[Unpublished paper. Presented November 5th 1992, Department of Psychology, University of Wales, Bangor; November 26-28, 1992, Conference of the Linguistic Society of Belgium on Conceptual and Linguistic Representation, Antwerp]

PHILOSOPHICAL FASHION AND SCIENTIFIC PROGRESS IN THE THEORY OF UNIVERSALS

Ullin T. Place

Abstract
Are universals (kinds) something over and above the things (their instances) of which they are kinds? Does the universe come already packaged into kinds of thing, or are the universals which the human and animal mind distinguishes simply the product of the mind's classificatory activity? Whether universals are mind-independent or mind-dependent, are the concepts human beings and other living organisms have of them innate or are they generated wholly or in part by some kind of learning process? In either case, what assurance do we have that our conceptual scheme does not seriously misrepresent the way things are, as Kant puts it, "in themselves."

While the tides of philosophical fashion have flowed backwards and forwards between the poles of this debate ever since the time of Plato and Aristotle, it is argued that there is now some reason to think that the current tide which appears to be moving away from platonism and nativism and back towards conceptualism and empiricism may be taking us towards a permanent scientifically-based resolution of the problem. This solution, if that is what it is, gives due weight to both innate factors and learning at the biological level and to social construction at the level of human linguistic communication. It sees Darwin's principle of variation and natural selection as operating as much in the ontogenetic development of our conceptual scheme as in its phylogeny, and as providing the assurance we need that, in B. F. Skinner's words, it takes

> account of the natural lines of fracture along which behavior and environment actually break. (Skinner 1938 p.33).

1. The Problem of Universals

No cognitive capacity is more fundamental than the ability to identify the ‘universal’ or kind to which a particular encountered in sense perception belongs. For more than two thousand three hundred years since Aristotle challenged Plato's account of the matter, philosophers have argued about the nature of universals and of how human beings and other living creatures acquire such knowledge of them as they do.

As is well known, Plato held that universals exist as abstract objects independently both of their instances and of human conception. As such, they exist unchanged from eternity to eternity in a realm inhabited exclusively by such objects. Human beings have knowledge of these universals. For without that knowledge they would not be able to recognize a particular as an instance of a kind and thus use the correct word to describe it. Moreover, since their ability to do so presupposes that the recognizer already possesses the concept of the universal in question, it follows, so Plato thought, that knowledge of universals is innate, acquired prior to birth from direct acquaintance with the realm of abstract universals.

There is some doubt as to how far Aristotle's position stands in polar opposition to Plato's on this issue.[1] What is certain is that those medieval philosophers of whom William of Ockham is probably the best known who undoubtedly *did* take the opposite view, appealed to Aristotle as the authority for their position. This position is sometimes known as nominalism, because it was thought that, on this view, there is nothing to a universal other than the human propensity to apply a common name to a range of resembling particulars. As a subscriber to the doctrine, I prefer the term 'conceptualism', partly because I believe that there is much more to classifying things into kinds than applying a common label to them, but partly because I believe, contrary to Aristotle and Locke, that animals do this too, and do it without the benefit of language.

2. Concepts in Platonism and Conceptualism

Conceptualism then, is the doctrine that universals are mind-made, that they are products of the generalization and discrimination abilities of the individual living organism. For the platonist, universals or kinds are to be distinguished from *concepts* which are the mind's ability to recognize instances of them. For the conceptualist, universals and concepts are one and the same thing. A universal just *is* a particular way of classifying particulars and nothing more.

This suggests that, for the conceptualist, before there were living organisms to classify things, there were no universals and hence no instances of them. But this is a misconception. On this view, to say that a universal or kind of thing exists is systematically ambiguous as between

(a) the claim that the universal in question has instances, that things of that kind exist,

and

[1] For the view that Aristotle was a full blown conceptualist, see Lloyd (1981) and Frede and Patzig (1988). For the view that he held a position intermediate between conceptualism and platonism in which universals exist as abstract objects independent of human conception, but only as embodied in their instances, see Fine (1980) and Tweedale (1987).

(b) the claim that some living organism has that concept, i.e. the propensity to classify particulars in that way.

Clearly, saying that a universal exists in sense (a) is very different from saying that it exists in sense (b). The fact that a universal exists in sense (b) is no guarantee that instances of it exist, i.e., that it exists in sense (a). The human conceptual scheme is littered with uninstantiated universals from mythical beasts to the various utopias of political ideology. Moreover, the time scale over which a universal can be said to exist in the two senses can be very different. For example, according to current physical theory, the universal 'quark' has had instances and hence has existed in sense (a) since the big bang, since the beginning of time as we know it. But the concept 'quark', the universal in sense (b) has been around only since 1964.

3. Abstracting Universals from Resemblances between Particulars: A Problem for Conceptualism

Just as platonism is committed to the view that our conceptual knowledge, our knowledge of universals is innate, so conceptualism is wedded to the idea that universals in their capacity as concepts are acquired by the process of learning, known as 'abstraction'. In abstraction, it is claimed, the organism learns to group together the objects, events or states of affairs it encounters on the basis of the ways in which they resemble one another. Doubts about this process whereby the organism learns to abstract universals from encounters with particulars have plagued conceptualism from the very beginning. Plato himself pointed out in the *Parmenides* (Plato, 1961) that if things resemble one another they must resemble one another in some respect and that already presupposes that they are instances of the same universal. In other words, you cannot notice a respect in which a group of objects resemble one another, unless you already possess the concept of the feature or respect in which they resemble one another. Take for example, the figures in the left hand box of Figure 1. Anyone looking at this group of objects can immediately tell, even without looking at the contrasting group of *Non-Dax's* on the right, that a *Dax* is a circle, with one dot inside the circle and another outside and adjacent to it. But your ability to make that judgment, to notice the respect in which these objects resemble one another depends on your already having the concepts of 'circle', 'dot', 'inside', 'outside', and 'adjacent'. And where do these concepts come from? Are these too learned by abstraction, by noticing the respects in which circles, dots and cases of something being inside or outside something else resemble one another? But how can you do that unless you already have *those* concepts.

