Walker, S. F. (1992) A brief history of connectionism and its psychological implications. In Clark, A. and Lutz, R. (eds) *Connectionism in Context*. Berlin: Springer-Verlag. 123-144

A brief history of connectionism and its psychological implications

Stephen F. Walker (**Home page**)

NB This file presents the essential content of the Chapter published in Clark and Lutz (1992), but not in the same format (i. e. this is not the actual Chapter as printed).

It would be possible to be fairly brief by defining connectionism as computational theories of neural networks — although it is conventional to acknowledge pioneers such as Hebb (1949) and Rosenblatt (1962) most of the impetus behind this form of connectionism has built up in the last 10 years (Hinton and Anderson, 1981; McClelland *et al*, 1986: cf Anderson and Rosenfeld, 1988 and Papert, 1988). My intention here is however to look at a selection of what might be called precursors to connectionism, mainly psychologists between 1850 and 1950 who espoused views which I shall claim had significant elements in common with the current approach.

The aspects of modern connectionism which have precursors include empiricism, associationism and materialism. Empiricism emphasises the role of culture, education and life experiences as determinants of human abilities and proclivities, while associationism identifies pairwise links between individual elements of experience, either subjective or behavioural, as the main process of such psychological change. Materialism is the theory that mental phenomena are physically determined, and more particularly that the mechanisms of the mind are identical to those of the brain. Strongly associationist theories of human perception and cognition were developed by the 18th philosophers Hume and Berkeley, and in the first half of this century the associationism within the theories of behaviourists such as Watson (1913) and Hull (1943) was widely influential before being generally abandoned. It tends to be the critics of connectionism who point to the characteristics it shares with such previous, presumptively flawed, conceptions. Fodor and Pylyshyn (1988) for instance have as a major theme their gnawing sense of deja view in putting forward arguments which seemed conclusive against the Connectionism of Hebb, Osgood and Hull twenty or thirty years ago (p. 49 e. g. see Fodor, 1965). Going further back they fear that adopting current connectionist premises will lead to a psychology not readily distinguishable from the worst of Hume and Berkeley (p. 64). In agreeing with such critics of recent connectionism that it appears to recapitulate many earlier arguments I do not wish in general to imply that it is therefore necessarily mistaken: rather to suggest that, insofar as new connectionism has virtues, both the weaknesses and the strengths of earlier versions deserve re-examination. A particular issue which has yet to receive much attention from the new wave of its proponents is that of the practical implications of connectionism: as a paradigm of the acquisition of knowledge and behaviour change it could have something to say about techniques of psychotherapy and educational practice. Earlier connectionists such as Thorndike (1905) and Hull (1943) were certainly convinced of this, and though they may often have been wrong it is conceivable that their efforts to find applications for connectionist theory could be instructive.

At the outset however, it should be acknowledged that there is a good deal new about the new connectionisms. First, connectionism is now sometimes a synonym for Parallel Distributed Processing: few earlier versions had explicit theories about how networks of connections would work, although appealing to the numerosity of available brain connections is hardly novel. Second, connectionism can now rely on the greatly increased availability of computational analogues and simulations as practical demonstrations of the powers of arrays of connections, as opposed to vague speculations. A previous empirical base for theories about how the effects of experience could be mediated by changed connection weights was in experiments on animal learning. There are those today who are sceptical about the claims made for the capacities of simulated neural nets, but these provide a source of evidence very different in kind from that previously appealed to - networks so far proposed may have limitations for instance in their capacities for learning the past tense (Pinker and Prince, 1988), but they have certain obvious advantages over the laboratory rat for modelling this sort of human cognitive capacity. Finally, some earlier versions of connectionism were strongly Behaviourist, stressing connections between identifiable stimuli and responses, which was a severe theoretical restraint. Papert (1988; p. 9) still sees connectionism as behaviourism in computer's clothing but makes it clear that it is the empiricism in connectionism he is most concerned with — neo-connectionism is empiricist in as far as it is about learning rules, but not behaviourist in the sense having an excessive concern with behavioural measurements.

Given these differences it may seem to be stretching a point to include empiricist philosophers and animal learning theorists in the same category as the San Diego PDP group. I therefore attempt to show below in synoptic accounts of earlier work that there are indeed putative similarities to modern connectionist approaches, before reviewing the relationship between them and considering the extent to which recent developments mean that old reservations about connectionism can be safely withdrawn.

Connectionist assumptions in earlier psychologies

Spencer's Principles of Psychology (1855/1899)

There is no shortage of associationists, empiricists, and materialists, in various combinations, before the 19th Century. Apart from Hume and Berkelely as associationists, in the 18th century Hartley (1749) can be used as an early example of associationism tied to connections between "brain vibrations", and de la Mettrie's *Man a Machine* (1748) was a thoroughgoing materialist version of associationism. Spencer is of greater interest partly because of the emphasis he placed on physiological detail and biological evolution (however inaccurately) but also because he had a strong influence on both psychologists and physiologists (see Boakes, 1984). The datedness of Spencer's physiology for the modern reader may be balanced by the fact the Sherrington (1906) still quoted him as an authority whose ideas were being confirmed and Pavlov (1927; p. 9) credited him with the initial idea of interpreting behavioural reactions as neural reflexes (both Sherrington and Pavlov won Nobel prizes for physiological work on reflexes). Hughlings Jackson, the English neurologist mentioned by McClelland *et al* (1986, p. 41) in the context of PDP, frequently quotes Spencer and made extensive use of his notion of neural hierarchies (Jackson, 1888/1931).

Spencer's neural nets

For Spencer the "Data of Psychology" started with the structure and function of the nervous system, and all subjective mental phenomena, including those of reasoning as well as perception, were to be related to neural mechanisms (Education, Ethics, Sociology and so on being analysable in terms of similar basic principles). Examination of the facts of "minute nervous structure" indicated branching and intermingling processes. The vague and general deduction is that "nervous communication" within a matrix of cells *changes* as a function of its activity (vol II; p. 545). All intelligence, from its lowest to its highest forms, can be interpreted in terms of associations between successive psychological states (I. p. 425) which are driven by "the strengths of the connexion between the internal states" (I. p. 409). The most basic laws of association apply to "immediately connected changes, and small groups of changes". (I. p. 408). Spencer thus comes very close to saying that "*knowledge is in the connections*" (Rumelhart and McClelland, 1986a, p. 132). The strength of the connections varies with use by something analogous to the Hebb rule. But perhaps the most striking aspect of Spencer's treatment is his emphasis on the importance of "concatenations of changes" (I. p. 408) and his explicit reference to neural networks (e. g. "network of their connexions" vol I, p. 544; he more frequently used the synonym "plexus" for network, e. g. vol I pp. 538-553 passim). The analysis of how networks might operate was of course superficial, but Spencer went so far as to sketch a diagram of the most basic neural net of much the same kind as may be found in more recent treatments (see Figure I).

Localization.

Despite relying on interactions between networks in higher and lower centres, Spencer was firmly in favour of localization of function, as "the law of all organization whatever" (I. p. 571- 3), while denying the phrenological view that defined faculties could reside in local cortical regions. His argument for what we would call distributed representation was partly in terms of the overlapping and entanglement of different networks (I. p. 574) and in terms of "diffused localization" in any individual net (I. p. 362) There is an explicit rejection of the "grandmother cell" concept of perception: "No one excited fibre or cell produces consciousness of an external object" — a different collection of cells may be involved depending on slight differences of presentation. The mechanical analogy appealed to is that of a piano with 100 keys. Five keys struck together yield over 75 million combinations, 50 keys a thirty figure number, and we are invited to consider the possibilities presented by the excitement of clusters of retinal elements — not as an example of the combinatorial explosion, but as an illustration of the relatively large number of perceptual experiences it would be possible to relate to a relatively small central network of cells (vol I, p. 562-3).

Limits of experience.

