Sector F Lambda Complex

The fall of lambda calculus 04.06.2018

Contents

Preface

"Don't be afraid of talking nonsense! But you must pay attention to your nonsense." — Ludwig Wittgenstein

This bits will perhaps only be understood by those who have themselves already thought the thoughts which are expressed in it, or similar thoughts. It is therefore not a guide or text-book. Its objective would be attained if it afforded pleasure to one who read it with understanding.

How far my efforts agree with those of other hackers I will not decide. Indeed what I have here makes no claim to novelty in points or details; and therefore I give no sources, because it is indifferent to me whether what I have thought has already been thought before me by another.

If this bits have value it consists in that thoughts are expressed in it, and this value will be the greater the better the thoughts are expressed.

I have not embellished with swelling or magnificent words, nor stuffed with rounded periods, nor with any extrinsic allurements or adornments whatever, with which so many are accustomed to embellish their works; for I have wished either that no honor should be given it, or else that the truth of the matter and the weightiness of the theme shall make it acceptable.

Here I am conscious that I have fallen short of the possible. My powers are insufficient to cope with the task, may others come and do it better.

Realistic manifesto

— Naum Gabo and Antoine Pevsner, 1920

We proclaim: For us, space and time are born today. Space and time: the only forms where life is built, the only forms, therefore, where art should be erected.

States, political and economic systems, die under the push of the centuries: ideas crumble, but life is robust; it grows and cannot be ripped up, and time is continuous in life's true duration. Who will show us more efficient forms? Which great human will give us more solid foundations? Which genius will conceive for us a legend more elating than the prosaic story that is called life?

The fulfillment of our perception of the world under the aspects of space and time: that is the only goal of our plastic creation.

And we do not measure our work by the yardstick of beauty, we do not weigh it on the scales of tenderness and feeling. The plumb line in hand, the look accurate as a ruler, the mind rigid as a compass, we are building our works as the universe builds. This is why, when we represent objects, we are tearing up the labels their owners gave them, everything that is accidental and local, leaving them with just their essence and their permanence, to bring out the rhythm of the forces that hide in them.

1. In painting, we repudiate color as a pictorial element. Color is the idealized and optical face of the objects. The exterior impression is superficial. Color is accidental and has nothing in common with the internal content of bodies.

We proclaim that the tone of bodies, that is, their material substance absorbing the light, is their sole pictorial reality.

1. We deny the line its graphic value. In the real life of the bodies, there is nothing graphic. The line is only an accidental trace that humans leave on objects. It has no connection to essential life and to the permanent structure of things. The line is a merely graphic, illustrative, decorative element.

We acknowledge the line only as the direction of static forces that are hidden in the objects, and of their rhythms.

1. We disown volume as a plastic form of space. One cannot measure a liquid in inches. Look at our real space: What is it if not a continuous depth?

We proclaim depth as the unique plastic form of space.

1. We disown, in sculpture, mass as a sculptural element. Every engineer knows that the static forces of solids, their material resistance, are not a function of their mass. Example: the rail, the buttress, the beam . . . But you sculptors of any trend and any nuance, you always cling to the old prejudice according to which it is impossible to free volume from mass. Like this: We take four planes and we make of them the same volume that we would make with a mass of one hundred pounds.

We thus restore to sculpture the line as direction, which prejudice had stolen from it. This way, we affirm in sculpture depth, the unique form of space.

1. We repudiate: the millennial error inherited from Egyptian art: static rhythms seem as the sole elements of plastic creation.

We proclaim a new element in plastic arts: the kinetic rhythms, which are essential forms of our perception

of real time . . .

Art is called upon to accompany man everywhere where his tireless life takes place and acts: at the workbench, at the office, at work, at rest, and at leisure; work days and holidays, at home and on the road, so that the flame of life does not go out in man.

Factor analysis

Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved latent variables called factors. For example, it is possible that variations in six observed variables mainly reflect the variations in two unobserved (underlying) variables.

Factor analysis searches for such joint variations in response to unobserved latent variables. The observed variables are modelled as linear combinations of the potential factors, plus "error" terms. The theory behind factor analytic methods is that the information gained about the interdependencies between observed variables can be used later to reduce the set of variables in a dataset.

Factor analysis performs a maximum likelihood estimate of the so-called loading matrix, the transformation of the latent variables to the observed ones, using expectation-maximization (EM).

Advantages

Both objective and subjective attributes can be used provided the subjective attributes can be converted into scores.

Factor analysis can identify latent dimensions or constructs that direct analysis may not.

It is easy and inexpensive.

Disadvantages

Usefulness depends on the researchers' ability to collect a sufficient set of product attributes. If important attributes are excluded or neglected, the value of the procedure is reduced.

If sets of observed variables are highly similar to each other and distinct from other items, factor analysis will assign a single factor to them. This may obscure factors that represent more interesting relationships.

Naming factors may require knowledge of theory because seemingly dissimilar attributes can correlate strongly for unknown reasons.

Expectation Maximization

The EM algorithm is used to find (local) maximum likelihood parameters of a statistical model in cases where the equations cannot be solved directly. Typically these models involve latent variables in addition to unknown parameters and known data observations. That is, either missing values exist among the data, or the model can be formulated more simply by assuming the existence of further unobserved data points.

Finding a maximum likelihood solution typically requires taking the derivatives of the likelihood function with respect to all the unknown values, the parameters and the latent variables, and simultaneously solving the resulting equations.

The EM algorithm proceeds from the observation that there is a way to solve these two sets of equations

numerically. One can simply pick arbitrary values for one of the two sets of unknowns, use them to estimate the second set, then use these new values to find a better estimate of the first set, and then keep alternating between the two until the resulting values both converge to fixed points. It's not obvious that this will work, but it can be proven that in this context it does, and that the derivative of the likelihood is (arbitrarily close to) zero at that point, which in turn means that the point is either a maximum or a saddle point.

