5
How We Make Categories and Constructs

The Mahayana branch of Buddhism aspires to a state of spiritual perfection where we perceive things just as they are, without the aid of categories or labels. In this state, which is called Nirvana, "the illusory distinctions of the discriminating mind are left behind."[1] In Nirvana,

> ... there is nothing but what is seen of the Mind itself; where there is no attachment to external objects, existent or non-existent; where ... there is an insight into the abode of reality as it is; where, recognising the nature of mind in itself, one does not cherish the dualism of discrimination ...[2]

I reserve judgment on whether the neural networks within us can lead us ultimately to a path of Neur-vana. But if so, that will probably happen in a life past this one, if such a life exists. For smaller-scale spiritual growth in a complex world full of both pain and pleasure, where decisions need to be made in limited time, we must make categories in order to live intelligently. Without categorization, a new person we encounter wouldn't be recognizable as a fellow human being, so we would miss her or his potential as a friend, acquaintance, or colleague. A fruit we see in the supermarket, being slightly different from any other we have eaten, wouldn't remind us of the delicious taste of fruits we have eaten before.

There is even a school of thought which holds that categorization is our *primary* cognitive function. The neuroscientist Gerald Edelman, for example, believes categorization precedes perception rather than following it, and is the basis of our "construction of reality."[3] Edelman's position seems to me an extreme one, since children have the capacity for simply enjoying sensory events as they happen. But clearly the categories we create influence our perception of individual objects and events as much as the other way around.

The limits on our discernment mean that we are unlikely to reach the Buddhist ideal in real life. We will always "see through a glass darkly and not face to face" (*I Corinthians* 13:11-13). But we have a lot of control over *how* we choose our categories. We can make categories that are rigid, putting people or objects in the same "boxes" regardless of the situation. This includes, for example, always classifying people by skin color even when that isn't relevant. Or else we can categorize in ways that shift in emphasis as the situation changes. In classifying foods, for example, an apple is like a grape if one is classifying by general taste characteristics, but not if one is classifying by size.

The capacity to shift categories is one of the keys to our better mental and spiritual selves. Clearly, this capacity is at the heart of our advanced cognitive capabilities. Less obviously, it is also at the heart of our ability to be compassionate.[4] This is because we need to shift categories in order to see people who are different from us in obvious ways, such as gender, skin color, or ethnicity, as similar to us in their feelings and aspirations. If our categorizations are too rigid, we fall easily into stereotypical conceptions of genders, races, or ethnic groups.

So in order to understand how stereotypes form, we need to understand how we develop categories in the first place. The experimental psychologists Carolyn Mervis and Eleanor Rosch found that when humans encode a category, they create in their minds a generic member (*prototype*) of that category.[5] When subjects are tested as to whether particular objects belong to that category, they say "Yes" soonest when tested with members of the category (*exemplars*) that are close to the prototype. For example, a sparrow is recognized as a bird sooner than an ostrich is, because for most English speakers the mental picture of a generic "bird" is closer to a sparrow than to an ostrich.

The psychologists Michael Posner and S. W. Keele studied categorization further by inducing their subjects to make artificial categories out of dot patterns they saw.[6] Posner and Keele found that if a subject was first taught some general pattern of dots and then given random variations on that pattern, the prototype the subject formed was close to the average of these variations. After training, the same subject was taken away from the scene, and then shown some more dot patterns. The subject then recognized the prototype faster than any of the actual exemplars even if he or she had not actually seen the prototype (see Figure 5.1).

