2

Cognitive Processes and Communication Networks: A General Theory

Joseph Woelfel
Department of Communication
State University of New York at Buffalo

This chapter demonstrates the applicability of the network concept for the description of both cognitive and social processes. Since a network may be considered a set of interrelated nodes, they may be classified according to the characteristics of the nodes and the relationships among them. Historically, networks have been categorized as *social* networks, when the nodes were specified as individuals or higher level systems; *cognitive* or *neural* networks, when the nodes were considered words, symbols, or any other abstract representation; and *communication* networks, when the relationship among the nodes involved the transfer of information rather than affect, power or social prestige. This chapter attempts to integrate these perspectives.

I. COGNITIVE PROCESSES AND COMMUNICATION NETWORKS

The concept of "network" has emerged as a central notion in several disciplines in the last two decades. Even this brief time has been enough to show the broad applicability and significance of the network concept. Networks not only appear everywhere -- in families, groups, organizations, societies, computers, satellites, radio and television stations, among neurons within the brain and even among the concepts, words and terms that make up semantic networks -- but they appear to be a central, controlling concept wherever they are found.

To be sure, in each of the special areas within which the network concept has been studied, special circumstances and purposes require different ways of conceiving networks. But the concept of network itself is of such abstract generality that one may speak about networks independent of their context.

At the most basic level, networks consist of sets of interrelated nodes. Networks may therefore be classified according to the characteristics of the nodes and the characteristics of the relationships among the nodes. This chapter is concerned with *communication networks*, which may be defined initially as *sets of nodes whose state is at least partly a function of the states of other nodes in the set.* A network is a communication network, then, if the state of any of its nodes changes as a function of the perturbation of the state of another node in the set. Clearly human interpersonal and mediated networks are communication networks under this definition, since the state of any individual's thinking and feeling at a given moment is at least partly a function of information received from other individuals in the network. But neural networks also fit within the definition, since the state of excitation of any particular neuron at a point in time is a function of the activation values of other neurons in the network.

Communication networks play a particularly important role in the social sciences, since they provide the substrate of cognitive processes on both the individual and social levels. On the individual level, neural networks provide the organic substrate on which individual human thought and feeling take place, while on the social level, interpersonal and media networks provide the mechanism which gives rise to collective cultural experience.

While the social sciences have recognized a fundamental relationship between social structure and culture since their beginnings, no precisely articulated mechanism for this relationship has been widely accepted. Recent advances in the understanding of neural networks and their mathematical simulations, however, have provided a profound basis for reexamining this relationship. This chapter presents a general theory which considers cognition, whether individual or collective, to be an emergent, epiphenomenal property of communication networks. More precisely, cognitive processes, whether individual thoughts or cultural processes, are defined as the changing patterns of activation of the nodes in a network. The fundamental hypothesis of this theory is that **the structure of cognition at any point in time is given by the collective state of the network underlying that cognition, and *cognitive processes* may be considered a function of the changing state of the network over time.**

For individual people, the network providing the substrate of cognitive processes is a neural network. Similarly, cultural processes may be considered a function of the changing state of the communication network of an organization or society over time. In an important sense, both individual and collective cognitive processes may be viewed as epiphenomena emergent from the pattern of activity in the underlying communication network, in the same way as the information on a television screen or theater marquee can be described as an epiphenomenon of the underlying electromechanical processes which give rise to them. Cognition and culture are defined here as emergent properties of their underlying communication networks.

II. THEORETICAL CONTEXT

In classical Western thought, the individual has always been primary, both in the sense of being most important and in the sense of coming first. Plato's analysis of Society in the *Republic*, for example, stresses the primacy of the individual by asserting there are inherently three classes of individual, men of gold, men of brass and men of iron, and derives its theory of social organization from this classification. Aristotle similarly advises that society organize itself around the kinds of individuals who inhabit it; if there is one best man -- a philosopher king, then he ought to be placed in charge. If a few best, they should rule in an aristocracy, and, in the absence of these people, rule by all qualified citizens in a democracy is appropriate. In both philosophers views, slaves, women, foreigners and the like are inherently unqualified to govern since they lack the power of judgement, and should not participate in the life of the state.

In the analysis of social action as well, the individual is the starting point. Aristotle's view has come down to our own time with little change: The starting point of action is a human individual who has a goal, which creates a *potential* for action. The activity of the individual consists of making actual the state or condition which was potential. In this "Western Model" (Woelfel, 1987), the individual is both the origin of social activity and the starting place for analysis of society. The individual person is an autonomous entity, who, while certainly affected by contact with society, nonetheless derives its identity and character from its inherent nature. In the absence of society, people would certainly be different, but they would still be people.

The classical Western view is also essentially a static model. Aristotle's model of action is one in which rest or inactivity is the "normal" state for an individual. The concept of a goal implies that an imbalance has arisen, which must be alleviated by action. Something has happened which moved the individual out of his or her "proper place" or condition, which has created a "potential" which drives that individual to seek to restore the proper condition. When the potential has been "actualized", the individual returns to a natural state of rest. Thus in the Western model, activity proceeds through a series of starts and stops from one "position" or state to another through delimited periods of action.

The priority of the individual and the understanding of motion as an intermediary, semi-real condition between two (real) states of rest lends the Western model a third important characteristic: the classical Western view is categorical and sharply bounded. Each individual is perceived as absolutely distinct from its environment and from each other individual. Each action is a distinct entity, having a beginning and an end; the action is initiated by a potential in a distinct, bounded individual, and ends with the achievement of the end state, where potential has been actualized. The classical Western model sees the world as a fairly jerky process, where sharply defined individuals and objects move discretely from state to state and from place to place.

