

**Please note**

The text in this file has been automatically extracted and may contain minor errors. For the original version please consult the paper copy held in the Swinburne Library.

# 1

## INTRODUCTION

|   |             |
|---|-------------|
| <b>INTRODUCTION</b>                       | <b>1-2</b>  |
| <b>STUDIES OF NATURAL INTELLIGENCE</b>    | <b>1-9</b>  |
| <b>CLASSICAL AND OPERANT CONDITIONING</b> | <b>1-10</b> |
| <b>ARTIFICIAL NEURAL NETWORKS</b>         | <b>1-12</b> |
| <b>SYNAPTIC LEARNING RULES</b>            | <b>1-12</b> |
| <b>TYPES OF FEEDBACK</b>                  | <b>1-14</b> |
| <b>No Feedback</b>                        | <b>1-14</b> |
| <b>Full Highly Specific Feedback</b>      | <b>1-15</b> |
| <b>Reinforcement Feedback</b>             | <b>1-16</b> |
| <b>INTRA-ELEMENT MECHANISMS</b>           | <b>1-17</b> |

## INTRODUCTION

It was only a **few** decades **ago** when it appeared that **the** development of digital computers, which reliably do exactly what they are instructed at very high speed, **might** solve most information processing problems. However, digital computers are now so powerful, compact, and affordable that the **task** of working out **and** expressing exactly what they should do is becoming an overwhelming burden, in all **but** the simplest or most modular of problems. In other words, the point is being approached where the task of programming a computer to solve **each** type of difficult problem is becoming impractical. What is needed **are** new types of computers which **learn** how to solve new **and** complex problems - "intelligent" computers.

With the **increasingly** liberal use of the term "intelligence" in recent times, **one** could be forgiven for thinking that it is well understood. However **this** is not the case. Any attempt to produce a specific definition of intelligence is fraught with peril, and usually serves well as **the basis** of vigorous debate. Nonetheless, the formulation of at least some minimum criteria for intelligence is required to clarify the objectives of man-made **systems** designed **to** produce intelligent behavior. This **can** also assist the process of understanding intelligence itself. As a relatively non-controversial starting point, consider the following general definition: ***Intelligence is the capacity to acquire and apply knowledge.***

**This** very general definition has several more specific implications. A memory capability is required in which **to** store & heacquired knowledge so that it **may** later be applied. The capacity to acquire knowledge implies some **degree** of sensory (**or** receptive) capability and a definite capacity to learn or adapt. The capacity to apply knowledge implies some ability **to** respond (or react) to circumstances appropriately by utilising **relevant** knowledge. Furthermore it is impractical, **if** not impossible, for any form of intelligence with finite computational resources to acquire or apply **all** possible knowledge from a **complex** environment. Therefore, intelligence will **usually** require some means of selectively acquiring that knowledge which, if applied, **may** be of assistance in the pursuit of its goals. This implies that an intelligent system also needs to

be receptive to **some** expression of its **goals**, and be capable of **utilising** this information **to** help it selectively acquire and apply useful knowledge. Such goal information may be provided in the form of an explicit input, or **may** be **implicitly** expressed in the functional design of the system. Thus, from the general **definition** above it can be inferred that: ***Intelligence requires memory, learning, sensory and goal input, goal directed behavior, and response output.***

Naturally occurring intelligence, as supported in its various biological forms, exhibits all of the above basic characteristics. The uninitiated **might** therefore deduce that "Artificial **Intelligence**" (AI) refers to a field of study with the general **aim of** developing man-made systems which exhibit all of the **same** basic characteristics. However, even a cursory survey of the behavior of systems developed under the AI guise will indicate that this is not so. Indeed, any system **with** characteristics which vaguely resemble **any** of those of natural intelligence **seems** to qualify as an AI **system** - though this observation, in itself, need **not** necessarily be taken **as** a criticism of AI work in general, as:

(i) AI research need only produce useful cost-effective systems to be intellectually and **commercially** justifiable. In other words, man-made systems need not actually be intelligent to be useful, even if they are classified as AI. However, an increasing awareness of the **limitations** of such AI **systems** is motivating a resurgence of interest in natural intelligence, in the hope that more generally applicable **and** more powerful mechanisms **will** be **developed**.

