Sector F Lambda Complex

The fall of lambda calculus 14.05.2018

Contents

Preface

Don't for heaven's sake, 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 text-book. Its object 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 written makes no claim to novelty in points of detail; 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.

I will only mention the great works of Robert Virding I owe in large measure the stimulation of my thoughts.

On the other hand the truth of the thoughts communicated here seems to me inconsistent and robust. And if I am not mistaken in this, then the value of this work secondly consists in the fact that it shows how little has been done to solve this problems.

One of the factors that make the life of self-regulating systems difficult at times is that they have blind spots where elements working with them are incapable of adequately understanding what is going on.

No natural language is entirely complete or self-sufficient.

The same is true of systems and, in fact, has been since "language" is vital to the transmission of information in a system; but the inconsistency, problem, bugs can enter systems in multiple ways.

To cope with these "blind spots" we have to make use of another vocabulary or system outside the one in which the doubt exists.

By going to another means of communication, such as finding a picture. The strategy in these solutions is the same.

The incompleteness of the system or language is compensated for by resorting to something outside of the system, by using another language or another system.

If this bits has a value it consists in two things. First that in thoughts are expressed, and this value will be the greater the better the thoughts are expressed. Here I am conscious that I have fallen short of the possible. Simple because my powers are insufficient to cope with the task, may others come and do it better.

Abstraction

In trying to understand what is happening around us we are faced with a fundamental problem.

There are just to many facts, details and items requiring consideration.

The brain operates in a completely different regime from the one that most statisticians are familiar with.

- It has about 10^14 parameters and we only live for about 10^9 seconds.
- Synapses are much cheaper than experiences so it makes sense to throw a lot of synapses at each experience.

In approaching any situaton, the system trying to understand it, does not attempt to gather all information. Instead it selects certain facts and searchers for others.

This selection of some items and ignoring of others is a process of abstraction.

It is the abstracting form a real or if you will empirical situation the tings seemingly most important to deal with.

Model your challenge

Taking the abstracted elements, a character with the flat tire begins to connect them into a pattern.

Better yet, he weaves them into a model of the confronting situation, which we can use both to understand his plight and figure out what to do about it.

The parts of this model would probably include, among other things, the flat tire, the image of the spare in the trunk, the telephone, the service station, a forthcoming business meting, etc.

A second model would contain the telephone, the service station, and the repairman there.

Finally, it concludes that it will call a cab and leave his wife to deal with the flat tire as best as she can.

These are extraordinarily elementary models, but they serve a very practical purpose.

With them the main character in our illustration can see the likely consequences of various courses of action.

We can find out these things by doing them directly by actually handling the tire and observing that we get dirty, or by calling the repairmen and waiting for him and learning that it takes too long.

In the age of big data; big models are good.

- For any given size of data, the bigger the model, the better it generalizes, provided you regularize well.
- This is obviously true if your model is an ensemble of smaller models.
- Adding extra models to the ensemble always helps.
- It is a good idea to try to make the data look small by using a big model.

By using the model, however, we can make some reasonable predictions about what will occur and therby accept or reject the choices open to us.

Overview

In this process of abstraction and model building we deliberately select a few items, ignore may others, and then place the items chosen in a particular relationship to one another.

In doing so we are intentionally ignoring facts or relationships that can influence the type of situation under study.

The problem it to select the most meaningful elements and relationships and dropout the rest.

Those who use abstraction skillfully know well that they neither have all the facts nor have considered all the relationships bearing on the outcome of what they are analysing.

We do not use the abstractions from one situation in another setting without carefully examining the fit.

Neither do we expect a model to handle all aspects of a situation.

We shall be dealing with many abstractions and models, not with the intention of exactly mirroring the real world but with the objective of clarifying our perception of its most essential features.

Abstractions and models are mechanisms for economizing both time and effort, but like any tool they must be used within their limits.

Predictive modeling

Our emphasis is on people in organizations where managers have to make decisions even through their knowledge and concepts may be wrong or inadequate.

As generalizations are made about what takes place in organizations, we shall often be talking about the decision processes that organizational members engage in or the sequence of factors which lead to a particular behavior.

We talk about the possible actions a person may take on recepipt of an order from his superior, which he thinks is improper and unwarranted.

We talk about choossing amount the courses of accepting the order competly, protesting, protesting and raiisng an alternative, or leavng the organization.

It sometimes appers as if there were a basic assumption that people consicosly bring our all these possibilities and rationally weigh the pros and cons of each.

Does the individual at times carry on the same process unconsicusly that on other occasions he perhaps conducts consciously? We hardly know...

We are faced with the fact that people sometimes do things and later say, I never thougth that I would act like that under those circumstances.

In saying this the individual indicates that he saw other opportunities and that, through some process unobservable to him he decided among them and chose on that came from elsewhere that his own conscious thought.

The distiontion between conscious and onconscious thought are by no means easy to determine, and for our purposes, it is not usually necessary to make them.

Considerng the development of the field of Artificial Intelligence at the moment, it seems reasonable to aconduct our analysis at the level at which both machines and humans do make decisions, wihout taling into account whether the choices and decisions processes are consicious or not.

