School of Psychology, Birkbeck College

Course PSYC044U (Psychobiology II.) WEEK **12**

May 10, 2007

LEARNING IN NEURAL NETWORKS/ CONNECTIONISM.

For other Notes see the Easter Handout for Summer Term Lectures

(**NB** “Parallel Distributed Processing”, “PDP”, “Connectionism”, “Neo-connectionism”, “New connectionism, “Neural Networks” and “Neural Network Simulations” can be used almost synonymously. The terms refer to theories about, and demonstrations of, the effects of training systems in which large numbers of simple processing units interact only via positive or negative connections between them)

Essay

Thorndike called himself a connectionist — is this just a co-incidence, or can comparisons be made between modern accounts of neural networks and previous theories of animal learning?

Very basic points

- There is a field called “Connectionism” which has developed very rapidly over the last 25 years.
- It arises from computer simulation of idealized networks of neurons.
- Its importance for present purposes comes from the fact that the simulations **learn** to perform certain tasks, rather than being programmed to do them in a way known in advance.
- The learning mechanisms used have some similarity to those proposed by some animal learning theorists in that learning takes the form of **strengthening connections** – or associations – between the idealized neurons.
- It is similar to early kinds of stimulus-response theory (Watson, Thorndike, Hull, Spence) in assuming that large numbers of simple associations, formed according to a few completely general rules, will be able accomplish relatively complex cognitive tasks — including those required for human cognition.

Further notes

The theme of the earlier draft notes is similarities and differences between recent connectionist theories and associative theories of animal learning. Before getting on to this we should consider the main thrust of new connectionist theories, which is to give accounts of specifically human cognitive processing (using Rumelhart and McClelland, 1986a, on past-tense learning, as an example).

<u>An illustrative quotation.</u>

> *“Connectionism is ‘in’. Not since the Dark Ages of the pre-Chomskyan era have we seen so much interest in associationist models of human thinking. Streaming forth from their banishment in the Skinnerian dungeons are dozens of detailed computational models based on the new language of networks, nodes, and connections.” (from McWhinney, B. and Leinbach, J. ,1991)*

This is from a paper on simulations of past-tense learning, which was one of the topics in the 1986 two volume work which attracted the most intense criticism (Rumelhart and McClelland, 1986a). Without going into any detail, it is possible to see from the claims and stated goals of this 1986 chapter why it attracted, and still attracts, so much attention.

1. The fact that the acquisition of English as a first language includes a stage at which children make errors by supplying regular past tense endings for irregular verbs they had initially used correctly (e.g. “goed”, “comed”, or “camed”), and can generate a regular past tense for an invented word, had been used to support than children make use of *explicit inaccessible rules*, which they discover through the use of a special purpose innately given *language acquisition device*.

2. Rumelhart and McClelland (1986a) ended up by directly challenging this for the past tense in particular and all other language processing more generally. —

> "We have shown that a reasonable account of the acquisition of the past tense can be provided without recourse to the notion of a 'rule' as anything more than a *description* of the language..... The child need not figure out what the rules are, or even that there are rules."
>
> "We view this work .... as a step toward a revised understanding of language knowledge, language acquisition, and linguistic information processing in general!

3. Many of the details used by Rumelhart and McClelland (1986a) are not relevant to this overall conclusion since they have been changed in subsequent simulations (e.g. McWhinney and Leinbach, 1991; Plunkett and Marchman, 1991; Plunkett and Juola, 1999; Joanisse & Seidenberg, 1999, 2005).

BASIC POINTS IN RUMELHART AND MCCLELLAND (1986A) ARE:

i. The system **learns the past-tense forms of English verbs by being exposed to pairs of inputs and correct outputs**.

ii. After learning the model makes sensible responses to **novel** verbs (irregular and regular).

iii. During the course of its learning, it is claimed that the model simulates the stages of children's learning: it starts with similar performance on a small number of high- and low- frequency verbs, then goes through a phase of incorrect regularizations, then reaches a high level of performance on both regular and regular verbs.

iv. The pattern of errors made by the model during the phase of incorrect regularization was similar to that of preschool children, in that verbs ending in */t/* or */d/* tended to be treated as the "no change" type, while verbs not ending in *t/d* were predominantly regularized.

v. There is a list of stimulus inputs to correspond to the base form of verbs, coded in a special (and artificial way) as a list of numerical values.

vi. There is a corresponding list of specially coded outputs which represent the correct past tense form of both irregular and regular English verbs.

vii. There are connections linking the input units to the output units. But there is no "innate" help given to this network, since all connections are set to zero initially. (In other examples they might be set initially at random).

viii. Note that the model does not decode speech input or produce motor output but only associates artificial codes which abstractly represent verbal material. (The perceptual and motor tasks required for real-life language may be *more* difficult than what this network accomplishes).

ix. Nevertheless, it is clearly a radical new proposal that phenomena of human language can arise from a system which can be characterized as a set of learned associations between input units and output units. Not surprisingly, such proposals have attracted criticism.