(ii) AI research **which** now only deals with fragments of intelligence, may at a later date conceivably contribute to a synthesis of **more** complete intelligent behavior. Indeed, it is usually a deliberate decision to divide the difficult problem of **understanding and** reproducing intelligence into many smaller (and ideally separate) **problems** in the **expectation** that these will prove less intractable. This approach is currently typified by largely isolated studies into the nature of pattern recognition, knowledge representation and application, learning, memory, **goal** seeking behavior, motor control, and sensorimotor control.

However, despite considerable effort, progress in AI aimed at producing intelligent behavior has been less than impressive, suggesting that an

alternative type **of** approach to reducing the complexity of the intelligent behavior to be produced **may** be more appropriate. This is the avenue explored in the research described herein.

The scale of the problem is reduced here by decreasing the level of intelligent behavior to be modelled, while still maintaining all of the basic characteristics of intelligence. With this **approach**, intelligent systems are then seen as comprising (usually less) intelligent elements. The general **aim** of the research described herein is to develop one such type of intelligent element.

The specific nature **of** the intelligent element to be developed **may** be **characterised** by determining the following:

- (i) Inter-element specifications. These define the number of inputs and **output(s)** of the element, and the type of **signal** they convey.
- (ii) Behavioral performance. This defines how the element **output(s)** respond to **various** important sequences and/or combinations of input conditions. In this case, the behavioral performance **will** be a detailed specification of the intelligent behavior required of the element.
- (iii) Intra-element mechanisms. These refer to **the** type **of** internal **signals** used, and **the** types of ways in which they **may** interact.

In seeking **to** develop a new AI element **capable** of capturing some of the behavioral advantages of natural intelligence, an **obvious source** of **data** for **all** three of the types of specifications outlined directly above are the natural systems themselves. All of these types of design constraints **for** the intelligent element under development here have been extracted **from** aspects of natural **systems**. This is not **because** natural **systems** necessarily represent **the best** type of solution, but because **they** still represent the only available comprehensive working solution.

Regarding inter-element **specifications**, the architectural constraints of biological neural networks are invoked. A complete system is therefore composed of multiple processing elements, interacting extensively via many simple scalar signals. These inter-element signals convey **only** amplitude information, which varies over time. They are therefore "non-symbolic" in **an**

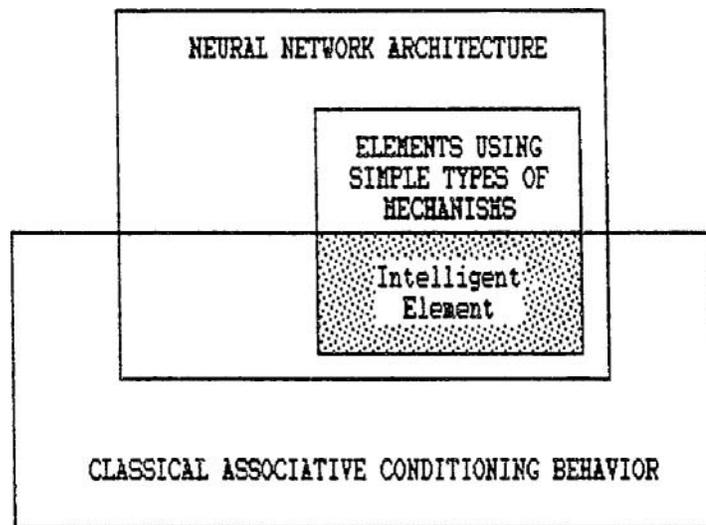
explicit sense. A neural network architecture was chosen primarily because it is demonstrably capable of supporting sophisticated intelligent behavior (**e.g.**, biological brains). As for the number of inputs and outputs, the minimum required **to** exhibit all of the basic characteristics of intelligent behavior are used. One output for response, one goal input, and **several** sensory inputs are sufficient **for** this purpose.