His concern with the evolution of neural organization meant that Spencer rejected extreme versions of empiricism - the "experience hypothesis". Pre- established relations exist in the nervous system "answering to relations in the environment", and more complex kinds are evident in the gradual changes in cognitive ability seen during childhood, which are "more attributable to the completion of cerebral organization, than to the individual experiences" (I. p. 469). Spencer would not therefore have had any difficulty bringing anti- empiricist views such as those associated with Piaget and Chomsky into his system. (It is all the more regrettable that it is in this context that he tends to appeal, unnecessarily, to Lamarkian evolutionary mechanisms, and inherited racial differences)

William James's Principles of Psychology

William James's *Principles* has been reprinted innumerable times since 1890, most recently in the Harvard paperback edition of 1983. As he had used Spencer's work with the same title as a textbook for many years it is not surprising that James' *Principles* share with Spencer's the major features of interpreting subjective mental phenomena in terms of brain activity, and describing the brain activity in terms of neurally mediated associations. There is a substantial chapter on "Functions of the Brain" (beginning with the frog's brain and going on to human aphasia and visual agnosia), and neural physiology is referred to in the context of more cognitive topics, not only mental association but also willed action (James 1890/1983 p. 1190).

Figure 1. ***Connectionist diagrams***

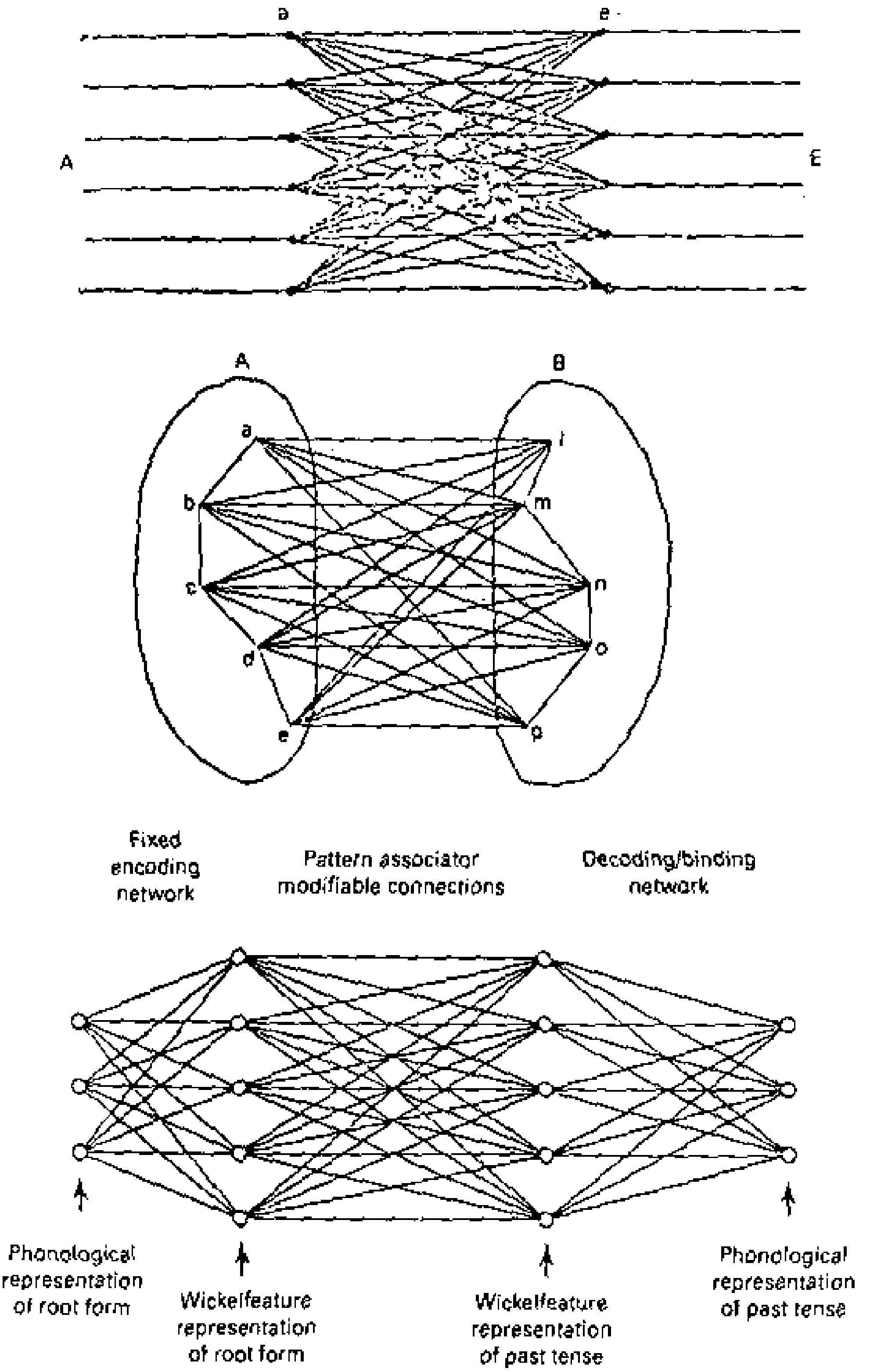

Fig. 8.1. Connectionist diagrams. *Top*: from Spencer (1855/1899 Vol. 1, p. 527) in a section on "Genesis of simple nervous systems", used to show that even in a simple ganglion there may be many modifiable "lines of connection" between afferent (A) and efferent (E) fibres. *Middle*: from James (1890/1983, p. 537) used to illustrate compound association between one idea (A – remembering the end of a dinner party) and another (B – afterwards walking home through a frosty night). *Bottom*: from Rumelhart and McClelland (1986b, p. 222), a model which learns to produce output codes corresponding to the past tense of verbs when presented with phoneme strings coding the verb roots. The central pattern associator is the primary focus of the model, which in practice had 460 inputs of phonological triples, and 460 similar outputs.

from Walker (1990)
1992

James and the Hebb rule.

For present purposes, the most notable example is the very clear assertion that all psychological phenomena of association (involving similarity and contrast, as well as contiguity) derive from the "law of neural habit". This is stated as follows:

> "*When two elementary brain-processes have been active together in immediate succession, one of them, on reoccurring, tends to propagate its excitement into the other*" (James, 1890/1983, p. 534)

In modern connectionist circles this sort of thing is referred to as a "Hebb rule" with similar wording:

> "When unit A and unit B are simultaneously exited, increase the strength of the connection between them. " (McClelland *et al*, 1986; p. 36)

Hebb's own version was more explicitly neural, with an axon of cell A gaining additional efficiency in being able to fire cell B if it "repeatedly and persistently takes part in firing it" (Hebb, 1949; p. 62). James had the distinction of formulating in 1880 the "law of forward direction" of neural conduction, but his detailed theory of neural associations relied on a concept of "drainage". A cell when activated discharged a current, and thereby drained cells behind it in a circuit of some form of tension: this drainage process was responsible for making the path between the two cells more "pervious". (James, 1890/1983; p. 1186). Hebb's more abstract notion that a "metabolic change takes place in one or both cells" which increases the ability of the first to fire the second, is less of a hostage to fortune. (Hebb, 1949; p. 62: see Lynch, 1986 for recent hypotheses on metabolic mechanisms of synaptic modifiability)

James and distributed representations.

James was generally persuaded of coarse localization of sensory functions in the cerebral cortex. But 19th century associationists typically considered psychologically realistic associations to be between compounds of distributed elements. James followed many in referring to Hamilton, an otherwise obscure Scottish philosopher, who defined all association as between constituent elements and the wholes of which they are part. (Termed "redintegration": cf the "auto associator" in Rumelhart and Zipser, 1986, p. 161). James felt that ordinary associations did not require a great deal of redintegration, but, in a confusing section on the association between two individual thoughts A and B, inserts a diagram which be interpreted as involving pattern association between pools of distributed micro-features in the modern manner (Fig 1.)

Serial associations.

An example of James's theorizing at the level of individual neurons is given in Figure 2. Here the behaviour to be explained is the reciting of the alphabet and the three sensory cells shown correspond to hearing the sounds of the letters A B and C and the motor cells to the activity of saying the letters. This is of course extremely crude early connectionism.