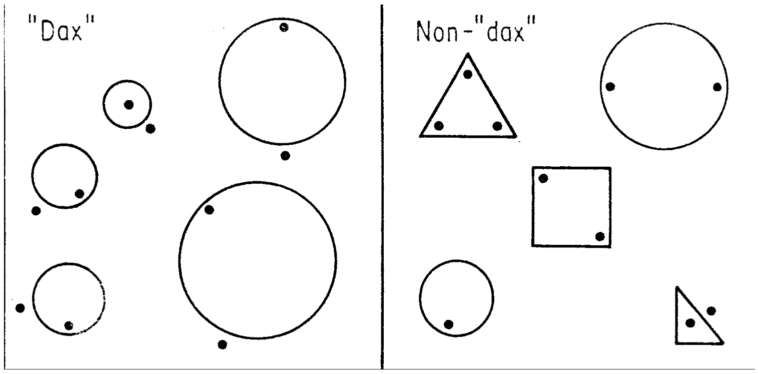

Figure 1. Samples of "dax" SD's and "dax" SΔ's *(From Kimble and Garmezy, 1963, after Smoke, 1932).*

4. Ephemeral Concepts as a Problem for Platonism

It is not difficult to see how arguments like this drive one back towards platonism. But platonism is an equally uncomfortable position. It seems plausible enough when applied to the grander universals such as the Platonic Trinity, the Good, the True and the Beautiful. It even makes reasonable sense when applied to what [has] been called 'natural kinds', things like the constituents of matter, the chemical elements, important compounds like air and water and the various species of living organism. But when it is applied to the more ephemeral of human artefacts, it collapses into absurdity. Take for example the universal 'yoyo'. A yoyo, when such things first appeared in the 1930s, was a brightly painted circular block of wood approximately 2½" in diameter and 1" thick with smoothed edges so as to fit comfortably into the palm of the hand with a deep groove cut all the way round the circumference leaving an axle in the center to which is attached a piece of string about 2' long with a loop at the end furthest away from the axle. By winding the string around the axle and slipping the loop over a finger it is possible to cause the yoyo to drop down and rise up again to the hand more or less indefinitely. Today yoyos, made now of coloured plastic, are still sold as a child's toy. In the 1930s there was a craze for such things. It affected not just children, but even grown men and women could be seen going around playing with their yoyos. The phrase "going up-and-down like a yoyo" has become part of the English language.

Now if you are a platonist you are required to believe not only that the universal *yoyo* has existed since the beginning of time, but that the concept *yoyo*, the knowledge of that universal, is and always has been part of the innate endowment of the human species.

The case of the yoyo is not just an example of an ephemeral concept of the kind which makes platonism appear silly, it is also a parable for the way in which current opinions on philosophical issues such as this are a matter of fashion, rather than genuine intellectual progress such as we are accustomed to find in the natural sciences. Nevertheless, in between the shifting tides of fashion that have blown and continue to blow this way and that between the two poles of platonism and conceptualism/nominalism, it is possible to discern, over the past two hundred years or so, a measure of progress. That progress, such as it is, has been marked by four major landmarks:

(1) Kant's conceptualist skepticism,

(2) Darwin's theory of evolution by natural selection and its behaviorist aftermath,

(3) the cognitive/computational revolution of the 1960s and 1970s, and

(4) the selectionist/connectionist revolution of the 1980s and 1990s.

5. Kant's Conceptualist Skepticism

Despite their insistence on the mind-dependent character of our conceptual scheme, the medieval conceptualists, such as Ockham, never questioned the assumption that our concepts follow what, in our own day, B. F. Skinner has called "the natural lines of fracture along which behavior and environment actually break" (Skinner, 1938 p. 33).[2] They were sustained in this common-sense realist view by the authority of Scripture, the Church and Aristotle. Descartes, despite his skepticism with respect to the existence of the external world, was a platonist who was likewise protected by the argument from divine beneficence from seriously contemplating the possibility that our divinely implanted conceptual scheme might be an elaborate deception. It was left to Kant (1781/1787/1929) to take that possibility seriously. He tried to answer Descartes' skeptical doubts about the existence of the external world, by arguing that our conceptual scheme leaves us with no option but to construe our experience as an encounter with a world of objects extended in three dimensions of space and one of time. But if, as the conceptualist maintains, that conceptual scheme is *itself* a product of the mind, there is no way we can ever know whether or not it represents things to us as they really are "in themselves."

[2] A similar metaphor expressing the same idea which is to be found in recent philosophical literature, but whose provenance is unknown to me, is that of "carving nature at its joints."