The behavioral criterion corresponds to basic associative conditioning, of a type which appears to be fundamental to most adaptive biological nervous systems. **More** specifically, it addresses the surprisingly complicated field of **classical** conditioning, which despite extensive study is still only **partly** understood. The general field of classical associative conditioning was chosen **as** the source from which to select a large number of behavioral constraints, mainly because it incorporates all of the **basic** characteristics of intelligence discussed above. These behavioral constraints have been carefully chosen to correspond to that which might conceivably be produced by a single neuron, or small group of neurons. That a single neuron could conceivably exhibit such complex behavior is suggested by the demonstration of basic classical associative conditioning behavior by presynaptic facilitation in the marine mollusk **Aplysia** (**e.g.**, Carew, Abrams, Hawkins, and Kandel, 1984).

And **finally**, the intra-element mechanisms invoked to **realise** the functions of these units are restricted **to** relatively simple types of interactions, such **as** summation, multiplication, **and** accumulation of internal scalar quantities, which can be achieved by the chemical and electro-chemical processes underlying **neural** function, (or analog electronic hardware). Note, however, that this is combined with a preparedness to **invoke**, where required, relatively complicated combinations of such simple types of interaction.

Thus, in summary, the research described herein lies at **the** intersection of three areas of consideration depicted in **Figure 1-1**. It may at **first appear** that the voluntary **imposition** of such constraints upon an AI system's behavior and design could only make the task of its **development** all the more difficult. However, in the case of the research described herein, the reverse **was** found to be true. This is essentially because the bodies of knowledge

related to all three types of constraints (behavioral, inter-element interaction, and intra-element mechanisms) are very **much** incomplete, and the expertise required to develop such a system is not **yet** established. It is not really possible to determine, in advance, a complete behavioral specification for the operation of a specific system component that is **known** to exist, or which is known to be able **to** be implemented. Nor is it possible to fully specify the characteristics of **the** mechanisms capable of producing the desired behavior. Therefore, all of the constraints together help produce a more complete specification of the initial route to **take** through a vast terrain of possible system development paths.



**FIGURE 1-1.** Relationship between *the three recognised* areas and the object of the research described herein - the intelligent element.

The object of the research described is a new artificial neural network element referred to as the **Associative Conditioning Element (ACE)**. As will be demonstrated using both computer simulation results and theoretical argument, ACE differs from other neuronal models of classical conditioning (e.g., Barto and Sutton, 1985; Klogf, 1987) in its robustness, its mechanistic and behavioral complexity, and the degree to which its behavior corresponds with empirical results.

This document, after some preliminary material of an introductory nature, progressively develops the conceptual and then the mechanistic detail of ACE. Not until the final chapters does a complete concept of ACE emerge. To help provide some specific indication of where the research documented is headed, a complete schematic **diagram** of ACE is provided below in Figure 1-2. It is evident **that** although ACE is only a single type of element designed to operate within a highly distributed neural network system architecture, it is in **itself** a system of considerable complexity.

ACE comprises multiple CS (Conditioned Stimulus) input channels, and a common output stage upon which **all** of them converge. **All** CS input channels **are** mechanistically identical, **and** so only a **single** representative CS input channel is illustrated interfacing with the output stage in Figure 1-2. **ACE's** single **output** signal (OUT) indicates **when, and to** what extent, US (Unconditioned Stimulus) INPUT activation has learned to be expected by each activated CS **INPUT** signal. OUT therefore has the amplitude and temporal characteristics required to generate a CR (Conditioned Response), though subsequent (and as yet undeveloped) elements would learn the qualitative nature of the CR. Each CS input channel contains the CS Trace Circuit illustrated in the **lower** section of Figure 1-2, which generates a sustained "trace" of prior CS INPUT activity suitable for acquisition and performance of an appropriately timed CR, and the Neural **Multiprocess** Memory Model illustrated in the upper section of the CS input channel in Figure 1-2, which **utilises** past experience to appropriately regulate the impact of current experience upon the expectation of the US by the CS. **A** decision was made to sequentially document and develop the performance and operation of each of the main **subsystems** in **relative** isolation, **before** later integrating them to complete ACE **and** reveal those aspects of **ACE's** behavior requiring their cooperative interaction. This approach clearly reveals the independent capabilities of **each** subsystem, and therefore facilitates their possibly separate **utilisation** or further development in future systems. Although devoid of **detailed** explanation at this point, Figure 1-2 provides a **useful** reference point for future chapters, as it indicates where each separately developed subsystem fits into ACE as a whole.