Criticisms of Connectionist claims

There have been many detailed and lengthy attacks on the claims made by Rumelhart *et al* (1986) and others subsequently. (e.g. Fodor and Pylyshyn, 1988; Pinker & Ullman, 2002, 2003). For present purposes they can be condensed to the two points examined by Kaplan *et al* (1992):

1. "Connectionism is merely a naive, computerized revival of behaviourism."

2. “Connectionism models are fundamentally associationist in nature, and this severely limits their cognitive potential.” (pp 91-2)

Kaplan *et al* endorse the first criticism (“connectionism equals behaviourism”) in the case of standard input-output nets like that of Rumelhart and McClelland (1986a) or those which are more elaborate only in having an intermediate layer of “hidden” units between inputs and outputs;

“..... past tense mapping must be implemented as a direct mapping from stem to past-tense form. This direct mapping means [that standard network models] are actually complex, parallel versions of a traditional stimulus-response (S-R) model.” (p.94)

“When a stimulus is presented to the input layer, a response follows immediately which is fully determined by that stimulus. This property, the complete determination of each response by the characteristics of the immediate stimulus, is a more accurate description of the behaviourist position, and it is true of ... networks regardless of whether or not they contain hidden layers.” (p.95).

Kaplan *et al* do not contest the claim that connectionist models are associationist, but “argue that this has improperly been considered a disadvantage only because the power and ubiquity of association in cognition has been underestimated.” (p.92). They suggest however that connectionist models will need to include mechanisms that correspond to “higher-level processes” such as abstraction and cognitive maps.

Quinlan (1991) has a short section on **New connectionism and human reasoning** (pp 262-5) in which he reviews the criticism that connectionist networks cannot exhibit the s*ystematicity* which is characteristic of the human understanding of sentences, and some forms of animal cognition (Fodor and Pylyshyn, 1988). The term “systematicity” is related to the concept of rule-learning, and Quinlan uses the example of the difference in rule-learning ability that apparently exists between corvids and pigeons, as discussed by Mackintosh (1988), who concluded that “associations alone do not generate rules.”

Thus, stimulus-response theories of animal learning (Thorndike, 1898; Hull, 1943; Spence, 1937) and direct input-output neural network models, have been subjected to the same kind of criticism, that they do not capture cognitive processes such as abstraction, rule-following and the use of cognitive maps.

The criticism is particular acute for the case of human language —

> “An overall impression gained from reading the new connectionist literature is that regardless of the seeming complexities of human language processing, *a few general principles of learning* and processing will suffice. Framed thus, it is easy to see why rather caustic comparisons have been drawn between new connectionism and old behaviourism.”
>
> (Quinlan, 1991; p. 193: my italics)

- Connectionists have not addressed in any detail the question of why, if language processing can be simulated using a few general learning principles, it remains unique to the human species even though very extensive training has been given to supposedly high-powered neural network systems in the shape of chimpanzees (**see Week 10**). The conventional view is of course that human language is based on special purpose, built-in mechanisms which do not depend on experienced associations and which deal with aspects of language such as grammatical rules, either as a result of evolutionary processes (Pinker and Bloom, 1990) or for some other reason (Piatelli-Palmarini, 1989). This debate is still continuing, e.g. see Marslen-Wilson and Tyler (1998, 2003, 2007), Joanisse (2004), Joanisse and Seidenberg (1999, 2005), Harm & Seidenberg (1999), Tikkala, A. (2000), Hutzler et al., (2004), Ullman et al., (2005), Desai et al. (2006), Newman et al. (2007) and Nicoladis et al. (2007).
- But the debate is not restricted to language, as the contrast between Tolmanian and stimulus-response accounts of animal learning demonstrates. For human psychology, connectionist approaches currently have a strong presence in theories of cognitive development (Berthier et al., 2005; Colunga & Smith, 2005; Elman, 2005; Mareschal & Johnson, 2002; Thomas & Karmiloff-Smith, 2002, 2003; Westermann et al., 2006, 2007) and are occasionally applied to a variety of other areas, such as

intelligence (Garlick, 2002) perceptual processing (Gurney, 2007) and issues in personality and social psychology (Queller & Smith, 2002; Van Overwalle & Jordens, 2002; Read & Urada, 2003)