Figure 2 — *William James's motor circles*

moment later, the effect of M^a's discharge comes back by the afferent nerve and re-excites S^a, *this latter cell is inhibited* from discharging again into M^a and reproducing the 'primordial motor

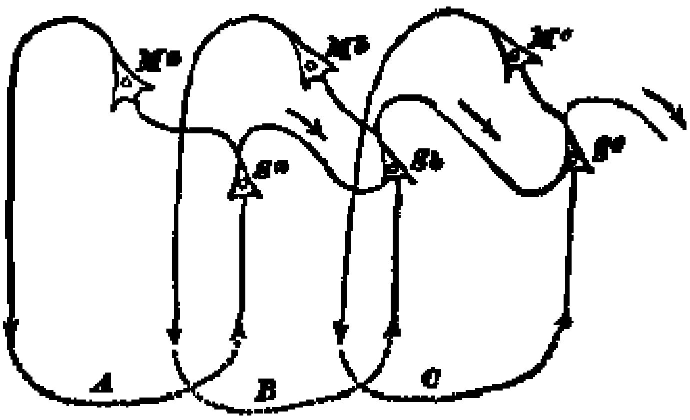

FIG. 90.

circle' (which in this case would be the continued utterance of the letter A), by the fact that the process in S^b, already under headway and tending to discharge into its own motor associate M^b, is, *under the existing conditions*, the stronger drainage-channel for S^a's excitement. The result is that M^b discharges and the letter B is pronounced; whilst at the same time S^c receives some of S^b's overflow; and, a moment later when the sound of B enters the ear, discharges into the motor cell for pronouncing C, by a repetition of the same mechanism as before; and so on *ad libitum*. Figure 90 represents the entire set of processes involved.

Inhibition and Prudent Conduct.

A surprising lack of hesitation in switching back and forth from social to cellular theory is indicated by the presence of Figure 2 in a section on "The Education of the Will". Early in James' book references to inhibition are in a physiological context, and he affirms in the chapter on Will that it operates at a cellular level for stopping motor processes and also in the "drainage" mechanism of association: "Inhibition is therefore not an occasional accident; it is an essential and unremitting element of our cerebral life" (pp. 1185-6). However, more broadly inhibiting irrelevant impulses is part of "training to moral and prudential conduct" and it is very clear that James, along with many of his contemporaries, saw inhibition as socially desirable for constraints on individual impulsivity in general and for self-discipline in respect of alcoholic or sexual over-indulgence in particular (pp 1143-1150). The concern with social controls is also apparent in the theory that higher centres typically operated by inhibiting lower centres (p. 768). It is clearly absurd to conflate neural and psychological inhibition, but it would also be a technical mistake to react against this by denying that inhibition exists as a neural phenomena, as did Hebb (1949, pp. 208-215) - though James may have emphasised it for the wrong reasons, inhibitory relations between cells are of course an accepted feature of modern neurophysiology.

Good and Bad Habits.

An even more extreme example of interaction between sociological and physiological theory occurs for the concept of habit. This is initially introduced by James as an obvious example of sensory-motor reflexes, concatenated for more complex behaviours and an organic and naturalistic product of repetition. However, there is soon a transition to "the ethical implications of the law of habit....... Habit is thus the enormous flywheel of society, its most conservative agent. It alone is what keeps us all within the bounds of ordinance, and saves the children of fortune from the envious uprisings of the poor" (pp. 124-5; one trusts this is not a theme that will be re-capitulated). Habits are personal, professional and intellectual, as well as political, but they are nevertheless all neurally based: "The great thing, then, in all education, is to make our nervous system our ally instead of our enemy" (see James, 1890/1983 pp. 113-130).

There is thus no difficulty in establishing that there have been precedents for a strategy of interpreting all psychological data that come to hand in terms of theories of neural activity. James is representative of prolific army of authors with broadly similar views, many of whom he quotes at length, who wrote volumes with self-explanatory titles such as "Mental Physiology", "The Physiology of Mind" and "The Brain as an Organ of the Mind".

Thorndike's Connectionism

Thorndike was a student of James', who called himself, and was called by others, a connectionist (Thorndike, 1949; Hilgard and Bower, 1975). He represents a point at which it can be said, in current terms, theories of neural association became sub-symbolic rather than implementational. The term "implementational" is used by critics (e. g. Fodor and Pylyshyn, 1988), who do not wish to deny that the brain contains neurons and neuronal connections, but who dispute the claim that this is relevant to psychological theories: thoughts may be implemented by neurons but the logical structures of mental symbol manipulation are supposedly independent of this fact. Connectionists such as Smolenskly (1988) argue that even if the physiological details are ignored, theories of cognition must be based at the "sub-symbolic" level: the mechanisms of thought being finer grained that any introspectively accessible units.

The connectionism of Spencer nor James was implementational in that it supported rather than disturbed the mentalist cognitive architecture of their contemporaries. And James' "Stream of Thought" is not far distant from Fodor's "Language of Thought" (1987) - introspective data on beliefs, desires and narratives were operating on symbolic level with plenty of syntactic and semantic structure, and the neural theories can be seen as simply providing explanations of the mechanisms which allow for this.

With Thorndike, however, it looks as though neural connections are becoming a substitute for, rather than an mechanism of, ideational processes. This is unequivocal in the case of the theory of animal learning he began with (Thorndike, 1898). He performed experiments in which cats acquired skills at getting out boxes in which they were confined by pressing catches or pulling loops, and concluded (erroneously, according to more recent authorities) that the skill took the form of direct connections between incoming sense impressions and the impulse to perform the required response. The cats were not supposed to have the sort of beliefs and desires that Fodor (1987) attributes to his Greycat, and not even an association between the learned response and its consequence of getting out of the box (and being allowed to eat fish). Technically Thorndike proposed a teaching rule, not unlike back propagation, by which the positive outcome strengthened or "stamped-in" connections between an immediately preceding behaviour and stimulus input present at the time. (He referred to this as the "Law of Effect")

Thorndike's concern with fine-grained mechanisms of association is apparent in his *Elements of Psychology* (1905), which is notable for its comprehensive neurophysiology (there are 57 plates and figures showing axons and dendrites of varying cell types and so on) and also for its fairly early use of the term "synapsis" for the point of connection between neurones (p. 141). The numerosity of the neurons themselves is emphasized, and the student is warned against the notion that individual neurons correspond to individual ideas.

Thorndike went on to become an educational psychologist at Columbia, and extrapolated connectionist principles to human cognition. "The situation- response formula is adequate to cover learning of any sort, whether it comes with ideas or without, conscious or unconscious" (Thorndike, 1931, p. 132). The consequence of this was that repetition and drill, with objective and quantified tests of performance became recommended educational techniques. "Frequency of a connection" become an experimental variable, and Thorndike's most widely read publication was undoubtedly *The Teacher's Word Book of 30,00 Words* (Thorndike and Lorge, 1944), an expanded version of one for which he had counted up the frequency of occurrence of 20,000 words in commonly available written material (word-frequency remains a widely used variable in experiments by cognitive psychologists). The differing implications of James's and Thorndike's associationism is illustrated by the back-handed compliments in James's foreword to Thorndike's *Elements of Psychology* (1905). In the book there were many paragraph headings, numbered exercises, and questions and directions in small print designed to elicit appropriate responses from the reader. These do not today seem especially remarkable. But James clearly disapproved of such connectionist didactic methods and "every other up-to-date device for frustrating the natural movement of the mind when reading". Could it be, he wondered, that Thorndike had become "a high priest of the American text-book Moloch, in whose belly living children's minds are turned to ashes?"

Pavlov's Theory of the Cerebral Cortex

Pavlov's name is indelibly associated with conditioned reflexes of the simplest "knee-jerk" type. Most of the experimental work he reported involved the measurement of salivation, and his earlier studies were on gastric and pancreatic secretions. However he introduced the book published in English in 1927 by saying it represented 25 years of research on the physiology of the cerebral cortex (note its subtitle: his work was known much earlier, for instance from a lecture given at Charing Cross Hospital in 1906, and published in *Science* and *The Lancet*).

