MIT/LCS/TM-57

ON BATESON'S LOGICAL LEVELS OF LEARNING THEORY

Michael Levin

February 1975

TM-57

On Bateson's Logical Levels of Learning Theory

By Michael Levin

February 1975

Project MAC - M.I.T.

Preface

In the summer of 1974, I was fortunate to attend a class given by Gregory Bateson at the Naropa Institute. This course was a general introduction to Bateson's ideas and general way of thinking, and the readings were from his book of collected papers entitiled Steps toward an Ecology of Mind, [Bateson, 1972]. In Bateson's class, I experienced a profound reawakening of my interest in the natural and social sciences, and their relevance to our present situation.

We live in a time when intellect seems to be responsible for the rapid intensification of horrible conditions that degrade the quality of life, and threaten to destroy the human species, and perhaps much of the biosphere of our planet along with it. What excited me about Bateson's approach was that it offers a view that is intellectually sound and scientific rather than romantic or anti-intellectual, and yet leads toward a balance or ecology of mind, rather than an intensification of the war between the familiar dualities: man vs. nature, self vs. other, conscious vs unconscious, life vs. death, subject vs. object. Quoting from a public lecture:

> The cybernetic epistemology which I have offered you would suggest a new approach. The individual mind is immanent but not only in the body. It is immanent also in pathways and messages outside the body; and there is a larger Mind of which the individual mind is only a subsystem. This larger Mind is comparable to God and is perhaps what some people mean by "God," but it is still immanent in the total interconnected social system and planetary ecology.
>
> Freudian psychology expanded the concept of mind inwards to include the whole communication system within the body--the autonomic, the habitual, and the vast range of unconscious process. What I am saying expands mind outwards. And both of these changes reduce the scope of the conscious self. A certain humility becomes appropriate, tempered by the dignity or joy of being part of something much bigger. A part--if you will--of God.
>
> If you put God outside and set him vis-à-vis his creation and if you have the idea that you are created in his image, you will logically and naturally see yourself as outside and against things around you. And as you arrogate all mind to yourself, you will see the world around you as mindless and therefore not entitled to moral or ethical consideration. The environment will seem to be yours to exploit. You survival unit will be you and your folks or conspecifics against the environment of other social units, other races and the brutes and vegetables.
>
> If this is your estimate of your relation to nature and you have an advanced technology, your likelihood of survival will be that of a snowball in hell. You will die either of the toxic by-products of your own hate, or, simply, of overpopulation and overgrazing. The raw materials of the world are finite. [Bateson, 1972, pp. 461-462]

There is nothing especially new in this message. It can be found in most of the spiritual and prophetic traditions, and has been expressed by many poets, artists, psychologists, naturalists, ecologists and other scientists. To me it is the starting point for Buddhist meditation. What is unusual is to find such a view expressed in a scientific language without a moral attitude imposed from outside of it, and to have that science be information theory and cybernetics, the very cutting edge of current technological change in the world.

Coming home after that summer, I tried to communicate my enthusiasm to friends, some of whom are expert computer theoreticians. This was difficult, and at time I wondered if I had merely been infected with the ambiance of what Bateson was saying without having acquired sufficient understanding to make any use of it. Attempting to formulate for myself, intellectually, what it was that Bateson was about has proven to be extraordinarily difficult. At time I wondered if I had been swept away by a style without scientific substance. Yet I do not think that this is so, for despite my intellectual limitations, the perceptual attitude implied by Bateson's ideas has sunk deeper into my mind as a better way of making sense of the psychological and natural world.

This essay is my attempt to work with this material in a personal way. The subject matter is Bateson's model for learning theory set forth in an article entitled "The Logical Categories of Learning Theory" [Bateson, 1972, pp. 279-308]. After describing the model, which is quite fascinating in itself, I ask the question: "Why, starting with certain basic concepts, do we end up with this particular model rather than with some other one?" This question motivates the rest of th essay. Notice that is is an attempt to deal intellectually with certain ideas, not an attempt to fit a theory to some data. It is addressed to myself, Gregory Bateson, and to any interested reader. I have attempted to make it self-contained, although it would be useful to have a copy of Steps to an Ecology of Mind on hand. My collegues at Project MAC and the Artificial Intelligence Laboratory will have to be patient with or skip those explanations of familiar concepts that are too elementary for them, but I think that what I have to say is of interest to them also.

Michael Levin

Cambridge, Massachusetts
January, 1975

I Levels of Learning

The model proposes that learning in man and animals can occur on several different hierarchically arranged levels, each of which is on a higher order of generality that the preceding because it is outside the scope of the previous level, and in fact is learning how to learn on that previous level. The terms Zero learning, Learning I, Learning II and Learning III are used to name these levels the description of which follows:

Zero learning or non-learning is a response to information which is the same each time the information is presented. For example suppose that a factory worker learns that is is lunch time each time the noon whistle blows, so he stops working and eats his lunch. The same thing happens each time the whistle blows so there is no change in his response. It is somewhat questionable to call this "learning" at all, and therefore the term Zero Learning seems appropriate. (The first few days on the job, the worker learns that the whistle means lunch and acquires this habit. This is Learning I. What we are talking about now, however, is his doing the same thing every day after the habit has been established.) The primary purpose of mentioning Zero Learning is to catalogue some situations such that if they were to change, this would be evidence for some sort of higher learning.

Learning I is the category that includes most of the learning that is studied in animal experiments, most of the learning of specific academic skills, as distinct from what is learned by being in an academic situation, most manual skills such as learning to walk or drive a car, and most of the responses that must be made in all sorts of situations in order to function. Learning I is about something other than learning itself. Its content is explicit enough to be readily described by an observer, even if the expereince itself is non-verbal for the learner.