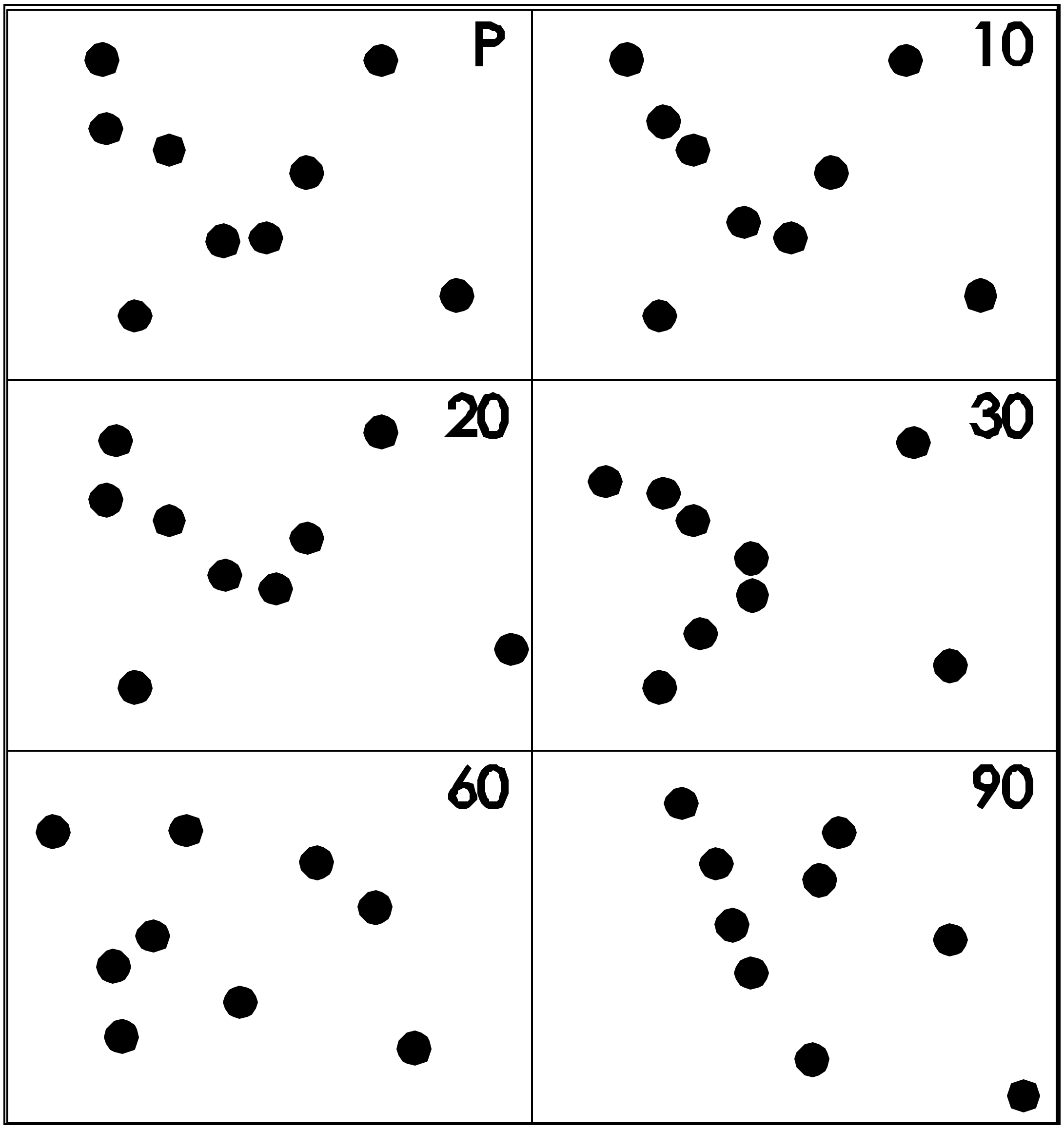

Figure 5.1. Prototype, labeled *P*, of a category of random dot patterns, and successively greater distortions of the prototype, labeled with increasing numbers from 10 to 90. (From Knapp and Anderson, 1984; reprinted with permission of the American Psychological Association.)

Categorization has been actively studied in model neural networks for over ten years.[7] Networks have been applied to a wide range of problems, for example, in cognitive psychology(such as recognizing male versus female faces); engineering (such as recognizing the shapes of specific machine parts in manufacturing applications); and medical diagnosis (such as recognizing visual patterns associated with specific types of cancer cells). An exciting area of current research involves expanding these networks to model how categorization influences and is influenced by other cognitive and social functions.

Personal Constructs

How might the categorization function be connected to other cognitive functions, such as context evaluation and planning? We need to understand such connections between functions in order to understand how we store knowledge *about* the different categories we have learned. To study this process, we can connect neural networks that perform categorization into larger aggregations with other networks that perform other functions. This type of construction is biologically realistic, because the brain is composed of multiple interconnected subsystems that are specialized for different tasks.

In a categorization network, there are nodes (see Chapter 2) that encode particular categories. Suppose the node of the network that encodes a category is connected to another node that encodes a specific property. Suppose further that members of the category are repeatedly observed to have that property. Then learning takes place: the functional connection ("synapse") between the category and property nodes gets strengthened. For example, if one learns to associate the category of "robins" with the property "red breast," some synapse between an idealized "robin category" node and an idealized "red breast" node becomes stronger. This is analogous to the association between bell and food that Ivan Pavlov conditioned in dogs.

Broadly speaking, we need to understand how to associate categories with significant feature properties (how we mentally encode, say, "robins have red breasts"). We also need to understand how we associate categories with positive or negative emotional values (how we encode "I like fishing" or "I hate broccoli").