KINDS OF LEARNING IN CONNECTIONIST MODELS

The basic distinction usually drawn between types of learning in connectionist model is between "**unsupervised**" "**supervised**" and "**reinforcement**" learning, (Quinlan, 1991 p. 53; Hinton, 1989). This distinction is related to the kind of **feedback** available to the systems as a consequence of its outputs (responses). There are however some gray areas between the categories because it is often possible to convert one kind of learning procedure into another (Hinton, 1989).

Unsupervised learning

Typically examples use Hebb rules of *association by contiguity*, and are able to capture regularities in repeated inputs. At the behavioural level an example is **habituation**, where a certain stimulus is repeated and comes to be recognized: there is no external feedback for a "right" or a "wrong". The connectionist equivalent is an "auto-associative network". In these the same pattern is presented at both the input and the output stages of a *pattern associator* (Quinlan, p.52; see overhead), and eventually the network can complete the pattern if only a partial input is given.

Simple kinds of Pavlovian conditioning can also be regarded as unsupervised learning: the connectionist equivalent is when a pattern associator is given pairs of different patterns at the "input" and "output" stages, and can subsequently reproduce the output pattern when given just the input. This is unsupervised in the sense of lacking external feedback for right or wrong responses (it is not sensitive to goals).

Supervised learning

A. The "Delta Rule" (Quinlan, 1991; pp. 55-6)

Supervised learning involves methods of changing the strength (or *weight*) of connections between input and output units that are more complicated that the Hebbian rule of contiguity (or "co-activation" — when two units are active at the same time the weight of the connection between them is increased.)

The underlying idea is a form of trial-and-error learning, where errors are systematically corrected. Usually starting at random, the network responds to an input, and its output (a list of numbers) is compared to the completely correct "target" output which it is supposed to learn. Then the **difference** between the actual output and the target output is quantified and used as an index of how much (and in which direction) the weights between input and output units should be changed. The delta rule applies when there are only direct connections between input and output

This type of supervised learning has an unexpected relation to Pavlovian conditioning, because the formula used by Rescorla and Wagner (1972) to successfully account for classical conditioning phenomena (especially blocking) is formally equivalent to a special case of the delta rule. (Sutton and Barto, 1981; see overhead). This is unexpected because Pavlovian conditioning is not supervised in the "right or wrong" sense, but understandable in that the Rescorla and Wagner (1972) principle is that the increment in associative strength to a conditioned stimulus on a given trial is proportional to the difference *between its current strength and the theoretical maximum for the signaled event. Another general similarity is that both the delta rule and the modern treatments of Pavlovian conditioning emphasize that the Hebbian rule of contiguity is insufficient.*

B. Back-propagation (Quinlan, 1991; pp. 56-8)

For present purposes the back-propagation method can be regarded as an elaboration of the delta rule for the purpose of supervising learning in "multi-layer" nets, where there is at least one layer of "hidden unit" which intervene between the input and output units. The important points for comparison with ideas derived from studies of animal (or human) learning are:

1. Back-propagation is very widely used in connectionist modelling.

2. As with the delta rule it is of the essence that the complete details of the target, or end-product of learning are provided to the system throughout the learning process.

3. No learning occurs when the system makes a correct output.

C. Problems with back-propagation, especially as a model for natural learning.

1. For connectionist modellers, back-propagation in multi-layer units is good since it can do things not possible for the delta-rule with direct input-output.

2. A major drawback in both theory and practices is that it provides very slow learning (Quinlan, 1991; p.69). There are many techniques for speeding it up, but it cannot be a model for "one-trial" learning in real life.

3. There is no theoretical guarantee that the system will not stop learning before it learns to produce the desired target (by "getting stuck in a local minimum"; Quinlan, p. 71). However, in practice it has been made to work for a wide variety of tasks by making networks bigger, and Hinton (1989) believes this is less of a practical problem than slowness in learning.

4. It is computationally convenient to go backwards though a network to change weights in a similar but reversed way of going forwards from input to outputs. **But** neurons are unidirectional, and thus this mechanism is not physiologically plausible in a direct way. There are plenty of reciprocal (forwards and backwards) connections between different brain structures, but the sorts of computations used mean that "the algorithm falls well outside of the realms of neurological plausibility."