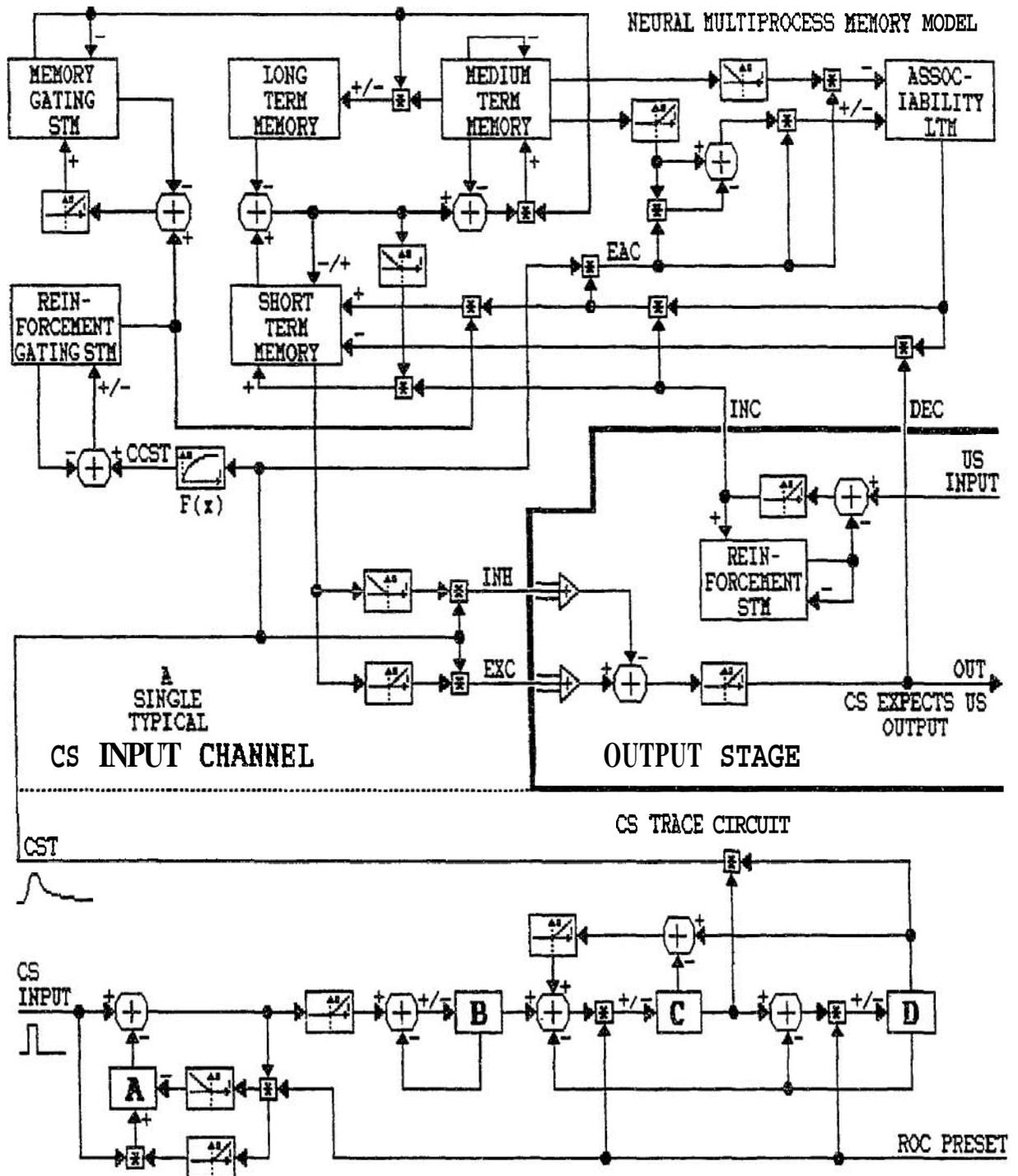


FIGURE 1-2. Complete schematic diagram of the Associative Conditioning Element (ACE). A single representative CS input channel is shown, comprising the Neural Multiprocess Memory Model (NMMM) and the CS Trace Circuit (CSTC). Also shown is the common output stage upon which multiple CS input channels converge.