The EM iteration alternates between performing an expectation (E) step, which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters, and a maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step.

In general, multiple maxima may occur, with no guarantee that the global maximum will be found. Some likelihoods also have singularities in them, i.e., nonsensical maxima.

Reductionist

"Man is evidently the most intelligent animal but also, it seems, the most emotional." — D.O Hebb

The reductionist; it seek a common ground with the physics, neurologist, and cognitive scientists, to show them how this bits relate to their problems.

This bits presents a theory of behavior that is based as far as possible on the physiology of the nervous system, and makes a sedulous attempt to find some community of neurological and psychological conceptions.

The effect of a clear-cut removal of cortex outside the speech area is often astonishingly small; at times no effect whatever can be found. Intelligence must be affected by any large brain injury, yet sometimes it seems not to be.

The level of intelligence-test performance is a function of the concepts a patient has already developed. Once developed, a concept is retained, despite brain damage that, if it had occurred earlier, would have prevented the development.

It has already been suggested that the essential need is to find out how to handle thought, and related processes, more adequately.

A sensory dominance of behavior. It is the idea that behavior is a series of reactions (instead of actions), each of which is determined by the immediately preceding events in the sensory systems.

Now the tradition in psychology has long been a search for the property of the stimulus which by itself determines the ensuing response, at any given stage of learning.

There are three points here: one responses are determined by something else besides the immediately preceding sensory stimulation. It does not deny the importance of the immediate stimulus; it does deny that sensory stimulation is everything in behavior, autonomous central process.

When the detailed evidence of neurophysiology and histology is considered, the conclusion becomes inevitable that the nonsensory factor in cerebral action must be more consistently present and of more dominating importance than reluctant psychological theory has ever recognized, the problem for psychology is no longer to account for the existence of set but to find out how it acts and above all to learn how it has the property of a consistent, selective action instead of producing the random-error distribution postulated by Hull (1943) in his "oscillation principle."

The central nervous system is continuously active, in all its parts, whether exposed to afferent stimulations or not. It is taken here as a working assumption that the EEG is correlated with neural firing.

Sensory activity is essential to the regulation of central neural firing but not essential to initiating it, the EEG is the summation of single sharp potentials, the result of actual cellular firing. The activity is not necessarily maintained by sensory activity, sensory processes, instead of supporting synchronous, rhythmic firing and large potentials in the EEG, have the opposite effect. They introduce irregularity and flattening of the electrical record. In the second place, large potentials, or "hypersynchrony," negate or may negate normal function (Jasper, 1941). That is, sensory activity breaks up hypersynchrony and makes for normal, coordinated, adaptive activity.

There are two radical modifications of earlier ideas transmission is not simply linear but apparently always involves some closed recurrent circuits; and a single impulse cannot ordinarily cross a synapse-two or

more must act simultaneously, and two or more afferent fibers must therefore be active in order to excite a third to which they lead.

In a single system, and with a constant set of connections between neurons in the system, the direction in which an entering excitation will be conducted may be completely dependent on the timing of other excitations.

The immediate objective is to show that "simple" perceptions are in fact complex: that they are additive, that they depend partly on motor activity, and that their apparent simplicity is only the end result of a long learning process.

One must decide whether perception is to depend (1) on the excitation of specific cells or (2) on a pattern pf excitation whose locus is unimportant.

It is notorious that attention wanders, and this is another way of saying that in perception any figure is unstable, one looks at this part of the configuration and that, and notices its corners or smooth contour, in the intervals between seeing the figure as a whole. In ordinary perception, moreover, the instability is far greater.

Identity is defined here as referring to the properties of association inherent in a perception.

The reference has two aspects: first, a figure is perceived as having identity when it is seen immediately as similar to some figures and dissimilar to others that is, when it falls at once into certain categories and not into others.

This similarity can be summed up as spontaneous association, since it may occur on the first exposure to the stimulus object.

Secondly, the object that is perceived as having identity is capable of being associated readily with other objects or with some action, whereas the one that does not have identity is recalled with great difficulty or not at all, and is not recognized or named easily.

Identity of course is a matter of degree and, as I shall try to show, depends on a considerable degree of experience, it is not innately given.

The real point at which he is driving seems to be that there are genuine differences of associability in different patterns.

Recognizability goes with selective similarity, or generalization: the figure that is readily remembered is also perceived as belonging to a particular class of figure, as remembered so.

With experience the perception of identity increases.

Thus identity is a matter of degree: readiness of recognition, and the extent to which generalization is selective.

Riesen (1947) has fully confirmed the conclusion that ordinary visual perception in higher mammals presupposes a long learning period.

The course of perceptual learning in man is gradually, proceeding from a dominance of color, through a period of separate attention to each part of figure, to a gradually arrived at identification of the whole as a whole: an apparently simultaneous instead of a serial apprehension.

Animal experiments and human clinical data alike indicate that the perception of simple diagrams as

distinctive wholes is not immediately given but slowly acquired through learning.

Receptor adjustment (head-and-eye movement) is the most prominent feature of visual perception whether in rat, chimpanzee, or man except in long-practiced habits.

The thesis is that eye movements in perception are not adventitious. They contribute, constantly and essentially, to perceptual integration, even though they are not the whole origin of it.

There is one main question: whether recognition, or a selective discriminatory response, require the excitation of specific neural cells or not.

Sensory equipotentiality can be coined for Lashley's "equivalence of stimuli" which is ambiguous.

"Equivalence of stimuli" has a double reference. It may mean only (1) that different stimuli can arouse the same response, (2) that it does not matter what sensory cells are excited in order to get a certain response, and this is interpretation.

