not be illuminated when the centre is, or else will be illuminated when the centre is not. In either case, the weights for the centre and the periphery of the connective field will diverge, one becoming inhibitory, and one excitatory. Since edges may appear in any orientation, neighbours in any direction may participate in this mutually inhibitory relationship, so that the cells which initially attempted to become brightness detectors tend to have centres which respond preferentially to the opposite illumination than the periphery. They become on-centre off-surround or vice-versa. They no longer respond very strongly to the overall illumination on the whole field, but do respond when their centre is near to an edge oriented in any direction, and respond most strongly to a spot concentric with their field. Thus non-recurrent inhibition is caused by the development of the analytic structure under the influence of recurrent inhibition.

Hierarchy of analysis

The analysis structure here proposed develops naturally from the statistical pattern of stimulation, and is based on the association between different inputs from the various receptors. As so far developed, the structure cannot handle interactive associations, where the association between A and B tends to occur if and only if C and D have happened together. Such interactions are the essence of what we call structure in perception. A line is detected because if cells A, B, C and D are spot detectors arranged in a straight line, then A and B tend to fire together if C and D do. A and B together mean no more than a spot, and there is very little tendency overall for any pair to fire together unless all do. If there were originally any such tendency, lateral inhibition would have changed the weights so as to remove it. Hence, higher order structure must consist of multiple associations among pairwise independent detectors.

The development of higher order structure depends on successive layers of analytic cells, each constructed more or less like the first layer that we have discussed. Initially random patterns of weights are associated with the connections from the next lower analytic layer. If a cell happens to fire in response to some pattern arriving from the lower layer, the connections that fired are made more excitatory, while those that did not are made more inhibitory. Now the most probable connections among elements are not spots, but lines of spots and groups of spots. In other words, we should see higher level features that are lines or edges, curves, angles, and so forth. Different ones of these features might appear at different levels in the hierarchy, depending on their statistical complexity in terms of the lower order units.

Timefactors

Thus far, we have been considering pattern detection as if patterns were individually presented at discrete moments in time. Of course, this is not what happens in everyday observation, although it may correspond with some experimental situations. The commonest observation is that patterns metamorphose into one another by changes in viewpoint or by movements of objects among themselves. The types of change are quite self-consistent, and must be important for the learning consistency in the world of the newborn. What is their effect on the analyser network that we have built?

When an object moves across the visual field, it successively excites analytic cells of the same kind, one after the other, in sequence. According to the analysis of network growth, if such motion happens often, the earlier cell will come to have a facilitatory effect on the later, since the earlier will fire as input to the later shortly before the later fires of its own accord. These neighbours should then come to have a mutually excitatory effect, whereas we have earlier claimed that neighbours tend have a mutually inhibitory effect or no interaction at all. However, these are neighbours of a particular kind, in that they tend to respond to the same sort of pattern elements, which would relatively rarely tend to be presented to both analytic cells at

once. An edge, for example, is rarely paralleled closely by another edge of the same kind. A bright-dark edge is often paralleled by a dark-light edge, to form a dark band, but seldom by another bright-dark edge. Among the thousands of edges that confront me at this moment, a search of a couple of minutes fails to reveal a single instance of two similar edges parallel to one another, though there are hundreds of parallel pairs of opposed edges (mainly formed by shadows on the edges of book pages). Hence one might expect neighbouring edge detectors of the same kind and orientation to have mutually facilitatory connections. They are related by the fact that one may substitute for another rather than by the fact that they tend to co-occur.

Substitution of one item for another over time is a very important effect in the development of the perceptual model. It is such substitution, yielding facilitatory effects, that permits higher level analysis cells to detect absolute and relative movement, and to select and classify. The higher level cell could detect one line in a wide field, like Hubel and Wiesel's complex cells (e.g. Hubel & Wiesel, 1968), in part because that line tends to move parallel to itself within the visual field and successively stimulates lower level cells with smaller receptive fields. By the same mechanism that gives rise to lateral inhibition, these smaller fields mutually facilitate and tend to fire together, so that particularly if the line is moving, one might expect wide-field detectors to be developed. This "selection" procedure is very important in separating out absolute from relative localization of items in the visual field. Relative localization yields structure in the world, whereas absolute localization derives only structure which includes the orientation of the eye. True, Taylor's theory adequately accounts for eye position, and indeed requires that it be known in order that behaviour may be properly oriented to the world. Nevertheless, it is encouraging that the same simple mechanism of associative patterning can probably account for simple effects of this kind at a rather early level of analysis.

