

CONDITIONS FOR VERSATILE LEARNING, HELMHOLTZ'S UNCONSCIOUS INFERENCE, AND THE TASK OF PERCEPTION

HORACE BARLOW

Physiological Laboratory, Cambridge, CB2 3EG, U.K.

(Received 10 August 1989; in revised form 1 March 1990)

Abstract—It is a mistake to consider perception and learning separately because what one learns is strongly constrained by what one perceives, and what one perceives depends on what one has experienced. I shall propose the hypothesis that perception is the computation of a representation that enables us to make reliable and versatile inferences about associations occurring in the world around us—that is, perception prepares the ground for learning. The statistical problem in learning is to determine whether a compound event such as “C followed by U” is a random co-occurrence or a significant association, for if it is the former it would be a mistake to pay any particular attention to C, whereas if it is the latter C is a conditional stimulus for U and a useful predictor for it. Now you cannot decide whether the association is random or not without knowledge of the prior probabilities of C and U: hence on my hypothesis when you perceive an object or event the representation must not only signal “it’s there” or “it’s happened”, but must also make evident (or rapidly accessible) the prior probability of what has been signalled. Furthermore it must do this for all the objects or events that can act as conditional stimuli, and this implies that the representative elements should be statistically independent (or approximately so) in the normal environment. Forms of coding that would do this, and the relationship with Helmholtz’s unconscious inference, will be discussed. These considerations imply that the task performed in perception has been overlooked both by learning theorists and by connectionists working on associative and adaptive networks. Coding for independence may be particularly important in understanding the developmental processes during the sensitive period: it may be the operation that leads ontogenetically-timed, activity-dependent, connections to imprint appropriate codes if the animal has experience, but inappropriate codes without experience.

Coding Conditioning Cortex Inference Helmholtz Prior probability Perception
Representation Sensitive period

THE RELATION OF PERCEPTION TO LEARNING

I decided to talk on this topic with some trepidation because Gerald knows his Helmholtz so much better than I do and does not, I suspect, trust a non-German speaker to get him right. However Helmholtz expressed himself with unrivalled clarity and the ideas I shall propose are directly descended from his well-argued proposal that percepts represent unconscious inferences, so I cannot avoid bringing him in and must risk Gerald’s criticisms.

My argument is, briefly, that to understand perception one must view it as a prologue to learning. Acquiring new knowledge of the world is among the most important things our brains do for us, and for most people at this meeting it is probably the most interesting thing it does. So I shall try out on you the idea that perception is the process of preparing a representation of the current sensory scene in a form that enables

subsequent learning mechanisms to be versatile and reliable.

I shall assume that learning is based on what we perceive, and that cerebral cortex is where the representation we perceive is computed, even though neither assumption is 100% certain: McCormick, Lavond, Clark, Kettner, Rising and Thompson (1981) and Yeo, Hardiman and Glickstein (1985) have shown that conditioning of the nictitating membrane response in the rabbit occurs in the cerebellum, and I am sure that many other forms of learning can occur without the learner consciously perceiving the sensory stimulus that is learnt. Nevertheless our perceptions certainly provide much of the information from which we learn, and the cerebral cortex must create the representations used for this purpose. This is a sufficient basis for my argument, though one should be aware that other types of representation and learning do occur.

Most people are familiar with the idea that representations vary in their completeness and accuracy, for these determine the results of tests of resolution, Weber fraction, and so forth. And the idea of a transformed representation, such as that provided by the coefficients of a Fourier transform, is now almost too familiar. I want to examine a completely different question, namely "what properties should a representation have in order to make it suitable for use by subsequent learning mechanisms?" It is often taken for granted that any complete representation would do, but this is not the case and I shall start by considering what information is needed simply to establish that an association exists. A model of efficient learning should allow access to all this information, but associative network models based on Hebbian synapses can only access some of it. Because of this they do not account for the versatility of learning, and it seems to me that this is the aspect that is most remarkable in higher mammals. We have an astonishing store of knowledge about the associative structure of the world around us, and can recognise changes very readily; I think Helmholtz's unconscious inference results from automatic access and use of this store, and it is this that makes perception so effective for learning.

These ideas also suggest a new hypothesis about the puzzling defects that result from deprivation of experience during the sensitive period.

THREE REQUIREMENTS FOR RELIABLY DETECTING PREDICTIVE ASSOCIATIONS

The formation of a conditioned reflex may be taken as a paradigm for the detection of a predictive association. To do this reliably the brain must determine that the conditional stimulus C precedes the unconditional stimulus U significantly more often than would be expected from the overall probabilities with which the two events have occurred in the past. This obviously requires knowledge of the occurrences of the sequence (U following C) and some means of estimating how often this happens, but it also requires knowledge of the past occurrences of U and C and estimates of the rates they have occurred.

One might question whether these three requirements must really be met, but I think it can be seen at once that predictive associations derived from less complete information would

be less reliable, and that an animal using an inefficient method would be at a disadvantage compared with one that made the correct computation: it would either detect fewer of the associations that were genuinely present, or it would attach importance to accidental associations, or it would make more errors of both kinds. Detecting predictive associations can bring enormous advantages, so it must be a very competitive business; of course no brain does the computation perfectly, but when considering the methods brains may use it is sensible to have in mind the requirements for the correct operation.