Categories, information about categories, and emotional evaluation of categories can all be lumped together as *personal constructs*, a concept defined in a book by the social psychologist George Kelly.[8] Constructs are the mental notions that each of us develop to make sense of the world we live in. They are *personal* because, whether the reasons are based on innate temperament, experience, or culture, each of us emphasizes different qualities in what she or he abstracts from the world. For example, the notion of "kindness" or "cruelty" might be particularly important to one person, whereas the notion of "bravery" or "cowardice" is more compelling to another. Kelly emphasized the need to study each individual as a complex system before making sense of how individuals interact in groups.[9]

Kelly noted that constructs are often *bipolar*, that is, coming in pairs of opposites. The person who places a high positive value on kindness will automatically place a high negative value on cruelty. Likewise, the person who places a high negative value on weakness will place a high positive value on strength. The way we encode such pairs of opposites involves the combined operation of the brain's systems for emotional evaluation and for sensory processing; these two systems correlate somewhat, respectively, with the "old mammalian" and "new mammalian" brains of Paul MacLean[10] (see Chapter 3). The amygdala, a part of the "old mammalian" limbic system (Figure 3.1), plays a key role in forming constructs by keeping track of the emotional values of sensory events.[11]

Bipolar constructs are sometimes modeled by a particular type of neural network has been developed that is specialized to encode pairs of opposites, whether of sensory percepts or emotions. This type of network is called the *gated dipole* and was developed by Stephen Grossberg[12] (link to gated dipole). The dipole gets its name because it is composed of a left and right "channel" representing two "polar opposites"; these opposite concepts may be pain or pleasure, satisfaction or frustration, presence or absence of an apple, or many other things. When a neural pathway coding one of these two concepts is turned off, this briefly turns on a pathway coding its opposite. This is a model of the tendency of people and animals to feel, for example, a sense of pleasure when a painful stimulus like electric shock is turned off.

The dipole employs a mechanism involving idealized chemical *neurotransmitters*, the substances that mediate the movement of electrical signals from one nerve cell to another. In the dipole it is assumed that transmitter gets *depleted* when it is used — the opposite of Pavlovian conditioning, where synapses or connections become *strengthened* with use. But depletion, in the sense of a signal getting weaker with repetition, also sometimes makes psychological sense. For example, if you look at a visual stimulus such as a light for a long time, after a while you become habituated to the light as part of the environment, and you no longer perceive it as strongly. This mechanism is used in the network, for example, to "deplete" the pathway representing pain so that pleasure is felt when the pain is removed.

How does this work? In general, at each node of a model neural network, there is a variable called *activity* which changes over time, as a result of interactions with other nodes and with stimuli from the outside world. This variable is not yet precisely defined but is usually thought of as analogous to the average electrical impulse frequency over some group of large numbers of biological neurons. If a network node encoding an emotion, for example, has high activity, this is taken to mean that the organism feels this emotion strongly. High activity in a node encoding a perceived object or class of objects, on the other hand, indicates such objects are perceived strongly.

The nodes in a gated dipole neural network are divided into two interconnecting "channels" encoding opposite concepts such as pain and pleasure. If one of the two channels has been active, neurotransmitters in that channel are more depleted than in the opposite channel. The two channels compete for activation, meaning that activity in one side tends to suppress activity of the other side. So if activity on one side of a dipole is turned off, this releases the other side from suppression and so leads to short-term activity of the other side. An example of this is feeling pleasure when something painful is suddenly ended. Figure 5.2 shows a graph of a typical time course of the activities of these pain and pleasure channels (or, more generally, channels that encode any pairs of psychological opposites). Note that after a painful input is shut off, there is high activity of pleasure nodes, indicating a pleasurable emotion.

The dipole mechanism is an abstract mathematical structure that can be used to model different types of "opposites," whether emotions or perceptions, for example. At the emotional end, the dipole can help explain why humans frequently jump back and forth between opposite feelings, such as elation and depression. At the perceptual end, it can be used to explain why novel features of the environment are often more noticed than features that are always present, because dipoles enhance the difference between "present" and "absent." Later we will see how the dipole can be used to model human emotional attachment to both novel and familiar objects in different contexts. In general, the gated dipole mechanism provides a theory of why, all else being equal, we tend to feel or perceive changes in the environment (whether in our internal states or an external visual scene) more than we feel or perceive aspects of the environment that stay the same.