5. Equally important is the fact that back-propagation is very implausible at the behavioural level for many tasks. In the first place it is clear that in both animal and human learning (see Week 13) learning does not take place by systematic alternation of errors, but by reward for, or practice of, *successful* responses and strategies. In the second place there are many situations in which it seems unlikely that perfect copies of the "target" behaviour are constantly available during learning. (E.g. are verb stems and their past-tense forms consistently paired during language acquisition: how could a rat compare erroneous routes with correct routes through a maze?)

Reinforcement learning

In these procedures external feedback is only given globally, to distinguish "right" from "wrong" outputs.

Hinton (1989) noted that there was a large literature on this topic "beyond the scope of this paper". I.e. not much use was being made of reinforcement procedures in connectionist simulations of learning. There is a technical problem called "credit assignment": if a reasonably large network produces a correct output which local connections are responsible? This is potentially solvable, and it is not clear that reinforcement methods could not in principle be made more use of.

The fact that more is made of the effects of reward and punishment in analyses of biologically "real" learning may be related to the involvement of motivational factors, which are not mimicked (so far) in neural network research. However, there is some interest in reinforcement learning in areas such as robotics (Dean, 1998; Coleman et al., 2005) and there was a special issue of the journal *Machine Learning* devoted to "Reinforcement Learning" (Kaelbling, 1996). Reinforcement learning may be used for practical purposes (Bingham, 2001; Franklin, 2007), or for simulating biologically realistic reward-related behaviours (Berthier et al, 2005; Hazy et al., 2006; Hampton & O'Doherty, 2007).

Main Sources — Animal Learning and Learning in Connectionist (Neural Network) Simulations (Week 12)

Lieberman, D. (1990/1993/2000) *Learning: Behavior and Cognition*. Belmont: Wadsworth. ("The Neural Network Solution": pp. 439-455 /1993 edition pp. 511-525; /2000, pp. 517-532)

Quinlan, P. (1991) *Connectionism and Psychology*. Harvester Wheatsheaf, Hemel Hempstead. Chapter 2 "Memory and Learning in Neural Networks" esp pp.51-56,.69-71, and 262-266. [152 QUI & AKCHN(Qui).]

Walker, S.F. (1990/1992) A brief history of connectionism and it psychological implications. *AI & Society 4*, 17-38. (TIED XEROX/SLC) Or 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. (BK library AKCHN [Cla])

References (Not normally required for further reading)

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition, 90*(2), 119-161.

Baxt, W. G., Shofer, F. S., Sites, F. D., & Hollander, J. E. (2002). A neural network aid for the early diagnosis of cardiac ischemia in patients presenting to the emergency department with chest pain. *Annals of Emergency Medicine, 40*(6), 575-583.

Becktel, W. and Abrahamsen, A. (1991) *Connectionism and the Mind: An Introduction to Parallel Processing in Networks.* Oxford, Basil Blackwell. (AKCHN).

Berthier, N. E., Rosenstein, M. T., & Barto, A. G. (2005). Approximate optimal control as a model for motor learning. *Psychological Review, 112*(2), 329-346.

Bingham, E. (2001). Reinforcement learning in neurofuzzy traffic signal control. *European Journal of Operational Research, 131*(2), 232-241.

Coleman, S. L., Brown, V. R., Levine, D. S., & Mellgren, R. L. (2005). A neural network model of foraging decisions made under predation risk. *Cognitive Affective & Behavioral Neuroscience, 5*(4), 434-451.

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review, 112*(2), 347-382.

Desai, R., Conant, L. L., Waldron, E., & Binder, J. R. (2006). FMRI of past tense processing: The effects of phonological complexity and task difficulty. *Journal of Cognitive Neuroscience, 18*(2), 278-297.

Elman, J. L. (2005). Connectionist models of cognitive development: where next? *Trends in Cognitive Sciences, 9*(3), 111-117.

Elman, JL, Bates, EA, Johnson, MH, Karmiloff-Smith A, Parisi, D. & Plunkett K. (1996) *Rethinking Innateness: A connectionism perspective on development*. London: MIT Press. (155.7 ELM in Bk Libary).

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

Franklin, J. A. (2006). Jazz melody generation using recurrent networks and reinforcement learning. *International Journal on Artificial Intelligence Tools, 15*(4), 623-650.

Garlick, D. (2002). Understanding the nature of the general factor of intelligence: The role of individual differences in neural plasticity as an explanatory mechanism. *Psychological Review, 109*(1), 116-136.