6. Darwin and evolutionary epistemology

Although it appears to have taken more than a century[3] before it was finally appreciated by philosophers, the answer to Kant's skepticism with regard to the correspondence between our conceptual scheme and the reality it depicts, was eventually provided by Darwin's (1859) theory of evolution by variation and natural selection. It is an implication of Darwin's theory that the survival and reproduction of complex free-moving living organisms, animals in other words, depends on their ability to change the spatial relations between themselves and other objects, including other organisms of the same and of different species, and so bring about the conditions necessary for that survival and reproduction. In order to do that the organism requires a nervous system whose function is to match the output to the current stimulus input on the one hand and the organism's current state of deprivation with respect to conditions required for its survival and successful reproduction on the other. Matching behavior to the conditions required for survival and reproduction is the function of the motivational/ emotional part of the system. Matching behavior to current stimulus input is the function of the sensory/cognitive part of the system. The sensory/cognitive system cannot perform its function successfully unless it can group inputs together in such a way that every actual and possible member of the category so formed is a reliable indicator of the presence of an environmental situation in which certain behavioral strategies will succeed, while others fail. In other words the survival and reproduction of an organism of this kind depend crucially on its having a conceptual scheme, a conceptual scheme, moreover, which reliably predicts what Skinner (1969) calls the "contingencies" or antecedent-behavior-consequence relations operating in its environment.

When applied to the phenomenon of concept formation, the principle of natural selection predicts

(a) that the practice of sorting stimulus inputs into kinds, classes or categories will play just as significant a role in the behavior of pre-linguistic organisms (animals and human infants), as it does in the linguistic behavior of human adults, and

[3] My colleague in the Department of Philosophy, University of Leeds, Dr. Harry Lewis (1979) cites Quine's (1969) paper 'Natural kinds' from his *Ontological Relativity and Other Essays* as the earliest use by a philosopher of the argument from evolution in the defence of realism in epistemology. Needless to say, other epistemological implications of Darwinism such as Peirce's pragmatism have been around for much longer.

(b) that, in order to yield accurate predictions of the consequences of behaving in one way rather than another in the current situation, the boundaries within which behavior generalizes and beyond which it does not, will tend to correspond rather closely to Skinner's "natural lines of fracture", in other words to the way the world really is.

Two objections are sometimes raised to this use of the Darwinian principle as a way of reconciling conceptualism and realism. One such objection is that there are many cases where the biologically adaptive response is to over-generalize in accordance with the maxim *Better safe than sorry*, rather than develop concepts that accurately follow the boundaries between one contingency or antecedent-behavior-consequence relation and another. One such case is that illustrated by Figures 2 and 3 both of which are taken from Tinbergen's (1951) book *A Study of Instinct*. Figure 2 shows a set of cardboard cutouts which produced escape reactions in birds belonging to various gallinaceous species (ducks and geese) when "flown" overhead above them. Those which produced the response are marked with a + sign. You will notice that the common feature that unites the effective stimuli is a short head with no visible neck and relatively long body. Now while it is true that all birds of prey have this characteristic shape, the same shape is also found in perfectly harmless birds, such as the swift (Figure 3). Observation shows, moreover, that young birds tend to react with the same escape/avoidance behavior to swifts overhead as they do to birds of prey. Here we have a case where the animal's responses generalize beyond the natural lines of fracture, where the animal's conceptual scheme fails to line up with the true antecedent-behavior-consequence relations.

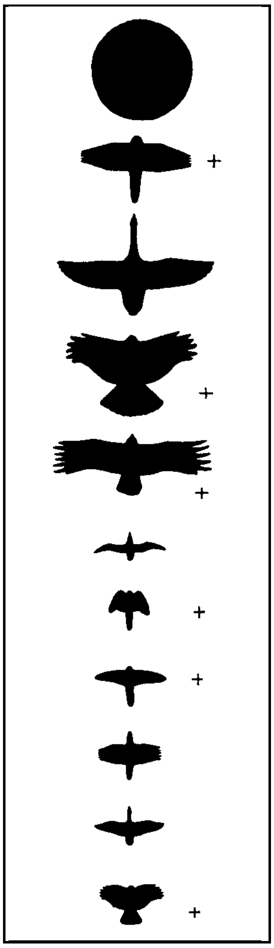

Figure 2. Models used for testing reactions to birds of prey. (*After Tinbergen, 1948*).

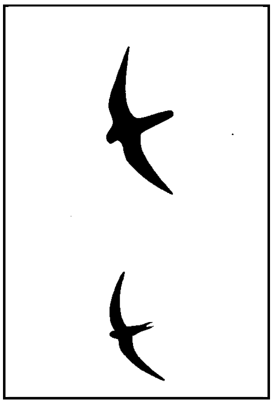

Figure 3. Flying hobby (above) and swift (below) (*From Tinbergen, 1951*).

However this kind of over-generalization is likely to be adaptive only in those cases where, as in this example, there is one contingency, one antecedent-behavior-consequence relation, which is of such biological importance that it overrides all others. But in what is arguably the more typical case where the organism must choose between two alternative responses, say between eating something and not eating it, depending upon what kind of a thing it is, the adaptive conceptual scheme will always be one that exactly follows Skinner's "natural lines of fracture."

The other objection that is sometimes raised against the use of Darwin's principle in support of a realist conceptualism is that, because of the different ecological niches they occupy and the different ways in which survival and reproduction are achieved, what is an adaptive conceptual scheme for one species or, in the human case, for one social group or individual will not be an adaptive way of carving up reality from the point of view of another species, social group or individual. Herbivores for example, can be expected to draw conceptual boundaries between different types of plant that will be of no concern to a carnivore, while carnivores will draw conceptual boundaries between different kinds of animal that will be of no concern to the herbivore. Similar differences are found between the conceptual schemes of human social groups using different technologies to exploit different environments, as is illustrated by Whorf's (1956) well known example of the five different kinds of snow distinguished by the Inuit peoples of the North American Arctic. What this shows, however, is not, as is sometimes supposed, that there are no right and wrong ways of carving up reality. What it shows is that there are indefinitely many different *right* ways of carving it up, depending on the particular motivational concerns of the species, the social group and the particular individual.