In the laboratory the classical conditioning experiments of Pavlov are examples of Learning I. A dog learns, after a certain number of trials, that when the bell rings in Pavlov's laboratory, this is followed by food. The dog anticipates the food by salivating, which it did not do before learning to associate the bell with food. Another example is Skinner's experiments with pigeons. In one of these experiments, a pigeon learns to peck at a button every time it wants a grain of food. We shall return to these examples later.

It seems to be characteristic of Learning I that it has a certain objectivity in that if a lesson is not learned correctly there is abundant feedback to make the organism learn about and correct its mistake. Humans and animals, except for those who are obviously injured or deficient, usually seem to have enough Learning I ability to handle their immediate situations.

If one may make a gross generalization, it is not at level I that the problems of the world seem to lie. For example, if a child does not learn to read, one first determines if the child is brain damaged or mentally deficient in which case there is very little that can be done. The child simply cannot master this Learning I situation. But in the majority of cases, we find a non-reader of normal or superior intelligence, and it is not easy to determine what the problem is. The subject is puzzling and full of controversy. It will not do to examine merely the mechanics of the reading lessons. A much larger experimental and emotional context must be considered, and the situation is no longer objective.

Learning II is learning about the contexts in which Learning I occurs. It is learning how to perceive the world, how to form meaningful gestalts from the background of sensory information, what to pay attention to. Learning II is the process of forming habitual value judgments. It is the source of what psychologists call character. Learning II is learning to be a particular personality, and therefore is not objective learning. The lessons of Learning II are not objective in the sense of being learned correctly or incorrectly and therefore not subject to simple revision by failure to coincide with external reality as seems to be the case with Learning I.

As an example of Learning II, let us consider in more detail the laboratory experiment mentioned earlier. Pavlov had placed a dog in a very special context which just happens to make the dog look damn stupid. It takes many trials before the dog learns to associate the bell with the food. Now anyone who is familiar with pets knows that they are much smarter than this. If you feed a stray once, it adopts you. It seems that a dog is not initially disposed to paying attention to bells and attributing meaning to them. Now suppose that a dog were put through a series of experiments, each of which was structured so as to be a learning of classical conditioning, i.e. a signal is given and then something happens which is already meaningful to the dog, and to which the dog already has a reaction. If the dog, generalizing from one experiment to the next, increased its rate of learning i.e., it learns a particular experiment faster than a naive dog would, this would be an example of Learning II. If the dog could verbalize its learning, this might be: "In this laboratory certain types of signals indicate that something is about to happen that concerns me directly. I can learn to be aware of these signals and make note of what follows them."

The character that the dog has learned could be called "fatalism" in some crude sense. There are people whose learning is such that they are always looking for premonitions or omens that portend some coming event over which they have no control except to prepare themselves for it. This is a very passive character. The pigeon in Skinner's experiment is likely to learn quite a different lesson on level II. This laboratory situation is called instrumental reward and involves much more active participation; in particular it involves exploring in order to learn what button is worth pushing. An equivalent human character would ask: "How can I make this situation work for me?" It has generally been recognized that effective learning involves asking the right question, and the question formulation (verbal or non-verbal) seems to lie on level II.

Learning II is characterized by a high degree of stabilization compared to Learning I. Learning II is frequently self-reinforcing, and it is generally not corrected by the environment because of its subjective nature. It is Learning II that makes the world of human affairs interesting, as well as providing most of the problems or knots (to borrow a word from R. D. Laing). It accounts for diversity, and explains why different individuals follow such different paths in their development, instead of becoming pretty much honogenized by living in the same culture, and it explains pathologoies of learning, that is, learning situations which are unstable, self-reinforcing, and if not interupted somehow, lead to disruption of the individual by insanity, illness or death.

It should be mentioned at this point that although this model is specifically about the learning that is done by an individual organism, it is closely related to other processes of adaptive change which can also be

regarded as forms of learning. Wherever there is life adapting creatively to emergingsituations, something closely analogous to individual learning occurs. The learning of a species is encoded in its genes, and the learning of a human culture is encoded in its preserved and maintained patterns of belief and relationship. One can even view an entire ecological system as learning to develop diversity, stability and balance.

We come now to Learning III about which the least can be said, although a lot may be speculated. Beacuse Learning II is non-objective and self-reinforcing and because it is largely unconscious, Learning III which is learning about the contexts of Learning II is likely to be extremely difficult. Bateson suggests that the concept of "self" is largely a collection of unexamined and contradictory pieces of Learning II. To go about examining these is dangerous business because it undermines the sense of self-existence, and generates tremendous fear and panic. Only a profound discontent with one's notion of self, combined with flashes of doubt as to its soiidity can motivate Learning III.

> When a man starts to learn, he is never clear about his objectives. His purpose is faulty; his intent is vague. He hopes for rewards that will never naterialize, for he knows nothing of the hardships of learning.
>
> He slowly begins to learn--bit by bit at first, then in chunks. And his thoughts soon clash. What he learns is never what he pictured, or imagined, and so he begins to be afraid. Learning is never what one expects. Every step of learning is a new task, and the fear the man is experiencing begins to mount mercilessly, unyielding. His purpose becomes a battlefield.
>
> And thus he has stumbled upon the first of his natural enemies: Fear! A terrible enemy--treacherous, and difficult to overcome. It remains concealed at every turn of the way, prowling, waiting. And if the man, terrified in its presence, runs away, his enemy will have put an end to his quest.
> [Castenada, 1968, p. 79]

Bateson describes Learning III as a deep reorganization of character that sometimes and unpredictably occurs in psychotherapy and spiritual disciplines, although this may be making too much of a thing of it already, since certain individuals seem to come by a bit of Learning III without benefit of any spiritual tradition, although not without experiencing a good deal of suffering.