Now consider the implications of these requirements: the site at which learning takes place must be influenced by the number of times C and U have each occurred previously, and also by the frequency of their joint occurrence in the correct temporal relation. Suppose these numbers are used to form estimates of the probabilities $P(C)$, $P(U)$, and $P(U.C)$ where (U.C) symbolises a compound event, namely the joint occurrence of U and C in the correct temporal relationship: then the coincidence is significant and not random if $N \times P(U.C)$ significantly exceeds $N \times P(U) \times P(C)$, N being the number of possible occurrences of (U.C) in the period under consideration. Questions naturally arise about the time scales over which these counts and estimates are made, for learning can occur in seconds, or may require years. It is probable that an efficient system would need to make estimates in parallel over several different times, and this is certainly a topic we need to know more about, but the logic of inferring that C predicts U requires some form of probability estimate so let us concentrate on this aspect.

Figure 1 shows in outline how Hebb (1949) suggested that these requirements might be met by a nerve cell. He postulated that joint pre- and post-synaptic activity strengthens synapses, and the main merit of this suggestion is that it identifies the synapse between the input for the conditional stimulus and the output neuron as the site where the conjunction of U and C produces lasting effects. This is widely accepted, even though we do not yet know just what these lasting effects are, nor even whether they affect the presynaptic terminal, post-synaptic mechanisms, or both. The unconditional stimulus U and its frequency of occurrence could also produce lasting effects at this synapse because it is assumed that U by itself fires the post-synaptic

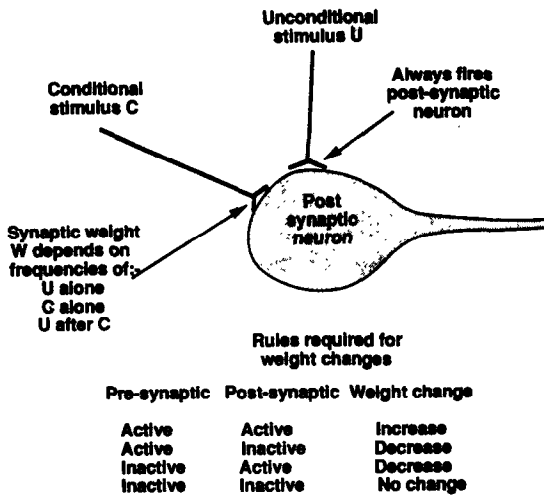


Fig. 1. Diagram showing Hebb's proposal that associations are detected and stored at the junction between synapses from afferents carrying information about the conditional stimulus on to a post-synaptic neuron. All the required information is available at this site, and his proposal is now widely accepted.

neuron, and that the membrane at the site of the modifiable synapse is depolarised when this occurs; again we do not know what these lasting effects are. The occurrence of C is obviously signalled at the pre-synaptic terminal, and again it could produce lasting effects there, or on post-synaptic mechanisms. But it is worth introducing immediately a rather different possibility for the way that $P(C)$ is computed and signalled.

Sensory messages often show habituation: they decrease in strength when a stimulus is repeated many times at short intervals. It is tempting to regard this as the means by which the prior probability of C is taken into account, strong signals with many impulses being given for rare events and weak signals with few impulses when the event signalled has happened frequently in the recent past. Although habituation occurs over time scales of seconds or minutes in the examples familiar from neurophysiological recording, one cannot exclude the possibility that much slower forms also occur, for they would be hard to observe in such experiments. One knows that sensory stimuli that have become familiar over much longer times are ineffective as conditional stimuli in learning experiments, which implies that $P(C)$ can be estimated over these longer times, so it is tempting to suppose that this is caused by much slower habituation mechanisms that we do not yet know about physiologically. This would fit well the notion I shall develop shortly that the provision of estimates of $P(C)$ for the

elements of the representation is an important part of perception.

Hebb only postulated an increase in synaptic efficacy with joint pre- and post-synaptic activity as specified in the top line of the table at the bottom of Fig. 1, but a decrease of the transmission across the synapse when C or U often occur separately is needed if the mechanism is to identify predictive stimuli correctly. It is also needed to model the extinction of a conditional response when reinforcement is withheld, or when reinforcement occurs too often without the conditional stimulus.

The most satisfactory development of the Hebb-type model is that of Sutton and Barto (1981), which includes suggested mechanisms for ensuring that C and U have the appropriate temporal relationship. Furthermore that paper showed the connection between this type of model, the learning theory of Rescorla and Wagner (1972), and the Widrow and Hoff (1960), L.M.S., or delta rule of adaptive networks. Here I want to go off on a different tack and consider how to make more versatile models based on the spirit of the Hebbian principle, for I think both learning theorists and those working on adaptive or associative networks have failed to consider some of the essential features that make this possible.