## STUDIES OF NATURAL INTELLIGENCE

The first **half** of the twentieth century saw the emergence of an enthusiastic attempt by (mainly) psychologists to unravel all natural intelligence by studying animals in contrived experimental paradigms. At the heart of this effort was classical reflex theory, **also** known as Stimulus-Response (or "S-R") theory, in which behavior is thought to consist of sets of **responses** to specific stimuli. Unfortunately, the success of this theory induced an unrealistic expectation **that** it would account for most aspects of behavior. When, in **the** latter **half** of this century it became apparent that the grand expectations would not be met, the entire field of study **was** largely discredited. However, while the **wildly** optimistic hopes for reflexology are now considered naive, its ability to **conceptualise** animal behavior in controlled experimental settings at a **general** level of explanation remains **largely** accepted. Reflexology is now seen in a more balanced perspective, as an important component of animal behavior, and **as** a possible **doorway** opening in onto the field of **neural** processing mechanisms. Theoretical **models** which include concepts of reflexology are now much more sophisticated, and integrate these concepts with other aspects such as drive level, expectation, selective attention, and higher level processing.

Researchers of **animal** learning have **historically** sought to reveal either **specialised** adaptive capabilities in a comparative study of different species, or to reveal functional relations between behavioral changes and environmental changes using a more analytic approach. **Romanes** (1882) and **Morgan** (1894) are cited by **Mackintosh** (1974) as pioneers of the former approach, seeking to produce evidence of mind in **animals** other than **humans**. **Morgan** went **further**, arguing that the apparent complexity of overt behavior may result from the operation of simpler underlying associative processes. Although initially continuing on from **Morgan**, **Thorndike's** (1898, 1911) main achievement was to appreciate the necessity for, and then to develop, the controlled "operant conditioning" experimental paradigm, to help reveal the nature of underlying associative processes. The operant conditioning experimental paradigm, **and** analysis of the results it produced, were further advanced by the productive work of **Skinner** (1938, 1966). The other main pioneer of controlled behavioral

experiments **was** Pavlov (1927), who found it necessary to investigate **associative** conditioning in order to further his physiological **analysis** of bodily functions and behavior, and in so doing developed the "classical conditioning" paradigm.

The analytic, controlled experimental study of learning now overshadows its predecessor, and has done so for some sixty years. While the varied environmental pressures upon different animal species may have resulted in the emergence of differences in learning processes, the apparent similarities are now far more compelling.

### CLASSICAL AND OPERANT CONDITIONING

Both classical conditioning and operant (or instrumental) conditioning are most clearly defined in terms of experimental procedure, **as** their underlying mechanisms are both **less** easily separated, and poorly understood. In a classical conditioning experiment, **the** experimenter arranges a contingency between presentation of a stimulus to the subject, and the subsequent delivery of reinforcement. In an operant experiment, a contingency is arranged between the subject's behavior (**i.e.** a response) and subsequently delivered reinforcement. Those who first appreciated this clear operational distinction between stimulus-contingent and response-contingent reinforcement suggested that differences in the underlying mechanisms may exist (**Konorski** and **Miller**, 1937; **Schlosberg**, 1934, 1936, 1937; Skinner, 1935, 1937, 1938). Eater, the dominant trend was to assume that the operational distinction signified **essentially** no underlying mechanistic difference (**Rescorla** and Solomon, 1967). However, it now seems that the earlier view **was** probably more correct.

The issue is further complicated, because the scheduling of **stimulus-**contingent or response-contingent reinforcement does not necessarily ensure that the subject experiences only this contingency. In fact, it is very **difficult** to design either classical or **operant** procedures in which some **mixture** of the two types of contingencies is not experienced by the subject. However, there is a firm basis for rejecting the notion of a single mechanism which **completely** accounts for both types of contingency. Instances of classical conditioning

exist, **such** as omission schedules, in which an operant contingency cannot account for learning, suggesting that a classical conditioning mechanism is required (**Mackintosh**, 1983, p. 32). Conversely, instances of operant conditioning exist for which a **classical** conditioning mechanism cannot account, suggesting that an **operant** conditioning mechanism is required. Mackintosh (1983, p. 98) argues that both types of conditioning **utilise** similar types of mechanisms in the formation of their respective associations, but that these are translated into performance **using** different mechanisms.