During the earliest days of the social sciences the Western model determined the shape of current theories of human behavior and social organization. Once again, Western analysts began with the individual. Thomas Hobbes began by postulating the existence of individual human beings with goals and desires, prior to society, whose goals ought to result in continuing conflict and warfare. The organization of individual human activity into relatively functional, cooperative societies had to be explained, and Hobbes explained it by assuming a central state with a monopoly of force -- a central police agency which prevented individual combat insofar as it could. Other early social scientists posited different mechanisms, but with few exceptions all accepted the notion that individual human beings with goals and plans preceded the origins of society. While scholars differed about the character of the "state of nature;" that is, the condition of people before the formation of society, they all agreed that they would have been persons, with minds, goals, beliefs and attitudes.

III. INTERACTIONIST THEORY

During the 20th Century, however, alternative views began to emerge. The interactionist position, for example, as exemplified by G. H. Mead, took the position that the society preceded the individual, both logically and chronologically. In this view, individual human

beings with "minds", beliefs, attitudes, goals and plans are not possible without society, and, indeed, it is the interaction of the individual with society which brings about that individual's mental organization. Regardless of the accuracy of detail in this theory, it is fundamentally important for its reversal of the roles of individual and society: in the classical western view, the individual creates the society. In the interactionist view, society creates the individual (Mead, 1932).

Along with the change in focus from individual to society, the interactionist model (as its name attempts to emphasize) focuses its attention on *process* rather than state. The development of the individual is an ongoing process, not simply an interval between birth and completion of the individual. In Mead's model, the individual is *being generated continually* by interaction with the society; should the interaction cease, the individual's self would not persist. Along with its deemphasis of the individual, interactionist theory points attention away from structure and toward process. In this view, structure emerges out of process.

The interactionist view also tends to blur the sharp, categorical distinctions between individuals, their environment and each other. In the interactionist model, the self of the individual consists of the set of relationships between self and objects, including other persons. The notion of the environment and society are included within the concept of self, which, indeed, makes no sense without such a merging. Although emerging out of a Western context, the interactionist view has a more organic character, in which the objects of experience are only distinct in light of the active perception of the individual. In this view, "objects" are simply arbitrary subsets of the experiential field which have been designated as having a unitary character for some purpose in some situations. In the same way, the distinction between a single self and its social context, including significant others and interaction partners in the social situation, is also strained within the interactionist model, which sees the definition of the situation emerging out of the *communication* or *interaction* of multiple selves in the social situation. Thus cognitive processes should not be thought of within this model as the mere intersection of the individual cognitive processes of the several participants, but are rather a joint product of all interactants which is continuously generated throughout the interaction.

Interactionist theory provides the background for the general theory of communication networks presented here. Each individual can be seen as a node in a communication network consisting of significant others and more casual contacts, and the character of the individual's self concept is a function of the values of these other nodes at any given moment. Similarly, the generalized cultural conception of the world of experience, as represented by the set of self conceptions of the individual members, is itself a function of the changing patterns within the overall social communication network. Neither the individual self concept and the overall culture are autonomous, but both are being generated continually by communication processes in the social communication network. They are epiphenomena emergent from the underlying communication processes within the social communication network.

A. Networks as Mechanism

Although quantitative empirical research based on interactionist principles is not unknown (Sewell, Haller & Portes, 1969; Woelfel & Haller, 1971; Picou & Campbell, 1975), for the most part, interactionist scholars have considered the dynamic, emergent, ongoing process by which the self is continually generated from symbolic interaction between individuals and others to be too "volatile and evanescent" to be subject to quantitative methodologies. In fact, it is fair to say that interactionists have had some difficulty getting beyond the initial interactionist insight, and one looks in vain for a "generative mechanism" by which the self is continually created. It is probably also fair to say that many, and perhaps most, interactionists believe that no such deterministic mechanism exists.

On the contrary, this chapter attempts to show that several very simple quantitative properties of simple networks produce effects remarkably like those described in interactionist theory. In particular, several quite simple networks can exhibit properties of pattern storage, recognition and recall that closely match interactionist descriptions of the process by which individual human beings symbolize and become aware of the objects of their experience, including themselves, and develop attitudes and orientations toward those objects through interactions with others.

B. Pattern Representation in Networks of Nodes

The foundational concept in the present theory is the concept of *communication*, which refers to the *changing distribution of energy in space as a function of time*. Communication in its most fundamental sense, as we define it here, means flow of energy. These flows are in general time dependent energy fields. There is no concept of intention or purpose implicit in this definition of communication; it is understood simply as a transfer of information or energy by whatever means.

The region at which two or more flows of energy intersect is defined as a *node*. Within this theory, the *state* of any node is a function of the flows which define it. If the energy fields which intersect to define a node are one dimensional (as the flow of electricity through an ideal one dimensional wire), then the node resulting from the intersection will be zero dimensional, or a point. If the energy flows are dichotomous, that is either on or off, then the node will take on only discrete values. If the energy fields are continuously variable, then the node can take on any positive real value; if the fields may vary in sign, the node may take on any real value positive or negative. If the fields are n-dimensional, then the node will be a diffuse n-dimensional region whose value will be a function of its coordinates in n-space.

In general, a set of energy fields may intersect to generate multiple nodes of various configurations, each of which will be a time-dependent energy field. The set of these intersecting energy fields at any moment will define a *network*, and the set of nodes resulting from the interactions will represent the "pattern" which the network represents at that moment. The "pattern" which results from the particular activation values of the nodes at any time can, of course, be random, but it could also represent any information structure whatsoever.