MAKING THE MODEL MORE VERSATILE

In Fig. 1 it was assumed that the conditional signal C was already known and the problem was simply to find whether or not it predicted the unconditional stimulus U. The natural way of extending it would be by adding conditional inputs in parallel, as shown in Fig. 2; the post-synaptic neuron might then be conditioned to respond to many alternative stimuli such

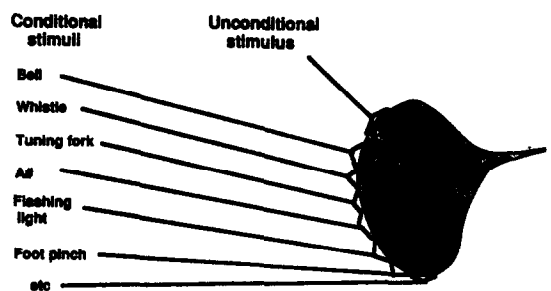


Fig. 2. Possible extension of Hebb's scheme to make learning more versatile. One difficulty is to provide afferents of word-like specificity; another is that all these possible conditional stimuli would have to be known in advance in order to be wired in.

as "bell", "whistle", "tuning fork", "G #", "flashing light", "foot-pinch" etc. But for learning in an advanced mammal it is totally unrealistic to assume prior knowledge of all possible conditioning stimuli, and it must be the versatility of our learning, based in my view on our perceptual capacities, that makes us pre-eminent in this way. We must therefore seek other ways of extending the model that will enable a large number of previously unknown conditional stimuli to be used, bearing in mind the requirement for estimates of $P(C)$ for all such stimuli if learning is to be done efficiently.

As well as the assumption that all possible conditional stimuli are known in advance there are other things wrong with Fig. 2. We have labelled the inputs to a neuron with words, but the fact that an appropriate word exists does not make it reasonable to postulate that the dog's brain has an input line with appropriate selectivity. Our facility with words has tricked us into making this step, but it isn't a simple one, and labelling the lines as in Fig. 2 just evades the problem.

The model is inadequate in yet other ways, for it would not show any initial generalisation of conditioning from one stimulus to others sharing similar qualities, nor is there any obvious mechanism for subsequent narrowing of the class of effective stimuli. But although this extension of Fig. 1 does not provide what we are looking for, it does illustrate the enormous gulf that exists between simple cellular models of learning and the real thing. Perhaps it also points to the nub of the problem, namely how to create a representation whose elements would give the same versatility as is provided by labelling the input lines of Fig. 2 with words.

George Boole (1854) thought that his logical functions composed with logical variables satisfactorily formalised the relation between a word and the set of sensations that it symbolised. It might be reasonable to suppose that the inputs to the nerve cell of Fig. 2 correspond to individual items of sensation, so to approach the versatility of word-labelling we need to find a learning system that can use logical functions of its input lines; such logical functions could then be analogous to words. Of course no-one has ever claimed that there is a word for every possible logical function of a set of sensations, so we need not make the impossible demand that our learning model be capable of using every possible logical function of its input lines as a conditional stimulus, but it should be

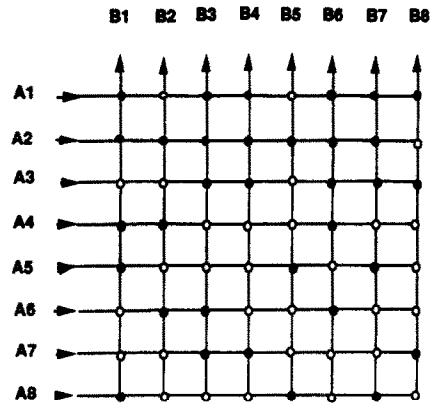


Fig. 3. An associative net (from Longuet-Higgins, Willshaw & Buneman, 1970). The input lines (A1–A8) run horizontally; the output lines (B1–B8) run vertically. (●) and (○) represent synapses that have or have not been turned on. Four associations have been recorded: A1, A2, A3 with B4, B6, B7; A2, A5, A8 with B1, B5, B7; A2, A4, A6 with B2, B3, B6; and A1, A3, A7 with B3, B4, B8. This is an efficient way of storing associations between input and output vectors, provided the probability of any input line being on is not too high, but it is not clear how the prior probabilities of input vectors could be taken account of, and this is necessary for efficient association formation.

able to use a reasonable number of them: for instance it should be able to use as a conditional stimulus some at least of the possible conjunctions of inputs. This is precisely what is claimed for associative nets, so we must take them seriously as candidates for achieving versatility.

ASSOCIATIVE NETWORKS

Figure 3 shows a neural network which associates inputs on the horizontal A lines with outputs on the vertical B lines. In discussing the problem of constructing "... an associative information store which can learn to associate very many pairs of conditional and unconditional stimuli ..." Longuet-Higgins, Willshaw and Buneman (1970) claimed that this provided an "entirely satisfactory" solution. But there are snags.

As we have seen, to do the job of detecting associations properly one should have access to the prior probabilities of the possible conditional stimuli, for without this one cannot tell whether an association is random or genuine; hence in this case the mechanism should have access to the prior probabilities of all the input vectors it can use, not just the probabilities of their components. It is perfectly reasonable to suppose that each synapse should have access to the prior probability of firing of its presynaptic input, but it is impossible to obtain the

probability of a vector from the probabilities of its components unless there are statistically independent. The inescapable conclusion is that, if associations with conjunctions of inputs are to be discovered efficiently, these inputs must be statistically independent of each other under normal conditions of use. Note that independence does not ensure that the prior probabilities of vector inputs are made use of; it merely prevents this being impossible.