Several references have been made with the intent of this guide to provide conceptial tools for analysis. As ith any other tool models, abstractions and generalizations are useful only when within their limitations.

Fall of lambda

Logic programming is the proposal to implement systems using mathematical logic.

Perhaps the first published proposal to use mathematical logic for programming was John McCarthy's Advice Taker paper.

Planner was the first language to feature "procedural plans" that were called by "pattern-directed invocation" using "goals" and "assertions". A subset called Micro Planner was implemented by Gerry Sussman, Eugene Chariak and Terry Winograd and was used in Winograd's natural language understanding program SHRDLU, and some other projects.

This generated a great deal of excitement in the field of AI. It also generated controversy because it proposed an alternative to the logic approach one of the mainstay paradigms for AI.

The upshot is that the procedural approach has a different mathematical semantics based on the denotation semantics of the Actor model from the semantics of mathematical logic.

There were some surprising results from this research including that mathematical logic is incapable of implementing general concurrent computation even though it can implement sequential computation and some kinds of parallel computation including the lambda calculus.

Classical logic blows up in the face of inconsistent information that is becoming more ubiquitous with the growth of the internet.

We have pass the midst of a huge paradigm shift and live now in a massive concurrent world full of new buzz words like IoT, cloud and edge computing, web and micro services and multi-core computer architectures living mostly in every computer and consumer device.

This change enables a new generation of systems incorporating ideas from mathematical logic in their implementation, resulting on some reincarnation of logic programming. But something is often transformed when reincarnated!

Actor model

Actors are the universal primitive of concurrent digital computation. In response to a message that it receives, an actor can make local decisions, create more Actors, send more messages, and designate how to respond to the next message received.

Unbounded nondeterminism is the property that the amount of delay in servicing a request can become unbounded as a result of arbitration of contention for shared resources while still guaranteeing that the request will eventually be serviced.

Arguments for unbounded nondeterminism include the following:

There is no bound that can be placed on how long it takes a computational circuit called an Arbiter to settle.

- Arbiters are used in computers to deal with the circumstance that computer clocks operate asychoronously with input from outside, "e.g, keyboard input, disk access, network input, etc..."

- So it could take an unbounded time for a message to sent to a computer to be received and in the meantime the computer could traverse an unbounded number of states.

A limitation of logic programming

In his 1988 paper on early history of Prolog, Bob Kowalski published the thesis that "computation is controlled deduction" which he attributed to Pat Hayes.

Contrary to Kowalski and Hayes, Hewitt's thesis was that logical deduction was incapable of carrying out

concurrent computation in open systems because of indeterminacy in the arrival order of messages.

Indeterminacy in concurrent computation

Hewitt and Agha [1991] argued that: The Actor model makes use of arbitration for determining which message is next in the arrival ordering of an Actor that is sent multiple messages concurrently.

For example Arbiters can be used in the implementation of the arrival ordering of an Actor which is subject to physical indeterminacy in the arrival order.

In concrete terms for Actor systems typically we cannot observe the details by which the arrival order of messages for an Actor is determined. Attempting to do so affects the results and can even push the indeterminacy elsewhere.

Instead of observing the internals of arbitration processes of Actor computations, we await outcomes.

Physical indeterminacy in arbiters produces indeterminacy in Actors. The reason that we await outcomes is that we have no alternative because of indeterminacy.

According to Chris Fuchs [2004], quantum physics is a theory whose terms refer predominately to our interface with the world. It is a theory not about observables, not beables, but about 'dingables' we tap a bell with our gentle touch and listen for its beautiful ring.

It is important to distinguish between indeterminacy in which factors outside the control of an information system are making decision and *choice* in which the information system has some control.

It is not sufficient to say that indeterminacy in Actor systems is due to unknown/unmodeled properties of the network infrastructure. The whole point of the appeal to quantum indeterminacy is to show that aspects of Actor systems can be unknowable and the participants can be entangled.

The concept that quantum mechanics forces us to give up is: the description of a system independent from the observer providing such a description; that is the concept of the absolute state of a system. I.e, there is no observer independent data at all.

According to Zurek [1982], "Properties of quantum systems have no absolute meaning. Rather they must always characterized with respect to other physical systems."

Does this mean that there is no relation whatsoever between views of different observers? Certainly not. According to Rovelli [1996] "It is possible to compare different views, but the process of comparison is always a physical interaction (and all physical interactions are quantum mechanical in nature)".

Information theory

One sector, semantics, concerns it-self with the meaning of symbols and words.

Another sector, syntax, is concerned with the systematic properties of communication, that is, the relations between the symbols used.

In communications such as the written or spoken words, syntax is concerned with the relationships among letters forming words or the relationships among words in a sentence, grammar.

A third part of communication is concerned with pragmatics and focuses on the relationship between communication and purposeful action.

Each of the approaches to computer communication represents an enormous field of study with many subdivisions.

The central question here is cern in with a communication system. Much of the communications that occur in organizations primarly involve people, and many times the specifics illustrations of the general system pertain to communications among individuals.