7. Platonism and conceptualism within a Darwinian framework

While Darwin's principle can be effectively invoked to defeat Kantian skepticism with respect to the correspondence between our conceptual scheme and reality, it does not tell us how that correspondence is achieved. Indeed, if it is interpreted purely as a theory of phylogenetic development, it can be taken to support a nativist and hence platonist view of the origin at least of those aspects of the human conceptual scheme which are invariant across all human languages and cultures and can, therefore, be plausibly represented as part of the genetic endowment of the species as a whole. Consequently, the advent of Darwinism has not prevented the tides of philosophical fashion from continuing to flow between the two poles of platonist nativism and conceptualist empiricism.

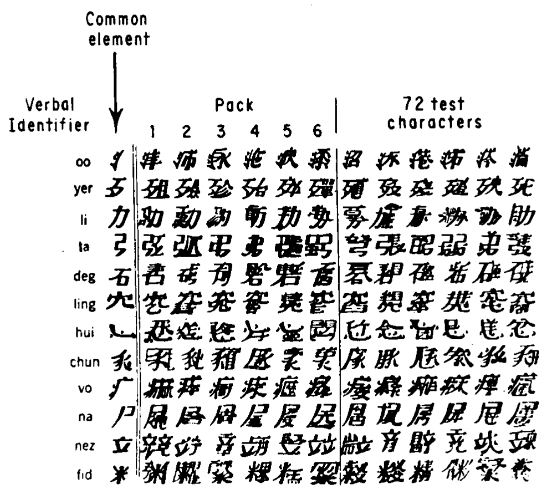

Figure 4. The 144 Chinese characters used by Hull (1920) to study concept acquisition.

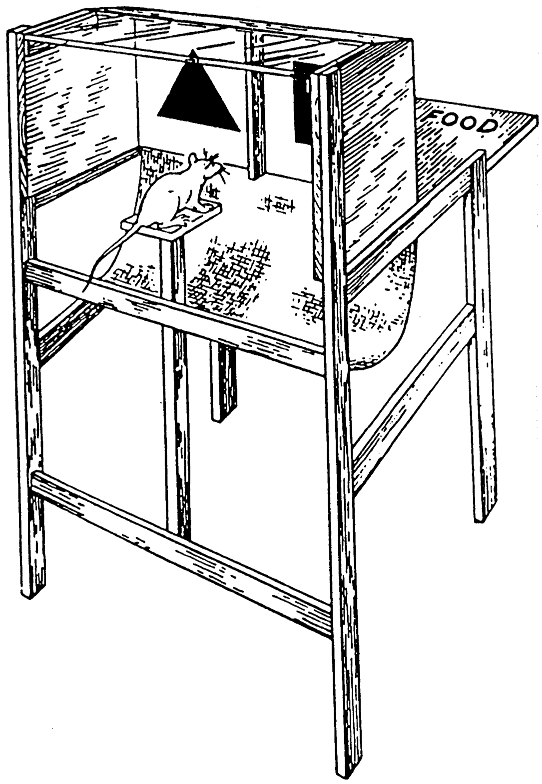

Figure 5. The Lashley Jumping Stand *(From Munn, 1950, after Lashley 1938).*

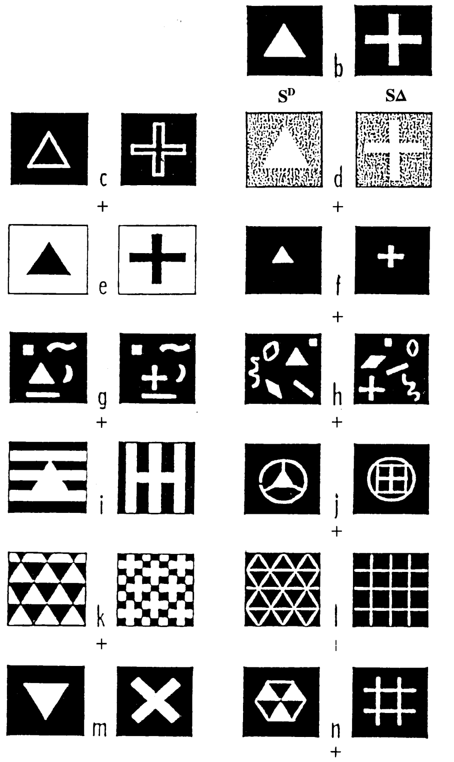

Figure 6. Pattern discrimination learning. Training stimuli b. Generalization to pairs + *(Munn, 1950, after Lashley 1938).*

Earlier this century when behaviorism in one form or other was the dominant orientation within psychology and to a lesser extent within philosophy, the conceptualist-environmentalist view prevailed. Typical of this period is the tradition of experimental studies in human concept formation which begins with Clark Hull's (1920) classic study of non-Chinese speaking subjects learning to recognize Chinese characters containing the same radical (Figure 4), which is readily assimilated to animal studies of pattern discrimination learning as illustrated by Lashley's equally classic experiments (Figures 5 & 6). Here the implication is that our conceptual scheme is learned, learned by the process of generalization and discrimination learning described in the case of classical or respondent conditioning by Pavlov (1927) and in the case of instrumental or operant learning by Skinner (1938, Chapter Five).

8. The cognitive revolution

Beginning in the 1950s three factors combined to move the tides fashion back towards a nativist theory of concept acquisition and a platonic theory of universals. One factor was the advent of the ethological approach to animal behavior which emphasized species-specific innate releasing mechanisms such as those illustrated on Figures 2 and 3 rather than the learned discriminations studied by the behaviorists. Another was Chomsky's (1959) critique of the behaviorist theory of language acquisition and his subsequent (Chomsky 1965) espousal of a nativist alternative. A third and most potent of the three was the so-called 'cognitive revolution' inspired by the adoption of the serial-digital computer as a model for the functioning of the 'mind-brain.'