Note also that although ignorance of prior probabilities makes *efficient* learning impossible, this does not imply that these networks cannot learn at all. The proofs that nets and perceptions converge on the right solutions (e.g. Rosenblatt, 1959; Minsky & Papert, 1969; Longuet-Higgins et al., 1970; Anderson, Silverstein, Ritz & Jones, 1977) are not cast in doubt, but because they cannot have access to prior probabilities they will be slow and error-prone compared with efficient methods of association detection that make use of this additional information.

I think we need to separate the mechanism that makes $P(C)$ available for all usable conditional stimuli from the learning mechanism that changes connectivity when there is evidence for a genuine association between C and U . The first mechanism corresponds roughly to perception, and I think it is this that is responsible for the versatile behaviour of higher mammals. The second corresponds to the simplest forms of learning, and on this view there is nothing very surprising in the suggestion that this is equally good in all subhuman vertebrates (Mcphail, 1982).

If this distinction has anything to it the name Rosenblatt (1959) chose for his associative network, the *perceptron*, is singularly unfortunate, for the insight it gives does not apply to perception but to learning.

To summarise, learning an association is a definable statistical task and it is possible to specify what is required to do it efficiently. The difficulty that arises is to make accessible the prior probabilities of all input patterns that can act as conditional stimuli; this seems to require the elements of the representation should be independent of each other in the normal environment, and this is not a problem that the current generation of associative networks tackle. Without it I think they do less than half the job: they may model learning, but they do not begin to model perception. There are, however, some other early ideas that are relevant to this problem.

SHOUTING FOR ATTENTION

Thirty-one years ago Oliver Selfridge (1959) proposed a learning model called *Pandemonium* (Fig. 4). It had *computational demons* tuned to detect features of a stimulus, each of which shrieked with an intensity dependent on how closely the actual stimulus resembled the feature to which it was tuned; a set of *cognitive demons* combined these shrieks with weights adjusted according to the resemblance of the set of shrieks to a paradigm such as a letter of the alphabet, and a decision demon then selected the loudest shrieking cognitive demon as the recognised letter. I think the key element here is that the demons shriek with a loudness that is supposed to indicate directly the importance of their message for the next stage of processing; they not only signal, they also attract attention. Since I am arguing that the prior probability of the current scene is one factor that determines its importance for association formation we need a *Probabilistic Pandemonium* in which the shrieks signal definite attributes of the stimulus as in Selfridge's model, but their loudness has a probabilistic interpretation: they signal how unexpected the occurrence of an attribute is on the evidence given by past history and the current presence of other attributes. When an unconditional stimulus U arrives, it is the demons which have just shrieked loudest that should be searched for possible conditional stimuli, for their low prior probability means that the expected number of co-occurrences with U is also low and there is therefore likely to be a genuine new causal factor behind the coincidence.

Another idea relevant to this line of thought is the *novelty filter*, as described, for instance, by Kohonen (1984). Such a device learns the set of usual images entering its input, and is able after a time to suppress those that have previously occurred. This idea has an initial appeal, but one does not want to suppress completely the non-novel inputs; what is needed is a filter that makes novelty salient, but which continues to transmit an image that is usable for ordinary purposes. Selfridge's model used fixed feature filters, and there is plenty of scope for ontogenetically determined structure in the connections of sensory pathways in addition to the plastic component for which there is evidence, as described below (p. 1568). But the point of interest here is the kind of outputs a model of perception should have in order to enable