The research described **herein** focusses upon **classical** conditioning, **primarily** in order to limit its scope to manageable proportions. The development of a robust mechanism supporting classical conditioning should **also** facilitate subsequent attempts to account **for** operant conditioning. In addition, most aspects of classical conditioning can be demonstrated using a single neural network element, whereas multiple elements may be required to demonstrate most aspects of operant conditioning.

Much of the technical vocabulary currently used in association with classical conditioning was established by **Ivan Pavlov** (1927). **Pavlov's** prior Nobel prize winning research (1904) into the physiology of the digestive system led **him** to investigate the processes underlying the anticipatory salivation exhibited by **dogs**. When meat powder (the unconditioned stimulus, or US) was placed **directly** onto a dog's tongue, it **reliably** salivated (the Unconditioned Response, or UR) without prior training. In contrast, the sound of a ringing bell initially caused an orienting reaction, but no salivation. However, if the **ringing** bell was then closely followed (or "reinforced") by presentation of **the** US on several successive trials, it acquired the salivation response. Subsequent presentation of the sound of the ringing bell (the Conditioned **Stimulus**, or CS) alone **then resulted** in salivation (the Conditioned Response, or CR). This nomenclature is used frequently throughout this text. Note, however, that while this simplified account of Pavlov's early work with **dogs** still serves well **as** an introductory example of classical conditioning, it does not reveal the many **complexities** which **have** since been clearly demonstrated. These will be referred to at the appropriate junctures in later chapters.

## ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (**ANNs**), sometimes referred to simply as neural networks, are a broad **class** of information processing systems with an internal architecture resembling the highly distributed **and** interconnected **organisation** of biological networks. However, that may **often** be **about as** far as the similarity to biological systems goes. The nature of the ANN processing elements is usually very different to that of **biological** neural hardware, typically being very simple analogues of neural and synaptic function. **The** type of behavior **ANNs** are designed to produce **also** bears varying degrees of correspondence to the behavior of biological networks. **ANNs** are thus not, in general, **models** of biological neural networks.

**The** highly distributed architecture common to both artificial and biological **neural** networks is becoming increasingly attractive to computational theorists, partly because it implicitly subdivides a large **computationally** intensive problem **into** many smaller (and often similar) components. Such subdivision facilitates implementation of a system on distributed parallel hardware, which in turn facilitates the real-time solution of highly complex problems. The distributed architecture of neural networks **also** encourages the development of new distributed methods of solving problems.

## SYNAPTIC LEARNING RULES

**Perhaps**, the **single** most enticing capability of biological neural systems is **their** general ability to learn, or adapt, in such a **way as** to **improve** their performance. Even before the synaptic doctrine appeared, Herbert Spencer postulated in 1862 that the ability of **one** cell to excite another could change **as** a **function** of prior activity, and **that** this was the **basis** of memory (Levy and **Desmond**, 1985, p. 105). Later, the famous Hebb **synapse** model (Hebb, 1949) emerged, in which synaptic efficacy changed in proportion to the product of **presynaptic** input and postsynaptic neural output signals. However, since both signals are non-negative, the efficacy of the Hebb synapse can **only** increase, which is now usually regarded as inadequate. Recognition of this apparent deficiency has spawned numerous varieties of Hebbian, **pseudo-**

Hebbian and anti-Hebbian learning rules, **all** of which are still relatively simple. The former, more Hebbian-like **learning rules**, supplement Hebbian increases in synaptic efficacy with **some** type of rule enabling decreases, while the latter also deviate in their conditions for increasing synaptic efficacy. The debate regarding whether **Hebbian** learning **occurs** within biological **neural** networks lingers on even **today**, with some **claiming supporting** empirical evidence (Singer, 1985), and others claiming an absence of such evidence (**Camardo**, Siegelbaum, and **Kandel**, 1984).

The current resurgence of activity in **ANN research**, which is largely dominated by Hebbian-derived learning rules, prompted **the** following comments from Allen **Selverston** (1988):

"For biologists, the hyperbole surrounding the promise of **artificial** networks being able to duplicate **mental** functions produces **a** certain feeling of déjà vu. Such **computational** schemes have come **and** gone with a period of about fifteen years **and** it is premature **to** depict the **present** upswelling of activity as anything more or **less** than previous attempts to **explain mentation**." (Selverston, 1988, p. 109).