The reason why the adoption of this model favors a nativist theory of concept acquisition is that, even [in] its idealized form, the Universal Turing machine (Turing 1937), the only input to which a serial-digital computer can respond is a pattern of digital pulses on a tape or disk presented to the device's reading head. The device's ability to do what this pattern of pulses tells it to do depends on both the data to be computed and the algorithm describing the computational transformation to be effected being translated into a pattern of digital pulses which conforms strictly to the "machine language" or "machine code." The ability of the device to respond appropriately to data and instructions formulated in this way is fixed by the way it has been "hardwired" at its initial construction. Consequently, on the hypothesis that the brain is a device of this kind, we have to suppose not only that the brain has its own "machine language", Fodor's (1975) "language of thought", but that the brain's ability to respond appropriately to "sentences" formulated in the language is likewise "hardwired" into that organ in the course of foetal development. Furthermore, as Edelman (1987) has pointed out, the model also requires that the input into the brain from the environment *via* the sense organs be neatly pre-packaged into "categories" corresponding to the different kinds of thing to whose presence in the environment the organism is thereby alerted. In other words, not only does the model require that the brain have an ability to "read" sentences in the language of thought, it must also have an innate ability to convert sensory input into such sentences in a way that accurately represents the universal or kind of thing whose presence is signalled by the current input. Evidently, such an innate ability could not have evolved, if the environment to which the brain is thereby "hardwired" to respond were not itself neatly pre-packaged into universals in the way that the platonic theory proposes.

9. The Selectionist/Connectionist revolution

For some twenty years between the mid-1960s and the mid-1980s the notion that the brain is a kind [of] serial-digital computer, and with it the nativist theory of concept acquisition and the platonic theory of universals, permeated the thinking of a whole generation of linguists, philosophers, psychologists and neuroscientists, not to mention practitioners in the new fields of artificial intelligence and cognitive science to which the model gave birth.

Now, the tide of fashion has turned once again. Since the mid-1980s two developments have been taking place both of which threaten the previously unquestioned assumption that there is no serious alternative to a computational theory of the functioning of the brain. The two developments are the *selectionist* movement in neurobiology represented by Edelman's (1987) *Neural Darwinism* and the *connectionist* movement in artificial intelligence represented by Rumelhart, McClelland and the PDP Group's (1986) *Parallel Distributed Processing*.

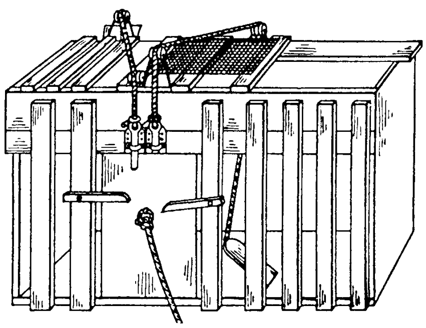

Figure 7. A Thorndike Puzzle Box

Selectionism

The term "selectionism" has been proposed by Palmer and Donahoe (1992) to describe a theory which extends the application of Darwin's principle of *variation and natural selection* to *any* kind of evolutionary or developmental process mainly, but not exclusively, in biology. On this view, natural selection is no longer restricted, to the process of phylogenesis whereby new species evolve over periods of geological time.

The idea of extending the principle of natural selection to processes of ontogenetic development within the individual organism has been familiar to students of animal behavior, since it was first demonstrated by E. L. Thorndike (1898) in his study (Figure 7) of cats learning by random trial and error to escape from a puzzle box. In this case, the population within which variations develop and which are then subject to a process of selection in which the successful survive and the unsuccessful are discarded is a population of behavior patterns or response tendencies. The animal begins by thrashing around in a more or less random fashion in the vain attempt to escape and get at the food outside, until by chance it hits on the response which unlocks the door and allows it to escape. In subsequent trials this successful response is gradually selected in preference to other unsuccessful responses which initially have a higher probability of occurrence, until eventually it is emitted immediately the animal is placed in the box. This is the process which B. F. Skinner (1981) calls "selection by consequences." As we shall see, the same process is called "learning by error-correction" or "supervised learning" by the connectionists.

For many years, Thorndike's selectionist analysis of the process of trial and error learning was the only application of Darwin's principle to a process of *ontogenetic* rather than phylogenetic development, in other words to a developmental process within the individual rather than one affecting the species as a whole.[4] The first selectionist analysis of a process of ontogenetic development *inside* the body seems to have been Jerne's (1955; 1967) selectionist analysis of the immune response. This idea was further developed by Edelman (1973a; 1974) and extended (Edelman, 1973b) from immunology to the processes of ontogenetic development within the nervous system. This culminated in Edelman's (1987) book *Neural Darwinism* which is the particular application of the selectionist principle which concerns us here.

In *Neural Darwinism* Edelman uses the selectionist principle in order to reclaim for neurobiology the high ground of theoretical neuroscience. Bogged down in the anatomical, biochemical and electrophysiological minutiae of the individual neuron and its synaptic junctions with other neurons, neurobiology had allowed this high ground to be increasingly usurped by the new disciplines of artificial

[4] In social anthropology, E. E. Evans-Pritchard's (1940) book *The Nuer* presents what is, though its Darwinian inspiration remains unacknowledged, a selectionist analysis of the socio-economic organization of that Southern Sudanese people.

intelligence and cognitive science which had grown up in the wake of the computer revolution and the adoption of the wholly non-biological S-D computer as a model for the brain.