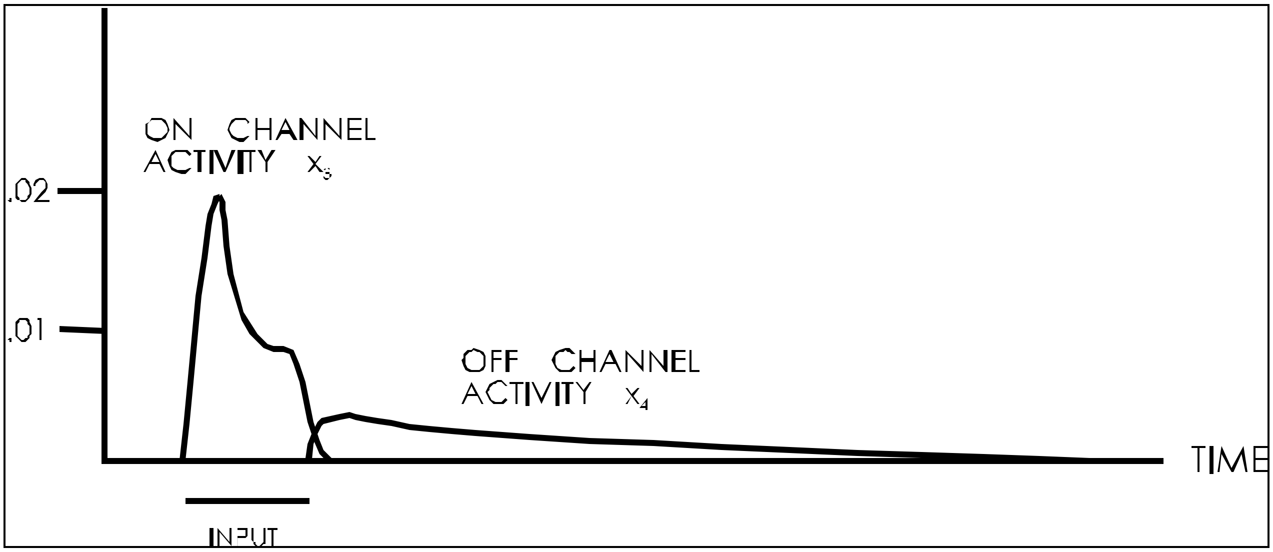

Figure 5.2. Typical time course of the channel activity variables of the gated dipole network. (Adapted from Levine and Prueitt, 1989, with permission of Elsevier Science Publishing Company.)

Let's try to go back and relate this neural mechanism to forming stereotypes. A person who forms stereotypes about a racial, ethnic, gender, or professional group clearly attaches importance to a construct, in George Kelly's term, such as "black," "Irish," "female," or "social worker." This leads to sharp emotional contrasts, or "dipoles," in the person's mind between members and nonmembers of the group. In extreme cases, such a construct is what Kelly calls *preemptive*: one of the categories to which a person or object belongs assumes such importance that it crowds out perceptions of the same person or object that are unrelated to the category. For example, if "talking with a Southern United States accent" is a preemptive construct in someone's mind, it may go with a certain mental picture that doesn't include a high level of education. Hence, the person with the construct may have trouble believing that a person with such an accent can also be a mathematician. Likewise, a person could find it hard to believe that a Jew can also be a baseball star, or a librarian can also be a beauty queen. This type of beliefs can be summed up as *nothing but* beliefs: a Southerner is "nothing but" a Southerner, et cetera. Even the most tolerant of us often hold such beliefs, unconsciously, as part of our common nonsense.

Like most pieces of common nonsense, stereotypes and preemptive constructs aren't completely figments of the imagination. Rather, they are usually distortions of what the stereotyping person, or someone else that person has talked to, has actually observed. Stereotypes often result from confusing *statistical* properties of a group with *individual* properties of a member of that group. Or else they may result from confusing *dynamic* properties of a person, which are true at some instant of time but subject to change, with *permanent* properties that are true all the time. The difference between dynamic and permanent categorizations will be developed a little later. We now discuss the difference between statistical and individual categorizations.

Probabilities and Certainties

We are always hearing statements about group differences in specific mental or physical abilities. In the case of gender differences, for example, we often throw around in casual conversation statements like "men have better spatial ability than women" and "women are more nurturing than men." Some gender differences in cognitive abilities have been verified in experimental studies as statistically significant.[13]

However, the usual quantitative definitions by scientists and social scientists of what is "statistically significant" still allow for a lot of overlap between groups and individual variation within groups. For example, Camilla Benbow and Julian Stanley did a study which made them conclude that males have more innate mathematical ability than females.[14] They gave the mathematics scholastic aptitude test (usually given to American high school students) to precocious seventh and eighth grade students. In one year of Benbow and Stanley's study, the average score was 436 for the boys and 404 for the girls, with 3.2 percent of the boys and 0.9 percent of the girls scoring over 600. Yet the highest scoring girl in their study made a score of 760, much higher than most of the boys' scores.

Benbow and Stanley wrote that they tested relatively young children in order to eliminate the possibility of the two sexes having had unequal formal mathematical training. However, it is well known that gender-specific socialization, via subtle signals from parents and other adults, begins at birth, and that could account for some or all of the difference. So to conclude from Benbow and Stanley's statistically significant differences that men *must be* genetically superior to women in mathematical ability is unwarranted.*

Not all stereotypes about gender differences favor men. Many social psychologists, such as Alice Eagly, Judith Hall, and Martin Hoffman, have found statistically significant gender differences in empathy or nonverbal communication, in which women tend to do better.[15] But it would be unwarranted to conclude from these findings that women must be genetically superior to men in empathic or intuitive ability (and Eagly, Hall, and Hoffman themselves didn't make that conclusion!) In fact, recent work in the laboratory of William Ickes suggests that women are only better than men at intuiting the feelings of others under some situations, and there is no significant gender difference in other situations.[16] Specifically, women perform better when told in advance that the test is one of their empathic abilities, because of their felt need (at least in American culture) to live up to the image of being skilled in this area. Also, women perform better at decoding emotions from facial expressions, but not from other types of nonverbal cues or from conversational cues.