"While the use of the mammalian brain as a **model** for computer engineers may sound **plausible** and even somewhat glamorous, the fact is **that** we really know very little about the brains detailed microcircuitry. In addition, neurons with grossly **oversimplified** physiological properties, and synapses whose **main** function seems to be only blind obedience to Hebbian learning rules, have come to be **the** foundation of this new approach." (Selverston, 1988, p. 109).

Selverston (1988) **goes on** to **list** numerous **cellular** and synaptic properties which are generally ignored by **ANN** researchers. Since they are primarily concerned with overall behavioral performance, ANN researchers seek to introduce only the minimum component complexity required, **as** additional complexity **would** only unnecessarily complicate an already difficult task, and consume **additional** processing resources. The main problem, is that no one **really** yet **knows** precisely which cellular and synaptic properties **are** functionally critical, and which are superfluous. Furthermore, the **still**

rudimentary **physiological** understanding of neural function (let alone neural network function) precludes the use of a biologically correct and complete neural element in ANN systems.

The abundance of behavioral data available **from** studies of associative conditioning, while still far from fully understood, seems to be a more **reliable** and direct source of data to aid in the design of ANN elements than does the limited detailed physiological **data** of neuron function currently **available**. In any case, **it** is an animal's behavior which is directly shaped by reinforcement (**discussed** further below), and not the internal means by which it produces its behavior. Many differing mechanistic solutions **are** probably capable of **generating** the required behavior. Indeed, different and sometimes overlapping hardware solutions **to** common behavioral problems **exist** both within an individual's nervous system, and among different species.

It is therefore **not** necessarily a **mistake for** ANN neuronal models to generally ignore most specific **cellular** and synaptic properties of biological neurons. However, the already apparent complexity of real neurons should at least suggest to ANN researchers that elements of greater complexity than are currently being utilised may be required to produce the desired adaptive behavior.

## **TYPES OF FEEDBACK**

ANNs may be subdivided into three main categories, on **the basis** of the type of feedback required **to** guide learning. The three categories of feedback are none, full, and reinforcement feedback. Each is discussed in turn below,

### **No Feedback**

Those ANNs which require **absolutely** no **feedback** include **Hopfield** networks (Hopfield, 1982), Boltzman machines (Ackley, Hinton and Sejnowski, 1985), **and the Cognitron** and Neocognitron (Fukushima, 1975, 1988). **The first** two **utilise** adaptive equilibrium processes in which symmetrically interconnected elements interact to enable the total network to reach **some** minimal energy state, or

"cost function". However, these types of systems have not been shown to be relevant to the learning mechanisms underlying biological neural network function (Klopf, 1987, p. 99). Fukushima's network models automatically form "clusters" of similar spatial pattern inputs, according to an inherent (built-in) measure of similarity. These clusters can sometimes provide sufficient behavioral specificity, because "similar" input patterns usually require similar behavioral responses. However, the environment is ultimately the best determinant of what degree of discrimination is required in various circumstances, and should therefore play some role in guiding the formation of such clusters.

Barto comments that "Unsupervised learning is more accurately regarded as supervised learning with a fixed, built-in teacher", and that "A supervised system is in fact more adaptive than is an unsupervised system because it forms clusters in order to solve problems posed to it by environmental contingencies rather than to solve a problem of its own." (Barto, 1985, p. 239). As such, the "unsupervised" learning exhibited by systems which utilise no environmental feedback, is really less powerful than that of systems which utilise such feedback.

#### Full Highly Specific Feedback

ANNs of this type require detailed specification of the desired output on every learning trial. The most popular of these types of systems in recent times are those utilising "back propagation" (Werbos, 1974, 1988; Rumelhart, Hinton, and Williams, 1986). In brief, these systems provide a means of propagating the specific difference between desired and actual output response of the output layer, back through the intermediate layers so that they can adjust their performance in a manner which reduces the output error signal. These types of networks tend to be algorithmic in nature, usually reaching an optimum solution quite reliably, for particular types of prespecified problems.