Of most obvious relevance to current connectionism is Pavlov's reliance on hypothetical neural processes of excitation and inhibition ("the currency of our systems is not symbols, but excitation and inhibition": Rumelhart and McClelland, 1986a; p. 132). Pavlov was eventually far more concerned with the topological dynamics of these processes than with connections per se, which he tended to take for granted. He fleetingly referred to telephone switchboards, pointing out that he could call his laboratory directly from home by his private line — a fixed connection, or go though the switchboard, which would involve a new connection. Conditioned or "*connection reflexes*" he saw as the formation of new paths in the cerebral cortex, an unproblematical development of "Bahnung": used to refer to facilitating effects between spinal reflexes (Pavlov, 1927; pp. 25-6). Others of course (in particular Konorski: 1948, 1967) in the Pavlovian tradition developed more detailed theories involving synaptic changes, and the existence of physically separate systems of inhibitory and excitatory synapses gives a more precise basis for the two physiological processes.

The nature of Pavlov's approach was, as he himself said, closely allied to that of Sherrington (1906), much of whose work was on spinal reflexes, where excitation had long been measured as activation of a reflex, often muscle contraction, with inhibition as its absence or opposite (in some cases sensory stimulation leads to muscle flaccidity). The essence of Sherrington's influence is the view that the isolated reflex arc, of a telephone connection type, is a misleading abstraction, even in the spinal cord, where individual reflexes interact, inhibiting or facilitating each other. Sherrington's emphasis was thus that even the spinal cord acts as an integrating whole. A major area of experimentation was in the spread or "irradiation" of spinal reflexes. Locally, there is an unsurprising rule of spatial proximity - stimulation of a particular sensory root elicits most easily reflexive movement innervated by its own motor root, next those corresponding to adjacent motor roots. There are also more distant interactions e. g. stimulation of a forelimb can elicit movements in the hind limb. Many results, and rules, went back some time, the work of Pfluger in 1853 being very widely cited. Sherrington's view was that "These so-called 'laws' of reflex irradiation were so generally accepted as to obtain a doctrinal eminence which they hardly merit" (1906, p. 164), but he continued to use the term.

There are thus historical reasons why Pavlov (1927) spent so much time discussing the spatial irradiation of excitatory and inhibitory processes in the cortex, although his hypotheses were usually closely tied to experimental observations. The clearest examples arose when mechanical tactile stimuli were applied to different points on a dog's body surface. If a light touch on a dog's front paw always occurs before it receives food, it comes to salivate in response to the touch there. It is also likely to salivate in response to touches at other places, in decreasing amounts according to distance from the front paw - these are accessory reflexes, produced according to Pavlov because excitation at the cortical point corresponding to the peripheral stimulation spreads out across the cortex (1927, p. 186). More complicated experiments using sequential combinations of tactile stimuli demonstrated greater interaction between excitatory and inhibitory stimuli when they are spatially closer together (Pavlov, 1927; p. 153 and 181). Responses to sound stimuli suggested that there was a also geometric projection of "the acoustic analyser" on the cortical surface and Pavlov adopted the general position of explaining experimental results in terms of topological relationships between points on the cortex, with wave-like phasic irradiation and concentration of both excitation and irradiation, over periods of seconds and minutes, being responsible for temporal variation in interactions between stimuli.

This led to his concept of the cortex as a "mosaic of functions...... integrated into a complex dynamic system" (Pavlov, 1927; Ch. XIII). All individual points of the cortex were presumed to interact. This is clearly in the same vein as "parallel distributed processing" and quite close to neo-connectionist treatments those such as that of Ballard (1986) which stress the topographical organization of cerebral cortex.

Implications of Pavlovian conditioning.

Pavlov sometimes appeared to extend his theories to human psychology in a very mechanistic way: "It is obvious that the different kinds of habits based on training, education and discipline of any sort are nothing but a long chain of conditioned reflexes" (1927, p. 395). However, although reductionist, his approach is flexible when evidence required it to be so. Even on the basis of experiments on dogs, he frequently stressed a distinction between "higher" and "lower" brain functions, discusses perception as a matter of analysis and synthesis in higher and lower centres.

The "knee-jerk" view of Pavlov is therefore misleading, and recent theories of animal conditioning propose that the process is often not a matter of reflex elicitation but a more cognitive "learning of relations among events" (Rescorla, 1978). When Pavlovian conditioning arrangements are used with human subjects, verbal knowledge of such relations may be more important than the direct effects of stimulus associations (Davey, 1988). Pavlov himself acknowledged that speech provides an additional "signalling system" which may replace the signalling functions of real events, and remained willing to entertain the importance of distinctions between conscious and unconscious factors in human thought (1927, p. 410). For physiological and metabolic reactions themselves though, involuntary Pavlovian processes can resist linguistic intervention: for instance, anticipatory nausea and taste aversions which are routinely experienced as side effects by cancer patients undergoing chemotherapy can be interpreted as conditioned responses, as can other aspects of environmentally influenced reactions to drugs (Cary and Burish, 1988; Stewart *et al*, 1984)

Watsonian and Skinnerian Behaviourism

Papert (1988) has suggested that connectionism now "promises a vindication of behaviourism" and it is therefore worth mentioning that the more radical behaviourists would have had no interest whatsoever in explaining cognitions in terms of brain-processes, since they were interested in neither - their point being that only overt and observable activities matter. Watson (1913; 1924), wishing to make psychology more objective and scientific, outlawed introspection and anything related to it: "thought processes are really motor habits in the larynx" (1913, p. 174). Human cognition was thus distinguished only as "language habits". Watson included chapters on the nervous system in his text- books, but everything psychological was kept peripheral: thinking was sub vocal talking, or other unobservable motor activity. Psychologists should be concerned solely with gathering and organizing data which would enable them to predict the response given the stimulus, or vice versa.

A somewhat similar position has been defended for many years by Skinner (1953; 1989), who may be said to treat not only brain science but also cognition as merely implementational with respect to behaviour. There remain some practical arguments in favour of objective empirical research (Watson made a fortune after retiring from academic life and going into advertising, with the philosophy of finding the stimuli which would produce appropriate consumer reactions), but these appear to be orthogonal to most of the issues raised by either old or new versions of connectionism.

However perhaps it is fair to say that recent connectionists seem to share the radical behaviourist's distrust of thought and the thinking self: "there is *no central executive*" (Rumelhart and McClelland, 1986; p. 134). And the problems that PDP systems are designed to solve are a version of the Watson type — how can we simulate a system which, given the stimulus, will produce the predicted response? (For pattern recognition, past-tense learning, and so on).

Hullian Stimulus-Response Theory

Clark Hull published an ambitious book in 1943, on the assumption that absolutely all aspects of psychology are generated from the same primary laws, and that "differences in the objective behavioural manifestations are due to the differing conditions under which habits are set up and function". He put forward a formal theory, based on equations which were intended to be empirically testable, which was widely influential — in the 1940's he was cited in 70% of papers published in major journals dealing with learning (Spence, 1952). Hull certainly qualifies as a connectionist, since he assumed that all learning was automatic, and a matter of "The strengthening of innate receptor-effector connections" (1943, p. 69). Many of the proposals embodied in his numerous equations in fact turned out to be empirically refutable, but is of some historical interest that his formula for the growth of stimulus-response habits can be seen as a primitive form of the generalized delta rule for back- propagation in neural nets. Sutton and Barto (1981, p. 155) examined in some detail the relationship between an equation introduced by Rescorla and Wagner (1972) and the rule for an adaptive element proposed by Widrow and Hoff (1960) and concluded that "the Widrow-Hoff rule is essentially identical to the Rescorla-Wagner equation"; and Rescorla and Wagner themselves introduced this as "a modification of Hull's account of the growth of ${}_{s}H_{r}$" (Rescorla and Wagner, 1972; p. 75). Versions of the relevant formulae are given below.