> Zen Buddhists, Occidental mystics, and some psychiatrists assert that these matters are totally beyond the reach of language. But, in spite of this warning, let me begin to speculate about what must (logically) be the case.
> [Bateson, 1972, pp. 301-302]

The preceding is a brief summary of the fruit of a certain method of thinking brought to bear on a very general problem. It is certainly not the result of classifying certain phenomena and observing that a hierarchical structure is supported by the data, although Bateson has had considerable experience observing sequences of human behavior in New Guinea, Bali, and Palo Alto among other places. What we now have to do is to examine the concepts used as premises for this way of thinking, and then see if we can follow the steps that lead to this result, so that we too can say that this

"must logically be the case." We shall keep in mind the question: Why this particular model rather than some other one? The next few sections deal with these concepts.

In observing the sequence I, II, and III just described, we notice a progression from a level of affairs taken for granted by most people and quantified over by the experimental psychologist, to the realm of human affairs generally found to be problematic and admitting much controversy and confusion, to the realm of mystical speculation. A study of these levels of learning may explain some of the confusion among the very different sorts of professional activity in the social sciences and social services. One of the key ideas to be discussed later is that confusion of levels of learning, or logical levels in general, is bad epistemology, and leads to trouble.

The experimental psychologist working in a laboratory is generally dealing with Level I phenomena. If he recognizes the limited context of his experimental results, he may simply respect wisdom and judgment in human affairs as being outside of his realm. He may also feel some contempt for the state of confusion of psychiatric theory as compared to the hard data that he works with because he lacks insight into why these affairs are so subjective and slippery. The therapist, on the other hand, may find experimental results irrelevant because he just cannot relate them to the problems that his clients bring to him. For lack of appreciation of the hierarchical structure of learning, he may ignore the known facts of learning theory and not consider them as a possible base for his own understanding of behavior. The worst mistake is to pretend that everything is workable with only one one level, and to suppose that one can deal with linguistics and social management as if they could be explained directly in terms of conditioning, instrumental reward and punishment etc. This is not a criticism of B. F. Skinner alone, although he stands as a conspicuous, extreme and brave example. Most of economics and political science has always made precisely this mistake.

II Learning as Change - The Metaphor of Motion

The first concept that we consider is to regard learning as change and compare it with physical motion. In Newtonian physics, we have the terminology (i) position, (ii) velocity, (iii) acceleration, and (iv) jerk. Starting with the first term, position of an object in space as given, each term after this is the derivative or rate of change of the preceding one. If you have not studied calculus, there is still no problem in understanding what had just been said. If we use the example of a car travelling along a straight and level highway, then the odometer tells you its position (especially if you have a trip odometer that can be set to zero at the beginning of the road), and the speedometer tells you the velocity, or the rate at which the position is changing. If the car is standing still, the speedometer reads zero. Acceleration is the rate of change of velocity. If the car, starting up from a dead stop acheives a velocity of 60 mph in 10 seconds, it is accelerating. If it then proceeds for a while maintaining 60 mph, there is no acceleration, although there is velocity. If the brakes are then applied, there is negative acceleration or deceleration. The term "jerk" applied to change in acceleration is fairly obvious.

To apply the metaphor to learning theory, we have to see learning as change. The person (or animal or species or culture or environment) who has learned is different from what it was before it learned. In the classical experiment of Pavlov, a dog initially does not salivate when a bell is rung, but after it has learned to associate the bell with food, then it does salivate when the bell is rung. At a certain time the dog responds to a certain stimulus (the ringing of a bell) by not salivating, and on a later occasion it responds to the same stimulus by salivating. The assumption of repeatability or sameness of the two stimuli is the critical point here. Since no two moments in time are ever exactly the same, we have to examine the validity of this "sameness". What the experimenter meant is that all conditions that he had considered relevant to the experiment has been rendered the same to the best of his ability. The dog was placed in the same familiar harness each time, the dog was hungry each time, the experimenter repeated his own motions as nearly as he could etc. The idea was to reduce all observable differences except for the one difference of the dog having experienced previous instances of this same routine. What is crucial here is the entire context in which the bell was rung, and at this essential point, I quote Bateson:

> The conventional assumption that context can be repeated, at least in some cases, is one which the writer adopts in this essay as a cornerstone of the thesis that the study of behavior must be ordered according to the Theory of Logical Types. Without the assumption of repeatable context (and the hypothesis that for the organisms which we study the sequence of experience is really somehow punctuated in this manner), it would follow that all "learning" would be of one type: namely, all would be zero learning. Of the Pavlovian experiment, we would simply say that the dog's neural circuits contain "soldered in" from the beginning such characteristics that in Context A at Time 1 he will not salivate, and that in the totally different Context B at Time 2 he will salivate. What previously we called "learning" we would now describe as "discrimination" between the events of Time 1 and the events of Time 1 plus Time 2. It would then follow logically that all questions of the type, "Is this behavior 'learned' or 'innate'?" should be answered in favor of genetics.
>
> We would also argue that without the assumption of repeatable context, our thesis falls to the ground, together with the whole general concept of "learning." If, on the other hand, the assumption of repeatable context is accepted as somehow true of the organisms which we study, then the case for logical typing of the phenomena of learning necessarily stands, because the notion "context" is itself subject to logical typing. [Bateson, 1972, pp. 288-289]

We continue to apply the metaphor of physical motion by noting that since Learning II is learning about Learning I, it implies a change, very likely a speed-up of Learning I, and possibly a change in the direction or quality of Learning I. So it is appropriate to compare Learning II with acceleration.