The simplest node can take on only two values along a single dimension, which may be described for convenience as "off" and "on." The value taken by a node at any point in time is called its "activation value." The set of values taken by any set of nodes at a given moment can be defined as a "pattern". "Communication" in this restricted model may be defined as the transfer of all or part of the activation value of any node(s) to any other node(s).

Like any system, a network may be partitioned arbitrarily so that a subset of the original network is defined as the "environment" relative to the other remaining part. This partitioning may be wholly heuristic, and done solely for the purpose of ignoring the internal properties of the portion of the network defined as the environment. This concept of arbitrary partitioning is particularly important in the case of social networks, where each individual person may be considered a node in an organization and each organization may itself be considered a node in a larger social network. The individual himself/herself may be partitioned into a set of neural networks.

Often the level of communication among an arbitrary set of neurons within a single individual may be small or zero while the communication between neurons in one individual and another (albeit mediated by electromagnetic forms of transmission other than typical neural mechanisms) may be substantial. In this (quite common) case, the communication network does not reside wholly within a single individual, but rather may exists across a set of individuals. This at least gives rise to the possibility that the intelligence of such a network may not reside solely in each of the individuals, but rather might be considered a property of the interpersonal network taken as a whole.

The network (considered at whatever level of aggregation) may communicate with its environment through weights or links from the environment to nodes within the network. Nodes which receive information from links to the environment are defined as "input nodes", and nodes which pass information through links to the environment are called "output nodes." Nodes which have no direct connection to the environment are typically called "hidden nodes." In neural networks, the function of a node as input, hidden or output is usually fixed by biological or programming factors, but in social networks, individual nodes may play each of these roles under different circumstances.

Input nodes receive information from the environment in the form of signals which alter their level of activation. In the general case, such signals can take on a wide variety of forms ranging from "simple, signed numbers of limited precision" to "...arbitrary symbolic messages to be passed among...units" (Rumelhart & McClelland, 1986, p. 132), but they all represent communication as defined above, that is, transfer of information or energy from one node to another.

The function by which the activation value of a node is related to an incoming signal is called the "activation function". For a binary node, this function may be as simple as a binary threshold, so that the value of the node is set "on" if the input signals exceed a given threshold level, and off otherwise. For nodes whose activation values may be multivalued, activation functions may be more complicated, particularly when the activation values may also be multidimensional, but the binary representation provides a sound starting point for initial understanding.

A single binary node can encode a pattern consisting of one bit of information. As the number of nodes in a set increases, the amount of information which can be encoded increases. For a network whose input nodes are binary, information received from the environment may be represented as a pattern of ones and zeros displayed over the input nodes. Thus, when a network receives information from the environment, it does so by encountering a signal at each input node at each point in time. Those nodes whose input signals exceed the threshold value will be activated, while others will remain off. The pattern of nodes which are activated constitutes a pattern which represents the pattern of signals at that point in time. The changing pattern of activations over time represents processes in the environment of which the network is "aware."

The number, arrangement and character of the input nodes, along with the character of the activation function, determines what kinds of pattern the input system will be able to represent. Most literature on neural networks and parallel data processing models considers only one dimensional (vector) arrays of one dimensional binary or continuous nodes, since the underlying model for the node in these areas is generally the neuron or the switch. In the more general case we consider here, nodes may themselves be networks whose activations may be highly multidimensional.

A one dimensional (vector) array of binary input nodes can record the presence or absence of a set of features. Figure 1 shows a vector of nodes, each of which represents a letter of the alphabet.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
X	0	X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	X	0	0	0	0	0	0

Neurons represented as a vector. Neurons marked with "X" are activated, all others are off. This pattern could represent "ACT", CAT", "TAC", "TCA", "CTA", or "ATC", since no sequence information is encoded in this pattern. Multiple occurrences of the same letter cannot be encoded within this scheme.

Figure 1. A One Dimensional Locally Encoded Network

The nodes marked "A", "C" and "T" are on, which indicates that the network recognizes the presence of those letters (features) in the environment. The one dimensional array of nodes, however, cannot encode the sequence of those features, so the pattern encoded in Figure 1 might represent "CAT", "ACT", or any of four other sequences of

letters. A two dimensional array of binary input nodes can keep track of not only the presence or absence of features, but also their sequence. Figure 2 shows a two dimensional (matrix) array of input nodes. As in Figure 1, each column represents a letter of the alphabet, but each row represents an ordinal position in a time sequence.

The pattern of activations shown in Figure 2 represents the English sentence "HELLO, SPOT". Higher dimensional arrays can represent correspondingly more complicated patterns[1].

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
0	0	0	0	0	0	0	X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	X	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	X	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	X	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	X	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	X	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	X	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	X	0	0	0	0	0	0

Input neurons represented as a matrix. Columns represent letters, rows represent ordinal position in a sequence. This network of neurons encodes the phrase "HELLO SPOT".

Figure 2. A Two Dimensional Locally Encoded Network

C. Distributed encoding

Both the models in Figure 1 and Figure 2 represent examples of "local encoding", in which each node represents one feature. A model which encodes a single feature as a pattern of activations among several nodes embodies what is called is called "distributed encoding", and can store considerably more information in a given number of input nodes.

[1] As suggested earlier, in the more general case of multidimensional energy flows, it is possible to find arrays in which the nodes are themselves multidimensional. A multidimensional node can represent more than one value simultaneously. This is quite a common case particularly in social networks, where nodes are typically themselves networks on another level of analysis. Like unidimensional nodes, multidimensional nodes might take on binary, multivalued or continuous values,or a combination of these. In general, the higher the number of nodes in a network, the higher the number of dimensions each node can encode, and the higher the number of values per dimension the node can take on, the more complex the patterns the network can represent.