Hull (1943; pp. 115-120)

$$\Delta H = f(M - H) \qquad (1)$$

Rescorla and Wagner (1972, p. 75)

$$\Delta V_a = x_a\, y_r\, (M_r - V_{ax}) \qquad (2)$$

Rumelhart et al, 1986a; p. 322)

$$\Delta_p W_{ij} = nI_{pi}\, (T_{pj} - O_{pj}) \qquad (3)$$

Hull wrote out equation (1) in a more elaborate form, but his procedure for calculating the strength of a habit was very straightforward. A maximum or target value (M) was selected - the theoretical significance of this is that its level was determined by the effectiveness of the teaching agent. The habit starts at zero, but on every occasion when it is strengthened (when the conjunction of stimulus input and response output is followed by motivational effects) the increment to the connection is a constant fraction (f) of the *difference* between the previous value (H) and the maximum (M).

This was partly a matter of convenience for Hull: he took observed response habits to correspond to connection strengths "hidden within the complex structure of the nervous system" and the equation was based on behavioural data.

Rescorla and Wagner (1972) were also concerned to give an account of behavioural results in Pavlovian conditioning paradigms. Hull's equation is for one stimulus and one response. A wide variety of phenomena are observed when combinations of stimuli are studied in animal learning experiments. Of relevance here is a reliable empirical result, known as "blocking", seen if one stimulus, say a buzzer, B, is paired with food many times before a combination of B plus A (say a light) are thereafter joint signals for the food — the light if subsequently tested separately has little or no behavioural effect (Mackintosh 1983). Equation (2) was designed to accommodate this result. The constants x and y are fractions which determine learning rate for a given stimulus "A" the event "R" it predicts, and can be ignored. There is a terminological change in that measured behaviours are attributed to the "associative value" of stimulus A (V_a) rather than to a habit. Its maximum is determined by the predicted or "teaching" event "R". But the major alteration to Hull's formula is that associative increments are proportional to the difference between the maximum (M_r) and the strength of all stimulus elements present at the time (V_{ax}). Thus if one stimulus element, X, already at the maximum, always accompanies A, nothing will happen to A and this takes care of the blocking phenomenon. The details of the relationship between the modified Hullian equation (2) and the rule for an adaptive element proposed by Widrow and Hoff are explained fully by Sutton and Barto (1981). Clearly the superficial similarity is simply that incremental learning is defined as proportional to the difference between a current value and some target. It is for that reason that the rule is often referred to as a "delta" rule (Rumelhart et al. 1986b, p. 53)

Equation (3) is the generalized delta rule used by Rumelhart *et al* to show that multi- layered networks with hidden units can learn to accomplish, albeit extremely laboriously, the task of producing the correct output when presented with the set of stimuli for the 'exclusive or' relationship, and the others which Minsky and Pappert (1969) proved to be impossible for single-layered perceptron networks. It is used to modify all the connection weights in a given network by "back propagation" of the difference between an actual output vector and a desired target output. The similarity between this and the simpler equations is that incremental learning is again defined as proportional to the *difference* between a current value and some target. It is for that reason that the rule is referred to as a "delta" rule (Rumelhart *et al*, 1986; p. 53).

The implications of defining a learning rule in terms of the difference between current and target values are clearly more far-reaching in the case of the generalized delta rule than in Hull's relatively straightforward growth function, but the formal similarity adds poignancy to the fact that Hull would certainly have been in sympathy with the attempt to model complex behaviour by mathematical treatments of the consequences of strengthening multiple connections.

Hull also provides a link between Pavlov's connectionism and hardware simulations, since he was explicitly in favour of constructing machines which could mimic psychological phenomena (Hull, 1930, 1937). His own efforts in this direction were technologically limited, but he collaborated in two - reasonably successful - attempts to simulate the rudimentary phenomena of Pavlovian conditioning with electrical circuitry. The first used a serial and parallel arrangement of mechanical switches which allowed current from a battery to charge up a set of capacitors which subsequently showed topographical generalization (Kreuger and Hull, 1931) while in the second (Baernstein and Hull,1931) 'stimuli' fed the heaters of mercury-toluene regulators, the contacts of which were arranged so that only forward conditioning was possible (empirically accurate under some but not all circumstances).

Difficulties in Hullian theory and mediating responses.

Many of the principles adopted by Hull turned out to be capable of empirical disconfirmation. He was wedded to a single teaching principle - the reduction of drives related to biological needs - and this is neither necessary nor sufficient. Both excitatory and inhibitory learning were tied to response performance, but the detection of events, or their absence, is sufficient for learning in many instances (Mackintosh, 1974; Walker, 1987).

A more general and serious problem for the whole approach is that it is impossible to account for even fairly ordinary forms of animal behaviour in terms of directly testable stimulus-response connections. Generations of experimental psychologists have rigorously tested the capacities of laboratory rats for negotiating mazes: at first sight the data suggest they possess structured knowledge of the location of objects in space ("cognitive maps"), which are consulted in pursuit of goals, rather than a collection of fixed habits, since they take short-cuts, detour around barriers, and avoid revisiting recently depleted food locations (Tolman, 1932, 1948; Olton, 1979). A simple example which exercised Hull (1934) was the behaviour of rats trained always to run some distance passed a closed door, and then come back to it, before it was opened to allow them to proceed towards a goal. It is not particularly surprising that under some conditions rats waste time trying to scratch their way through the door as they go past it, and perform the novel act of going through on the first pas if it is left open. But neither of these behaviours is directly predictable from Hull's first principles, in which the underlying mechanisms are entirely frequency- sensitive and response based.

Hull therefore developed an account of the "hidden processes" (1930, p. 511) which intervene in human cognition, and when "the principle of frequency must be overridden" (1934, p. 134) in humbler creatures. In order to retain the stimulus- response format, knowledge, purpose and expectancy were attributed to internal "pure stimulus acts", these being the loci of associations other than between receptor and effector, and the naturalistic basis of symbolism and thought (1930, p. 517). For human linguistic processes fleeting references were made to "sub-vocal speech pure stimulus acts" (Hull, 1952; p. 151). For animal behaviour pure stimulus acts were typically internal anticipations of goals ("fractional antedating goal reactions": Hull, 1952; p. 151). Also in the context of novel or geographically informed performances of rats Hull elaborated something called the "habit-family hierarchy" to accommodate systematicity and apparent inference: "It is believed, for example, that the habit-family hierarchy constitutes the dominant physical mechanism which mediates such tests of truth and error as organisms employ -- that it provides the basis for a purely physical theory of knowledge" (Hull, 1934, p. 40).

The domain of stimulus-response theory.

The claims made by Hull and others for stimulus-response theory were extremely sweeping. Its principles were tested on animal behaviour for experimental convenience, on the assumption that laws established under controlled "test- tube" conditions could be extrapolated more widely. Hull led a large team at the Institute for Human Relations at Yale, and left most of the extrapolating to others, but was very explicit that his formal principles should be regarded "as a general introduction to the theory of all the behavioural (social) sciences" (1943, p. *v*), and held up the eventual goals of "a really effective and universal moral education" (1943, p. 401), based on a scientific theory of moral behaviour of an explicitly Benthamite kind (1952, pp. 340-3). These claims now of course seem quite wildly over-ambitious, but it is of interest to note that associates and admirers of Hull proceeded to explore the implications of his assertions in two opposite directions.

The initial applications of Hullian ideas to wider issues in social and clinical psychology were to some degree implementational: this is most obvious in the early attempts to provide associationist accounts of phenomena associated with Freudian theory. The tendency (apparent also in Watson, 1925; e. g. p. 209) was to take on trust some rudimentary version of Freud's theories about both the origin of human mental disorders and the appropriate therapy for them, and then put forward interpretations of the ideas in stimulus-response terms. For instance Mowrer (1939) examined Freud's book on anxiety published in the U. S.

in 1936, derived some behavioural predictions, and went on to test these by measuring rats' avoidance of artificial stimulus associated with electric shocks (Mowrer, 1940). Miller (1944), in a chapter on conflict, refers both to psychoanalytic therapy and to the work of Lewin (1935) which was on field theory, later the foundation of Gestalt therapy. The main content of the chapter was discussion of data obtained by measuring the force with which harnessed rats pull against a spring when placed at various distances from a goal at which they had been both fed and shocked. Under these circumstances hungry animals move closer to the anxiety-provoking goal than others, and those with stronger avoidance tendencies stay further away from it. This evidence was used to deduce "from stimulus-response principles" that there was a plausible basis for the "Freudian contention" that strong fear may indicate the presence of strong approach tendencies. Also, reducing behavioural avoidance paradoxically increases manifest anxiety, if the animal as a consequence goes closer to the source of conflict, and this, Miller speculates, may be the basis for the observations of therapists that steps taken to reduce fears sometime appear to make patients more anxious (1944, pp. 439-40).