Channeling Shannon

Claude E. Shannon originally developed his model of communications while working on problems of telephone communications.

Note that analytical and highly mathematical rather than empirical, these ideas have been the foundation for many concepts grouped under the title of "Information Theory" the important aspects of these concepts are that they have been found application in analyzing not only problems in mechanical systems with links and leavers and in social systems with human beings.

Shannon's model is a general model of communication useful wherever there is a flow of information.

The block labeled Sender, identifies the place, person, or equipment, which produces the message and sends it out.

For the message to move between sender and receiver requires some way of its being conveyed, a Channel.

Channels can take many forms; a simple illustration is telephone wires. Channels may also be radio waves, light waves, sound waves, pieces of paper, etc.

We are concerned hare not only with the various forms of channels but also the ways in which organizations create them or facilitate or impede their creation.

Channel capacity

What could be described as common sense in the context of information theory?

The concept of error-free transmission suggests that since there is always some type of noise in a communications system there is always some error in the communication which we can at least hope to reduce to a minimum level.

One of the interesting factors Shannon's analysis has shown is that this is not the case.

It is possible if certain conditions are met, to achieve error-free transmissions.

To explain this necessitates the introduction of the idea of channel capacity.

Although difficult to compute, it is relatively easy to imagine that any channel will have some upper or maximum limit on the amount of information it can transmit.

Through mathematical analysis Shanon has shown that it is possible to achieve error-free transmissions by the use of proper coding provided if the capacity of the channel is not exceeded.

Let us note two things about this:

1. the code is developed by the group during a process of two-way communication,
2. everyone involved hears the proposal and agrees to it.

They all learn the code and are able to use it with maximum effectiveness.

Development of codes or special languages is, as we have noted, one of the ways of improving communications, but these take time to perfect, and they are only effective when both sender and receiver understand them.

One aspect of two-way communications or feedback is that is extremely useful in developing adequate special languages and having them learned quickly by all parties involved.

Second, an observer may note that in a two-way communication the speaker in describing the subject looks up at the group, and even though they say nothing he pauses, apparently makes a decision, and then goes back to repeat his point.

What he is doing is getting a response from the general appearance of his audience that lets him know whether or not he "got through".

Feedback, then, is a process of two-way communication.

Two principal aspects of interest are:

1. the response it gives the sender regarding the understanding the receivers have of his message, and
2. the development and learning of special languages.

Flow of information

Shannon's model is a general model of communications useful wherever there is a flow of information.

The block labeled source or sender identifies the place, person or equipment which produces the message and sends it out.

For the message to move between sender and receiver requires some way of its being conveyed, a channel or medium protocol.

Protocols and devices are not perfect ways to convey messages.

They are the best we have but are subject to innumerable disturbances, many of them unpredictable, which distort confuse, or even block and therefore interfere with the accuracy of the transmission of information.

This interference is called noise and is comprised of anything moving the the same space apart from the actual message signals wanted by the sender.

To be realistic, we have to take into account not only the noise of the channel, protocol, layers, datasets, but also of the sender and the receiver.

Hence, as we listen to a radio, we are concerned not only with the static generated as a result of the electric storm outside, but also with the background buzz or hum which comes from the thermal disturbances in the receiver or perhaps from the broadcasting station.

Reducing errors

Noise in a communications system is not only a nuisance but also the source of error.

For example, during an electrical storm, a sudden flash of lightning may cause so much loud static on a communications line that it becomes impossible to hear a portion of the conversation.

The listener will probably ask to have the message repeated.

This simple illustration includes the way on type of noise can interfere with communications and the basic way of reducing error due to noise: repeating part of the message, or as it called, redundancy.

Redundancy is a repetition of any information contained in a message.

This definition includes both redundancies applied to specific disturbances, as in the illustration, and those build into language and used regularly in communication.

Error-free transmission

Common-sense analysis suggests that since there is always some type of noise in a communications system there is always some error in the communication which we best hope to reduce to a minimum level.

Shannon's analysis has shown that this is not the case, it's possible if certain conditions are met, to achieve error-free transmissions.

To explain this necessitates the introduction of the idea of protocol, process, device or channel capacity.

Although difficult to compute, it's relatively easy to imagine that any process will have some upper or maximum limit on the amount of information it can transmit, by Shannon's analysis has shown that it is possible to achieve error-free transmissions by the use of proper coding provided the capacity of the channel is not exceeded.

Coding information

Essentially coding is transforming the message into a special form for transmission.

In a sense it is a special language which enables is to convey more information in an individual communicative act.

In this way the transmission rate of information can be increased.

The problem is that the original message must be translated into the code, and the process of coding takes time.

Needless to say our objective could have just as well been defined as transmission with 2 per cent or 10 per cent error.

Shannon's concepts enable us to think in terms of precisely defined transmission error with one point being zero error.

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.

Encoding and decoding

Introduction of the idea of code and the process of coding a message makes it apparent that the sender really includes several components.

Since all communication is carried on in some form of language, all communications must have some process of encoding, that is, of translating information into the specific form in which it is going to be transmitted.