A set of 7 binary nodes is sufficient to encode any of the ASCII characters; a 50 by 7 matrix of binary nodes can encode the English sentence "The quick red fox jumped over the lazy brown dog," -- or any other string of fifty ASCII characters -- including capitalization and punctuation. The set of 5 X 10^9 brains, each of which consists of 10^{11} neurons, can and do encode all the cognitive structure of all the minds on earth at any given moment.

D. Sparse Distributed Encoding

When very large numbers of nodes are available,a procedure called *sparse distributed encoding* is possible (Kanarva, 1988; Keeler, 1988). In both local encoding and distributed encoding, each element of a pattern is represented by one and only one node. But in sparse distributed encoding, each individual element of a pattern is encoded by setting all the nodes within a specific "neighborhood" (e.g., within a region circumscribed by a circle or hypersphere of a given radius) to a specific value. It is possible, within this model, for pattern elements to overlap, so that a set of nodes are activated for an arbitrary pattern "A", and an other set of nodes are activated for an arbitrary pattern "B", with some overlap between the activated nodes for the two patterns.

To recall a pattern stored in a sparse distributed model, an input stimulus activates an arbitrary node; all the nodes within the arbitrary radius are then "polled" and the results "tallied" or "thresholded"; if the majority of the responses represent pattern "A", then A is recalled; if the majority represent pattern "B", then B is recalled.

Sparse distributed encoding is relevant to the process whereby cultural information is encoded in society. Essentially identical or parallel information is stored in many individuals close to each other in the same communication networks (in the same "neighborhood'), either because they communicate with each other and spread the activations among themselves, or because they are in structurally equivalent locations (Burt, 1987).

In a social network, each individual is him/herself a network of substantial size, and can thus be thought of as a multidimensional real valued node. Social networks, which can consist of many individuals, thus have a total number of nodes many times this figure.

Networks consisting of arrays of multidimensional real-valued nodes can store very complicated patterns. Humans and organizations represent nodes which can produce highly multidimensional real-valued outputs. As an illustration of the pattern storage and recognition capacity of networks of this type, a group of students heard the following paragraph read aloud:

> I have a very small bedroom with a window overlooking the heath. There is a single bed against the wall and opposite it is a gas fire with a gas ring for boiling a kettle. The room is so small that I sit on the bed to cook. The only other furniture in the room is a bookcase on one side of the gas fire next to the window -- its got all my books on it and my portable radio -- and a wardrobe. It stands against the wall just near to the door, which opens almost directly onto the head of my bed. (Johnson-Laird, 1983)

Afterward, each of them was asked to report their estimates of the distance between each of the 13 objects mentioned in the paragraph and each of the others, which represents a highly multidimensional output of 78 real valued numbers per person (node). The numbers for each distance were entered into the Galileo Version 5.4 computer program (Woelfel & Fink, 1980), which averaged[2] the distances over nodes (respondents) and generated the coordinates of the 13 objects in space. The picture generated from this exercise is consistent with the room in the text read to the students. Random splits of the data show the same room, as should be expected. Thus the network of individuals recorded the pattern of the entire paragraph within a single aloud reading -- and could reproduce it accurately -- even though none of the individuals reported being able to picture or draw the room.

This illustrates two important characteristics of information processing networks: first, a network is capable of encoding very complicated patterns of information very quickly and of retrieving it accurately. Second, it shows clearly that the information encoded is not a property of the individual nodes, since, in the example given, none of the individual nodes (individuals) reports being able to picture or recall the overall geometric structure of the room. Rather the pattern exists only in the network of nodes considered as a whole. Moreover, in the present example as well, it is possible to see that the information is stored in such a distributed and redundant way that the pattern can be retrieved from a reasonably sized random subset of the nodes. This characteristic makes it clear how extensive networks such as organizations, nation states and cultures can retain complex information patterns such as attitudes, beliefs and values over generations even when many or even all individual nodes are lost to the system due to immigration, death and other factors. Most importantly, however, this example makes it clear that *the network as a whole exhibits emergent pattern storage and retrieval capabilities that go beyond those of the component nodes.*

E. Pattern Association

Clearly, the combination of possible activation values for a large number of nodes makes it possible for networks to represent virtually any pattern conceivable. It is also possible, given appropriate patterns of communication among the various nodes, to associate any pattern with any other. The model presented up until now has considered only sets of nodes each of which communicates with the environment, and none of which communicates with each other. Theater marquees and television screens are examples of

---

[2] Averaging the values is a very simple but common function for numeric outputs which can be viewed as analogous to thresholding. More complicated functions, such as log transforms, trimming, and the like, are often used, as are other measures of central tendency, but the concept of an aggregate pattern which has meaning for a collection of nodes while essentially uninterpretable based on outputs from only a single node remains the same.

this class of network. But while the patterns they can encode can be very elaborate, they are passive copies of the environmental input and exhibit essentially no internal processing. Nodes may, of course, communicate with each other at various levels. The channels through which nodes communicate have been called variously "links", "connections", "weights" and other terms, and those terms will be used here as synonyms. These weights may in general take on any real value[3], and are meant here to represent the proportion of the activation level of any node that will be transmitted to another node to which it is connected by that channel. Thus the weight $w_{i,j}$ represents the proportion of the of the j_{th} node that will be communicated to the i_{th} node.

How a node will respond to the inputs it receives from those nodes which communicate with it is determined by its "activation function." The activation function determines how a node will combine the various signals it receives from all those nodes connected to it. The actual array of potential activation functions is infinite, but they may be described in general from simpler to more complicated functions.