Others of Hull's colleagues made more extensive surveys of psychoanalytic phenomena, from a not unsympathetic standpoint (Sears, 1944; Sears *et al*, 1957) and sometimes psychoanalysts themselves were interested in the possibility of support from these sources (French, 1933, 1944: more recently there have been suggestions that psychoanalysis might ally itself with aspects of AI - Turkle, 1988). Thus there was some intermingling of associationist and experiential and clinical theories, despite Hull's ambitions for an eventual grand synthesis. Freud is in some ways a special case, since his emphasis on unconscious processes was congenial to behaviourists, and his background as a neurologist led to elements of biological determinism and even neural connections in his own writing, especially early on (see Sulloway, 1979, especially for Freud's links with Spencer via Hughlings Jackson: as late as 1920 he was still referring to the eventual replacement of psychological terms by physiological ones).

Nevertheless, it was perhaps predicable that stimulus-response theorists would be among those wishing to jettison Freud along with all other mentalisms, in favour of more rigorous and testable extrapolations from basic principles. Eysenck (1952, 1986) has been a persistent critic and Wolpe (1952, 1958, 1978, 1981) developed one of several behavioural techniques of therapy from formal analyses along explicitly Hullian lines (Walker, 1984)

Comparisons of old and new connectionism

Modern connectionism should clearly be judged on its merits, and it would be unfair to saddle it with the sins and omissions of previous theories. But on some issues the similarity of current controversies to earlier instantiations is stark, while on others there are relatively unexplored precedents.

Neural plausibility

It is no great matter that current connectionists cannot claim priority for the idea that brain processes are involved in cognition, but worth pointing out that in some ways Spencer, Jackson, Sherrington and others had a relatively sophisticated view of brain processing - because they started from the position that the central nervous system was a complex product of long evolution, which contained qualitatively different sub-systems. Connectionism is sometimes physiologically extremely implausible at the neuronal level (Crick and Asanuma, 1986) and sometimes deliberately so (Smolensky, 1988). Questions asked at earlier stages which still need answering have to do with relations of structure to biological (not necessarily cognitive) functions and therefore with the heterogeneity of psychological processes. These are motivational as well as perceptual, and reflexive as well as voluntary. How, and why, does one inhibit a cough? This still seems a reasonable question to ask.

Thought and the Thinking Self

The major argument is over the "cognitive architecture" — whether causal determinants and theoretically valuable laws should be about incremental connections between small scale sub-conceptual units or about the complex properties of large scale holistic representations, whose relation to neural physiology can be safely ignored.

Earlier connectionists foundered because the structured nature of behaviour, even animal behaviour, often appears to require interpretation in terms of larger scale, frequency insensitive, psychological units. The success of strict connectionism along conditioned reflex lines in accounting for biological phenomena has always been extremely limited: Pavlov had to resort to arguments about higher forms of synthesis and there is a catalogue of similar additions to theories which started off with simple incrementing connections. Associative phenomena themselves now tend therefore to be attributed to beliefs and cognitive maps, and expectancies and goals (Walker, 1985; Dickinson, 1988; Mackintosh, 1988). Neoconnectionist models have already rapidly moved towards complex and higher-level units under similar constraints. Zipser (1986) for instance, in order to tackle the problem of organized spatial knowledge, uses "view field units" which recognize landmarks. This is interesting and useful, but illustrates that larger units may tend towards becoming no longer sub- symbolic, having identifiable functions for pragmatic reasons, despite the connectionist principle that functions should be widely distributed among simple small-scale units.

From a longer historical perspective it looks as though cognitive science is re-enacting old psychological battles about introspection. Spencer and James, along with other pre-20th century associationists took psycho-neural parallelism largely for granted. Though associations were neurally implemented, the evidence for them could be comfortably subjective. James is more famous for his "Stream of Consciousness" than for his Hebb rule; Spencer distinguished subjective experience as being only serial, by comparison which physiological changes which are always both serial and parallel. They would thus have had scope for talking about symbol manipulation, a language of thought or "cognitive penetrability" — the determination of behaviour at the level of beliefs and goals (Fodor, 1987; Pylyshyn, 1985). Earlier empiricists had been even more explicit about subjective intentionality. For Locke abstract reasoning was related to language and pure associations a source of unnatural aberrations. For James Mill associations were subject to the guidance of the will. Thus irrespective of whether there are supposed to be underlying associations, or whether these are causally related to brain mechanisms, it has usually been found necessary to consider also regularities at a semantic and/or subjective level. Introspection remains flawed as a methodology, being unreliable and fickle at the best of times, and in many important cases (as Fodor, 1983, pointed out for input modules, and Freud for personal conflicts) incapable in principle of supplying necessary information. But behaviourist accounts of human psychology exclusively at the level of sub-conceptual regularities have proved equally flawed, and have usually required their adherents to make major concessions to sustain any degree of psychological plausibility.

When Rumelhart *et al* (1986b) grafted onto associative networks some proposals about goal direction in thinking, and the usefulness of self-instruction, they were therefore following well-worn paths. One has to say however that these additions appear to leave PDP theorists still in a somewhat impoverished state, by comparison with many of their associationist predecessors.

Empiricism

This is the oldest battle of all, and shows few signs of being finally won or lost. The recent origins of PDP are in "random self-organizing networks" and its goal frequently seems to be to account for perception with the minimum of innate pre-conditions. Rumelhart and McClelland (1986a, pp. 139-142) point out that neurally based models are well placed to account for innate knowledge, where is it necessary, but in practice attention is given to learning paradigms that require a minimum of help, and there is little comparably systematic on what formally needs to be innate in nervous systems, or what actually biologically is. This could easily change, since neuroscience provides data (Livingston and Hubel, 1984; Zeki and Shipp, 1988) and some connectionists (e. g. Ballard, 1986; Strong and Whitehead *in press*) use it. Not that connectionists need to abandon empiricism — while there may be compelling reasons for anticipating that the real nervous system will be found to contain genetically determined arrangements which allow for the accomplishment of complex cognitive tasks, particularly in visual perception and in the early acquisition of spoken language, there will always remain features of human adaptability which require additional kinds of explanation. Neo-nativists still appear to be riding on the errors of behaviourism: Fodor and Pylyshyn (1988) say that both connectionism and their classical architecture are neutral on Empiricism/Nativism and then a few pages on castigate connectionism for being "the Empiricist idea that all learning consists of a kind of statistical inference". But neither connectionists nor empiricists *have* to say that (though some do). Several behaviourist accounts of language acquisition and use were demonstrably inadequate, but the conclusion that everything important in human psychology is therefore innate is an extreme reaction. A subsidiary role for neo-connectionism could therefore be to inject some empiricist realism into post-Chomskean theories of human cognition.

Practical implications of connectionism

Purely nativist theories of cognition have the disadvantage of being wildly implausible in the context of everyday life. Few nativists are confident enough to advise that the education of the human infant is unnecessary; and even those convinced childhood is a period during which selections are made among pre-formed possibilities have difficulties in extending this notion to later education. For instance, Piateli-Palmarini (1989) begins by advocating on Chomskean grounds that the scientific use of the term "learning" be entirely abandoned, but ends by acknowledging that it would need to be retained for such minor matters as the acquisition of knowledge concerning algebra or history. Connectionists have the advantage that life experience usually counts for something: motor skills require practice; kind and length of schooling affects educational performance; and many forms of psychological therapy have effects on some patients. Thus a test of the degree to which any connectionism is merely a neutral kind of materialism (mind depends on the brain, and the brain is a connectionist machine) is whether it makes any predictions, both about experimental results and desirable social interventions. By and large modern connectionism fails this test, but it may be that this is because there has been a delay in the examination of the consequences of theoretical positions only recently put forward. Both connectionists and their opponents therefore deserve to be questioned on their recommendations as to the optimal forms of the conditions of useful experiences.