Taking an idea and converting it into words in normal conversation is an encoding process, but so is the act of a dude filling in a form using the appropriate numbers and symbols so that the proper material may be drawn from stock.

Needless to say if the information is in code form, there must also be a decoding process in the receiver that extracts the information the code contains.

Selection

In addition to encoding, any sender carries out a process of selection.

There are numerous reasons for this.

One of these is to keep the amount of information going through a system at a reasonable level to avoid overloading the system.

Another is to reduce noise by preventing other than desired messages from entering the system.

Another type of selectivity occurs when information wanted by some members of an organization is kept from entering the communication system or when information important to one individual is introduced into the system by him even though it may not really be what others in the system need.

Selectivity also occurs at the receiving end.

Channels

Thus for our mental illustrations have suggested that a channel is nothing more than a simple link, such as a telephone/network wire, between two people/machines.

In organizations, however, channels are frequently not that simple.

As we have noted, a channel may have parties on either end who are both senders and receivers.

There may be a number of connected links in a network so that a message starting at one end of the system has to pass through several people/machines or transmission units in getting to the other end.

Sometimes links carry only certain types of messages, sometimes only in one direction, the conceptual model discussed identified the following elements:

a selector and encoder embodied in the sender, a system of communication links which forms a network called the channel and actually transmits the information, and steps of decoding and selecting information for use which occurs in the receiver.

The movement of information through a system is influenced by two principal factors, its capacity and the noise within it.

If the capacity of the system is not exceeded, the accuracy of the transmission of information is a function of the noise, which, in turn, can be combated by redundancy, which uses up more of the channel capacity, and by coding.

Coding permits more information to be communicated but at a cost of reducing the independence of

choice of the persons involved and of bringing about a greater delay in communication because of the necessity of encoding and decoding.

Artificial neural networks

Ego, Id, and Superego are concepts that help one to see and state important facts of behavior, but they are also dangerously easy to treat as ghostly realities: as agents that want this or disapprove of that, overcoming one another by force or guile, and punishing or being punished.

We can take for granted that any theory of behavior at present must be incomplete.

Modern psychology takes completely for granted that behavior and neural function are perfectly correlated, that one is completely caused by the other. There is no separate soul or life-force stick a finger into the brain now and then and make neural cells do what they would not otherwise.

Our failure to solve a problem so far does not make it unsolvable. One cannot logically be a deterministic in physics and chemistry and biology, and mystic in psychology.

All one can know about another feelings and awareness is an inference from what he say and does from his muscular contractions and glandular secretions. These observable events are determined by electrical and chemical events in nerve cells.

Mind can only be regarded, for scientific purposes, as the activity of the brain, and this should be mystery enough for anyone; besides the appalling number of billions of cells and even more appalling number of possible connections between them.

The problem of understanding behavior is the problem of understanding the total action of the nervous system, and vice versa.

There is much more certainty in the study of the electrical activity of a well-defined tract in the human brain. One can discover the properties of its various parts more or less in isolation; but that part may have properties that are not evident in isolation, and these are to be discovered only by study of the whole intact brain.

The central problem with which we must find a way to deal can be put in two different ways. Psychologically, it is the problem of thought: some sort of process that is not fully controlled by environmental stimulation and yet cooperates closely with that stimulation.

Physiologically, the problem is that of the transmission of excitation from sensory to motor cortex; the transmission may be a very complex process indeed, with a considerable time lag between sensory stimulation and the final motor response.

The links are complex and we know practically nothing about what goes on between the arrival of an excitation at a sensory projection area and its later departure from the motor area of the cortex.

In general the bridge can be described as some comparatively simple formula of cortical transmission having all the finer structure of a bowlful of porridge.

What is the neural basis of expectancy, or of attention, or interest?

Any frequently repeated, particular stimulation will lead to the slow development of a "cell-assembly", a diffuse structure comprising cells in the cortex, capable of acting briefly as a closed system, delivering facilitation to other such systems and usually having a specific motor facilitation. A series of such events constitutes a "phase sequence" the thought process. Each assembly action may be aroused by a preceding assembly, by a sensory event, or normally by both. The central facilitation from one of these activities on the next is the prototype of "attention".

It is proposed also that there is an alternate "intrinsic" organization, occurring in sleep and in infancy, which consists of hypersynchrony in the firing of cortical cells. But besides these two forms of cortical organization there may be disorganization.

The theory is evidently a form of connectionism, one of the switchboard variety, the connections server rather to establish autonomous central activities, which then are the basis of further learning. In accordance with modern physiological ideas, the theory also utilizes local field processes and gradients. It does not, further, make any single nerve cell or pathway essential to any habit or perception.

We want to explain certain clinical facts. To really do so, we must find an anatomical and physiological

understanding of what is known psychologically as a concept; and we must me able to deal with its relation to perception and to learning.

And with that, we land right in the middle of the generalized problem of explaining mammalian behavior. What is a concept, if it is not a conditioned response? What is perceptual learning? And so on. Before such questions can be answered, psychological theory must have a new base of operations.