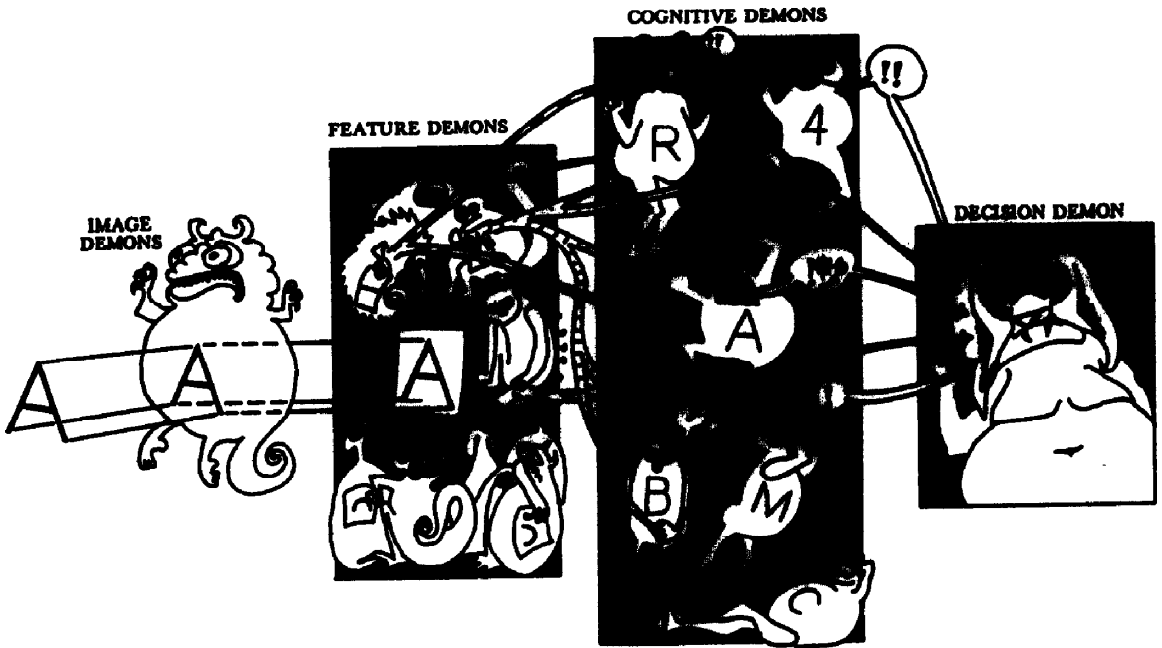
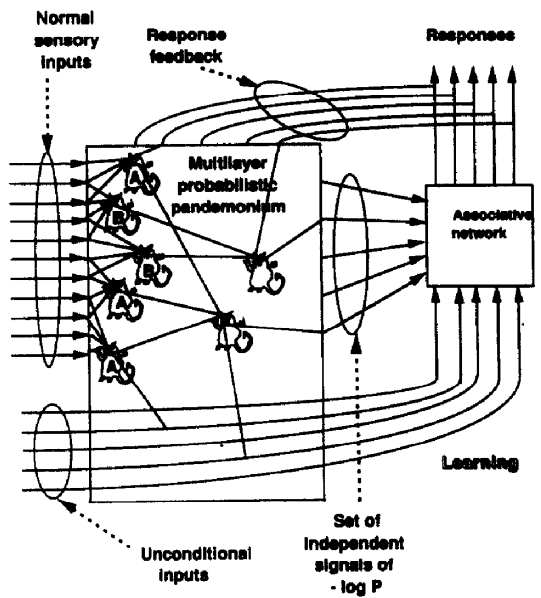


Fig. 4. Selfridge's *Pandemonium* (1959), as depicted by Lindsay and Norman (1977). This scheme for pattern recognition has many interesting adaptive aspects, but the feature of interest here is that each demon shrieks with loudness determined by the fit of the data to the patterns they represent, so they signal the importance of their messages as well as the presence of the pattern. In a probabilistic pandemonium the shrieks would be proportional to $-\log P$, where P is the probability of occurrence of the feature the demon detects.

associative nets to be more versatile and efficient at learning; this does not depend on the preformed or plastic nature of the connections that precede the output.

Probabilistic Pandemonium

Figure 5 is a somewhat fanciful diagram illustrating the complexity of *Perception* compared with *Learning*. The perceptual outputs which act as inputs to the associative process are logically binary, that is they signal the presence of a particular feature of the input when they fire, but let them also be capable of a graded discharge in which the number of impulses depends upon the probability P of the feature being present, based upon the past history of excitation and also upon the other features that are present in the current input. Furthermore let us assume that this graded response approximates to $-\log P$, so the perceptual output neuron shrieks loudly when its feature is unexpectedly present, softly when it is present but in circumstances such that this is not surprising. Finally let us suppose that each of these perceptual output elements fires independently of the others as long as the system is in an environment to which it had adapted. I'll say a little more about how this might be achieved later, but first



Perception

Fig. 5. Diagram illustrating the complexity of perception and the simplicity of learning. The task of providing a representation whose elements are independent, and of signalling $-\log P$ for all of them, is much more difficult than the simple associative task of learning. Note, however, that the two are probably not completely separable, for the results of the associative process are likely to influence the demons of the probabilistic pandemonium through feedback.

look at how desirable the output of such a perceptual representation would be to act as the input to the next, associative, stage.

The impulse frequency on any active perceptual output line gives $-\log P$, where P is the probability of the event or happening signalled by that line; in other words it signals how unexpected the feature or event is, given the past history of its occurrence and the other sensory events that are occurring. As we have seen, this is just what is needed in order to assess whether the co-occurrence of this event with another event U is random or not. But this representation does more than provide this for each single output line: because the outputs are statistically independent, the probability of a combination of them is the product of their individual probabilities, so the prior probability of all the outputs together *or of any subset of them* is simply the sum of the number of impulses each is firing.

This idea of a probabilistic pandemonium is obviously informal and preliminary and is based on the fact that likelihoods can be multiplied or their logs added. The use of likelihoods to decide statistical problems goes back to Fisher (1925), and Kullback (1978) relates likelihoods to modern information theory and develops the methods further. If one considers a plausible physiological realisation, the requirement that the demons shriek with loudness proportional to $-\log P$ would not be too hard to approximate, for any mechanism of adaptation or habituation discounts frequently repeated events and thus leads to something like the desired "unexpectedness" signal. The requirement for independence is harder.