The term "jerk" applied to a change in acceleration is not a part of the classical vocabulary of physics, but its use has become more accepted at least since the age of jet aircraft. One hears of experiences such as

"sudden enlightenment" in Zen that suggest the appropriateness of comparing Learning III to jerk.

At this point, I would like to pose some false starts, because they may suggest the necessity of the ideas that follow. The question remains: Why this hierarchy?

(i) If this sort of study permits us to understand Learning III from a theoretical viewpoint, why are we not then involved in Learning IV?

(ii) More realistically, if our studies allow us to understand Learning II, then why is this not Learning III?

(iii) An anthropologist is doing his field work which consists of living next to people of a foreign culture and studying its patterns. Let us call this Learning I. Suppose that he is able to learn efficiently because in the classroom work that preceded the field work he learned how to go about this. Since his courses taught him how to go about his own learning, would they be Learning II?

(iv) Bateson relates [Bateson, 1972, pp. 276-277] the story of a porpoise who learned to "do something different each time." This he regards as Learning II because the idea of "something different" could not be learned in any one lesson, but had to be inferred from a sequence of training incidents. But if you ask a person to do "something different" he can oblige you promptly if he pleases. Why does Learning II become merely fufilling a request in the context of verbal communication? How does learning differ in verbal and non-verbal situations?

III Logical Types or Why Hierarchies?

> "...Just look down the road, and tell me if you can see either of them"
>
> "I see nobody on the road," said Alice.
>
> "I only wish I had such eyes," the King remarked in a fretful tone. "To be able to see Nobody! And at that distance too! Why, it's as much as I can do to see real people, by this light!"
>
> . . .
>
> "Who did you pass on the road?" The King went on, holding out his hand to the Messenger for some more hay.
>
> "Nobody," said the Messenger.
>
> "Quite right," said the King; "this young lady say him too. So, of course, Nobody walks slower than you."
>
> --Through the Looking Glass--
> --Lewis Carroll--

Language and logic always invoke the danger of paradox, and all formal attempts to have a language or logic free of paradox result in some sort of hierarchical typing structure. The one that Bateson continually refers to is Bertrand Russell's theory of logical types, which has had a profound effect on logic within the last century. [Frankel et.al., 1973]

When the human species made the jump to linguistic communication, the power of speech was increased enormously but the potential for paradox was introduced. Animal speech consists mainly in the use of signals with fixed meaning.[1] The range of human speech is increased enormously by (i) the use of modes and particularly negation, and (ii) the ability to talk about speech. Compare the ease with which one can pantomime "He Jumps" with the difficulty in pantomiming "He does not jump" or "She says that he jumps." Not jumping is not the same thing as sitting or standing. It is not any particular physical act. Linguistically, it is a step removed from physical description. Also, note that "She says that he jumps" is not of the same order as "She uttered certain words." The sentence refers to the meaning of a certain speech not the act of speaking. Putting these two linguistic capabilities together, we can some up with the sentence "Everything I say including this statement is not true." This is known as Epimenides paradox.

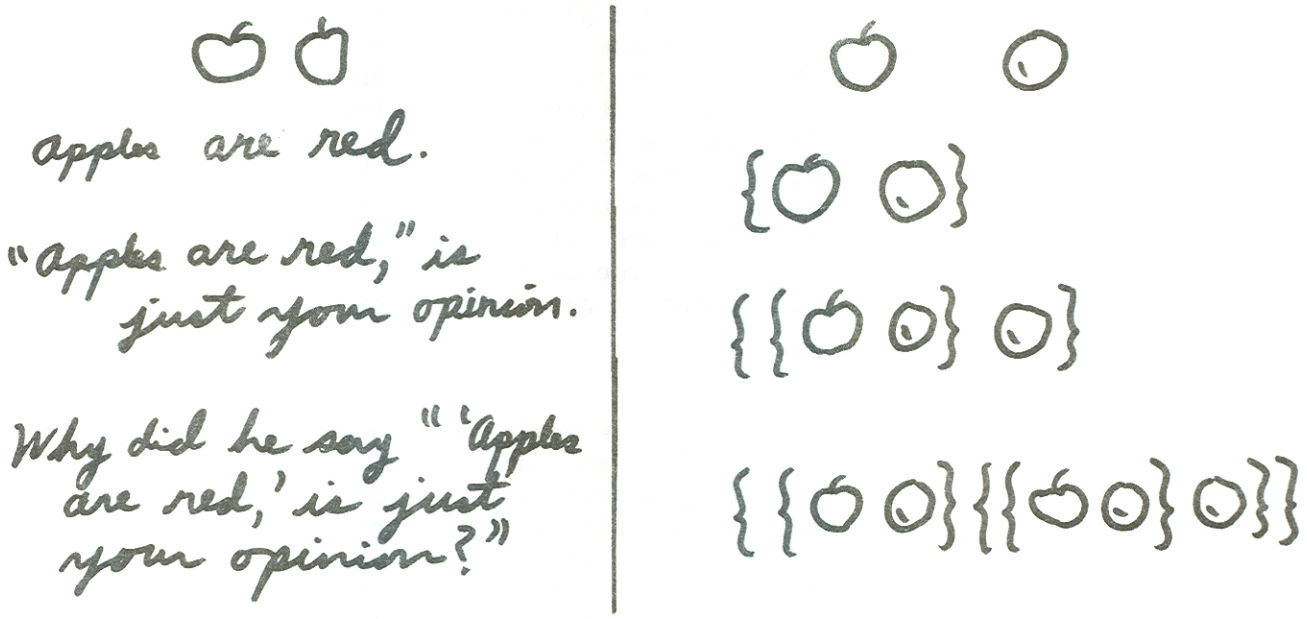
It has been stated that there are basically two types of paradoxes, the linguistic of which Epimenides' is an example, and the set theoretic which is exemplified by Russell's paradox: Any collection of physical or abstract entities that we care to consider as being itself an entity is a "class". Now consider the class of all classes that are not members of themselves. Call this class S. If S is a member of itself, then it is not a member of S because we said that S contained only classes that are not members of themselves and nothing else. Therefore S is not a member of S, but in this case, it is a class which is not a member of itself, and therefore it is a member of S."