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

Perceptual structural memory trace

The assumption we must accept is that the memory trace, the basis of learning, is in some way structural and static; a learned discrimination is not based on the excitation of any particular neural cells. It is supposed to be determined solely be the pattern, or shape, of the sensory excitation; the pattern and not the locus of stimulation is the important thing, has developed a theory of electrical fields in the brain which control cerebral action.

The mnemonic trace, the neural change that is induced by experience and constitute "memory", is not a change of structure. The idea that the trace is structural but diffuse, involving, that is, a large number of cells widely spaced in the cortex, physiologically but not anatomically unified.

Thus the only barrier to assuming that a structural change in specific neural cells is the basic of memory lies in the generalization of the perception of patterns. Man sees a square as a square, whatever its size, and in almost any setting. These are concrete, undisputed facts of behavior. They have been interpreted as meaning that perception is independent of the locus of excitation; and this interpretation has been accepted as inescapable. The result is an awkward dilemma for theory, since, as we have seen, it is hard to reconcile an unlocalized afferent process with a structural localized mnemonic trace.

The attack mode

How are we to provide for perceptual generalization and the stability of memory, in terms of what neuron does and what happens at the synapse? We must suppose that the mnemonic trace is a structural change; the difficulty, in suooising it, is a conflict with the idea that only the pattern, and not the locus of sensory stimulation, is important in perception.

That paragraph puts the theoretical approach of this monograph. We propose to go over the evidence again on perception and show that it is not what it seems. We do not know that pattern is everything, locus nothing. Perception does depend on exciting specific parts of the receptor surface; mnemonic trace and perceptual generalizations need no longer to be at odds with one another; a physiological meaning of attention (or set, or expectancy) shows up.

Learning in Perception

The immediate objective is to show that "simple" perceptions are in fact complex; 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,
2. on a pattern of excitation whose locus is unimportant.

Current (1949) opinion seems tacitly to have accepted the Gestalt argument that the only tenable assumption is the second of these possibilities.

The theory presented here is diametrically opposed to this aspect of Gestalt theory, and is based on assumption 1, that a particular perception depends on the excitation of particular cells at some point in the central nervous system.

The perception of identity depends on a series of excitations, the perception is additive, a serial reconstruction though very rapid and "unconscious" for the normal adult. Simple figures do not always act as wholes, innately. But it is undoubtedly true that they sometimes do so in one respect, in the figure-ground relationship: so this property of a perceived figure is to be distinguished from others, in which summation and learning are important.

Accordingly, the following conceptions are distinguished:

1. a primitive, sensory determined unity of a figure is defined here as referring to that unity and segregation from the background which seems to be a direct product of the pattern of sensory excitation and the inherited characteristics of the nervous system on which it acts,
2. a nonsensory unit, affected by experience defined as one in which the boundaries of the figure are not fixed by gradients of luminosity in the visual field. Affected by experience and other nonsensory factors, and is not inevitable in any perception. In general terms, the nonsensory figure occurs in perception whenever the subject responds selectively to a limited part of a homogeneous area in the visual field.
3. the identity, also affected by experience, of a perceived figure.

In contrasting the primitive and the nonsensory figure, however, one need nt imply that a perceived figure must be either of one or the other kind. They are rather two extremes, and in most perceptions both sensory and nonsensory factors affect figure-ground organization.

Even commoner in everyday perception is the perceived entity in which both sensory and nonsensory factors cooperate.

Success in discrimination and generalization evidently depends on one's finding a differentiating characteristic between two figures, or one that is common to a pattern already known and the new one which is presented in the testing situation. No one today would argue that perceptual organization is wholly acquired; there is some innate organization. But this of course does not show that the organization is entirely innate.

There is always a possibility that perception has a partly innate, partly learned organization; and that besides the figure that has a "primitive unity" there are "nonsensory figures" in which experience has an important role.

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 contours, in the intervals between seeing the figure as a whole.

Identity in perception

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 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 all, and is not recognized or named easily.

Identity of course is a matter of degree and depends on a considerable amount of experience; it is not innately given, an approach to this conception, using the term "identifiability" and has proposed that identifiability promotes the formation of associations.

If we go a little further, it appears that the proposition is circular. Identifiability is not merely a perceptible difference of one figure from another when the two are side by side, but implies a remembrable difference, identifiability is clearly related with previous experiences, recognizability; and recognition is one form of association and associability affects the ocurrence of associations.

The real point at which we are driving seems to be that there are genuine differences of associability in different patterns. Also, more is involved in these differences than the number of trials necessary to

establish recognition; there are also the spontaneous associations referred to in speaking of similarity.

An irregular mass of color or a pattern of intersecting lines drawn at random has some coherence and unity, but one such figure is not readily recognized and distinguished from others when it is seen a second time, and generalization (or similarity) is not selective among a number of such stimuli.

We distinguish between the "so-called generalization" which means only a failure to observe differences and the generalization which involves perception of both similarities and differences. The amorphous figure, lacking in identity, is generalized in the first sense only.

A further illustration of these points is found in the development of identity in the perception of capuchin monkey faces by one who has seen no capuchin monkeys before.