INDEPENDENCE

Independence of the output signals implies that the presence of other outputs does not affect the significance (i.e. the prior probability) of any given one; positive or negative correlations among the sensory inputs thus have to be taken into account. If this approximate independence of the representative elements is achieved, then the proposed representation can show how unexpected a signal is, given the other sensory stimuli that are present. Two methods of approximating independence can be suggested.

Decorrelation

One method is to change the coordinate system used to represent sensory variables

(Barlow, 1989; Barlow & Földiák, 1989). It is a plausible process physiologically because it could be achieved by having mutually inhibitory connections between the outputs whose strength increases when these outputs are correlated, along the lines earlier suggested by Wilson (1975); thus it would do what ordinary lateral inhibition does, but instead of being fixed, the strengths of all the inhibitory interconnections would be increased until the outputs were no longer correlated. Instead of such a negative feedback process, decorrelation could result from regulated positive feedback between outputs, the strength of which diminishes when the feedback connections help to fire a neuron—anti-Hebbian positive feedback. Adaptation to patterns, and contingent adaptation as in the McCollough effect, are thought to result from such mechanisms.

Decorrelating networks of this sort could only handle small subsets of the sensory input, for it would be too vast a job to handle much of it at once. In addition, note that it only handles pairwise correlations, and would be insensitive to triples or larger groups of inputs that might be associated.

Minimum entropy coding

This goes about the task of obtaining a set of independent representative elements by imposing two constraints on the code: first it must be reversible, and second the entropy calculated from the probabilities of the representative elements must be as low as possible (Barlow, 1989; Barlow, Kaushal & Mitchison, 1989). This entropy is always greater than the true entropy of the output calculated from the probabilities of all the output states, unless the representative elements are completely independent of each other; hence by finding a reversible code that reduces the entropy calculated from the representative elements one can diminish the mutual dependencies between these elements. It is a much more general method than decorrelation, but although it has the flavour of a problem suitable for a synthetic neural net it is hard to image a real neural mechanism for generating these codes.

Multi-stage recoding

The above recoding methods are done in a single stage, though they could be repeated in the sort of way that the organisation of visual areas in the cortex suggests (Zeki, 1978; Barlow, 1981). By multi-stage recoding one might obtain

representations which decorrelated associations between orientation and colour, or between stereo and motion depth cues; illusions to be mentioned below suggest that such mechanisms are present.

At first one might think that such recoding mechanisms would make it impossible to keep track of the prior probability of newly generated pattern elements: how would one know the prior probability of an element that received inputs, some inhibitory and some excitatory, from motion, colour and disparity selective units at earlier levels? This is not a serious problem, because once the pattern selectivity of an element has been established its prior probability can be determined afresh, simply by waiting to see how often it fires. Admittedly this would be inefficient, because it would not make use of experience ante-dating the establishment of the pattern selectivity, but it would avoid the need for working out the probability of what might be a very complicated logical function from the probabilities of its components. Thus it seems quite possible that the repeated application of a principle as simple as decorrelation would lead to a representation that was a good approximation to that postulated in the Probabilistic Pandemonium.

Figure 5 shows some of the tasks perception must achieve in order to give associative networks the versatility that has been claimed for them. Of course the diagram only poses the problem, but I hope it suggests to neuroscientists and psychophysicists the important role perception plays in giving higher mammals, especially humans, their intellectual pre-eminence, and I hope it reminds connectionists how much pre-processing of sensory input is required before their adaptive networks will function efficiently.

THE ROLE OF PLASTICITY AND THE SENSITIVE PERIOD

I think there is good experimental evidence for two types of plasticity in the physiological mechanisms of perception. First there is the rapid adaptation or habituation that I have so far talked about, and which we think tends to make the representative elements uncorrelated in the recently experienced environment (Barlow, 1989; Barlow & Földiák, 1989). The illusions which provide the psychophysical evidence for such a process require adaptation times of the order of a minute or so to produce quite marked effects, and following such exposures the illusion

usually vanishes in a few minutes with normal use of the eyes. It is true that very long-lasting and powerful adaptation can produce effects persisting for days or more, but one cannot be sure this does not involve other mechanisms.

The other form of plasticity is that, long known to ophthalmologists, whose physiological basis was revealed by Hubel and Wiesel (1970) in their celebrated experiments on visual deprivation in kittens. In contrast with the first form of plasticity this requires a longer period to induce, it occurs mainly during a restricted sensitive period early in life, and the results persist indefinitely. Note in particular that plasticity in the sensitive period *increases sensitivity* to the inducing experience, whereas the other process *actively desensitises* the system to the adapting stimulus. Földiák (1989, 1990) has developed a network of elements that receive inputs through simple Hebbian synapses and interconnect with each other through anti-Hebbian synapses. Possibly the rapid anti-Hebbian decorrelating mechanism and the slower Hebbian mechanism of the sensitive period may work synergistically in the following manner.