Equipotentiality implies (1) that any other retinal cells, excited in a circular pattern, will elicit the same response with either left or right hand; (2) that the right hemisphere may be extirpated, and the left will be found then to have "learned" whatever the right did; and (3) that this transfer of learning from one set of cells, primarily excited, to other sets does not depend on an earlier experience that set up connections between them.

The question now is whether gradients and fields are the only mechanism of selective neural action or whether they are combined with an equally important mechanism of connections and specialized conduction paths.

Man or animal tends to perceive relative rather than absolute intensity, extent, or frequency.

Marshall and Talbot point out that the whole visual system, from receptors to the several layers of the cerebral cortex, must act to damp strong stimulations, amplify weak ones.

It is evident that the perception of one or two parts of a figure may be the clue to recognizing the whole.

D.O Hebb propose that the human capacity for recognizing patterns without eye movement is possible only as the result of an intensive and prolonged visual training that goes on from the movement of birth, during every moment that the eye are open, with an increase in skill evident over a period of 12 to 16 years at least.

During the continuous, intensive, and prolonged visual training of infancy and childhood, we learn to recognize the direction of line and the distance between points, separately for each grossly separate part of the visual field.

Growth of the assembly

Let us assume then that the persistence or repetition of a trace tends to induce lasting cellular changes that add to its stability.

The assumption can be precisely stated as follows: When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.

The general idea is an old one, that any two cells, or systems of cells that are repeatedly active at the

same time will tend to become "associated" so that activity in one facilitates activity in the other.

One cannot guess how great the changes of growth would be; but it is conceivable, even probable, that if one knew where to look for the evidence on would find marked differences of identity in the perceptions of child and adult.

To get psychological theory out of a difficult impasse, one must find a way of reconciling tree things without recourse to animism: perceptual generalization, the stability of memory, and the instabilities of attention.

Our problem essentially is to see how a particular sensory event can have the same central effects on different occasions despite spontaneous central activity.

Considering the association areas as made up of a population of transmission units, two factors must affect the length of time needed to bring all these units under control.

One is the number of controlling fibers leading from sensory areas into association areas. The second is the number of transmission units in the association areas themselves.

We can then regard the stage of primary learning as the period of establishing a first environmental control over the association areas, and so, indirectly over behavior.

The learning occurs when the events to be associated ca already command organized trains of cortical activity; in other words, when the environment has a control of association areas that can be repeated, so that the central activity is not at random and the stimulation can impinge on the same central pattern when the training situation is repeated.

The characteristic adult learning is learning that takes place in a few trials, or in one only.

So adult learning is typically an interaction of two or perhaps three organized activities; being organized, they are capable of a continued existence after cessation of the stimulation that set them off, which gives time for the structural changes of permanent learning to take place.

This organized activity of the association areas is subject to environmental control. To the extent that the control is effective, and re-establishes the same central pattern of activity on successive trials, cumulative learning is possible.

Adult learning is thus a changed relationship between the central effects of separate stimulations, and does not directly concern the precipitating stimulus or, primary, the motor response whose control is embedded in the central activity.

The facts already discussed have indicated that one-trial learning occurs only as the association of concepts with "meaning" having, that is, a large number of associations with other concepts.

But more: the perception of an actual object that can be seen from more than one aspect, and touched, heart, smelled and tasted involves more that one phase cycle.

It must be a hierarchy: of phases, phase cycles, and a cycle of series of cycles.

"Cycle" is of course temporal: referring not to a closed anatomical pathway but to the tendency of a series of activities to recur, irregularly.

The two ideas or concepts to be associated might have not only phases, but one or more subsystems in common.

Two concepts may acquire a latent "association" without even having occurred together in the subject's past experience.

Essay on mind

Mind is the central psychological problem, although it is no longer fashionable to say so, psychologists prefer to talk about "cognitive processes" instead.

They also, most of them, abstain from discussion of what those processes consist of and how their effects are achieved.

It is inaccurate-worse, it is misleading, to call psychology the study of behavior:

It is the study of the underlying processes, just as chemistry is the study of the atom rather than pH values and test tubes; but behavior is the primary source of data in modern psychology.

All science, from physics to physiology, is a function of its philosophic presuppositions, but psychology is more vulnerable than others to the effect of misconception in fundamental matter's because the object of its study is after all the human mind and the nature of human thought.

There is a well-developed specially called social psychology, which certainly sounds like social science; but social behavior can be considered from a biological point of view.

The idea that mind is a spirit is a theory of demonic possession, a form of the vitalism that biology got rid of a century ago.

Monism, the idea that mind and matter are not fundamentally different but different forms of the same thing: in practice in psychology, the idea that mental processes are brain processes.

Mind is the capacity for thought, consciousness, a variable state, is a present activity of thought processes in some form; and though itself is an activity of the brain.

What about the computer?

Doies it think, and if so does that make it conscious?

The mammalian brain is enormously more complex than any present AI. not only in the number of functional elements but also in its connections, the individual neuron frequently having synaptic connection with upwards of a thousand others.

The argument then is that a computer built on the plan of the mammalian brain, and of a complexity at least equal to that of the brain of the laboratory rat, might be conscious.

Nonsense about behaviorism

Another impediment to understanding in this field common failure to see how the behavioristic emphasis in psychology came about, together with failure to understand the meaning of that term, behaviorism, known in some circles as "cognitive computing" has become a term of abuse.

In 1913, Watson did two things, that must be distinguished if we are to understand his position.

He proposed a general method for psychology, and he began the development of a theory to agree with

it.

The theory was soon found to be defective, but it was not stupid and it had the immediate value of stimulating the research that led to its refutation and at the same time added to our understanding of human beings.