If we decide that the abstract entities that mathematicians study are linguistic constructs rather than pre-existing things, then there is really only one kind of paradox. Note that the common elements of both paradoxes (which were proposed over an interval exceeding two millennia) are (i) negation as in "not true" and "not a member" and (ii) self reference as in "including this statement" and "member of itself." A language that does not have negation, or does not have the ability of statements to be about statements is an impoverished language. On the other hand, negation combined with statements about themselves leads to paradox. The standard and only known formal solution to this problem is hierarchy. If level I statements can only refer to non-linguistic facts (i.e. the "real" world), if level II statements can refer only to non-linguistic facts and level one statements, and if level III statements can only refer to non-linguistic facts and levels I and II statements, etc., then no statements in the hierarchy can be self-referent. Similarly with classes: if level I classes contain only things that are not classes such as apples or human interactions, and level II classes contain only things that

---

[1] Bateson insists that animal communication is always abstract because it expresses relationship such as dependency or hostility rather than being able to name a specific object such as "glass of milk." To me, this seems to miss the point that the animal speech is very specific in that it refers only to the immediate situation, and only from the point of view of the subject. The abstractness of human verbal communication is that it can be at any place, at any time, from any point of view, and express many modalities such as truthfulness, intention, probability, contingency, desirability, necessity and appropriateness. Much of the frustration of speaking in a foreign language that you only know slightly is knowing how to talk only about things and situations rather than about meanings. The writings of Benjamin Whorf on this subject are interesting.

are not classes and level one classes etc., then no class can be a member of itself, and the means of constructing Russell's paradox is blocked.



It should be admitted at this point that although the use of hierarchical ordering is the only feasible means to avoid paradox, there is no particular agreement among philosophers, logicians or linguistics as to whether this state of affairs is natural, desirable, self-evident, expedient, as-hoc, etc. Mathematical logicians spend a considerable portion of their time wi' problems of consistency and paradox. Many of the important theorems of logic are only a hairs-breath away from paradox. Yet it is generally felt that while paradox is within the logicians's department, it is something that the applied scientist need never worry about. A more realistic attitude toward paradox is to recognize that human language, unlike formal systems, does not generally adhere to strict logical typing, and for good reasons.

> But I do not believe that consistency is necessary or even desirable in a developing intelligent system. No one is ever completely consistent. What is important is how one handles paradox or conflict, how one learns from mistakes, how one turns aside from suspected inconsistencies.
>
> A famous matematitician, warned that his proof would lead to a paradox if he took one more logical step, replied "Ah, but I shall not take that step." He was completely serious. A large part of ordinary (or even mathematical) knowledge resembles that in dangerous professions: when are certain actions unwise. When are certain approximations sage to use? When do various measures yield sensible estimates? Which self-referent statements are permissable if not carried too far?
> [Minsky, pp 76-77]

This brings us out of the neat world of the logician into a world of real risks. Bateson has taken this matter one step further by seriously suggesting not only that paradox is a necessary risk and leaving it at that, but that paradox is the essential mathematical model for the pathologies of mind (individual, societal, biological, ecological), as well as for humor.

Of course in a behavioral context what we shall see is not a logical paradox but inappropriate and self-reinforcing behavior.

The importance of correct logical typing in the behavioral sciences is illustrated by Bateson's example of exploratory behavior in rats. [Bateson, 1972, pp 281-282] Approaching or avoiding a specific object is an item of rat behavior, while exploratory behavior is a class of such items, and therefore one logical type higher. Experimenters have worked out formulas for predicting the way a rat will approach or avoid an object under certain conditions of reward or punishment. But these formulas do not predict correctly how exploratory the rat will be. They deal with Learning I where the message might be: "If I touch this thing, then I get a shock." but not Learning II where the message based on a sequence of such experiences may no longer be negative: "I am able to discover by approaching them which things give me shocks."

Now consider how this distinction is obliterated in the usually imprecise discussion over "strict" and "permissive" upbringing of children. If we confuse levels, and merely associate "punishment" with "strict", and "reward" with "permissive", and the distinction between strict and permissive as merely a summation of how much reward and how much punishment, then we miss the point that either a strict or a permissive sequence may generate either the level II lesson of consistent meaning resulting in self-confident autonomous behavior, or the second level lesson of arbitrariness resulting in confusion and despair. A famous consequence of this sort of theorizing is Bateson's double bind theory of schizophrenia [Bateson, 1972, pp 194-278], which has been used extensively by R. D. Laing and others [Laing, 1969 and 1971]

IV A Cybernetic Epistemology

> I know that you believe you understand what you think I said, but I am not sure you realize that what you heard is not what I meant.