The first is the simple linear function, in which all inputs to a given node are summed, and that node then outputs a signal which is the sum of all its inputs. Simple linear networks can have substantial information storage and retrieval capacities, but cannot produce internal representations of environmental patterns that differ from those in the environment, nor can they perform complex inferences, such as the "exclusive or" relation. Included within the class of linear networks is the *perceptron*, which was studied extensively by Rosenblatt (1962) and Minsky and Papert (1969) who first demonstrated the limitations of inference inherent to the linear two layer network.

A second common activation function is a simple step function, in which a node outputs a given value if the inputs to it sum to more than a given threshold[4]. Even such a simple rule as this introduces important nonlinearity into a network which makes it capable of generating internal representations of external patterns which are not simple linear combinations of external signals, and thus substantially increases its inferential capabilities. Non linear networks can solve problems like the "exclusive or" relation (Rumelhart, et. al., 1986, pp. 318-362, McClelland & Rumelhart, 1988, Chapter 2). The step function, however, is not everywhere continuous, which causes mathematical difficulties for some learning algorithms.

[3] Sometimes, particularly in the case of social networks, precise data about the actual weights or connection strengths is not available to investigators, and so a considerable literature exists in which the connections between nodes are discussed and analyzed as if they were binary. Whatever measurement difficulties might be encountered in any empirical situation, however, this practice is clearly inadequate for the investigation of intelligent, self referencing networks, since these are sensitive to very small variations in weights. In the case of the SPOT and ROVER programs discussed below, rounding the weights at the third decimal place results in serious deterioration of performance.

[4] The concept of a threshold function is appropriate particularly for social networks, where a communication from one or more nodes may activate another; e.g., "Please call me if anyone calls", or "If there are too many complaints, contact Quality Control".

A third commonly used activation function is the logistic function, sometimes referred to as a "sigmoid" function, because its shape when plotted resembles an integral sign:

$$a_{pj} = 1/(1+e^{-net_{pj}}) \tag{1}$$

where:

a_{pj} = the activation of the j_{th} node for the p_{th} pattern, and
net_{pj} = the net input to the j_{th} node for the p_{th} pattern from all input nodes.

The logistic function is particularly useful since it provides the nonlinearity and increased inferential capacity of a step function, but is a continuous differentiable function. This is particularly important in supervised learning or "back propagation" models, since these require that the differences between the pattern output by a network and the desired or "target" pattern be expressed as a continuously differentiable function of the weights so that the weights may be changed to produce the correct output (Rumelhart, et. al., 1987,pp. 318-362).

Each of these activation functions establishes the activation value of the node solely as a function of the inputs from other nodes, but more complicated models can take into account the present absolute or relative activation value of the node. These considerations produce another family of activation functions such as "competitive learning", in which nodes already highly activated are more likely to be further activated for a given level of input than those not so highly activated or "resonance", in which sets of interconnected nodes, once activated, will tend to maintain each other's activation levels (Grossberg,1978).

Activation functions can take into account variables other than the set of inputs from other nodes and the activation value of the node itself. *Time* is perhaps the most common such variable, and is usually included to model a decay function such that the node loses a proportion of its activation as a function of time. This decay functions as a "restoring force" which tends to return nodes to their "resting activation levels" as a function of time (Grossberg, 1978; McClelland & Rumelhart, 1988, pp. 12-15).

Activation functions need not be deterministic. Several important models, such as the Harmony Model (Smolensky, 1986, pp. 194-281) and the Boltzman Machine (Hinton & Sejnowski, 1986, pp. 282-317) employ stochastic activation functions, in which the *likelihood* that a node will be activated is a function of the inputs to that node. Stochastic models may well be better representative of actual neural functioning, but are almost certainly more representative of the way inputs function to activate or fail to activate nodes in social networks than deterministic models, at least insofar as the great complexity of input patterns in social networks usually precludes complete measurement of the total net input to any node.

F. Information Processing and Network Structure

The weights, along with the activation functions for each node, make up the structure of the network and determine the patterns of flow of information through the network. These flows in turn determine the process by which a network receives information from the environment, constructs an internal representation of that information, and outputs a response.

Figure 3 shows a hypothetical network consisting of six nodes representing the words "Cat", "Dog", "Barks", "Howls", "Meows", and "Purrs".

Input = "Meows"

		Cat	Dog	Barks	Howls	Meows	Purrs
	Cat		-.8	-.9	.2	.8	.9
	Dog	-.8		.9	.3	-.8	-.7
	Barks	-.9	.9		.5	-.3	-.9
	Howls	.2	.3	.5		-.2	-.1
+1	Meows	.8	-.8	-.3	-.2		.8
	Purrs	.9	-.7	-.9	-.1	.8	
		on	off	off	off	on	on

Figure 3. Spreading Activation Network 1

Each of the nodes may take on the value "0" (off), or "1" (on). The nodes are connected to each other by weights which represent their relative "closeness" in the network.[5] They communicate with each other by a simple threshold rule: the signal sent from any node i to any node j equals the product of the activation value of i and strength of the connection between i and j. Thus the total signal received by any node j will be the sum of the signals received from all the other nodes, or

$$anet_i = \sum_{j=1}^{N} w_{ij} a_j \qquad (2)$$

The way a node responds to the set of signals it receives is determined by its activation function; in this case we adopt the rule that the node will be activated if the

[5] In the present example, the weights are essentially the correlations between frequencies of occurrence of the various words. Thus "Meows" and "Cat" tend to "go together", with a weight of .8, while "Meow" and "Dog" have a negative coefficient of -.8.

sum of its input signals is positive; otherwise it will be turned off, or

$$\begin{array}{ll} \text{if } x > 0, & a_i = a_i + 1 \\ \text{if } x = 0, & a_i = \text{unchanged} \\ \text{if } x < 0, & a_i = a_i - 1 \end{array}$$

Following this rule, we assume the network receives the input "Meows" from its environment (i.e., the node which represents "Meows" has been activated.) This sets the activation value of "Meows" at +1, and the activation values of the other nodes at 0. Multiplying the weights in each column by the activation values of the corresponding rows, then summing within each column shows that the activation of the node "Meows" will "spread" to the nodes "Cat" and "Purrs", setting their activations to 1, but will leave the nodes "Dog", "Barks" and "Howls" off.