During the sensitive period pathways that are active become permanently connected, but the decorrelating mechanism (if it works the same during the sensitive period as in adult life) influences which pathways are active and can thereby determine the permanent pattern of connections that is established. To caricature the suggested process, decorrelation ensures that different commonly occurring patterns of sensory stimulation each stimulate different cortical neurons, because if they failed to do so commonly occurring patterns would cause correlated outputs from two or more neurons. The Hebbian mechanism then ensures that the pattern of connections made to an activated neuron comes from the inputs that successfully activated it; thus the decorrelation mechanism helps to determine the pattern of connectivity that is permanently laid down.

If the animal is deprived of experience during the sensitive period the Hebbian process presumably still operates, but there are no correlations in the spontaneous maintained activity of the inputs resulting from patterns in the outside world, so there is nothing for the decorrelating mechanism to work on. The consequence will be a somewhat disordered pattern of connectivity guided only by the ontogenetic mechanisms, without any adaptation to the

statistical characteristics of the normal sensory input. This seems a very intriguing possibility that might reconcile the conflicting views about mechanisms, consequences, and purpose of the sensitive period (Movshon & Van Sluyters, 1981), and it could turn out to be a model for the influence of experience on connectivity elsewhere in the brain.

UNCONSCIOUS INFERENCE

The link between the suggestions I have made and Helmholtz's views about unconscious inference or induction will be obvious to anyone familiar with his writings, but one cannot take it for granted, even with this audience, that everyone has read the *Treatise on physiological optics* (Helmholtz, 1925) from cover to cover, and Vol. iii is the part most likely to have been skipped. Warren and Warren (1968) have collected together Helmholtz's main writings on perception, and I think it is surprising that his views about unconscious inference are so rarely quoted or discussed. In his own writings one can perhaps detect a note of disappointment that the philosophers, whom he always took seriously even when he disagreed with them, seem to have dismissed the idea of unconscious inductive inference for the apparently trivial reason that induction is a process necessarily conducted with the conscious use of words.

Helmholtz argued, using examples from a wide field, that our percepts have a status analogous to the conclusions that are drawn by the process of inductive inference, the sole difference lying in the use of words to express major premiss, minor premiss, and conclusion in the latter case. I think his argument is correct, and it is a major inspiration for the views developed here. Thanks to Fisher (1925) and his followers the logic of induction is now understood very much better than in Helmholtz's day, and I have tried to use this understanding to draw conclusions about the operations that must necessarily occur in perception if it is to do what Helmholtz said it did, so let us examine what he said more closely.

A very simple example of the analogy he draws is given in Table 1, which shows how straightforward syllogisms following the acceptance of the major premisses lead to the conclusion that Caius is mortal, or that there is a luminous object in the temporal visual field. Helmholtz used this example to explain why one refers the excitation to the nasal visual field even

Table 1.

Inductive inference	Perception
<i>Major premiss</i> All men are mortal	Stimulation of the temporal retina always results from luminous objects in the nasal field
<i>Minor premiss</i> Caius is a man	The temporal retina is being stimulated
<i>Conclusion</i> Therefore Caius is mortal	Therefore there is a luminous object in the nasal field

when it is actually caused by mechanical pressure on the temporal retina. However to "see" retinal stimulation in the position in the visual field that normally causes such excitation seems such a straightforward phenomenon that one is initially unwilling to attach much importance to it, let alone to call it an inference. But when you realise that a few days wearing inverting glasses changes the "always" condition in the major premiss, and correspondingly changes the position to which the excitation is referred, then it becomes difficult to call it anything else. Stratton (1897) published his account of the effects of wearing inverting glasses just after Helmholtz had died and long after he had formulated his arguments, but I do not believe he made any clear reference to Helmholtz's theory, even though it is hard to conceive any more dramatic verification of its predictions. Of course it is still a mystery how the major premiss is changed as a result of experience, and how most of the complex inferential structure of perception is preserved, becoming adapted to the new conditions simply by this change in the accepted major premiss; perhaps we can dimly foresee a day when the hallowed subject of logic will be recognised as an idealisation of physiological processes that have evolved to serve a useful purpose.

The extent to which our perceptions depend upon normal experience as a reference point does not need emphasising to this audience, but it comes as a continual surprise (even to me) to realise how the principle applies to minute associative details, as proved by a host of illusions which I can do no more than mention. Motion and tilt after-effects; the McCollough effect; micropsia with accommodative effort induced by minus lens or drugs; ditto with the convergence effort induced by prisms; the "toy-town" effect induced by an unnaturally large

range of disparities in a stereo scene; the reverse apparent motion when stereo parallax is present but the motion parallax expected from a movement is absent. All these illusions and many more can be explained as inferences that become false through the failure of a previously valid perceptual major premiss of the type shown in Table 1. They show what an astonishingly deep knowledge of the normal patterns of associated activation our visual system possesses and automatically uses.

The decorrelation model for approximating independence of the representative elements of perception would provide a means of storing this information in the form of the "anti-Hebbian" coefficients of interaction required to decorrelate. At the moment we know that cortical neurons show pronounced adaptation or habituation to patterned stimuli, but we cannot be sure that this interpretation of pattern adaptation is correct. The attempt to prove or disprove the hypothesis might bring some new life to cortical neurophysiology for it might bring out a common feature of cortical processing—the detection of new causal factors in the environment from the new associations they cause among cortical afferents.