A factual theoretical development which have changed the study of mind and behavior as radically as genetics changed the study of heredity; have all been the product of objective analysis, that is to say: behavioristic analysis.

Thorndike, and later Skinner, on learning and reinforcement, Binet, and later Piaget on the development of though in childhood, kOhler on insight, Lashley o perception, Tolman on spatial orientation (the cognitive map), Beach on instinct as an aspect of intelligence (or vice versa), Lewin on cognitive motivation, Broadbent on the channeling of perception, and all from D.O Hebb's laboratory, showing the dependence of mind and thought on a close relation with the environment.

Unfortunately for Watson, the facts were wrong, at best incomplete, but it was not only in 1938, 25 years later, that this was finally established by Lorente de N0, and only in 1940 that the changed situation was brought to the attention of the psychological world by Hilgard and Marquis (1940), in a brief reference to the possibility of the brain's holding input before transmitting it to the muscles.

Meanwhile, psychologists had been attacking Watson's theory by experimental means, beginning with Hunter's (1913) demonstration of delayed response in Harvey Carr's laboratory at Chicago, which showed that raccoons and children possessed a kind of cognitive function that Watson's theory denied them.

"Behaviorism" is not synonymous with any particular theory of behavior; Lashley and Toldman both called themselves behaviorists, though each of them spend the better part of his career showing the inadequacy of the particular theory of behavior that Watson had proposed.

Programmed learning is not a substitute for teaching, but a valuable tool for the teacher's use where learning is concerned.

To talk of behaviorism as bind incompetence is ignorance or prejudice or both. It is certainly not a mark of fellowship.

There are things going on in ones mind that are not introspectable at all.

The role of introspection in psychological research was challenged in George Humphrey's book Thinking (1951) showed that classical instrospectors though were describing a sensation were really describing the external event or object that had given rise to the sensation.

He generalize his convincing and has not been refuted:

"We perceive objects directly, not through the intermediary of 'presentations', 'ideas', or 'sensations'.

Similarly, we imagine objects directly, not through the intermediary of images, though images are present as an important part of the whole activity" (p. 129)

What one is aware of in perception is not a percept but the object that is perceived; what is given in imagination is an illusory external object, not an internal mental representation called an image.

The great 20th century change was made by Adrian on spontaneous firing, in 1934, and Lorente de N0 on holding, in 1938.

In 1920 the cerebral cortex was solely a variable transmitter: in the common analogy, a mere switchboard.

Watson had full scientific warrant for denying any purely internal activity such as ideas or imagery, set or delayed response (in which excitation is held for short periods) or perception in the form of an elaboration of sensory input.

Today it is apparent that the nervous system is fully capable of handling ideation and creative thought. But there is hard core of learning theorist who, denying any interest in neural function, still conspicuosly limit their research and discussion to those features that Watson's neurophysiology could comprehend.

Other psychologist, many of the, are set in a different theoretical posture, but still one that was determined by the long battle against Watsonian theory.

They too deny any interest in neurology and still be deeply influenced by an earlier set of neurological ideas.

Parallelism

The theory that mental events and brain events run side by side, perfectly, correlated but not causally related: in the old analogy, like two clocks that stay perfectly in step but not because either influences the other.

It has been highly regarded as a way of avoiding commitment to an interaction of mind and body, or even worse, identifying them, while recognizing how closely influences the other.

It has been highly regarded as a way of avoiding commitment to an interaction of mind and body, or even worse, identifying them, while recognizing how closely they are related.

Parallelism says that the actors in the theater, representing anger and fear did so with no guidance from their conscious minds; whatever thought there may have been in those entirely separated minds, the bodies functioned on the stage as self-programmed robots.

In the two-clock comparison with parallelism, the two clocks are separate entities by virtue of their reparation in space; if in addition to being identical in function.

They also occupied the same space, as mental activity and brain activity appear to do visual, auditory, verbal fluency, verbal comprehension, and so on each relating to particular parts of the brain, they would be on clock.

The objective evidence tells me that something complex goes on inside my head, I conclude therefore that something else is active also.

Reductionism

Useful tool but handle with care.

- D.O Hebb

Properly done, reductionism does not substitute neurophysiology for psychology; when it assumes that pain or ecstasy consists of neural firing it recognizes the reality of pain and joy at the same time; it does not try to explain them away.

It is the method of theoretical analysis followed resynthesis, whose validity depends strictly on wheter the result accords with the psychological as well as the neurological evidence.

What has given reductionism a bad name is the conclusion, after a the

oretical analysis of a mental variable has been made and it is "reduced" to some pattern of neural activity, that mental process in effect no longer exists.

Obviously this is all nonsense; when a complex is reduced, theoretically, to its component parts, the whole still exists.

Anxiety must be a pattern of firing of neurons in the limbic system, but the pattern is as real as the individual neurons.

En engineer designing a bridge must think at several levels of complexity.

His conceptions of the bridge as a whole is very molar, in terms let us say of a center span, two side spans, two piers, and two abutments or (if you are from CR a single bailey thing and call it done).

When the engineer tuns to the design of the center span he begins to think in terms of lower-order units such as steel beams, rivets or welding, and masses of reinforced concrete. However, these items are still very molar.

Are engineer if asked would say that a steel I-beam is just a special arrangement of atoms or of electrons, neutrons, and so forth.

At this level of analysis there is indeed nothing but atoms or atomic particles.

But there are other levels of analysis; from the point of view of a practical man, all this stuff about atoms may be fine in theory but then it comes to bridge-building it is no more than theory.

At this level of analysis the I-beam is an elementary unit, obviously real and no fiction.

Reality now is steel and concrete.

So from one point of view, reality is the atom, and the steel beam being a convenient way of dealing with large numbers of atoms in a particular pattern; while from another, the steel that is heavy and cold and resistant to distortion is reality, and atoms are theoretical items only.