Epistemology, to a philosopher, is the study of how we know what we know. A cybernetic epistemology is an epistemology based on a scientific concept of how a living being must function. This model assumes that the organism must get its knowledge of what is happening by selecting and interpreting the data reported by its sensory organs both external and internal and that this happens in a context of interaction between the organism and the environment. Thus: (i) what the organism sees in the enviornment affects its own actions, and (ii) its actions, in turn, have an effect on the environment. The situation is one of feedback, or co-regulation of the organism and environment, such that one cannot be understood without the other. The concepts that we consider in this section are information and feedback.[1]

---

1. Much of what is said here is quite similar to the point of view of the anarchist pholosopher and social critic Paul Goodman as expressed in the book Gestalt Therapy which he co-authored with Fritz Perls and Ralph Heferline. What is an anarchist doing amidst all this talk of hierarchy? It is important to distinguish hierarchical levels of authority in a centralized institution from hierarchically perceived levels of meaning in the world of phenomena. Both Goodman and Bateson lead one to an appreciation of the spontaneous and decentralized nature of intelligence.

Bateson defines information as a difference that makes a difference. A signal or message may contain very little physical energy, but because it is perceived, amplified, and acted upon by a living organism (or even an electronic organism), the result might be quite a large difference --- a small electric light may prevent a train from crashing, a whispered message may change a persons life, one man nailed to a cross changed two millennia of religious thought.

Whenever we consider information we must take into account its context. A signal may be meaninful only in a certain context or it may mean different or even opposite things in different contexts. In the game of twenty questions, each question sets up a context in which "yes" or "no" has a specific meaning. The key signature of C Major is a context in which the note G is dominant. A modulation changes that context. (The context of the last two sentences is tonal music theory.)

This brings us to the issue of how an organism selects and makes use of messages from the environment. Living things generally seem to be purposeful, which suggests the cybernetic concept of feedback. The purpose of a central heating system is to maintain a house at a stable temperature. The system consists of a furnace and heat distribution system, and a sense organ called a thermostat which compares the air temperatue in the living room with an ideal temperature. If the room is too cold, it turns on the furnace, and when the room is warm enough it turns it off. This system is extremely simple compared to a biological system that is regulating many variables at once such as body temperature, oxygen supply, physical balance, nutritional need, sexual need and physical safety -- all interacting in complex ways and at various levels of immediacy.

A simple feedback mechanism such as the thermostat and central heating plant is an instance of Zero Learning. The action of the thermostat is appropriate for maintaining the correct temperature, and its performance does not change. The same could be said of a dog that salivates each time it hears the bell once its learning has been completed. The salivating is them an appropriate part of the dog's nutritional feedback system. In each case of Learning Zero we find a feedback loop part of which is a sensory organ. The entire loop provides a context in which the sensory organ perceives information and other parts of the system act meaningfully on the basis of this information to maintain some sort of equilibrium. We would not classify a system consisting of a push-button and a doorbell as a Zero Learning system because considered by itself it contains no feedback and therefore no purpose except in the larger context of people ringing at doors and other people answering the door.

Zero Learning implies the existence of feedback loops and contexts in which information is meaningful. Learning I must then be the creation of such situations. But this implies that there is some intelligence that is capable of creating and modifying Zero Learning situations. This intelligence must know how to go about building circuits usine presently uncommitted hardware or neural capacity. But it must also be able to measure how well such systems work and know how to correct them. This itself a feedback loop, but on a higher level. The intelligent learner (i) is part of a feedback system that created feedback systems, (ii) is operating in a context that contains the contexts of the learned responses, (iii) has purposes that provide the purposes for the individual learned responses. These three are really isomorphic. This Learning I intelligence must perceive the world as

sequences of events, and the way it does this will be determined by its purposes. This is already too dualistic a statment, for the evidence that it perceives must be the way in which it interacts, and the way it interacts is its purpose, so these are really the same thing.

Postulating a learner (which is not different from a capability to learn because we do not want to put little men inside men) raises the question of its origin and maintainance. Let us consider this generically. An organism may be genetically equipped to respond on several levels: (i) A vegetable response is simple to respond to each situation in predefined way. A plant is capable only of Zero Learning. This is a subjective statement, but mostof us donot perceive a plant as being more intelligent than its seed. (ii) Even the simplest animals seem to learn something from their environment, but the range of perceptions that they pay attention to and the purposes of these are preprogrammed. Such organisms do not appear to us to have personality. (iii) Higher animals are capable of Learning II. We probably would include most mammals in this category. Learning II implies modification of what is taken into account in perceiving the world as meaningful. This is much more problematic than Learning I, although it is also much more powerful. A viable organism must bave individual survival, reproduction, and social interaction among the sources of motivation of its Learning I. Learning II introduces the possibility that an organism can be mistaken in purpose as well as in fact. It might learn to be antisocial or self-destructive. So there must be some balance between innate and acquired purpose. In the human species, the possibilities for acquired purpose is much larger than in other species.

This line of thought suggest that Learning II must have some stability to it as a matter of necessity. The individual ought not to learn anti-survival too easily. But when we examine Learning II as being the epistemology for the Learning I system then we find a mechanism predicting this stability. Learning II is the process of evaluating the lessons of Learning I. It is the feedback of the appropriateness of the individual lessons. If I am persistent, then either I am persistent enough so that it pays off, or else I quit too soon and conclude that I was not persistent enough. If I am dependent, then I interpret personal interactions in terms of establishing a dependence relationship. If this does not work then I figure out what went wrong or go somewhere else rather than question my basic motive. Character traits seem to have a life of their own. Success in any type of learning encourages more of the same. Any genuine freedom from such self-reinforcing bonds can come only from a level of intelligence beyond that of individual purpose.

> But in any freedom from the bondage of habit must also denote a profound redefinition of the self. If I stop at the level of Learning II, "I" am the aggregate of those characteristics which I call my "character". "I" am my habits of acting in context and shaping and perceiving the contexts in which I act. Selfhood is a product or aggregate of Learning II. To the degree that a man acheives Learning III, and learns to perceive and act in terms of the contexts of contexts, his "self" will take on a sort of irrelevance. The concept of "self" will no longer function as a nodal argument in the punctuation of experience. [Bateson, 1972, p. 304]

Learning III thus implies a sane epistemology. As Bateson points out in the passage quoted in the introduction to paper, this involves seeing the illusory quality of the division between self and other.