Confusion of statistical with individual differences is a common nonsense that can lead to a pernicious conformity. This comes from a perception that individuals who differ from the average of their group are in some way abnormal for the group. The contributions to society by such "abnormal" individuals, many of whom are likely to be unusually creative, tend to be undervalued.

For example, the educational psychologist Carol Gilligan described an experiment in which college students enrolled in a psychology course were asked to write fictional stories about possible personal relationships. She found a sex difference in the amount of violence in these stories: "Of the eighty-eight men in the motivation class, 51 percent wrote at least one story containing images of violence, in comparison to 20 percent of the fifty women in the class ... ."[17] Gilligan's finding is statistically significant, though it doesn't address whether the root causes of the sex difference are hormonal or cultural. But it's possible

---

* Benbow and Stanley drew such a conclusion partly because they assumed that the SAT measures a "native aptitude" for mathematics that is largely separable from training. This is itself a controversial position; I disagree with that stance intuitively but haven't studied it deeply enough to form an educated opinion.

for the unwary to draw the wrong lesson from it. A reader could conclude that the 49 percent of men who didn't include violence were "unmasculine" or that the 20 percent of women who did were "unfeminine." Since college students are often anxious about whether they are normal, it wouldn't surprise me if at least one of those gender-atypical students saw the results of that study and felt anxious about his or her *own* gender identity!

Too much focus on statistical differences makes individuals who don't fit the average of their group invisible and often discourages them from pursuing areas of potential success. I believe this is one reason, for example, that African-American and Hispanic minorities in the United States have been slow to increase their numbers in many of the high-paying professions. The shortage of role models from the group they identify with makes them feel "abnormal" among other doctors, lawyers, or professors who are mostly not of their ethnic group. This feeling of abnormality deters many of them from continuing on their path to advancement, as well as discouraging friends from the same ethnic groups who hear about their experiences.

Dealing effectively with minority invisibility requires a change in general social attitudes. It demands that we try to reduce the common nonsense that tends to make people feel "different" if they differ from the average of a group in some superficial way. So if we want to overcome the resistance of prejudiced people to accepting minority members and the resistance of minority members themselves to making an effort to enter professions, society needs to improve its attitude toward "oddball cases" of all sorts. This can be helped by groups of "oddballs" banding together for mutual emotional support. An example is the networks of American and European female business executives and engineers that have grown in the last twenty years.

Inhibitions against engaging in group-atypical behavior are not confined to races or genders. They also affect the scientist who wants to speculate about afterlife and mental healing, or the seventy-year-old who wants to hike on the Pacific Crest Trail. The attitude behind these inhibitions boils down to George Kelly's idea of a preemptive construct.[18] That is to say, a person who is a scientist is *nothing but* a scientist, a person who is elderly is *nothing but* elderly.

If we want to design normative theories about how people *should* behave, though, we need to remember that preemptiveness sometimes is good. If you want to hire a person to put on a new roof, for example, it is common sense to consider the person as *nothing but* a good roofer (as long as he or she works as promised and doesn't overcharge you) and not worry about whether he or she is Jewish or Republican or wears dental braces.

What function of our brains makes us sometimes confuse statistical properties with individual properties? It is an offshoot of our mental perception of probabilities. For our adjusting to the real world, it is important to distinguish between frequent and rare events, and our brains do so. For example, a team of researchers on brain wave (EEG) patterns, Jean-Paul Banquet, Mark Smith, and Wilfried Guenther, reviewed some of their EEG data showing that brain wave patterns evoked by common and rare stimuli are significantly different.[19] Moreover, this difference in responses to common versus rare stimuli is less in people with some mental illnesses, such as depression, than in mentally healthy people.

To be inaccurate in acting on probabilities of events, like the depressed patients studied by Banquet and his colleagues, is a serious pathology. But there is an opposite pathology which is just as serious: to confuse *improbability* with *impossibility.* This leads to cynicism about improving society or human nature for the better, because only average or prevailing human tendencies are considered and not unusual people or rare moments of heroism by ordinary people. Such people, often ignored by the media, live out novel combinations which are the true common sense: family values without intolerance of minorities; life style creativity without "anything goes" amorality; ecological consciousness without rejection of technology; feminism without male-bashing. These are the people who can honestly say, in the words of Robert Frost:[20]

> Two roads diverged ... and I —
> I took the one less traveled by
> And that has made all the difference.

Tendencies are Not Unbreakable

Just as statistical tendencies of a group do not apply to every member of the group, statistical tendencies *within an individual* do not apply to that individual at all times and places. The compassionate view, the idea that "what you get is more than what you see," means that we should strive to see a person who is violent or cowardly *most of the time* as having the potential to act otherwise. This is a religious ideal: Jesus' approach to corrupt public officials and prostitutes, Martin Luther King's approach to white racists, even Luke Skywalker's approach to the dying Darth Vader, were based on seeing under evil behavior the potential for love and goodness that was largely unexpressed.