Two animals seen side by side are obviously different in details, but the inexperienced observer is not able easily to remember one selectively. Also, all capuchin monkeys at this stage look alike; the "so-called generalization" occurs. With experience the perception of identity increases. Similarity is still perceived between animals, and cofusion between some animals is still possible; but there is a marked change in the perception.

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

This bits has been meant to establish the conception of "identity" as an important property of perception which should be kept carefully distinct from the "unity" of the perceived figure (as well as from its "meaning"). Unity may be innately determined, an immediate property of sensory dynamics, whereas identity is dependent on a prolonged experience.

Learning to perceive

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

The shortest time in which a subject approximated to normal perception, even when learning was confined to a small number of objects, seems to have been about a month.

It is possible then than the normal human infant goes through the same process and that we are able to see a square as such in a single glance only as the result of complex learning.

Gellerman (1933) reports that chimpanzees and two-year-old children recognized a triangle that had been rotated through 120 from the training position, but (in the one protocol that is given) responded selectively only after a head rotation; and persistent head rotation continued in the later discriminations.

When a simple figure such as square, circle, or triangle, subtending a retinal angle of from 2 to 10 is fixated at one point it tends in a second or so to become almost amorphous except near the point of fixation.

The factors involved are evidently complicated; it will be found, for example, that with a large fogure merely imagining eye-movements (of following the contours) will restore definition of the figure. Also, these "imaginary" eyemovements, or subliminal activations of the motor system, occur more frequently and are less easy to control in looking at a smaller than at a larger figure, and it is hard to be sure that the size of the figure is unimportant. But this at least seems definite, that a stable, clear, and effective perception of circle or square is more possible with eyemovement than without.

The point is not that eyemovements are essential to perception by a sophisticated observer (nor, in the following paragraph, that they are completely necessary for an image); but that the perception is definitely clearer, more effective, with them than without.

This is really an evident fact. It is to be interpreted in the light of all evidence, showing that the perception of square or circle is slowly learned and depends originally on multiple visual fixations.

Evidence found on a one-sided loss of vision by monkeys on extirpation of the opposite frontal eyefield, a

cortical motor area for head-and-eye movement. The most significant and striking observation was startle by the monkey when an object was passed from the blind side into the seeing side, at the moment of passing the midline.

Now the question is what the motor cortex can have to do with visual perception unless perception intimately involves a motor activity, liminal or subliminal.

Perception of even a simple object involves a "phase sequence." This is a chain of central cortical events with motor links. Although the motor activations may be subliminal and do not always produce overt response, their role is essential in any perception.

Animal experiments and the human clinical data alike indicate that the perception of simple diagrams as distinctive wholes is not immediately given but slowly acquired through learning.

The process of "successive part reinforcement," as an aid to perception, exists at the same time as an essential unity of the whole; and theory must provide for the additive process, with its motor elements, as well as for the primitive unity. 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.

The labeling trouble

The basic trouble with the labeling approach of scene analysis is that it is too limiting and stultifying a paradigm for vision, in much the same way that resolution is for deduction.

The fundamental principle of resolution that **(not A)** and **(A or B)** together imply B, is occasionally useful.

But attempting to make a uniform resolution proof procedure, to mechanize deduction in a way that cannot be very sensitive to hints, hunches, and a wide variety of higher level knowledge about the particular domain in question is a cul-de-sac.

Views

A major theme running through our approach, and one that we shall try to justify, is that the representation of visual objects should be translated into 3-D language as soon as possible.

Thus we distinguish between that we call a VIEW of an object, and an underlying 3D model of that object.

The VIEW is a peripheral mini-world in which low level visual routines can leave assertions about what they see outside.

The VIEW knows about the possible different directions away from the viewer, which directions are near which other directions, and it has a crude knowledge about distance away from the viewer. (It seems familiar to the camera in a 3-D engine)

Closely associated with this world are other mini-worlds that can describe two-dimensional shape, movement, colour, and disparity between e.g the images on two cameras leaving their information bound to the names of directions in the view.

The view is thus like a canvas on which these features are painted to be studied and interrogated by more central routines.

The final ingredient of the view is the CUE. One form of CLUE is a particular combination of predicates, which can be bound to the view and which suggests the presence of a particular 3-D structure.

Clues

Such CLUES are like masks that act as triggers to specific central 3-D representations, and cause them to ask particular questions to try to verify that they represent is actually visible.

3-D models are arranged into a number of mini-worlds: the first concern descriptions of primitive 3-D shapes, and later ones, the elaboration of these into representations of the objects with the system has to deal.

Each 3-D model has its set of triggers, e.g the CUES from the VIEW mini-worlds, and a body of code that it will try to execute if it is triggers, ranging from low resolution visual CUES, through high resolution clues that detect a particular detail, to clues arrived at from a guess at the whole object. Notice that triggers are simply special ways of using an important kind of knowledge, akin to the role that features play in a systematic grammar for natural language (Halliday 1968).

3D models are designed to suicide as early as possible, if they are inappropriate: and should be regarded as specific suggestions about how to see the information that caused their activation.