While an external teacher can play an important role in natural intelligence at higher cognitive levels, it does not usually do so at fundamental cognitive

levels. Furthermore, the overall gradual refinement of response by explicit comparison with desired response can only **allow** a system to **attain** a prespecified **response** to particular inputs. There is no scope here for the system to discover **new** improved responses on its own, or to adjust to novel situations for which a **prespecified** response cannot be provided.

### **Reinforcement Feedback**

This type of system requires some feedback from the environment, but does not require an external teacher. Explicit desired-response feedback is replaced by a less specific reinforcement feedback signal. In the context of operant conditioning the reinforcement feedback is a **simple scalar** signal which indicates the degree of **success** of the system's response, without providing **any** specific information regarding how behavior **should** be altered to improve performance. It is the **responsibility** of **the** system to determine such details. These **systems** tend to be heuristic in nature, providing reasonably **good** adaptive performance in new and changing circumstances, **where** algorithmic **optimal solutions** are unable to be predetermined.

In the case of classical conditioning, a reinforcement feedback system is **phylogenetically** predisposed to associate **particular** stimuli (USs) with particular and **generally** appropriate responses (URs), and then to learn to produce (normally similar) responses (CRs) following perception of previously neutral antecedent stimuli (CSs). Hence, classical conditioning more directly extracts information from the reinforcement feedback **signal** concerning the type of response to be produced, than operant conditioning, but less fully exploits the potential to produce successful new types of response. In **the** context of **classical** conditioning, the reinforcement feedback signal **may** appear to **act like** an external teacher, though no explicit active **external** teacher actually exists - only the **contingency-rich** external environment.

Reinforcement feedback systems operate in a manner most analogous to natural intelligent systems. In fact, it is the very **same** reinforcement feedback described here that is used to guide learning in the associative conditioning experiments which provide the behavioral specifications for the research

described herein. As a consequence, **theories** of classical associative conditioning (e.g., Konorski, 1948, 1967; **Rescorla** and Wagner, 1972) are also theories of reinforcement feedback systems. Similarly, **neuronal** models of classical conditioning (e.g., **Barto and Sutton**, 1985; **Klopf**, 1987), including ACE, are **also** AI models of reinforcement learning at the single element level. A small scale network of reinforcement feedback elements has also been shown to produce reasonable results using reinforcement feedback (**Barto, Sutton, and Anderson**, 1983).

### **INTRA-ELEMENT MECHANISMS**

Neural functions are mediated by chemical, and electrochemical processes. These differ considerably in nature from the electronic processes underlying digital computer component operation, and can appear to be simpler, at least to **the casual** observer. However, this simplistic view is at odds with even the current partial knowledge of **intracellular** processes, which is revealing an increasingly complicated network of chemical interactions (e.g., **Carew, Abrams, Hawkins, and Kandel**, 1984). **The** spatial interconnection pathways **which** form the substrate of most neural interaction, and which already make **good** use of all three spatial dimensions, are augmented by an impressively diverse **set** of chemically discrete interconnections which share physical pathways. This enables a very complex system **to** occupy a tiny volume, by reducing the number of **physical** structures required to implement discrete spatial pathways which tend to consume more **space**. In addition, the relatively slow rates of change of accumulated substances, due to both slow production, transport, and consumption mechanisms, can become a critically **important** aspect of intraneural function. Hence the mechanistic complexity of a **biological neuron** is even greater than **its** already intricate physical structure **alone** might suggest.

**The** basic building blocks from which complex neural mechanisms are formed tend to be dominated by operations approximating analogue multiplication and summation, which affect rates of production, **accumulation**, consumption, or transport of chemical quantities. **Also**, the use of **cumulative** quantities which cannot physically **take** on negative values introduces **non-linearities** similar to rectification, in which only positive values are allowed. However, as **alluded** to

above, such **simple** types of interaction may be **combined** in relatively sophisticated **local** circuits to produce much more sophisticated behavior.

**In** summary, mathematical models of intra-neural mechanism would be dominated by operations, which although **simple** in themselves, **may** be combined to form very sophisticated systems. The **same** approach is **adopted** for the development of the intra-element mechanisms documented herein, in the expectation that it **may** help guide **the** research along a type of path which is demonstrably capable of supporting **classical** conditioning behavior.