Summary: Relation to the behavioural theory of perception

The behavioural theory of perception as published in 1962 suffered from an assumption about the conditioning processes which yielded perception. The "equivalence classes" that yield object constancies were built in by *a priori* assumption as to what the newborn infant "knew" about the world. If these assumptions were omitted, then the number of stimulus components available for conditioning and which had to be segregated and grouped by conditioning became impossibly large. The attempt of this paper is to show how segregation, property analysis and classification could be carried out without behavioural reinforcement.

The analytic structure developed here is not expected to serve as an accurate model of a nervous system, but I hope that it can serve as a guide to the potentialities of some of the effects which do seem to occur in the system. Whether or not lateral inhibition comes about in the manner described, it still may have the function of diversifying the stimulus analysis, of selecting the optimum transform both for the general descriptions of the world and for the description of each individual stimulus, and of permitting fine discriminations in the neighbourhood of important (i.e. common) norms (Gibson, 1950) of the world. The use of the transform as a model of the analytic structure has much to recommend it, particularly the reduction of information from a distributed to a concentrated form which deals with common properties. How complete the real transforms of the nervous system are is a very open question. There seems to be considerable evidence (e.g. Pollen et al, 1971; Blakemore et al, 1970; Carter & Henning, 1971; Breitmeyer & Cooper, 1972) that the visual system can act as a Fourier transformer if presented with stimuli for which the Fourier transform is appropriate; whether it does so under any other circumstances is moot. The postulated redundancy of the transform system would permit the use of the Fourier transform (a complete transform) when appropriate and completely different transforms for other signals.

The transform components that emerge from the analytic structure are relatively few in number, and each has a meaning that is well-defined terms of the structure of the real world. Volkmar and Greenough (1972) have shown that the complexity of dendritic branching and hence the analytic structure in the cortex of the rat was increased by increased complexity of the visual environment in which they were reared. Presumably the rats reared in the complex environment produced more different structures in their transform operations. The meaning of these structures for the individual is a question that depends on the

needs of the individual. Statistical theory can say nothing on such a topic.

It is here that the stimulus elements produced by the analytic structure enter the domain of operant conditioning described by Taylor's theory. A contemporary application of the Karhunen-Loève transformation to improve dramatically an automatic classification scheme has been given by Atal (1972). Grossberg (1972) has shown how a scheme akin to that described here can account for operant conditioning at a cellular level. Miller et al (1972) have shown further that single cell activity in an auditory discrimination task can depend on behavioural reinforcement. In the light of these and other works, Taylor's theory can be regarded as a framework for a viable contemporary theory of perception.