Instead of using the S-D computer as a model for the way the brain functions, Edelman uses the computer to *construct* a more biologically acceptable theory of how it operates. This model is based on accurate neuroscientific knowledge of the microstructure of the brain, using variation and natural selection as a unifying, synthesizing principle. The principle of variation and natural selection is applied not just to the growth and decay of individual neurons and their synaptic junctions with other neurons, but, more important for our present purpose, to the process whereby the organism learns *ab initio* to classify stimulus inputs according to the different "categories" of object or event in the environment whose presence is signalled by those inputs. As Edelman puts it, the theory

> was formulated to explain how perceptual categorization could occur without assuming that the world is pre-arranged in an informational fashion. (Edelman 1987, p. 4)

Connectionism

Edelman's selectionism totally repudiates the S-D computer model of brain function. This repudiation is more qualified in the case of the second of the two movements which are threatening to undermine it, the connectionist movement in cognitive science and artificial intelligence. This is understandable, given the commitment of those disciplines from their inception to exploring and exploiting the S-D computer model. The source of connectionism is a belated realization on the part of those working in artificial intelligence (AI) of something that should have been obvious from the beginning, namely that the S-D computer is a device invented by the human brain to perform quickly and efficiently those symbolic computation tasks which even the most intelligent human brain performs slowly and inefficiently. Such devices, while performing computations way beyond the capacity of a human mathematical genius, are slow and inefficient when it comes to performing tasks such as learning to recognize

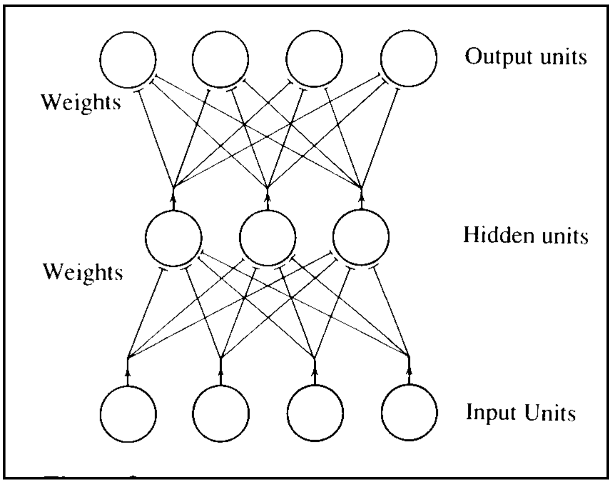

Figure 8. Schematic model of a three-layered network (*From Churchland and Sejnowski, 1989*).

complex patterns in a variety of different contexts which both the animal and human brain perform quickly and efficiently.

A connectionist network is a device in which semi-conductor units or *nodes*, as they are usually called, are linked together in the form of a network in the same way that neurons are linked together in the "grey matter" of the central nervous system (Figure 8). Like the S-D computer, a connectionist network is an information-processor which transforms a sensory input at one end into an appropriate output at the other. But whereas, in the S-D device, the output is computed by following a sequence of steps prescribed by a set of symbolically formulated instructions, in the case of a connectionist network the nature of the output is determined simply by the way in which the pattern of stimulation is transformed as it passes through the network of synaptically connected nodes. The nature of this transformation is determined partly by the size and complexity of the network but mainly by the so-called 'weights' of the individual synaptic connections between one node and another. The *weight* of a synaptic connection is a dispositional property of the connection whereby the firing of the node on the pre-synaptic or input side of the connection contributes either to the excitation or the inhibition of firing in the node on the post-synaptic or output side of the connection.

By assigning a set of initial weights to the individual synapses, a network can be given an 'innate' predisposition to respond in a particular way to inputs of a particular kind. But, unlike the S-D computer's 'hardwired' predisposition to respond to rules and instructions formulated in the machine code, these predispositions are always susceptible to modification through changes in the synaptic weights brought about by subsequent learning experiences. It is this process whereby weights are changed either up or down each time a particular connection is activated or inhibited which gives the connectionist network its distinctive functional property, that of acting as a *pattern discrimination learning device*.

Despite incorporating certain features that have no known counterpart in the living brain (Reeke and Edelman 1988), it is obvious that a connectionist network is much more like the living brain than is a conventionally constructed S-D computer, not only in its structure which is deliberately modelled on that of the brain, but also in its functional properties.

The most striking functional property of a connectionist network is its ability to learn to discriminate patterns in a variety of different settings by the process of trial and error-correction. This gives

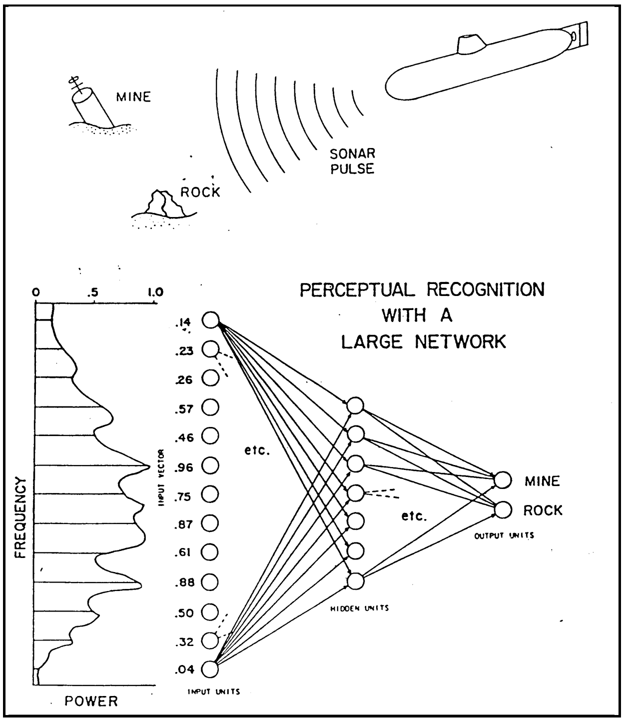

Figure 9. (*From Churchland, 1988, after Gorman and Sejnowski*).