For different modes of thought, different realities: "reality" referring evidently to the mode of being that one takes for granted as the starting point of thought.

Even if we could identify the part played by every of the 10 or more billion neurons in the brain, the human mind of the scientist is obviously incapable of thinking of the whole activity in such terms.

It is not possible to follow the varying patterns of the firing of these cells as individual units.

Reductionism is not a means of abolishing psychological entities and processes but a way of learning more about them.

Connectionism

Understanding through the interplay of multiple sources of knowledge. It is clear that we know a good deal about a large number of different standard situations. Several theories have suggested that we store this knowledge in terms of structures called variously: scripts, (Schank, 1976), frames (Minsky, 1975), or schemata (Norman & Bobrow, 1976; Rumelhart, 1975). Such knowledge structures are assumed to be the basis of comprehension. A great dial of progress has been made within the context of this view.

However, it is important to bear in mind that mostly everyday situations cannot be rigidly assigned to just a single script. They generally involve an interplay between a number of different sources of information.

Representations like scripts, and schemata are useful structures for encoding knowledge, although we believe they only approximate the underlying structure of knowledge representation that emerges from the class of models we consider in this bits.

Parallel distributed processing

A number of different pieces of information must be kept in mind at once. Each plays a part, constraining others and being constrained by them. What kinds of mechanisms seem well suited to these task demands? Intuitively, these tasks seem to require mechanisms in which each aspect of the information in the situation can act on other aspects. simultaneously influencing other aspects and being influenced by them. To articulate these intuitions, we and others have turned to a class of models we call Parallel Distributed Processing (PDP) models. These models assume that information processing takes place through the interactions of a large number of simple processing elements called units, each sending excitatory and inhibitory signals to other units.

How a pattern learns

So far, we have seen how we as model builders can construct the right set of weights to allow one pattern to cause another/ The interesting thing, though. is that we do not need to build these interconnection strenghts in by hand. Instead, the patterns associator can teach itself the right set of interconnections through experience processing the patterns in conjuction with each other.

A number of different rules for adjusting connection strengths have been proposed. One of the first, and definitely the best known is due to D. O. Hebb (1949) Hebb's actual proposal was not sufficiently quantitative to build into an explicit model. However, a number of different variants can trace their ancestry back to Hebb. Perhaps the simplest version is:

When unit A and unit B are simultaneously excited, increase the strength of the connection between them.

A natural extension of this rule to cover the positive and negative activation values allowed in our example is:

Adjust the strength of the connection between units A and B in proportion to the product of their simultaneous activation.

In this formulation, if the product is positive, the change makes the connection more excitatory, and if the product is negative, the change makes the connection more inhibitory. For simplicity of reference, we will

call this the Hebb rule, although it is not exactly Hebb's original formulation.

It turns out that Hebb rule as stated here has some serious limitations, and, to our knowledge, no theorists continue to use it in this simple form. More sophisticated connection modulation schemes have been proposed; All these learning rules have the property that they adjust the strengths of connections between units on the basis of information that can be assumed to be locally available to the unit. Learning, then, in all of these cases, each connection without the need for any overall supervision. Thus models which incorporate these learning rules train themselves to have the right interconnections in the course of processing the members of an ensemble of patterns.

We already have noted Hebb's contribution of the Hebb rule of synaptic modification; he also introduced the concept of cell assemblies, a concrete example of a limited form of distributed processing, and discussed the idea of trace of activation within neural networks. Hebb's ideas were cast more in the form of speculations about neural functioning than in the form of concrete processing models, but his thinking captures some of the flavor of parallel distributed processing mechanisms.

Cognitive Science or Neuroscience?

One reason for the appeal of PDP models is their obvious "physiological" flavor: They seem so much more closely tied to the physiology of the brain than are other kinds of information-processing models. The brain consists of a large number of highly interconnected elements which apparently send very simple excitatory and inhibitory messages to each other and update their excitations on the basis of these simple messages.

The connectionism framework

It is useful to begin with an analysis of the various components of our models and then describe the various specific assumptions we can make about these components. These are eight major aspects of parallel distributed processing model:

- 1. A set of processing units
- 1. A state of activation
- 1. An output function for each unit
- 1. A pattern of connectivity among units
- 1. A propagation rule for propagating patterns of activities through the network of connectivities
- 1. An activation rule for combining the inputs impinging on a unit with the current state of that unit to produce a new level of activation for the unit
- 1. A learning rule whereby patterns of connectivity are modified by experience
- 1. An environment within which the system must operate

1. A set of processing units

Any parallel activation model begins with a set of processing units. In some models these units may represent particular conceptual objects; in others they are simply abstract elements over which meaningful patterns can be defined. When we speak of distributed representation, we mean one in which the units represent small, feature-like entities.

A unit's job is simply to receive input from its neighbors and, as a function of the input it receives, to compute an output value which sends to its neighbors.

The system is inherently parallel in that many units can carry out their computations at the same time.

Within any system we are modeling, its useful to characterize three types of units: input, output, hidden.

Input units receive inputs from sources external to the system under study.

The output units send signals out of the system.

The hidden units are those whose only inputs and outputs are within the system we are modeling.

2. The state of activation

We need a representation of the state of the system at time T. This is primary specified by a vector of N numbers, a(t), representing the pattern of activation over the set of processing units.

It is the pattern of activation over the set of units that capture what the system is representing at any time.

It is useful to see processing in the system as the evolution, through time, of a pattern of activation over a set of units.

Different models make different assumptions about the activation values a unit is allowed to take on.

Activation values may be continuous or discrete. If they are continuous, they may be unbounded or bounded.

If they are discrete, they may take binary values or any of a small set of values.

3. Output of the units

Units interact.