It is also a political and social ideal. Many people who have a negative personality trait can discover characteristic settings in which that trait disappears and may even be replaced by its opposite. This type of behavioral change with context has begun to be modeled in neural networks. We need to create such settings whenever possible, striving toward a "user-friendly" society based on bringing out the best talents in everyone rather than pitting people against each other to see who is superior.

Cortical Versus Limbic Categorizations?

Paul MacLean discussed the different information processing capacities of the limbic system (his "old mammalian" brain) and the cerebral cortex (his "new mammalian" brain).[21] In speaking of those two parts of the brain plus the "reptilian" brain, he states "the three basic brains might be regarded as biological computers, each with its own special form of subjectivity and intelligence, its own sense of time and space, and its own memory, motor, and other functions."[22] His statement of neural "separatism" contains some exaggeration, but his point that the limbic system processes information somewhat independently of the cortex is still valid. Specifically, limbic processing is more subjective and less influenced by semantic and rational organization.

MacLean attributed some of the most pernicious preemptive constructs, such as those that can lead to paranoia about other races, genders, and ethnic groups, to "limbic" information processing.[23] People are selectively categorized on the basis of those attributes which we are most familiar with or which engender strong emotions, whether pleasant or unpleasant. He said "a primitive system of the brain with an incapacity for verbal communication may generate feelings of conviction that we attach to our beliefs, regardless of whether they are true or false."[24]

But MacLean's picture is incomplete because *hardening* our emotions into beliefs actually requires semantic and rational capacities as well! For example, I can feel a visceral dislike for a person, but translating my raw feelings into a careful statement of "I hate so-and-so *because* he or she is Peruvian, or a firefighter, or has a long nose" is an act of reason (even if used in a harmful way). It probably involves the orbital part of the frontal lobes, the area of the brain which integrates reason and emotion to form rules about which categories of events are pleasant or painful.

This example reminds us that reason and emotion both perform necessary and valuable functions, and both can be used (separately and together) for helpful or harmful purposes. The belief that one set of functions, or one brain area, is "superior" to another is common nonsense. After all, there are at least two forms of "acting emotionally": yielding impulsively to the emotions of the moment, or planning on the basis of long-term emotional needs. Calling the long-term planning capacity "new mammalian" is simplistic, because many reptiles can, for example, forage for food at times when they don't currently feel hungry but anticipate future hunger. But distinguishing the common sense of long-term emotional planning from the common nonsense of running with current emotions is vitally important. This is why emotion needs to be complemented by reason* (and the reverse), why the limbic brain was given what MacLean calls a "thinking cap."

So if we arrive at a good neural network model for categorization, we might conjecture that copies or variants of such a network are repeated in many areas of real brains. These areas include both the limbic system and many parts of the cerebral cortex. One of the more versatile categorization networks so far developed is the two-layer *adaptive resonance theory* (ART) network developed by Gail Carpenter and Stephen Grossberg.[25] In the ART network, nodes at one layer encode patterns of inputs to another layer.**

---

* Where does intuition fit into this puzzle? Acting on deep intuitive understanding is not the same as reacting to raw emotion. I believe that intuition is based on abstracting from incomplete information. This is usually coherent information from multiple sources, some of them visceral, some of them affective, and some of them semantic — forming into some kind of neural "laser beam." So intuition involves *all* of MacLean's "three brains."

** The actual structure of ART is shown at (link to ART Networks)

These patterns become prototypes for particular concepts and can be updated with experience; for example, your bird prototype may start out looking like a sparrow if that's the most common bird you see, but then after you see more robins, the prototype could become some average of a sparrow and a robin.

At the cerebral cortex, versions of ART or some similar network design could occur in the areas of cortex encoding a single sensory modality — sight, hearing, smell, or touch. Other versions of ART could occur in the cortical *association areas* (temporal, parietal, and frontal lobes) that integrate information from different modalities. In the association cortex, we expect the categorizations to be the most sophisticated because they involve information from many other cortical areas and, in some cases, the limbic system as well. For this reason, the networks in the association cortex should also be the most vulnerable to malfunction. In living animals, network malfunction can arise from neural tissue degeneration, or else from changes in the amounts of chemical transmitter substances at synapses between neurons.

A Tentative Neurochemical Theory

Enough knowledge has recently been obtained, and speculation indulged in, to suggest a very tentative theory for how chemical neurotransmission might be involved in categorization. This theory will also suggest which brain processes might differ in stereotyped categorization versus flexible categorization.

Several cautionary notes need to be added before I introduce my theory. First, the brain pathways discussed herein are far from completely worked out. Much experimental work still lies ahead regarding which of the known or unknown chemical transmitter substances are used at particular synapses, whether these synapses excite or inhibit electrical activity of the neurons to which they lead, and what other surrounding chemicals might influence their functioning. Second, due to the complexity of the neurochemical pathways involved, the same transmitter substance can have different effects on different regions of the brain. The clinical neuroscientists Daniel Weinberger, K. H. Berman, and B. P. Illowsky, for instance, found evidence suggesting that the effects of the transmitter dopamine in the limbic system might *inhibit* effects of the very same transmitter on the frontal lobes.[26] Third, finding a biochemical basis for a type of behavior doesn't imply that environmental and context effects on that behavior should be ignored. On the contrary, such an insight is a step toward understanding *how* the biochemistry of that behavior is influenced by interpersonal interactions. With all these caveats, I believe it is useful to start from somewhere to make sense out of the tremendous complexity of both the biochemical and cognitive mechanisms involved. If the theory itself can't be tested experimentally, it can at least be a step toward better theories that are testable.