it the ability to abstract universals from encounters with their instances, without needing to already possess the concept of what those instances have in common, as in the kind of concept-formation illustrated by the *dax* example (Figure 1 above). To illustrate this point take the following two examples, one taken from the connectionist literature and the other from that of human discrimination learning. The first (Figure 9) is the example of a network housed in submarine learning to discriminate between mines and rocks on the seabed by their sonar echo. It comes from the 1988 Edition of Paul Churchland's *Matter and Consciousness.* In this example the network begins by responding 'Mine' or 'Rock' more or less at random. But if, every time it does so, it receives a feedback message telling it whether its response is right or wrong, it gradually learns to eliminate the errors until it can eventually discriminate between the two with 100% accuracy. It learns this, however, *without* being able to provide *any* information about what it is about the sonar echo characteristic of mines which distinguishes it from that characteristic of rocks.

The other example (Figure 10) comes from Canfield (1941). It shows the different varieties of external genitalia found in day-old chicks, males on the left, females on the right. A swift glance at this feature is the only information which a skilled chicken-sexer has to go in making the economically vital determination of the sex of day-old chicks. Although chicken-sexers can make this discrimination with considerable reliability, they do not and could not learn to do so by studying a series of drawings such as these or even by being shown a variety of examples of both sexes as determined by an experienced sexer. They learn to make the discrimination in exactly the same way that Churchland's network does, by responding initially at random and then being told by an experienced teacher when the response is right and when it is wrong. No concept of the relevant distinguishing features ever emerges. Here we have a process of abstraction which escapes the objections that have been raised to this account of concept/universal formation since Plato's day.

One final point. In both these cases what ensures the accuracy, reliability and biological relevance of the discrimination is the error-correcting feedback supplied by the trainer. In the natural environment that error-correcting feedback is supplied by the immediate consequences of behaving in one way rather than another. Those consequences are interpreted either as an 'error' message or as a 'correct' message, depending on the motivational attitude of the organism to those consequences, in other words, on whether it likes them or dislikes them. If it likes the consequences of its behavior, its inclination to do the same thing again on similar occasions in the future will be strengthened. If it dislikes them, its inclination to repeat the behavior will be weakened. That is the Law of Effect, first formulated more than eighty years ago by Thorndike (1911), but traceable, in the form of the doctrine of 'psychological hedonism', back to Epicurus.

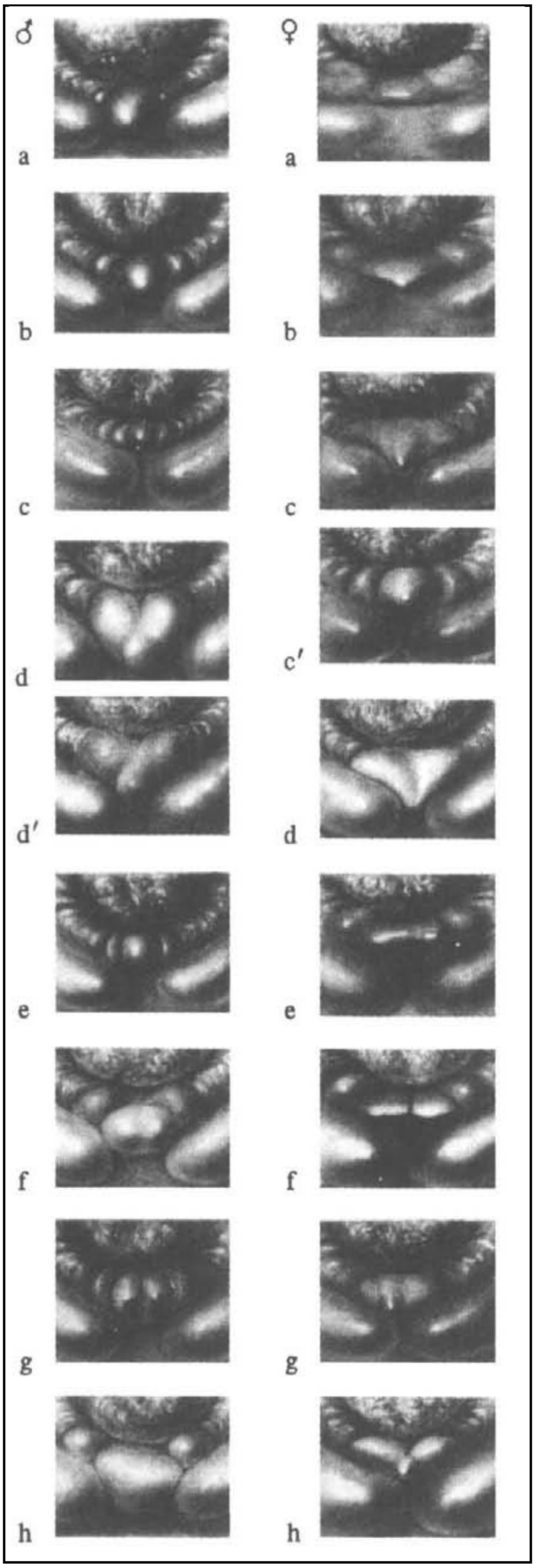

Figure 10. External genitalia of male (left) and female (right) day-old chicks (*Canfield, 1941*).