Gurney, K. (2007). Neural networks for perceptual processing: from simulation tools to theories. *Philosophical Transactions of the Royal Society B-Biological Sciences, 362*(1479), 339-353.

Hampton, A. N., & O'Doherty, J. P. (2007). Decoding the neural substrates of reward-related decision making with functional MRI. *Proceedings of the National Academy of Sciences of the United States of America, 104*(4), 1377-1382

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106*(3), 491-528.

Hartshorne, J. K., & Ullman, M. T. (2006). Why girls say 'holded' more than boys. *Developmental Science, 9*(1), 21-32.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2006). Banishing the homunculus: Making working memory work. *Neuroscience, 139*(1), 105-118.

Herd, S. A., Banich, M. T., & O'Reilly, R. C. (2006). Neural mechanisms of cognitive control: an integrative model of stroop task performance and fMRI data. *Journal of Cognitive Neuroscience, 18*(1), 22-32.

Hinton, G.E. (1989) Connectionist learning procedures. *Artificial Intelligence, 40*, 185-234.

Hutzler, F., Ziegler, J. C., Perry, C., Wimmer, H., & Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition, 91*(3), 273-296.

Joanisse, M. F. (2004). Specific language impairments in children - Phonology, semantics, and the English past tense. *Current Directions in Psychological Science, 13*(4), 156-160.

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences of the United States of America, 96*(13), 7592-7597.

Joanisse, M. F., & Seidenberg, M. S. (2005). Imaging the past: Neural activation in frontal and temporal regions during regular and irregular past-tense processing. *Cognitive Affective & Behavioral Neuroscience, 5*(3), 282-296.

Kaelbling, LP (1996) Special issue on reinforcement learning - introduction. *Machine Learning, Vol.22,* No.1-3, Pp.7-9

Kandel, E. R. (2001). Neuroscience - The molecular biology of memory storage: A dialogue between genes and synapses. *Science, 294*(5544), 1030-1038.

Kaplan, S. Weaver, M. and French, R.M. (1992) Active symbols and internal models: Towards a cognitive connectionism. In Clark, A. and Lutz, R. (eds) *Connectionism in Context*. Berlin: Springer-Verlag. 91-110. (TIED XEROX)

Kemp, N., & Bryant, P. (2003). Do beez buzz? Rule-based and frequency-based knowledge in learning to spell plural -s. *Child Development, 74*(1), 63-74.

Mackintosh, N.J. (1988) Approaches to the study of animal intelligence. *British Journal of Psychology, 79*, 509-25.

Mareschal, D., & Johnson, S. P. (2002). Learning to perceive object unity: a connectionist account. *Developmental Science, 5*(2), 151-172.

Marshall, C. R., & van der Lely, H. K. J. (2006). A challenge to current models of past tense inflection: The impact of phonotactics. *Cognition, 100*(2), 302-320.

Marslen-Wilson, W and Tyler, LK (1997) Dissociating types of mental computation. *Nature,* Vol.387, No.6633, Pp.592-594.

Marslen-Wilson, W and Tyler, LK (1998) Rules, representations, and the English past tense. *Trends in Cognitive Sciences,* Vol.2, No.11, Pp.428-435 Is: 1364-6613.

Marslen-Wilson, W. D., & Tyler, L. K. (2003). Capturing underlying differentiation in the human language system. *Trends in Cognitive Sciences, 7*(2), 62-63.

Marslen-Wilson, W., & Tyler, L. (2007). Morphology, language and the brain: the decompositional substrate for language comprehension. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 823-836.

McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: what does the evidence rule out? *Trends in Cognitive Sciences, 6*(11), 465-472.

McWhinney, B. and Leinbach, J. (1991) Implementations are not conceptualizations: Revising the verb learning model. *Cognition, 40,* 121-157.

Monaghan, P., & Shillcock, R. (2004). Hemispheric asymmetries in cognitive modeling: Connectionist modeling of unilateral visual neglect. *Psychological Review, 111*(2), 283-308.

Monaghan, P., & Shillcock, R. (2007). Levels of description in consonant/vowel processing: Reply to Knobel and Caramazza. *Brain and Language, 100*(1), 101-108.

Newman, A. J., Ullman, M. T., Pancheva, R., Waligura, D. L., & Neville, H. J. (2007). An ERP study of regular and irregular English past tense inflection. *Neuroimage, 34*(1), 435-445.