Knowledge about the failure of a 3-D model can act as a trigger to other 3D models, yet another kind of information that the system must be able to use. It is a basic philosophy of the system that the 3-D representations of the world are what the system "remembers", and what it tries to maintain consistent with the information bound to the VIEW.

Our insistence of using 3D models for the basic representation of objects does not preclude the use of catalogues of appearances of objects from different viewpoints.

Indeed we regards knowledge about appearances as indispensable kinds of CLUE.

Plans

It is helpful to introduce a tentative basic unit of organization of knowledge, called a PLAN! a 3-D model is a particular type of PLAN.

A plan consists of code that serves the following three purposes:

1. A trigger, that defines the circumstances in which the plan is run. The trigger defines the CLUES for the plan, and a given CLUE may depend on more than one circumstance. Furthermore, clues may be armed by particular circumstances, so that information that in one context fails to elicit a particular plan, in more suggestive circumstances will succeed.

2. A body of code that runs when the plan is triggered. The body can apply further tests to see if the plan is really applicable and otherwise record its impressions of the situation in which it is being executed. It can contain other kinds of information, like what to do if the plan fails, and how to acquire more information about something the plan ran successfully on.

PLANS are derived from a mixture of sources: from the "state vectors" of McCarthy (1964), from PLANNER-69, from an unpublished character recognition program that uses PLANS to represent the characters (Blomfied, Marr & Mollison 1972) and from ideas that come under the general heading of frame theory (Minsky 1973).

Finally, at the risk of offending purists, we would like to mention that scattered over clinical and neurophysiological literature are hints that the mechanisms of masks, clues, and underlying 3-D representations that we organized into small theory of recognition, may have fairly closely corresponding analogs in primate and in Human visual systems.

When we know in more detail the kinds of assertions that are useful for 3-D representations, it may be possible to formulate a succinct and concrete hypothesis about the kind of single unit response that one would expect from cells in the relevant occipito-parietal regions.

Hypothesis of this kind are extremely difficult, but not quite impossible, to formulate: they are however comparatively easy to test.

Schemas

TBD

Control theory

Systems are so extraordinarily complex that it is difficult to deal with them in any detailed fashion.

We are likely to find analyses which discuss the performance of organizational elements in handling the control process but relatively little discussion of how this is accomplished.

Control systems are to a considerable degree influenced by basic human organizational phenomena, principal among them are perception formation, motivation and communications.

The basic elements of the control process are few in number:

- Gathering of data about performance by a sensor;

- Comparison of performance with standard by a discriminator

- When a difference between performance and standard is noted, two courses of action are likely to be taking by a decision maker.

1. Undertake corrective action which will bring performance back into line with standards perhaps by selecting or developing a new process

2. Direct information about this difference to a goal setter for the establishment of new goals or the redefinition of old ones.

When these elements exists, there is control through self-regulation of whatever entity contains them.

The elements, however, may be allocated in a number of ways within the entity.

They may all be performed by one person or each element may be performed by a different person or for that matter, by different departments.

As the work of carrying the control cycle is divided further, coordination of this effort becomes more difficult.

Control as we have been seen it is one type of system, we observe how control has many of the same properties as systems in general; numerous systems interblocked in many ways.

One such way may be seen when there are hierarchies of systems paralleling progression of goals.

There are also systems within systems and loops within the loops. Since control is also a system, it can influence other systems and in turn be influenced by them.

Looking at control as a particular type of system make us appreciate how extraordinarily complex a topic it is, and at the same time gives us an extremely valuable approach to this subject.

Control systems

Control is a word that has a number of meanings.

We are concerned with control in relation to matching performance with necessary or required conditions to obtain a purpose or objective.

The essence here is on reactivity and integration of effort, required accomplishment of an end.

There are a number of different processes which produce control.

Control is concerned not only with events directly related to the accomplishment of major purposes but also with maintaining the organization in a condition in which it can function adequately to achieve these major purposes.

The control process

The study of control has received a great deal of attention by people in number of different fields. This has come about with the realization that control is basic to an extraordinarily wide array of things.

All though the mechanism of control are different in each of these situations, analysis shows the process is the same.

Therefore, people have begun to study control as a universal process, a topic which can be studied in its own right.

There is a data-gathering phase in which a device, in this case generally called a sensor, gathers information about a subject.

Then comes a comparing or examination phase when the desired output is compared with the standard by a unit called a discriminator.

After this comparison is made and, for example a property is noted to be too low corrective action is required; the process is turned on by a unit we will call a decision maker.

Once the process is running, the fist phases are repeated until the discriminator determines that the property has reached or exceeded the upper limit; then another action is required, and the process is shut off.

The basic steps in the control process identified thus far are: gathering data on performance, comparison of data with a standard, and taking corrective action if performance does not match the standard properly.

These phases are carried out by a sensor, discriminator, and decision maker.

Coordination

Having covered the essence of control let us now consider how several loops are connected to provide that synchronization of effort we think of as coordination.

Programmed links

If the process had given instruction to report immediately on completion of the task, this instruction facilitates linking the completed act with the next one. Through an information transfer, we call this a programmed link.