They do so by transmitting signals to their neighbors. The strength of their signal, and therefore they degree to which they affect their neighbors, is determined by their degree of activation. In vector notation, we represent the current set of output values by a vector, o(t).

In some of our models the output level is ，equal to the activation level of the unit.

4. The pattern of connectivity

Units are connected to one another. It is this pattern of connectivity that constitutes what the system, knows and determines how it will respond to any arbitrary input.

Specifying the processing system and the knowledge encoded therein is, in a parallel distributed

processing model, a matter of specifying this pattern of connectivity among the processing units.

We assume that each unit provides an additive contribution to the input of the units to which it is connected.

In such cases, the total input to the unit is simply the weighted sum of the separate inputs from each of the individual units.

That is, the input from all of the incoming units are t multiplied by a weigh and summed to get the overall input to that unit.

A positive weight represents an excitatory input a negative weight represents n inhibitory input.

The pattern of connectivity is very t. It is this pattern which determines what each unit represents.

One important issue that may determine both how much information can be stored and how much serial processing the network must perform is the fan-in and fan-out of a unit.

The fan-in is the number of elements that either excite or inhibit a given unit.

The fan-out of a unit is the number of units affected directly by a unit.

5. The rule of propagation

We need a rule which takes the output vector, o(t), representing the output values of the units and combines it with the connectivity matrices to produce a network input for each type of input into the unit.

In vector notation we can write network i(t) to represent the network input vector for inputs of type i.

6. Activation rule

We need a rule whereby the network inputs of each type impinging on a particular unit are combined with one another and with the current state of the unit to produce a new state of activation.

We need a function, F, which takes a(t) and the vectors **net**j for each different type of connection and produces a new state of activation.

In the simplest cases, when F is the identity function and when all connections are of the same type, we can write a(t+1)=Wo(t)=net(t).

Sometimes F is a threshold function so that the net input must exceed some value before contributing to the new state of activation.

7. Modifying patterns of connectivity as a function of experience.

Changing the processing or knowledge structure in a parallel distributed processing model involves modifying the pattern of inter-connectivity.

In principle this can involve three kinds of modifications:

1. The development of new connections.
2. The loss of existing connections.
3. the modifications of the strengths of connections that already exists.

8. Representation of the environment

It is crucial in the development of any model to have a clear model of the environment in which this model is to exist. In parallel distributed processing models, we represent the environment as a time-varying stochastic function over the space of input patterns.

We image that at any point in time, there is some probability that any of the possible set of input patterns is impinging n the input units, this probability function may in general depend on the history of inputs to the system as well as outputs of the system.

Each entity is represented by a pattern of activity distributed over many computing elements, and each computing element is involved in representing many different entities.

Sequential symbol processing

The obvious way to allocate the hardware is to use a group of units for each possible role within a structure and to make the pattern of activity in this group represent the identity of the constituent that is currently plating that role.

This implies that only one structure can be represented at a time unless we are willing to postulate multiple copies of the entire arrangement.

One way of doing this, using units with programmable rather than fixed connections is required.

The recursive ability to expand parts of a structure for indefinitely many levels and the inverse ability to package up whole structures into a reduced form that allows them to be used as constituents of larger structures is the essence of symbol processing.

It allows a system to build structures out of things that refer to other whole structures without requiring that these other structures be represented in all their cumbersome detail.

This is exactly what is provided by sub-patterns that stand for identity / role combinations.

They allow the full identity of the part to be accessed form a representation of the whole and a representation of the role that the system wishes to focus on, and they also allow explicit representation, of an identity and a role to be combined into a less cumbersome representation, so that several identity / role combinations can be represented simultaneously in order to form the representation of a larger structure.

By encoding each piece of knowledge as a large set of interactions, it is possible to achieve useful properties like content-addressable memory and automatic generalization, and new items can be created without having to create new connections at the hardware level.

Another hard problem is to clarify the relationship between distributed representations and techniques used in artificial intelligence like schemas, or hierarchical structural descriptions.

Existing artificial intelligence programs have great difficulty in rapidly finding the schema that best fits the current situation.

Parallel networks offer the potential of rapidly applying a lot of knowledge to this best-fit search, but this potential will only be realized when there is a good way of implementing schemas in parallel networks.

Level of analysis

This bits describe a general framework proposed by Marr for studying and understanding visual perception.

In this framework, the process of vision proceeds by constructing a set of representation starting from a description of the input image, and culminating with a description of three-dimensional objects in the surrounding environment.

The main motivation behind this model was the creation of invariant object representation for the purpose of recognition, which will be independent of the particular viewing direction and irrelevant details in the object's shape.

In a working 1973 paper, written with Carl Hewitt and titled Video Ergo Scio they make the following comment: "Our insistence on using 3-D models for the basic representation of objects does not preclude the use of catalogs of appearances of objects from different view points."

Our view is that both types of representations are required computationally, and both are likely to exist within the human visual system.

In 1971, Roger N. Shepard and Jaqueline Metzler made line drawings of simple objects that differed from one another either by a three-dimensional rotation or by a rotation plus a reflection. They asked how long it took to decide whether two depicted ibjects differed by a rotation and a reflection or merely a rotation. They found that the time taken depended on the three-dimensional angle of rotation necesary to bring the two objects into correspondence. One is led thereby to the notion that a mental rotation of sorts is actually being performed, that a mental description of the first shape in a pair is being adjusted incrementally in orientation until it matches the second, such adjustment requiring greater time when greater angles are involved.

What does it mean to see?

But what of explanation? The development of amplifiers allowed Adrian (1928) and his colleagues to record the minute voltage changes that accompanied the transmission of nerve signals. Their investigation showed that the character of the sensation so produced depended on which fiber carried the message, not how the fiber was stimulated.