My theory of categorization (see Figure 5.3) involves balance between two of the commonest neurotransmitter substances: *acetylcholine* and *norepinephrine* (also called *noradrenalin*). Acetylcholine (ACh) was the first neurotransmitter to be positively identified in the laboratory. Otto Loewi discovered in 1921 that it is the substance secreted by a nerve going from the brain stem to the heart (the *vagus* nerve) and its release slows down the heart rate.[27] Soon thereafter, this same substance was found at junctions between nerves and muscles, and finally within the brain itself. In particular, ACh is released by several brain regions below the cortex, including a part of the midbrain (see Figure 1.1) called the *nucleus basalis*, and broadcast out to wide areas of the cortex and limbic system. So most scientists believe it has some role in modulating the basic cortical and limbic processing and memory storage operations. A hint for what that modulation might entail comes from work of Michael Hasselmo and James Bower.[28] These neurobiologists found that at least in one area of the cortex (the olfactory area), acetylcholine inhibits the strength of connections *within* that part of cortex, but has no effect on connections *to* that part of cortex from sensory processing areas. In cognitive terms, this means that acetylcholine tends to suppress recall of associations between previously stored memories, thereby allowing new memories to form from new outside inputs.

So a deficit in cortical ACh leads to mental "hardening of the arteries," not allowing new associations to form easily. One condition associated with too little ACh is Alzheimer's disease. Alzheimer dementia is progressive over many years and ultimately impairs all forms of memory, but in its mild stages it impairs new memory formation sooner than it impairs memory for earlier events.[29] Could a similar deficit — even if much less drastic than Alzheimer's and reversible through most of life — underlie tendencies to stay with emotional connections that were previously formed

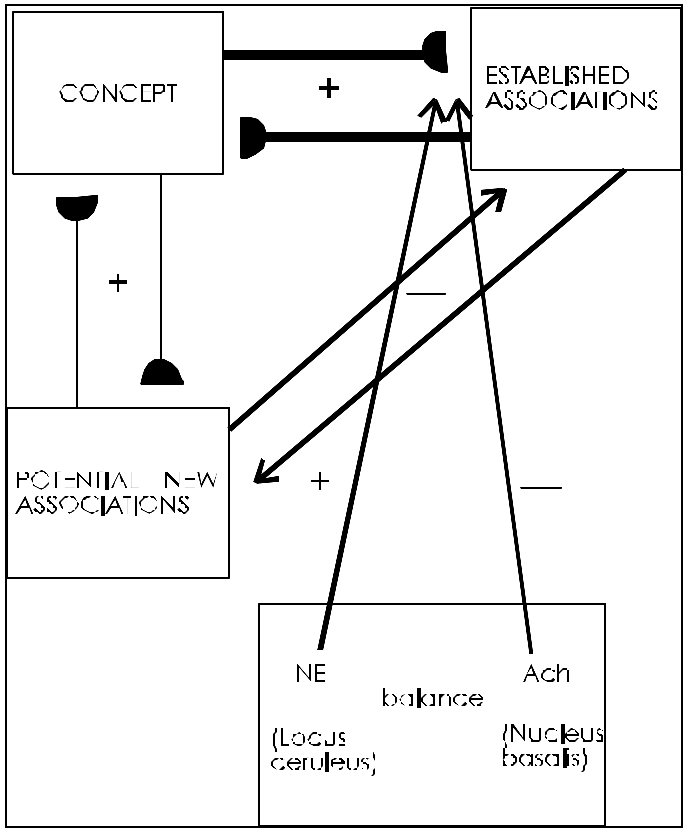

Figure 5.3. A speculative theory for the complementary roles of the two transmitters acetylcholine (ACh) and norepinephrine (NE). Arrows going to semicircles indicate that incoming electric signals modulate transmission at the receiving synapse. NE excites, whereas ACh inhibits, synapses corresponding to previously made associations. Representations of input forming new associations compete with those of inputs forming established associations.

even if they are no longer appropriate? Forming preemptive constructs when inappropriate may be an example of this, or at least closely related. In this case, the connections to cortex that ACh liberates are not from representations of outside inputs, as in Hasselmo and Bower's work, but from representations elsewhere in the brain of current emotional states. Perhaps the perseveration of people with frontal lobe damage on the Wisconsin Card Sorting Test (see Figure 4.1) is also implemented through the ACh system, say, via the "bias nodes" of that figure.