The supervisor node, of course may detect that something is wrong through another control cycle. It can then take corrective action by including or adding into this programmed link or perhaps by attacking on the more difficult problem of human apathetic attitudes and motivation his units could display.

Progression of goals

Organizations have progression of goals which result from a division of work.

A subdivide goal becomes the task of a process contained within a specialized organizational unit.

This nesting of goals is contained as part of the core organizational means-ends chain.

Needless to say, the hierarchy of control loops which are connected with the progression of goals may be handle in number of ways, regardless of how the elements are allocated, the important factor is that all elements must be provided for in some way. Hence, our model supplies an extremely useful tool in analyzing complex control situations by telling us what basic functions bust occur and in what sequence, even though initially we have no idea as to where or how they are executed in the organization.

Goals and feedback

The feedback loop containing information about organizational performance and conditions leads to definition of subunit goals or standards. It's important to show how a situation in one area could lead to modifications in a number of organizational units at higher levels.

This even result in reformulating the basic goals of organizations. Feedback is essential to adequate goal formation.

This leads us to an important point which, when stated without the type of development presented, appears to be a somewhat mystical statement; namely, organizations automatically take actions to bring about their continuity or to perpetuate themselves. Of course there are times when organizations do not change goals and do go out of existence. Many, however, shift to those goals feedback defines as possible; feedback leading to goal redefinition is one of the basic ways by which organizational survival is achieved.

Time discrepancies

Another problem that can exist with control is that the feedback of information on performance takes so long it became impossible to correct performance.

There is another aspect of time which should be mentioned, the time horizon of people in an organization gets progressively shorter as we come down in a hierarchy.

Hence, by taking into account the perceptions of organizations members along with the time during which work is being performed on a project we can begin to define a time period within which feedback must occur if adequate corrective action is to be taken.

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.

System identification

The art and science of building mathematical models of dynamic systems from observed input-output data. It can be seen as the interface between the real world of applications and the mathematical world of control theory and model abstractions.

The field of system identification uses statistical methods to build mathematical models of dynamical systems from measured data this also includes optimal design of experiments for efficiency generating information data for training such models as well as model reduction

A dynamical mathematical model in this context is a mathematical description of the dynamic behavior of a system of process in either the time or frequency domain

Sometimes pure feed-forward control without feedback is called "ballistic" because once a control signal has been send it cannot further adjusted any corrective adjustment must be by way of a new control signal.

In contrast "cruise control" adjust the output in response to the load that it encounters, by a feedback mechanism.

Grey box model

In mathematics, statistics, and computational modeling a gray box model combines a partial theoretical structure with data to complete the model,

The theoretical structure may vary from information on the smoothness of results, to models that need only parameters values from data or existing literature.

Feedback

Which adjusts the output to take account of how it affects the load, and how the load itself may vary unpredictably; the load is considered to belong to the external environment of the system.

Feed-forward

The element of control that passes a signal from a source in its external environment, often a command from a external operator, to a load elsewhere in its external environment.

A system which has only feed-forward behavior responds to its signal in a pee-defined way without responding to how the load reacts; its is in contract with a system that also has feedback.

Stages of vision

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 inflences 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.

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.

Dealing with conflict

Organizations that posses the capacity to deal adequately with conflict have been described as follows:

1. They posses the machinery to deal constructively with conflict. They have an structure which facilitates constructive interaction between individuals and work groups.

2. The personnel of the organization is skilled in the process of effective interaction and mutual influence (skills in group ledership and membership roles in group building and maintenance functions).

3. There is a high confidence and trust in one another among members of the organization, loyalty to the work group and to the organization, and high motivation to achieve the organization's objectives.

Confidence, loyalty, and cooperative motivation produce earnest, sincere, and determined efforts to find solutions to conflict. There is greater motivation to find constructive solution that to maintain an irreconcilable conflict. The solutions reached are often highly creative and represent a far better solution than any initially proposed by the conflicting interests.

The essence here is that out of conflict will come a new synthesis superior to what existed before and perhaps superior to any individual point of view existent in conflict.

Conflict, resting in part on different perspectives of what "ought" to be, is one of the avenues for opening new directions for the organization or one of the ways of moving in new directions. This is not only useful but also vital for organizational survival. The question, therefore, as we view conflict is not, "How to eliminate it?" but, "Is it conflict of such a type and within circumstances where it will contribute to rather than detract from organizational interest?"

Whether a conflict is good or bad for an organization, whether a conflict can be made useful for an organization, depends not so much on manipulating the conflict itself as on the underlying conditions of the overall organization. In this sense, conflict can be seen as;

1. a symptom of more basic problems which requires attention

2. an intervening variable in the overall organization to be considered, used, and maintained within certain useful boundaries.

Organizational adaptation frequently proceeds through a new arrangement developing informally, which, after proving its worth and becoming accepted, is formally adopted. The first informal development, however, may be contrary to previously established procedures and in a sense a violation or a subversion of them; or the informal procedures may be an extension of a function for internal political purposes.

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.