Input = "Howls"

		Cat	Dog	Barks	Howls	Meows	Purrs
	Cat		-.8	-.9	.2	.8	.9
	Dog	-.8		.9	.3	-.8	-.7
	Barks	-.9	.9		.5	-.3	-.9
+1	Howls	.2	.3	.5		-.2	-.1
	Meows	.8	-.8	-.3	-.2		.8
	Purrs	.9	-.7	-.9	-.1	.8	
		on	on	on	on	off	off

Figure 4. Spreading Activation Network 2

Figure 4 shows that activating the node "Howls", will also activate the nodes "Cat", "Dog" and "Barks"; Figure 5 shows that activating both the nodes "Barks" and "Howls" will also activate "Dog", but will leave "Cat", "Meows" and "Purrs" off.[6]

[6] A more thorough example would examine the results of the communication after more than the first step or "cycle" of the network. This exercise can produce a surprising amount of complexity very rapidly, particularly in real cases where finite speeds of communication determine the order in which nodes are turned on or off. The activation of the node "Howls", for example, turns on both "Barks" and "Cat". But since "Dog" and "Cat" are so strongly negatively connected, each turns the other off. If "Howls" communicates its activation to "Dog"
(continued...)

Input = "Barks and "Howls"

		Cat	Dog	Barks	Howls	Meows	Purrs
	Cat		-.8	-.9	.2	.8	.9
	Dog	-.8		.9	.3	-.8	-.7
+1	Barks	-.9	.9		.5	-.3	-.9
+1	Howls	.2	.3	.5		-.2	-.1
	Meows	.8	-.8	-.3	-.2		.8
	Purrs	.9	-.7	-.9	-.1	.8	
		off	on	on	on	off	off

Figure 5. Spreading Activation Network 3

This example shows clearly that communication among the nodes of the network produces an apparently qualitative change in the pattern recognition and storage capabilities of the network. When the nodes do not communicate, the network can represent a pattern of virtually any complexity when activated directly by the environment, but the complete input is required to produce the complete pattern. When the nodes communicate, however, the complete pattern can be produced with only a partial input. When a sufficient subset of the nodes in a stored pattern is activated, the activation of those nodes will "spread" through the links and in turn activate the rest of the nodes in the pattern.

It is worth emphasizing the fundamental role communication as it has been defined here plays in this process. A pattern is stored by "connecting" its elements together. Things that "go together" are "close". Nodes or elements in turn *communicate* their activation values to other nodes in proportion to their closeness in the communication network. If a node is "on", it will tend to transmit that "on-ness" to other nodes through the links between them, so that the "on-ness" will spread to other nodes which represent the other elements in the pattern. Similarly, if a node is "off", it will tend to communicate its "off-ness" to other nodes through the links between them. The entire pattern is encoded in the pattern of communication among the nodes as connections or weights, and can be recovered by the activation of any suitable subset of nodes.

[6](...continued)
before "Cat", "Cat" will not be activated. If it communicates its activation to "Cat" before "Dog", "Dog" will not be activated. This phenomenon is referred to as "hysteresis" (McClelland & Rumelhart, 1988, pp. 16-17).

G. Self Reflexive Networks

As the previous section showed, communication networks are capable of storing patterns of arbitrary complexity, and linking those patterns in such a way that the occurrence of one pattern gives rise to another, or the occurrence of part of a pattern recalls the entire pattern. Networks that behave like this may be capable of very sophisticated cognitive tasks, but could always be accused of "mindless" pattern matching activities. It is possible, however, for a network to be structured in such a way that it is self reflexive, taking its own past behavior into account in determining its future behavior.

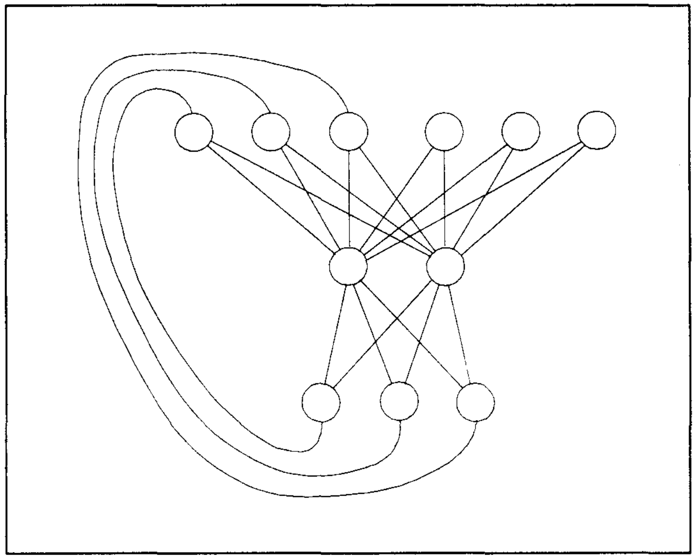

Figure 6. Self Reflexive Network with feedback

Figure 6 shows a network connected in such a way that its output is partially fed back to its input nodes. It consists of a layer of input nodes (at the top of the diagram, connected through a layer of hidden nodes at the center of the figure, to a layer of output nodes at the bottom of the figure, which in turn are connected to some of the input nodes. Such a network will receive its own previous activities as part of the input pattern which it attempts to match when finding an appropriate output pattern. Figure 6 represents only one possibility out of an infinite number of such structures. When the communication network consists of human individuals, a feedback architecture suggested by Jordan (1986) might be more appropriate. Jordan's model allows feedback not only from output nodes to input nodes, but from every node to itself as well; this is equivalent to a set of self reflexive nodes which is itself self reflexive as a structure.