But perhaps the most exciting development was the new view that questions of psychological interest could be illuminated and perhaps even explained by neurophysiological experiments.

Barlow (1972) then goes on to summarize these findings in the following way:

The cumulative effect of all the changes I have tried to outline above has been to make us realise that each single neuron can perform a much more complex and subtle task that had previously been through.

Neurons do not loosely and unreliably remap the luminous intensities of the visual image onto our sensorium, but instead they detect pattern elements, discriminate the depth of objects, ignore irrelevant causes of variation and are arranged in an intriguing hierarchy. Furthermore, there is evidence that they give prominence to what is informationally important, can respond with great reliability, and can have their pattern selectivity permanently modified by early visual experience.

This amounts to a revolution in our outlook. It is now quite inappropriate to regard unit activity as a noisy indication of more basic and reliable processes involved in mental operations: instead, we must regard single neurons as the prime movers of these mechanisms. Thinking is brought about by neurons and we should not use phrases like "unit activity reflects, reveals, or monitors thought processes," because the activity of neurons, quire simply, are thought processes.

This revolution stemmed from physiological work and makes us realize that the activity of each single neuron may play a significant role in perception (p.380)

This aspect of his thinking led Barlow to formulate the first and most important of his five dogmas: A description of that activity of a single nerve cell which is transmitted to and influences other nerve cells and of a nerve cell's response to such infliences from other cells, is a complete enough description for functional understanding of the nervous system. There is nothing else "looking at" or controlling this activity, which must therefore provide a basis for understanding how the brain controls behaviour. Barlow, 1972.

We shall return later on to more carefully examine the validity of this point of view, but for now lets us just enjoy it.

But somewhere underneath, something was going wrong. The initial discoveries of the 1950s and 1960s were not being followed by equally dramatic discoveries in the 1970s.

As one reflected on these sorts of issues in the early 1970s, it gradually became clear that something important was missing that was not present in either of the disciplines of neurophysiology or psychophysics. The key observation is that neurophysiology and psychophysics have as their business to describe the behavior of cells or of subjects but not to explain such behavior.

What are the visual areas of the cerebral cortex actually doing? What are the problems in doing it that need explaining, and at what level of description should such explanations be sought?

Gone are the ad hoc programs of computer vision; gone is the restriction to a special visual miniworld; gone is any explanation in terms of neurons except as a way of implementing a method. And present is a clear understanding of what is to be computed, how it is to be done, the physical assumptions on which the method is based, and some kind of analysis of algorithms that are capable of carrying it out.

The other piece of work was Horn's (1975) analysis of shape from shading, which was the first in what was to become a distinguished series of articles on the formation of images. By carefully analyzing the way in which the illumination, surface geometry, surface reflectance, and view-point conspired to create the measured intensity values in an image. If the surface reflectance and illumination are known, one can solve the surface geometry (see also Horn, 1977). Thus from shading one can derive shape.

The message is plain clear. There must exist an additional level of understanding at which the character of the information-processing tasks carried out during perception are analyzed and understood in way that is independent of the particular mechanisms and structures that implement them in our heads. This was what was missing, the analysis of the problem as an information-processing task. Such analysis does not usurp an understanding at the other levels of neurons or computer programs, but it is necessary complementary to them, since without it there can be no real understanding of the function of all those neurons. The important point is that if the notion of different types of understanding is taking seriously, it allows the study of the information-processing basis of perception to be made rigorous. It becomes possible, by separating explanations into different levels, to make explicit statements about what is being computed and why and to construct theories stating that what is being computed is optimal in some sense or is guaranteed to function correctly. The ad hoc element is removed, and heuristic computer programs

are replaced by solid foundations on which a real subject can be built. This realization, the formulation of what was missing, together with a clear idea of how to supply it, formed the basic foundation for a new integrated approach.

Level of analysis

According to David Marr, information processing systems must be understood at three distinct yet complementary levels of analysis - an analysis at one level alone is not sufficient.

1. Computational

The computational level of analysis identifies what the information processing system does (e.g.: what problems does it solve or overcome) and similarly, why does it do these things.

1. Algorithmic

The algorithmic level of analysis identifies how the information processing system performs its computations, specifically, what representations are used and what processes are employed to build and manipulate them.

1. Physical

The physical level of analysis identifies how the information processing system is physically realized (in the case of biological vision, what neural structures and neuronal activities implement the visual system).

Marr's notion of levels

David Marr (1982) has provided an influential analysis of the issue of levels in cognitive science, although we are not sure that we agree entirely with Marr's analysis it is thoughtful and can serve as a starting point.

Computational models, according to Marr, are focused on a formal analysis of the problem the system is solving, not the method by which it is solved.

It is the algorithm level at which we are concerned with such issues as efficiency, degradation of performance under arise or other adverse conditions, whether a particular problem is easy or difficult, which problems are solved quickly and which take a long time to solve, how information is represented, etc.

These are all questions to which psychological inquiry is directed and to which psychological data is relevant.

At the computational level, it does not matter whether the theory is stated as a program for a Turing machine, as a set of axioms, or as a set of write rules.

It doesn't matter how the information is represented as long as the representations is rich enough, in principle, to support computation of the required function.

The question is simple what function is being computed, not how is it being computed.

Marr recommends that a good strategy in the development of theory is to begin with a careful analysis of

the goal of a particular computation and formal analysis of the problem that the system is trying to solve.

There is still another notion of levels which illustrates our view. This is the notion of levels implicit in the distinction between Newtonian mechanics on one hand Quantum theory on the other.

Convolutional networks

Convolutional networks (LeCun, 1989), also known as convolutional neural networks, or CNNs, are a specialized kind of neural network for processing data that has a known grid-like topology. Examples include time-series data, which can be thought of as a 1-D grid taking samples at regular time intervals, and image data, which can be thought of as a 2-D grid of pixels.