CONCLUSION

I think we have neglected the important role that perception must play in providing a representation that promotes the efficient learning of predictive associations. Ignoring this necessary preprocessing of the input is a serious defect both in current learning theory and in work on adaptive networks. But the argument I've given here probably does not go far enough, for it still leaves perception and learning as two processes almost as separate as they seem to be in current thinking. In reality it is pretty certain that what we learn influences what we perceive; in other words the demons of the probabilistic pandemonium are not only sensitive to the statistical structure of the input, but must also be influenced by fear of flogging and hopes of bribery administered on the basis of the results they have delivered. A composite, multistage, process, exploiting direct instruction as well as the statistical structure of the input, might begin to model the astonishingly versatile and useful representation that real perception gives us.

REFERENCES

- Anderson, J. A., Silverstein, J. W., Ritz, S. A. & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, *84*, 413–451.
- Barlow, H. B. (1981). Critical limiting factors in the design of the eye and visual cortex. (the Ferrier Lecture 1980). *Proceedings of the Royal Society, London, B* *212*, 1–34.
- Barlow, H. B. (1989). A theory about the functional role and synaptic mechanisms of visual after-effects. In Blakemore, C. (Ed.), *Vision: Coding and efficiency*. Cambridge: Cambridge University Press.
- Barlow, H. B. & Földiák, P. F. (1989). Adaptation and decorrelation in the cortex. In Durbin, R., Miall, C. & Mitchison, G. J. (Eds.), *The computing neuron*. Mass.: Addison-Wesley.
- Barlow, H. B., Kaushal, T. P. & Mitchison, G. J. (1989). Finding minimum entropy codes. *Neural Computation*, *1*, 406–416.
- Boole, G. (1854). *An investigation of the laws of thought*. New York: Dover publications reprint.
- Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver & Boyd.
- Földiák, P. F. (1989). Adaptive network for optimal linear feature extraction. *International joint conference on neural networks 1989*, Washington, D.C. Vol. 1, pp. 401–405.
- Földiák, P. F. (1990). Forming sparse representations by local anti-Hebbian learning. *Biological Cybernetics* (In press).
- Hebb, D. O. (1949). *The organisation of behaviour*. New York: Wiley.
- Helmholtz, H. von (1925). *Treatise on physiological optics*. Translated from the 3rd German edition (1910), Southall, J. P. C. (Ed.). Washington: Optical Society of America.
- Hubel, D. H. & Wiesel, T. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *Journal of Physiology, London*, *206*, 419–436.
- Kohonen, T. (1984). *Self-organisation and associative memory*. Berlin: Springer.
- Kullback, S. (1978). *Information theory and statistics*. Gloucester, Mass.: Smith.
- Lindsay, P. H. & Norman, D. A. (1977). *Human information processing: An introduction to psychology* (2nd edn, p. 260). New York: Academic Press.
- Longuet-Higgins, H. C., Willshaw, D. J. & Buneman, O. P. (1970). Theories of associative recall. *Quarterly Review of Biophysics*, *3*, 223–244.
- McCormick, D. A., Lavond, D. G., Clark, G. A., Kettner, R. E., Rising, C. E. & Thompson, R. F. (1981). The engram found? Role of the cerebellum in classical conditioning of the nictitating membrane and eyelid response. *Bulletin of the Psychonomic Society*, *18*, 103–105.
- McPhail, E. (1982). *Brain and intelligence in vertebrates*. Oxford: Oxford University Press.
- Minsky, M. & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. Cambridge, Mass.: MIT press.
- Movshon, J. A. & Van Sluyters, R. C. (1981). Visual neural development. *Annual Reviews by Psychology*, *32*, 477–522.
- Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In Block, A. H. & Prokasy, W. F. (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

- Rosenblatt, F. (1959). Two theorems of statistical separability in the perceptron. In *Proceedings of a symposium on the mechanisation of thought processes* (pp. 421-456). London: Her Majesty's Stationary Office.
- Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In *Proceedings of a symposium on the mechanisation of thought processes* (pp. 3-16). London: Her Majesty's Stationary Office.
- Stratton, G. (1987). Vision without inversion of the retinal image. *Psychological Review*, 4, 341-360 and 463-481.
- Sutton, R. S. & Barto, A. G. (1981). Towards a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, 88, 135-170.
- Warren, R. M. & Warren, R. P. (1968). *Helmholtz on perception: Its physiology and development*. New York: Wiley.
- Widrow, G. & Hoff, M. E. (1960). Adaptive switching circuits. *Institute of Radio Engineers, western electronic show and convention, convention record, 1960, Part 4*, pp. 96-104.
- Wilson, H. R. (1975). A synaptic model for spatial frequency adaptation. *Journal of Theoretical Biology*, 50, 327-352.
- Yeo, C. H., Hardiman, M. J. & Glickstein, M. (1985). Classical conditioning of the nictitating membrane response of the rabbit (3 papers). *Experimental Brain Research*, 60, 87-98; 99-113; 114-125.
- Zeki, S. (1978). Functional specialisation in the visual cortex of the rhesus monkey. *Nature, London*, 274, 423-428.