A description or theory, however, is not the thing described. Learning III is obviously not a thing to be learned from a book or lecture. A line drawn on a map is not a journey, but it may help us in undertaking the journey provided we remember this distinction. This distinction may help to explain some of the false starts proposed at the end of section II. The hierarchy we are seeking here is not one of levels being "about" lower levels in the sense of "talking about". What we want is a hierarchy in which each level is a qualitative change in the source from which the next lower level was created.

V Trial and Error

Learning has often been characterized as "trial and error" experience utilized correctly (or incorrectly if we are to be concerned with the pathologies of learning.) But psychologists have not really taken this idea seriously enough to use it as a model. On the other hand, the biologoists have long used trial and error as the only model for explaining genetic evolution. It will be useful to start by considering evolutionary theory and then returning to learning theory. This has two advantages, one of which is that we thereby get to examine the parellel between these subjects that we have been hinting at all long, and the other is to point out that lacking a hierarchial theory, the psychologists and the biologists have been left hanging on two extremes of a dichotomy with no way to bridge the gap. The psychologists cannot regard learning as only trial and error, because he sees something more, namely, the intelligence of the individual, which he cannot ignore even though he cannot explain it mechanistically. To the biologist, on the other hand, nature is mere mechanism, and without an operating intelligence, which in this case would have to be God, evolution can only be explained by trial and error and the selection mechanism of survival, although anyone looking at nature wih a less prejudiced eye would have to admit that there seems to be a lot more intelligent design than can be accounted for in this manner.

We can presume that everyone has heard of the theory of evolution which proposes that random changes in genetic material are introduced by radiation or other natural causes. Since genetic material is self-replicating, if such a change occurs in egg or sperm, and if this then produces a viable organism that in turn reproduces, and if the change is toward increased viability, then we can expect it to increase among the general population. Without doubting that this in fact does happen, if this were the only explanation it leaves quite a few problems unresolve.

(1) The same milieu that gave rise to this theory saw the emergence of social theory founded on the premisethat the interests of the individual were inconsistent with the intersts of the society, therefore to avoid a situation called "the law of the jungle" it was necessary to have a social contract whereby the individual forfeited his inherent right to kill, eat, or plunder his neighbor, and a governemnt with police powers was created to enforce this contract, by violence if necessary.

Why doesnot the jungle itself have a similar problem? Specifically, suppose the one individual wolf was a mutant and violated the genetic taboo against attacking and eating members of its own pack? If this animal prospered in such a fashion and had numerous offspring, why would not its genes become prevalent. How does the wolf pack protect itself genetically from this sort of disaster without benefit of a government?

(2) Economics has been called "the dismal science." Why isn't ecology similarly a dismal science with most creature nearthe point of starvation, and most behavior limited to what is essential to survivial? Why is energy wasted on sexual display which may also leave the species exposed to predators? Why are elaborate and aesthetic flowers and fruits produced? Why is there art? In the field of human psychology, we have theories based on survival and instinct which consider art and religion to be expressions of pathology (Freud), and theories based on individuation or creativity (Jung, Maslow). The former are grim, and the latter are mystical and non-explanatory.

(3) When a complicated organ such as the eye is compared to the most sophisticated products of a technological society, it seems to be too much to believe that even a billion years of random alterations followed by selection based on survival could have created anything so complicated and intentional.

These are all questions that Bateson considers at some point, and we are not going to attempt answers here, although one might guess that recognizing logical levels in evolution might be an important part of such explanations. What is important to note is that the reason that we do not ask the same questions in learning theory as in evolution is that we are in the habit of seeing intelligence under the centralized authority of ego, but tend to ignore intelligence distributed throughout the world. It is a wise doctor who respects the wisdom of his patient's body as much as he respects what he learned in medical school. But the evidence is there to be seen:

(1) Compare the design of a bone with the design of a steel girder. The girder is not a solid steel bar. It has parellel members, and cross members, or braces. These members may in turn be similarly composite. But this is only achieved two or three layers deep. The bone, however, is structurally composite down to the level of individual coiled molecules correctly aligned to act as springs or shock absorbers. A bone that is broken and mends crooked has the intelligence to straighten itself out in line with the direction of stress. This involves complicated rearrangement of structural material, marrow, blood bessels, etc. The intelligence to do this is not coming from the brain.

(2) It is common to here DNA referred to as a "blueprint". But no blueprint can be realized as a completed house or airplane without a certain amount of intelligent interpretation. In the biological case, the situation is even more remarkable, It is not possible to take two blueprints of similar but slightly different airplanes, cut each in half, and form a viable blueprint by taping half of each together. Yet a process somewhat more random than this quite regularly produces the genetic material of a viable person, even when the parents are of different physique or race. It would seem that a considerable amount of intelligent adjustment of various interfaces would be necessary.

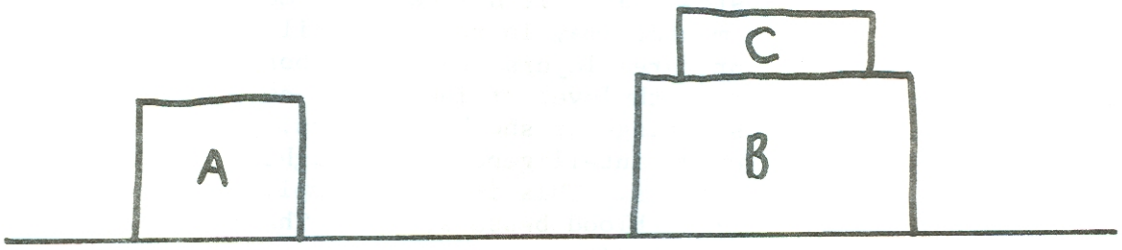
VI Artificial Intelligence

In the early days of Artificial Intelligence, Arthur Samuel wrote a famous computer program that played checkers and learned to improve its game. The program had several parameters that determined the relative importance of certain strategic considerations as the program evaluated the consequences of move sequences. Rather than discover himself what the optimal value of these parameters was, Samuel let the machine do this by analysing games. The machine definitely learned to play checkers better as a result of this analysis. Is this Learning I?