REFERENCES

ATAL, B.S. Automatic speaker recognition based on pitch contours. *J. Acous. Soc. Amer.., 1972, 52, 1687-97.*
BLAKEMORE, C, NACHMlAS, J. & SUTTON, P. The perceived spatial frequency shift; Evidence for frequency-selective neurones in the human brain. *J. Physiol,* 1970, 210, 727-750.
BLAKEMORE, C. & COOPER, G.F. Development of the brain depends on the visual environment. *Nature,* 1970, 228,477^78.
BRACEWELL, R. *The Fourier Transform and Its Applications.* New York: McGraw Hill, 1965.
BREITMEYER, B.G. & COOPER, L.A. Frequency-specific color adaptation in the human visual system, *Perception & Psychophysics,* 1972, 11, 95-96.
CARTER, B.E. & HENNING, G.B. The detection of gratings in narrow-band visual noise. *J.Physiol,* 1971, 219, 355-365.
DUIFHUIS, H. Audibility of high harmonics in a periodic pulse. *J. Acous. Soc. Amer.,* 1970, 48, 888-893.
EASTON, Thomas A. On the normal use of reflexes. *Amer. Scientist,* 1972, 60, 591-599.
FINKBEINER, Daniel T. *Introduction to matrices and linear transformations.* San Francisco: W.H. Freeman, 1966.
(iIBSON, J.J. Adaptation, after-effect and contrast in the perception of curved lines. *J. Exp. Psychol,* 1933, 16,1-31.
GlBSON, J.J. *Perception of the visual world.* Boston: Houghton-Mifflin, 1950.
GROSSBERG, S. Neural expectation: Cerebellar and retinal analogs of cells fired by learned or unlearned pattern classes. *Kybernetik,* 1972, 10, 49-57.
HlRSCH, H.V.B. & SPINELLI, D.N. Modification of the distribution of receptive field orientation in cats by selective visual exposure during development. *Exp. Brain Res.,* 1971, 12, 509-527.
HARTH, E. & PERTILE, G. The role of inhibition and adaptation in sensory information processing. *Kybernetik,* 1972, 10, 32-37.
HUBEL, D.H. & WIESEL, T.N. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology,* 1968, 195, 215-243.
LINDSAY, P.H. & NORMAN, D.A. Human information processing. New York: *AcademicPress,* 1972.
MILNER, Peter M. *Physiological Psychology.* New York: HoIt Rinehart & Winston, 1970.
MILLER, J.M., SUTTON, D., PFINGST, B., RYAN, A., BEATON, R., & GOUREVITCH, G., Single cell activity in the auditory cortex of Rhesus monkeys; Behavioural dependency. *Science,* 1972, 177, 449-51.
NEISSER, U. *Cognitive Psychology.* New York: Appleton, 1967.
POLLEN, D.A., LEE, J.R. & TAYLOR, J.H. How does the striate cortex begin the reconstruction of the Visual World? *Science,* 1971, 173, 74-77.
RATLIFF, F. *Mach Bands,* San Francisco: Holden-Day, 1965.
ROSENBLATT, F. *Principles of Neurodynamics.* Washington D,C.: Spartan, 1962.
SCHAFER, R.W. A survey of digital speech processing techniques. *IEEE Transaction on Audio and Electroacoustics,* 1972, AU-20, 28-35.
TAYLOR, J.G. *The Behavioral Basis ofPerception.* New Haven: YaIe University Press, 1962.
TAYLOR, MM. Visual discrimination and orientation. *J. opt.Soc. Amer.,* 1963, 55, 763-765.
TURNBULL. CM. Some observations regarding the experiences of behaviour of the BaMbuti pygmies. *Amer. J. Psychol.,* 1961,74, 304-308.
VOLKMAR, F.R. & GREENOUGH, W.T. Rearing complexity affects branching of dendrites in the visual cortex of the rat. *Science,* 1972, 176, 1445-1447.
VON HEKESY, G. *Sensory Inhibition,* Princeton University Press, 1967.
WATANABE:,S. *Knowingatui(iuessing.* Ncw York; Wiley, 1969.
WATANABE, S. Karhunen-Loeve expansion and factor analysis. *In* Transactions of the Fourth Prague Conference on Information Theory, Statistical Decision, Functions, Random Processes. Academia Publishing House, Czechoslovak Academy of Science, Prague 1967. p.p. 635-660. (Quoted by Atal, 1972).

COMMENT BY J. G. TAYLOR

When I started work on the behavioural theory of perception, I took a firm resolution that never, in any circumstances, would I treat perception, or any of its forms or aspects, as an independent variable. My aim was to explain perception, and I obviously could not do this by appealing to itself or any part of it, just as a physicist would not dream of trying to explain the viscosity of oil by appealing to a principle of viscosity.

But to one reared on Helmholtz, Fechner, Hering, Wundt and Titchener, with a sprinkling of Thorndike and

Watson, and a rather more generous dose of Freud, it was not easy to purge myself entirely of the philosophy I had absorbed. By the time I came to write the final version of *The Behavioral Basis of Perception,* I thought the process of purgation was complete, but in fact I was wrong. When I tried to deal with the astronomical size of the set I x O, that is, the Cartesian product of the set, I, of input elements and the set, O, of output elements, I fell straight into the trap I had determined to avoid. Instead of treating the retinal portion of the input set, I, as containing 2×10^8 elements, I postulated, quite arbitrarily and without supporting evidence, that it included among its elements a relatively small number of subsets of itself, as determined by patches of colour in the environment, and that those subsets could operate as units in the conditioning process. In fact, the only evidence I had for the existence of those subsets was subjective; they were elements of my own perceptual field. I had therefore violated my own determination never to treat perception as an independent variable in constructing a theory of perception; and did not even know that I had done it.