References

Canfield, T. H. (1941) Sex determination of day-old chicks, II. Type variations. *Poultry Science*, **20**, 327-328.

Chomsky, N. (1959) Review of B.F.Skinner's *Verbal Behavior*. *Language*, **35**, 26-58.

Chomsky, N. (1965) *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Churchland, P. M. (1988) *Matter and Consciousness*, Revised Edition, Cambridge, Massachusetts: MIT Press.

Churchland, P. S. and Sejnowski, T. J. (1989) Neural representation and neural computation. In L. Nadel, L. A. Cooper, P. Culicover and R. M. Harnish (eds.) *Neural Connections, Mental Computation.* Cambridge, Massachusetts: MIT Press.

Darwin, C. (1859) *On the Origin of Species by means of Natural Selection*. London: Murray.

Edelman, G. M. (1973a) Antibody structure and molecular immunology. *Science*, **180**, 830-840.

Edelman, G. M. (1973b) Molecular recognition in the immune and nervous systems. In F. G. Worden, J.-P. Swazey and G. Adelman (eds.) *The Neurosciences: Paths of Discovery*. Cambridge, MA: MIT Press, pp. 65-74.

Edelman, G. M. (1974) The problem of molecular recognition by a selective system. In F. J. Ayala and T. Dobzhansky (eds.) *Studies in the Philosophy of Biology*. London: Macmillan, pp. 45-56.

Edelman, G. M. (1987) *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books.

Evans-Pritchard, E. E. (1940) *The Nuer*. Oxford: Clarendon Press.

Fine, G. (1980). The one over many. *The Philosophical Review*, LXXXIX: 197-240.

Fodor, J.A. (1975) *The Language of Thought*. New York: Crowell.

Frede, M. and Patzig, G. (1988). *Aristoteles "Metaphysik Z": Text, Überzetzung und Kommentar*, 2 Vols., München: Beck.

Hull, C. L. (1920) Quantitative aspects of the evolution of concepts. *Psychological Monographs*, **28**, No. 123.

Jerne, N. K. (1955) The natural selection theory of antibody formation. *Proceedings of the National Academy of Sciences, U.S.A.*, **41**, 849-856.

Jerne, N. K. (1967) Antibodies and learning: Selection versus instruction. In G. C. Quarton, T. Melnechuk and F. O. Schmitt (eds.) *The Neurosciences: A Study Program*. New York: Rockefeller University Press, pp. 200-205.

Kant, I. (1781/1787/1929) *Kritik der reinen Vernunft*. First and Second Editions, Riga: Hartknoch. English translation by N. Kemp Smith as *Immanuel Kant's Critique of Pure Reason*, London: Macmillan.

Kimble, G. A. and Garmezy, N. (1963) *Principles of General Psychology*. New York: Ronald.

Lashley, K. S. (1938) The mechanism of vision. XV. Preliminary studies of the rat's capacity for detail vision. *Journal of General Psychology*, **18**, 123-193.

Lewis, H. A. (1979) The argument from evolution. *Proceedings of the Aristotelian Society*, Supplementary Volume LIII, 207-221.

Lloyd, A. C. (1981). *Form and Universal in Aristotle*, Liverpool: Cairns.

Munn, N. L. (1950) *Handbook of Psychological Research on the Rat: An Introduction to Animal Psychology*. Boston: Houghton-Mifflin.

Palmer, D. C. and Donahoe, J. W. (1992) Essentialism and selectionism in cognitive science and behavior analysis. *American Psychologist*, **47**, 1344-1358.

Pavlov, I. P. (1927) *Conditioned Reflexes*. English Translation by G. V. Anrep. London: Oxford University Press.

Plato (1961). *Parmenides*, *Theaitetos*, *Sophist*, *Statesman*. English translation with Introduction by John Warrington, London: Dent.

Quine, W. v. O. (1969) Natural kinds. In *Ontological Relativity and Other Essays.* New York: Columbia University Press, pp. 114-138.

Reeke, G. N., Jr. and Edelman, G. M. (1988) Real brains and artificial intelligence. *Daedalus,* **117**, 143-173.[5]

Rumelhart, D. E., McClelland, J. L and the PDP Research Group (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Volumes 1 and 2. Cambridge, MA: MIT Press.

[5] [Editor: This reference was missing from the original list. In 1988 there appeared three publications by Reeke and Edelman. I added the publication to the References, which I think is the most probable one used by UTP.]

Skinner, B. F. (1938) *The Behavior of Organisms*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1969) *Contingencies of Reinforcement*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1981) Selection by consequences. *Science*, **213**, 501-504.

Smoke, K. L. (1932) An objective study of concept formation. *Psychological Monographs*. **42**, No. 191.

Thorndike, E. L. 1898) Animal intelligence: an experimental study of the associative process in animals. *Psychological Monographs*, **2**, No.8.

Thorndike, E. L. (1911) *Animal Intelligence*. New York: Macmillan.

Tinbergen, N. (1948) Social releasers and the experimental method required for their study. *Wilson Bulletin*. **60**, 6-52.

Tinbergen, N. (1951) *A Study of Instinct*. Oxford: Clarendon Press.

Turing, A. (1937) On computable numbers with an application to the *Entscheidungsproblem*. *Proceedings of the London Mathematical Society*, **42**, 230-265.

Tweedale, M. (1987). Aristotle's universals. *Australasian Journal of Philosophy*, 65, 412-426.

Whorf, B. L. (1956) *Language, Thought and Reality*. New York: Technology Press, M.I.T. and Wiley.