H. Forming and Changing the Network Structure

There are two fundamental types of patterns implicit in the discussion up until now. The first are the patterns composed of the activation values of the nodes at a given moment, or the changing pattern of activations across an interval of time. These patterns, however complex they may be, are transitory and represent the image a network is displaying at a historical time. In a neural network, they represent the "thoughts" the network is exhibiting, while in a social network they represent the "currents of opinion" flowing through a culture at any moment. Changes in the patterns being represented by a network over time may be considered cognitive processes or cognitive change.

The second type of pattern is represented by the much deeper, more stable patterns represented in the set of weights in the network. These weights represent the relatively stable basic structure of the network and the set of relationships among the other patterns. They represent not what a network is "thinking" at a given moment, but what it has learned since its inception. While the activation values of the nodes at any given moment determine the pattern a network is displaying at that moment, the memory of all patterns known to the network is contained in the pattern of weights or communication patterns among the nodes. The processes by which these weights are formed and changed, therefore, are central to whatever intelligence a network might exhibit. Such changes might be defined as *structural.*

The first general type of processes by which the architecture of a network can be changed might be considered *causal* or perhaps *accidental* processes, that is, processes in which forces from the environment impose themselves on the network and its components without any deliberate goal in view.[7] Perhaps the most common such process is that suggested in broad outline by Hebb (1949): Weights among nodes which are simultaneously activated are strengthened.[8] The logic behind the Hebbian rule is quite straightforward: when a network is displaying a given pattern as a set of activated nodes, it can "remember" or store that pattern by connecting together the nodes which represent the pattern so that they tend to be activated as a unit. If the rule which "ties together" nodes which are simultaneously active is always automatically enforced, then such a network will tend to store any pattern it displays automatically. If the rule is repeatedly

[7] These forces include genetic forces which determine the structure of neural networks, geographic and climatic forces which influence social networks, as well as accidental forces which may alter or disrupt the structure of any network.

[8] Hebb did not provide an exact mathematical form for this rule, and minor variations in form are abundant. The weights in the examples presented in Figures 3, 4 and 5 are given as correlations among the occurrences of the items, or in the terminology developed here, correlations among the frequencies of activations of the nodes in the patterns.

enforced, then each time a pattern is displayed by a network the connections among its constituent nodes will be strengthened and the network's memory of the pattern will be reinforced. If the initial activation values of the nodes are set by inputs from the environment, then the Hebbian rule guarantees that such a network will develop a memory of the major features of its environment, and will selectively favor the recollection of those patterns most frequently presented to it by the environment.[9]

A network whose weights are determined solely by a variant of the Hebb rule or other causal mechanisms becomes "like" it's environment in a wholly passive way. A second set of processes by which network structure can be changed might be called *purposive* or *supervised* processes. One such alternative, which is itself a variant of the Hebb rule, has been suggested by Rumelhart, et al, (1987, pp 318-362).The essential feature of this "back propagation" model is the existence of a "target" pattern, that is, a pattern which is, for any arbitrary reason, considered to be the "correct" output pattern for a particular input pattern.

The target pattern represents a pattern of activation values of the output nodes of a network corresponding to the desired output. The difference between the pattern desired and the pattern actually output by the network can easily be defined as the difference between the activation values of the nodes observed and those expected by the pattern. These differences may be considered the *errors* produced by the network. These errors can of course be described as a function of the activations of the nodes, which can in turn be expressed as a function of the weights connecting the nodes. It is possible, then, to express the errors as a function of the weights. If the activation functions of the nodes are continuous (as is the logistic function typically used in back propagation networks), then the derivative of this function is defined everywhere on the function, and it is easily possible to modify the weights (usually by a quasi steepest descent algorithm) until the error is minimized[10]. Such a network can learn to produce a desired output pattern for a given input pattern. Connections between input and hidden nodes and hidden and output nodes are initially randomized, so that when the network receives an input pattern, the response it outputs is simply a random activation of the output nodes. The errors are then calculated as the differences between the actual values of the output nodes and the values associated with the correct pattern. By a quasi steepest descent algorithm, the weights of the connections among the nodes are then modified until all the correct response patterns

[9] The Hebbian rule has certain limitations which are well understood in the Parallel Data Processing community. Among the most important is the tendency to confuse similar patterns.Since similar patterns will be encoded into similar weights, presentations of partial information about highly covariant patterns will result in frequent confusions among the patterns (McLelland, et. al., 1987, p.38).

[10] Notice that this process occurs "backwards" through the network, beginning with the errors of the output nodes, then moving to the weights from hidden to output, then to the activations of the hidden nodes and then to the weights from input nodes to hidden nodes. This backwards sequence is the basis for the name "back propagation."

are associated with the appropriate input patterns.

These networks are "trained" by presenting them with lists of paired patterns. The first pattern in each pair of patterns is an "input pattern", and represents a given pattern of activation of the input nodes of the network. The second pattern in each pair represents the pattern of activation of the output nodes which is meant to be associated with that input pattern.

When the input pattern is initially displayed, the (initially random) connections between input and hidden nodes and hidden and output nodes causes a random pattern of activation of the output nodes. The values of this output pattern are subtracted from the values in the "target pattern" and the differences represent error. These errors can then be expressed as functions of the activations which in turn are expressible as a function of the weights. The derivative of this function is then calculated and the weights are modified, the input pattern is presented again and the process is iterated until the errors fall below a specified tolerance. This procedure is essentially an iterative non-linear multiple regression model which finds a set of weights which maps the pattern of input activation values onto the desired pattern of output activations.