If inner structure is a self-organized product of the reaction of local units to repeated input from the environment, what are the educational implications? Typically earlier connectionist and other empiricists have recommended in education: rote learning; structured teaching materials; regular practice and fast and detailed feedback about performance; and learning is expected to be slow and incremental. On the other hand if everything important is innate and/or internally highly structured then educational and therapeutic experiences should concentrate on the triggering of insight and understanding, with less concern for the properties of the external input. In fact Piagetian and Gestalt theorists also advocate structure in the environment but by contrast anticipate relatively discrete episodes of insight and the discovery of relationships, rules being internally formulated by active mental processes in the learner, rather than being explicitly provided by the teacher. For psychotherapy learning theorists have either recommended overt forms of emotional reconditioning, or attempted to analyse more elaborate interpersonal dynamics in the light of reductionist principles.

It remains to be seen whether computational connectionism will lead to pedagogical or therapeutic novelties. At first sight algorithms which require hundreds of iterations to master an "exclusive or" have little to offer to educational psychologists. But the beginnings of an educational theory are discernible in the chapter by Rumelhart *et al* (1986) where "*real*" symbol processing is identified with the internalization of externally provided representational schemes. This is revealing, since it concedes that there is such a thing as symbol processing. It is also traditionally empiricist in that the symbols are externally provided, rather than intuitively triggered. And none the worse for that — to suggest that mathematical notions will be assisted by a number system, more assisted by arabic than roman numerals, and potentiated by conventions for writing equations should be uncontentious. But is it connectionism? Not if this requires universally that high-level properties must emerge from the exclusively environment-driven interactions of local elements. Accepting higher-level regularities in symbol processes means accepting the charge that the lower-level interactions are sometimes implementational. It doesn't mean accepting that they are always or "*merely*" implementational.

Conclusions

Connectionists are no doubt best employed in exploiting new possibilities — in examining the mathematical and computational properties of idealized neural networks, and combining this knowledge with the latest developments in neuroscience. Acquaintance with the work of some of their conceptual ancestors suggests that translating the advances in knowledge thereby achieved into psychological and behavioural predictions will be less than straightforward. Successful computational theories may add justification to earlier speculations about the materialist and naturalistic basis of thought. But there is a wide divergence in implications previously drawn out from this common starting point. One must expect continued heterogeneity in the inferences made from connectionist assumptions. However if there is a lesson to be learned from the history of such assumptions it is that modern connectionists would be well-advised to avoid charges of creeping behaviourism by further exploration of Jacksonian hierarchies of neural control. There are neurological, psychological and clinical reasons for assuming important divisions between what is most voluntary and what is most automatic and between what Jackson called the propositional and non-propositional. This complicates but does not necessarily dilute a computational confirmation of the biological facts of brain evolution and the psychological facts of cultural development.

References

Anderson, J. A. and Rosenfeld, E. (eds) (1988) *Neurocomputing: A Reader*. London: MIT Press.

Baernstein, H. D. and Hull, C. L. (1931) A mechanical model of the conditioned reflex. *Journal of General Psychology, 5*, 99-106.

Ballard, D. H. (1986) Cortical connections and parallel processing: structure and function. *Behavioural and Brain Sciences, 9*, 67-120.

Boakes, R. A. (1977) Performance on learning to associate a stimulus with positive reinforcement. In Davis, H. & Hurwitz, H. M. B. (eds.) *Operant- Pavlovian Interactions*. Lawrence Erlbaum Associates: Hillsdale, N. J. , 67-97.

Carey, M. P. and Burish, T. G. (1988) Etiology and treatment of the psychological side effects associated with cancer chemotherapy: A critical review and discussion. *Psychological Bulletin*, 104, 307-25.

Crick, F. H. C. and Asanuma, C. (1986). Certain aspects of the anatomy and physiology of the cerebral cortex. In McClelland, J. L and Rumelhart, D. E. (eds). *Parallel Distributed Processing. Volume 2. Psychological and Biological Models*. London: MIT Press, 333-371.

Davey, G. (ed.) (1987) *Cognitive Processes and Pavlovian Conditioning in Humans*. Wiley: Chichester.

de la Mettrie, J. (1748/1912) *Man a Machine*. La Salle: Open Court.

Dickinson, A. (1988) Intentionality in animal conditioning. In Weiskrantz, L. (ed) *Thought without Language*, Clarendon Press, Oxford, 305-325.

Eysenck, H. J. (1952) The effects of psychotherapy: an evaluation. *Journal of Consulting Psychology, 16*, 319-324.

Eysenck, H. J. (1986) *The Decline and Fall of the Freudian Empire*. Penguin, Harmondsworth.

Fodor, J. A. (1965) Could meaning be an Rm?. *Journal of Verbal Learning and Verbal Behavior, 4*, 73-81.

Fodor, J. A. (1983) *The Modularity of Mind*. MIT Press: London.

Fodor, J. A. (1987) *Psychosemantics: The Problem of Meaning in the Philosopy of Mind*. London, MIT Press.

Fodor, J. A. and Pylyshyn, Z. W. (1988) Connectionism and cognitive architecture: a critical analysis. *Cognition, 28*, 3-71.

French, T. M. (1944) Clinical approaches to the dynamics of behaviour. In Hunt, J. McV. (ed.) *Personality and the Behaviour Disorders. Vol. I.* Ronald Press: New York, 255-268.

Freud, S. (1936) *The Problem of Anxiety*. Norton: New York.

Hartley, D. (1749/1834) *Observations on Man, his Frame, his Duty and his Expectations*. London: Tegg.

Hebb, D. O. (1949) *The Organization of Behaviour*. Chapman & Hall: London.

Hilgard, E. R. and Bower, G. H. (1975) *Theories of Learning* 4th edition. New York: Prentice Hall.

Hinton, G. E. and Anderson, J. A. (eds) (1981) *Parallel Models of Associative Memory*. Hillsdale, NJ: Erlbaum.

Hull, C. L. (1930) Knowledge and purpose as habit mechanisms. *Psychological Review, 37*, 511-525.
Hull, C. L. (1931) Goal attraction and directing ideas conceived as habit phenomena. *Psychological Review, 38*, 478-506.
Hull, C. L. (1934) The concept of habit-family hierarchy and maze learning. *Psychological Review, 41*, Part I 33-54, Part II, 134-152.
Hull, C. L. (1937) Mind, mechanism and adaptive behaviour. *Psychological Review, 44*, 1-32.
Hull, C. L. (1943) *Principles of Behaviour*. Appleton-Century-Crofts: New York.
Hull, C. L. (1952) *A Behaviour System*. Yale University Press: New Haven.
Jackson, J. H. (1888/1931) *Selected Writings. Volume One*, J. Taylor, ed. Hodder and Stoughton: London.
James, W. (1890/1983) *The Principles of Psychology*. Harvard University Press: London.
Konorski, J. (1948) *Conditioned Reflexes and Neuron Organization*. Cambridge University Press.
Konorski, J. (1967) *Integrative Activity of the Brain*. University of Chicago Press.
Krueger, R. C. and Hull, C. L. (1931) An electro-chemical parallel to the conditioned reflex. *Journal of General Psychology, 5*, 262-9.
Lewin, K. (1935) *A Dynamic Theory of Personality*. McGraw-Hill, New York.
Livingstone, M. S. and Hubel, D. H. (1984) Anatomy and physiology of a colour system in the primate visual cortex. *Journal of Neuroscience, 4*, 309-56.
Lynch, G. (1986) *Synapses, Circuits and the Beginnings of Memory*. Cambridge, Mass: MIT Press.
Mackintosh, N. J. (1974) *The Psychology of Animal Learning*. Academic Press: London.
Mackintosh, N. J. (1983) *Conditioning and Associative Learning*. Clarendon Press: Oxford.
Mackintosh, N. J. (1988) Approaches to the study of animal intelligence. *British Journal of Psychology, 79*, 509-25.
McClelland, J. L. , Rumelhart, D. E. and Hinton, G. E. (1986) The appeal of parallel distributed processing. In Rumelhart, D. E. and McClelland, J. L. (eds) *Parallel Distributed Processing. Volume 1. Foundations.* London: MIT Press, 3-44.
Miller, N. E. (1944) Experimental studies of conflict. In Hunt, J. McV. (ed.) *Personality and the Behaviour Disorders. Vol. I.* Ronald Press: New York, 431-65.
Minsky, M. and Papert, S. (1969) *Perceptrons*. MIT Press: London.
Mowrer, O. H. (1939) A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychological Review, 46*, 553-65.
Mowrer, O. H. (1940) Anxiety reduction and learning. *Journal of Experimental Psychology, 27*, 497-516.
Mowrer, O. H. (1960) *Learning Theory and the Symbolic Processes*. Wiley, Chichester.
Olton, D. S. (1979) Mazes, maps, and memory. *American Psychologist, 34*, 583-96.
Osgood, C. E. (1953) *Method and Theory in Experimental Psychology*. Oxford University Press: New York.
Papert, S. (1988) One AI or many?. *Daedelus*, Winter, 1-14.
Pavlov, I. P. (1927) *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. Dover: New York.