The name "convolutional neural networks" indicates that the network employs a mathematical operation called convolution. Convolution is a specialized kind of linear operation. Convolutional networks are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers.

Usually, the operation used in a convolutional neural network does not correspond precisely to the definition of convolution as used in other fields, such as engineering or pure mathematics.

Convolutional networks stand out as an example of neuroscientific principles influencing deep learning.

In convolutional network terminology, the first argument (in this example let's say the function x) to the convolution is often referred to as the *input*, and the second argument (the function w) as the *kernel*. The output is sometimes referred to as the *feature map*.

In machine learning applications, the input is usually a multidimensional array of data, and the kernel is usually a multidimensional array of parameters that are adapted by the learning algorithm. We will refer to these multidimensional arrays as tensors.

It is rare for convolution to be used alone in machine learning; instead convolution is used simultaneously with other functions, and the combination of these functions does not commute regardless of whether the convolution operation flips its kernel or not.

Discrete convolution can be viewed as multiplication by a matrix, but the matrix has several entries constrained to be equal to other entries.

In two dimensions, a doubly block circulant matrix corresponds to convolution. In addition to these constraints, convolution usually corresponds to a very sparse matrix (a matrix whose entries are mostly equal to zero).

Convolution leverages three important ideas that can help improve a machine learning system: sparse interactions, parameter sharing and equivariant representations.

Sparse interactions

Sparse interactions (also referred to as sparse connectivity or sparse weights). This is accomplished by making the kernel smaller than the input when processing an image, the input image might have thousands or millions of pixels, but we can detect small, meaningful features such as edges with kernels that occupy only tens or hundreds of pixels. We need to store fewer parameters, which both reduces the memory that computing the output requires fewer operations.

Parameter sharing

As a synonym for parameter sharing, one can say that a network has tied weights, because the value of the weight applied to one input is tied to the value of a weight applied elsewhere.

The parameter sharing used by the convolution operation means that rather thn learning a spearte set of parmeters for every location, we learn only one set.

Convolutions is thus dramatically more efficient than dense matrix multiplication in terms of the memory requirements and statistical efficiency.

Equivariant representations

Convolution is an extremely efficient way of describing transformations that apply the same linear transformation of a small local region across the entire input.

In the case of convolution, the particular form of parameter sharing caused the layer to have a property called equivariance to translation. To say a function is equivariant means that if the input changes, the output changes in the same way.

Implementing convolution

Other operations besides convolutions are usually necessary to implement a convolution network. To perform learning, one must able to compute the gradient with respect to the kernel, given the gradient with respect to the outputs.

Recall that convolution is a linear operation and can this be described as a matrix multiplication. The matrix is sparse, and each element of the kernel is copied to several elements of the matrix. This view help us to derive some of the other operations needed to implement a convolution network.

Multiplication by the transpose of the matrix defined by convolution is one such operation. This is the operation needed to back-propagate error derivatives through a conolution layer, so it is needed to train convolutional networks that have more than one hidden layer.

These three operations, convolution, backprop from output to weights, and backprop from output to inputs are sufficient to compute all the gradients needed to train any depth of feedforward convolutional network.

It is also worth mentioning that neuroscience has told us relatively little about how to train convolutional networks. Model structures with parameter sharing across multiple spatial locations date back to early connectionist models of vision (Marr and Poggio, 1976), but these models did not use the modern back-propagation algorithm and gradient descent.

Convolutional nets were some of the first working deep networks train with back-propagaion. It is not entirely clear why convolutional networks succeeded when general back-propagation networks were considered to have failed. It may simple be that convolutional networks were more computational efficient than fully connected networks, so it was easier to run multiple experiments with them, and tune their implementation and hyperparameters.

It may be that primary barriers to the success of neural networks were psychological (practitioners did not expect neural networks to work, so they did not make serious effort to use neural networks).

Convolutional networks provide a way to specialize neural networks to work with data that has a clear grid-structured topology and to scale such models to very large size. This approach has been the most successful on a two-dimensional image topology.

Capsule identities

So the idea of a capsule is a capsule is a vector thing it's got a fctor activity and capsules in one layer send information to capsules in the next layer the capsule in the next layer gets active if it sees a bunch of incoming vectors that agree, now of course the capsule in the next layer doesn't see just the output but it see the output multiplied by a weight matrix.

If those products agree, if it gets good agreement even if there is some outliers they will say hey I doung something and of course high-dimensional coincidence if you get six dimensional things to agree even if only dimension, each dmension only agreeto within ten percent the chance of a six dimensonal thing agreein is like one in a million, I mean if its sort of attempt of the normal long to disparity on each dimension then it's a millionth of the disparity of two random things so a high dimensional agreement is a really, really significant thing and is a much better filter than what we do at present which is you apply some weights see if you get above the threshold...

Obviously we know that if you stack that up and you train it by stochastic gradient decent it can do anything , but if you go back to basics its not in its nature to automatically be looking ot covariances between vectors and we want units where that's part of their nature.

The edge of a capsule

The edge of a capsule is a vector that represent the different properties of a thing, of an entity, now the entity might be a little fragment it might be something much bigger and if we won't doing vision it might be something else all together but the other aspect of capsules is we're going to try inside the network to get entities, a common neural network is not really committed to found multi-dimensional entities, where if you look at how people deal with the world they deal with the world in terms of objects that have properties.

We're going to understand the world in terms of entities and these entities are going to have properties and what a capsule is going to do, is its going to make a fundamental commitment.

If I'm a capsule I got a bunch of units and some of them are active at once then just the fact that were active together means they apply to the same thing, if you go sequential that is only do one thing at a time then you can bind together arbitrary things just by simultaneity, We want to do that at a low level two.