Such a lower than optimal level of ACh can sometimes be accompanied by a higher than optimal level of norepinephrine (NE). NE is also broadcast to the cortex and limbic system from a different midbrain area (called the *locus ceruleus*). This transmitter tends to be associated with a state of arousal, and deficits in this transmitter are associated with certain types of depression.[30] The novelist Aldous Huxley noted that release of norepinephrine, and the chemically related hormone adrenalin, is an effect, sometimes a pleasurable one, of intense states of anger. Huxley speculated that this kind of chemically induced arousal is one of the attractions of mass popular movements that follow dictatorial leaders or go off unthinkingly to war.

How might excess NE affect category and stereotype formation? The arousing effects of NE led the neural network theorist David Hestenes to suggest that this transmitter enhances responses to emotionally significant stimuli.[31] Consequently, NE could play a role in perseveration as well, by increasing the tendency to continue responding to attributes that have become emotionally significant, whether those attributes are relevant or irrelevant to current needs. For example, if an employer selecting among job applicants tends to attach emotional significance to a person's skin color, excess NE could enhance the skin color response even though it's inappropriate for applicant selection.

Huxley's insight also suggests that one inducement to form negative stereotypes about groups is the feeling of arousal that accompanies hate for a group. So if we want to overcome harmful stereotypes, we need to provide substitute norepinephrine releasers. The passion aroused by racial, gender, religious, ethnic, or professional hatred can't be suppressed by mild reason alone. It must be countered by generating a substitute passion for something socially desirable. This is what the philosopher William James meant by the "moral equivalent of war."[32] Some recent American Presidents, notably John F. Kennedy and Jimmy Carter, have tried to encourage such a passion for social service among the general public.[33] To generate the will to solve our massive global problems, we need this kind of passionate appeal not only from politicians, but also from religious leaders, psychotherapists, fiction and nonfiction writers, academics, and leaders in every other walk of life. I hope this book will do its part.

Another likely substitute norepinephrine releaser is the pursuit of pleasure. In modern Western culture, pleasure has a bad press, being associated with impersonal sex, for example, or abuse of ecstasy-inducing drugs. But shifting away from the common nonsense of intergroup hatred or violence requires that we adopt a more positive attitude toward all kinds of pleasure, including a variety of sexual arrangements as long as they don't harm the people involved. In fact Riane Eisler,[34] Marilyn French,[35] and many other contemporary writers (particularly feminist writers) advocate shifting toward pleasure or delight as a major organizing principle of society, rather than control or duty. At the end of this book, I will return to this theme in outlining a desirable future society.

In this chapter we have outlined the neural process of categorization. Now we will build on this and suggest some ways in which a neural network for categorization can be made dynamic. In such a network, different attributes of a pattern to be categorized are significant at different times, and the significant attributes are based on changes in "mood." One of these network architectures was originally developed to model some behavioral data on consumer preference (for soft drinks) that changed from controlled taste tests to an actual market. The dynamics of such a network are likely to have implications for how both categorizations and preferences can be shaped, by changes in either external environments or internal beliefs. This in turn has likely implications for education, psychotherapy, management theory, advertising, and social policy among many other pursuits.

REFERENCES

Chapter 5: How We Make Categories and Constructs

1 Burtt, 1955, 162.

2 Ibid, 163.

3 Edelman, 1989.

4 The clinical neuroscientist Paul Eslinger, in an April, 1997, lecture at Texas Christian University, related the “executive” functions of the brain’s frontal lobes to psychosocial skills. The executive function of cognitive flexibility he related to the social skill of empathy. See also Grattan, Bloomer, Archambault, and Eslinger, 1994, and Eslinger, 1998.

5 Mervis and Rosch, 1981.

6 Posner and Keele, 1970.

7 Anderson and Murphy, 1986; Carpenter and Grossberg, 1987; Rumelhart, Hinton, and Williams, 1986.

8 Kelly, 1955.

9 Ibid., 42-43.

10 MacLean, 1970.

11 Gaffan, 1992; LeDoux, 1991.

12 Grossberg, 1972a, 1972b.

13 Bleier, 1984; Halpern, 1992.

14 Benbow and Stanley, 1980.

15 Eagly, 1987; Hall, 1984; Hoffman, 19??.

16 Graham and Ickes, 1996.

17 Gilligan, 1982, 41.

18 Kelly, 1955.

REFERENCES

19 Banquet, Smith, and Guenther, 1992.

20 Frost, 1936.

21 MacLean, 1969, 1970.

22 MacLean, 1970, 339.

23 MacLean, 1969.

24 MacLean, 1987, 1237.

25 Carpenter and Grossberg, 1987.

26 Weinberger, Berman, and Illowsky, 1988.

27 Loewi, 1921.

28 Hasselmo and Bower, 1992.

29 Zec, 1993.

30 Samson et al., 1992.

31 Hestenes, 1992.

32 James, 1910/1984.

33 Carter, 1977; Kennedy, 1961.

34 Eisler, 1987, 1995.

35 French, 1985.