I. Communication Networks and Social Structure

These mechanisms gained from the study of neural networks and their mathematical abstractions make it possible to construct a theory of cognitive processes which is able to incorporate all the features symbolic interactionist theory considers essential to human cognitive activity, both at the individual and the collective level, yet based on a solid, empirical mechanism which is at once observable, quantifiable and testable. In summary, the theory holds that cognitive processes, whether individual or collective thoughts, attitudes, or beliefs, can be described as patterns of activations of nodes in an underlying network. The nodes may be linked to one another by communication channels of greater or lesser strength, so that the activation values of linked nodes tend to be communicated to those nodes to which they are connected. This "spreading activation" model means that stimulation of some subset of nodes which represent a pattern will result in the activation of the remaining nodes, recalling the entire pattern. The individual or society is thus able to record complicated patterns and recall them. These processes by which subsets of activated nodes communicate their activations in turn to other subsets of nodes represent "conversations" among the nodes, on a structural level, and among the concepts or patterns these subsets represent on the conceptual level. These conversations may be external, between patterns of nodes within the network and elements of its environment, or they may be entirely internal, among subsets of nodes within the network. In any case, these conversations represent one class of cognitive processes: those processes which result from the existing state of the communication linkages among the nodes.

The memory of the system, whether individual or collective, is a function solely of the weights or connection strengths among the nodes. These constitute the basic structure of the system, and can be changed in two ways. The first of these ways is "unsupervised", given by a variant of the Hebb rule, and by this mechanism, a system constructs patterns which are internal representations of the environment to which it is exposed. The second process is a "supervised" learning rule, in which outputs resulting from any input are compared against a template for the "appropriate" or desired output, and weights or connections are changed to minimize the discrepancy between actual output and the desired pattern. Each of these procedures represent structural change in the system. The processes which lead to changes in the communication linkages among the nodes in the network may be considered a second class of cognitive processes.

While the process by which patterns flow through the system may be referred to a kind of consciousness, it is clearly not a "self consciousness". Self consciousness can be modelled in this theory, however, by simple feedback loops between output nodes and input nodes, so that the system is aware of its own actions, and considers its own actions as part of the pattern to which it is responding.

References

Burt, R. S. (1978). Social contagion and innovation: Cohesion versus structural equivalence, *American Journal of Sociology*, 92:1287-1335.

Grossberg, S. (1978). A theory of visual coding, memory and development, in E. L. J. Leeuwenberg & H. F. J. M. Buffart (Eds.), *Formal theories of visual perception*, New York, Wiley.

Hebb, D. O. (1949). *The organization of behavior; a neuropsychological theory*. New York: Wiley.

Hinton, G. E., and T. J. Sejnowski. (1986). Learning and relearning in Boltzmann Machines, in Rumelhart, D. E., and J. L. McClelland, (Eds.), *Parallel Distributed Processing: Explorations in the microstructure of cognition*, Cambridge, MA, The MIT Press, 1986, pp. 282-317.

Johnson-Laird, P. N. (1983). *Mental models: Toward a Cognitive Science of Language, Inference and Consciousness*, Harvard University Press, Cambridge, MA.

Jordan, M.I. (1986). Attractor dynamics and parallelism in a connectionist sequential machine, *Proceedings of the Eight annual conference of the Cognitive Science Society*, Hillsdale, NJ, Lawrence Erlbaum Associates.

Kanarva, P. (1988). *Sparse distributed memory*, Cambridge, MA MIT Press.

Keeler, J. D. (1988). Comparison between Kanarva's SDM and Hopfield-type neural networks, Cognitive Science, 12, 299-329.

McClelland, J. L., and D. E. Rumelhart (1988). Explorations in Parallel Distributed Processing: A handbook of models, programs and exercises, Cambridge, MA, The MIT Press.

Mead, G. H. (1932). *Mind, Self and Society*, Chicago, U. of Chicago Press.

Minsky, M., and S. Papert (1969). *Perceptron*, Cambridge, MA, MIT Press.

Picou, J. S., and R. Campbell (Eds.). (1975). *Career patterns of minority youth*, Columbus, Ohio, Charles E. Merrill.

Rosenblatt, F. (1962). *Principles of Neurodynamics*, NY, Spartan.

Rumelhart, D. E., and J. L. McClelland, (Eds.). (1986). *Parallel Distributed Processing:*

Explorations in the microstructure of cognition, Cambridge, MA, The MIT Press.

Rumelhart, D. E., G. E. Hinton and R. J. Williams. (1986). Learning internal representations by error propagation , in Rumelhart, D. E. and J. L. McClelland, (Eds.), *Parallel Data Processing: Explorations in the Microstructure of Cognition*, Cambridge, MA, The MIT Press, 1986, pp. 318-362.

Smolensky, P. (1986). Information processing in Dynamical Systems: Foundations of Harmony Theory, in Rumelhart, D. E., and J. L. McClelland, (Eds.), *Parallel Distributed Processing: Explorations in the microstructure of cognition*, Cambridge, MA, The MIT Press, 1986, pp. 194-281.

Woelfel, J. and A. O. Haller. (1971). Significant others, the self-reflexive act, and the attitude formation process. *American Sociological Review*, 36(1), Feb..

Woelfel, J. and E. L. Fink. (1980). *The Measurement of communication processes: Galileo theory and method*, Academic Press, NY.

Woelfel, J. (1987). The Western Model, in D. L. Kincaid, *Communication Theory from Eastern and Western Perspectives*, New York, Academic Press.