We can make a good case by pointing out that the machine was discovering by trial and error the best choice with a specified context. The machine could not invent or modify any of the strategic parameters. It could only decide which ones were most useful for winning. After it optimized these parameters it could not learn any more.

On the other hand, we could argue that this is Zero Learning because no new situations are being created. Rather than acquiring any new response, the program was merely tuning up pre-existing ones. This case seems sort of borderline.

A much later artificial intelligence project, Gerald Sussman's, "Computational Model for Skill Acquisition" goes into the epistemology of learning in a much more thorough way. The program, called Hacker, tries to solve problems in the "blocks world" which is a table top containing toy blocks of various sizes, shapes and colors. A command is given such as: Put A on B. In order to solve this problem,



Put A on B.

Hacker must realize that this cannot be done until C is removed from B. Hacker has only one had, so C must be moved first. The attempt to move A first is an error. Hacker solves this problem and then remembers it so that in the future it does not pick up a block without first checking to see if the space to set it down is clear. This is the sort of thing that it learns although some of the situations are more complex and involve several goals that must be achieved in a specific order because some them mess up the results of

others. Hacker is always making notes to itself about why it does each thing. These reasons in turn have their reasons which are either to accomplish specific sub-goals, or to avoid interference between sub-goals, or prepare something so that something else can happen. But the regress of reasons always leads back to obedience to the command of the programmer, and Hacker never does anything except upon being ordered to do so. For this reason, I think that all the items in these sequences of reasons are of the same level and that Hacker only does Learning I.

This is an excellent illustration of the subjectivity of any definition of Learning, which was referred to earlier. Hacker is a computer program which is a deterministic process or an algorithm when looked at from the outside. So one is perfectly free to claim that no learning occurs. In each situation (which now has to be defined as a whole sequence of previous situations) it simply does what it is programmed to do. In fact, Hacker is not as good at solving blocks world problems as would be a much simpler program that just goes about it directly with some good heuristics and a minimum of exploration. Hacker's justification is as an epistemological model, and not as a real problem solver. Like Faust's homunculus, Artificial Intelligence is not yet ready to go out into the real world and survive. The justification for the value of a project such as Hacker is not how well it performs, but how good an epistemological model it is, which is a subjective matter based on an introspection of the program (and therefore known best only by its creator). This situation has led in the past to very sharp and derisive criticism of Artificial Intelligence, and the leaders of Artificial Intelliegence have responded by backing off from the original stance based on claims of performance and are now more concerned with claims of epistemological and psychlogical modeling.

Conclusion

On the whole, I feel that I have failed to accomplish what I set out to do, which was to explain the theory to myself and others. It is a very big job, and many of the sections feel rushed. Another problem was that I needed all the ideas at once to talk about any of them. But a more serious fault of the paper is that it does not always connect logically even when it says it does. Too many explanation that should have been deductive were in fact descriptive.

The question of levels of learning has been approached in several different ways. Certainly no definitive classification of learning has been established. This may never be possible. The value of this theory is not in allowing us to see the world as a theatre stage set with three distinct layers of backdrop, but rather to see the world itself in depth.

The claim that context of context and purpose of purpose are isomorphic descriptions is an important one. It amounts to saying that there is no meaning beyond the meaning or patterning of the world itself. It is a non-idealist or non-transcendental or non-theistic attitude. Bateson would say that mind is immanent in the world. Scientists would say that God is not necessary in order to explain the world (and then go out and create new gods). Buddhism would point to "emptiness" (shunyata) as a way of suggesting that the meanings of things are within themselves.

Many other things occurred to me while writing this paper for it leads out in all sort of directions. One is that this is merely a new way of expressing a very conventional wisdom. Another is that almost everything said here has an unspoked political implication. Another is the dryness of a merely intellectual approach, like skillful means without wisdom or male without female.

Bibliography

Bateson, Gregory. Naven, 2nd ed., Stanford University Press, 1958

_____. Steps to an Ecology of Mind, Ballantine Books, 1972.

Casteneda, Carlos. The Teaching of Don Juan, Ballantine Books, 1969

Fraenkel, Abraham, Bar-Hillel, Yehosua, Levy, Azriel. Foundation of Set Theory, North-Holland, 1973

Goodman, Paul -- see Perls

Laing, Ronald. The Politics of Experience, Pantheon, 1967

_____. Knots, Pantheon, 1971

Minsky, Marvin. "A Framework for Representing Knowledge," Memo No. 306, Artificial Intelligence Laboratory, M.I.T., 1974

Perls, Frederick, Hefferline, Ralph, and Goodman, Paul. Gestalt Therapy, Dell, 1951

Samuel, Arthur L. "Some Studies in Machine Learning Using the Game of Checkers," IBM Journal of Research and Development, No. 3, July, 1959, pp. 210-229

Sussman, Gerald Jay. "A Computational Model of Skill Acquisition," Artificial Intelligence Laboratory, AI TR-294, M.I.T., 1973

Whorf, Arthur L. Language, Thought and Reality, M.I.T. Press, 1956