Piatelli-Palmarini, M. (1989) Evolution & cognition: From ‘learning’ to parameter setting in biology and the study of language. *Cognition 31:*1- 44.
Pinker, S. and Prince, A. (1988) On language and connectionism. *Cognition, 28*, 73-193.
Pylyshyn, Z. W. (1985) *Computation and Cognition*. MIT Press, London.
Rescorla, R. A. (1978) Some implications of a cognitive perspective on Pavlovian conditioning. In Hulse, S. H. , Fowler, H. and Honig, W. K. (eds,) *Cognitive Processes in Animal Behviour*. Lawrence Erlbaum Associates: London, 15-50.
Rescorla, R. A. and Wagner, A. R. (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In Black, A. H. and Prokasy, W. F. (eds.), *Classical Conditioning II: current research and theory*. Appleton-Century-Crofts: New York, 64-9.
Rosenblatt, F. (1962) *Principles of Neurodynamics*. New York: Spartan.
Rumelhart, D. E. and McClelland, J. L. (1986) On learning the past tenses of English verbs. In McClelland, J. L. and Rumelhart, D. E (eds) *Parallel Distributed Processing. Volume 2. Psychological and Biological Models.* London: MIT Press, 216-271.
Rumelhart, D. E. and McClelland, J. L. (1986) PDP Models and General Issues in Cognitive Science. In Rumelhart, D. E. and McClelland, J. L. (eds) *Parallel Distributed Processing. Volume 1. Foundations.* London: MIT Press, 110-46
Rumelhart, D. E. and Zipser, D. (1986) Feature Discovery by Competitive Learning. In Rumelhart, D. E. and McClelland, J. L. (eds) *Parallel Distributed Processing. Volume 1. Foundations.* London: MIT Press, 152-93.
Rumelhart, D. E. , Hinton, G. E. and Williams, R. J. (1986) Learning internal representations by error propogation. In Rumelhart, D. E. and McClelland, J. L. (eds) *Parallel Distributed Processing. Volume 1. Foundations.* London: MIT Press, 318-364.
Rumelhart, D. E. , Smolensky, P. , McClelland J. L. and Hinton, G. E. (1986) Schemata and sequential thought processes in PDP models. In McClelland, J. L and Rumelhart, D. E. (eds) *Parallel Distributed Processing. Volume 2. Psychological and Biological Models*. London: MIT Press, 7-57.
Sears, R. R. (1944) Experimental analysis of psychoanalytic phenomena. In Hunt, J. McV. (ed.) *Personality and the Behaviour Disorders. Vol. I.* Ronald Press: New York, 306-332.
Sears, R. R. , Maccoby, E. E. and Levin, H. (1957) *Patterns of Child Rearing*. Row and Peterson, Evanston IL.
Sherrington, C. S. (1906) *The Integrative Action of the Nervous System*. Yale University Press: New Haven.
Skinner, B. F. (1953) *Science and Human Behaviour*. Macmillan: New York.
Skinner, B. F. (1989) The origins of cognitive thought. *American Psychologist, 44*, 13-18.
Smolensky, P. (1988) On the proper treatment of connectionism. *Behavioural and Brain Sciences, 11*, 1-74.
Spence, K. W. (1952) Clark Leonard Hull: 1884-1952. *American Journal of Psychology, 65*, 639-46.
Spencer, H. (1855/1899) *Principles of Psychology: Vols I and II.* London: Williams and Norgate.
Stewart, J. , de Wit, H. and Eikelbom, R. (1984) Role of unconditioned and conditioned drug effects in the self-administration of opiates and stimulants. *Psychological Review, 91*, 251-68.

Strong, G. W. and Whitehead, B. A. (1989) A solution to the tag assignment problem for neural networks. *Behavioural and Brain Sciences, 12*, 381-433.

Sulloway, F. J. (1979) *Freud, Biologist of the Mind.* Basic Books: New York.

Sutton, R. S. and Barto, A. G. (1981) Toward a modern theory of adaptive networks: expectation and prediction. *Psychological Review, 88*, 135-171.

Thorndike, E. L. (1898) Animal Intelligence: an experimental study of the associative processes in animals. *Psychological Review, Monograph Supplements, 2(8)*, 1-109. (Re-published, 1998, *American Psychologist, 53*, 1125-1127.)

Thorndike, E. L. (1905) *Elements of Psychology*. A. G. Seiler: New York.

Thorndike, E. L. (1931) *Human Learning*. Century: London.

Thorndike, E. L. (1949) *Selected Writings from a Connectionist's Psychology*. Greenwood Press: New York.

Thorndike, E. L. and Lorge, I. (1944) *The Teacher's Wordbook of 30,000 Words*. Teacher's College: New York.

Thorpe, W. H. (1963) *Learning and Instinct in Animals*. 2nd edition. London: Methuen.

Tolman, E. C. (1932) *Purposive Behaviour in Animals and Men*. Century: New York.

Tolman, E. C. (1948) Cognitive maps in rats and men. *Psychological Review, 55*, 189-208.

Turkle, S. (1988) Artificial Intelligence and Psychoanalysis: A new alliance. *Deadelus*, Winter, 241-268.

Walker, S. F. (1984) *Learning Theory and Behaviour Modification*. Methuen: London.

Walker, S. F. (1985) *Animal Thought*. Routledge & Kegan Paul: London.

Walker, S. F. (1987) *Animal Learning: An Introduction*. Routledge & Kegan Paul: London.

Watson, J. B. (1913) Psychology as the behaviourist views it. *Psychological Review, 20*, 158-177.

Watson, J. B. (1924) *Psychology from the Standpoint of a Behaviourist*. Lippincott, Philadelphia.

Watson, J. B. (1925) *Behaviourism*. Kegan Paul, Trench & Trubner: London.

Widrow, G. and Hoff, M. E. (1960) Adaptive switching circuits. *Institute of Radio Engineers, Western Electronic Show and Convention, Convention Record, Part 4*, 393-395.

Wolpe, J. (1952) Experimental neurosis as learned behaviour. *British Journal of Psychology, 43*, 234-68.

Wolpe, J. (1958) *Psychotherapy by Reciprocal Inhibition*. Stanford University Press.

Wolpe, J. (1978) Cognition and causation in human behavior and its therapy. *American Psychologist, 33*, 437-446.

Wolpe, J. (1981) Behavior therapy versus psychoanalysis: therapeutic and social implications. *American Psychologist, 36*, 159-164.

Zeki, S. and Shipp. S (1988) The functional logic of cortical connections. *Nature, 335*, 311- 317.

Zipser, D. (1986) Biologically plausible models of place recognition and goal location. In McClelland, J. L and Rumelhart, D. E. (eds) *Parallel Distributed Processing. Volume 2. Psychological and Biological Models*. London: MIT Press, 432-470.