Nicoladis, E., Palmer, A., & Marentette, P. (2007). The role of type and token frequency in using past tense morphemes correctly. *Developmental Science, 10*(2), 237-254.

Penke, M., & Westermann, G. (2006). Broca's area and inflectional morphology: Evidence from Broca's aphasia and computer modeling. *Cortex, 42*(4), 563-576.

Pinker, S. and Bloom, P. (1990) Natural language and natural selection. *Behavioural and Brain Sciences, 13,* 707-784.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences, 6*(11), 456-463.

Plunkett, K., & Bandelow, S. (2006). Stochastic approaches to understanding dissociations in inflectional morphology. *Brain and Language, 98*(2), 194-209.

Plunkett, K., & Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science, 23*(4), 463-490.

Queller, S., & Smith, E. R. (2002). Subtyping versus bookkeeping in stereotype learning and change: Connectionist simulations and empirical findings. *Journal of Personality and Social Psychology, 82*(3), 300-313.

Ralph, M. A. L., Braber, N., McClelland, J. L., & Patterson, K. (2005). What underlies the neuropsychological pattern of irregular > regular past-tense verb production? *Brain and Language, 93*(1), 106-119.

Read, S. J., & Urada, D. I. (2003). A neural network simulation of the outgroup homogeneity effect. *Personality and Social Psychology Review, 7*(2), 146-169.

Rumelhart, D.E. and McClelland, J.L. (1986a) 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. (1986b) 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

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

Thomas, M. S. C., & Karmiloff-Smith, A. (2003). Modeling language acquisition in atypical phenotypes. *Psychological Review, 110*(4), 647-682.

Thomas, M., & Karmiloff-Smith, A. (2002). Are developmental disorders like cases of adult brain damage? Implications from connectionist modelling. *Behavioral and Brain Sciences, 25*(6), 727-+.

Thorndike, E.L. (1905/1919) *The Elements of Psychology*. New York, A.G. Seiler. [in 'Early Texts' at Senate House]

Tikkala, A. (2000). A connectionist word production tool for Finnish nouns with a model for vowel harmony restrictions. *Computer Speech and Language, 14*(1), 1-13.

Ullman, M. T., Pancheva, R., Love, T., Yee, E., Swinney, D., & Hickok, G. (2005). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language, 93*(2), 185-238.

Van Overwalle, F., & Jordens, K. (2002). An adaptive connectionist model of cognitive dissonance. *Personality and Social Psychology Review, 6*(3), 204-231.

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. 91-110.

Westermann, G., Mareschal, D., Johnson, M. H., Sirois, S., Spratling, M. W., & Thomas, M. S. C. (2007). Neuroconstructivism. *Developmental Science, 10*(1), 75-83.

Westermann, G., Sirois, S., Shultz, T. R., & Mareschal, D. (2006). Modeling developmental cognitive neuroscience. *Trends in Cognitive Sciences, 10*(5), 227-232.

White, R. L., & Snyder, L. H. (2007). Spatial constancy and the brain: insights from neural networks. *Philosophical Transactions of the Royal Society B-Biological Sciences, 362*(1479), 375-382.

pp 159-161 (Rumelhart et al ,1986)

“Paradigms of Learning”

- Auto Associator. a set of patterns repeatedly presented and system “stores”it ... “a pattern is associated with itself”. Then when part is given, the whole original is retrieved **== *Habituation or perceptual learning***

- Pattern Associator. pairs are repeated. System learns that when one member is presented it is “supposed” to produce the other. “In this paradigm one seeks a mechanism in which an essentially abitrary set of input patterns can be paired with an abitrary set of output patterns. **= *Pavlovian conditioning, or stimulus-response motor habits***.

- Classification paradigm. this is a variant on the previous 2. There is a fixed set of categories into which the stimulus patterns are to be classified. There is a training session in which the system is presented with the stimulus patterns along with the categories to which each stimulus belongs. This is where the “perceptron convergence theorem is proved” **== *Discrimination Learning***

- Regularity detector. For a population of stimulus patterns each pattern S(k) is presented with a probability p(k). The system is supposed to discover statistically salient features of the input population. (cf Rescorla's conditional probability suggestions) and experiments on probability learning in animals. **== *probability learning and complex Pavlovian conditioning; discrimination of reward rates***

Skinner, B.F. (1957) *Verbal Behavior*.

p.119. “The most familiar examples of functional units are traditionally called words.”.....there eventually emerges a basic repertoire of smaller functional units also at the level of the word” - Child learns 'I have a doll' and 'I have a kitten' and later says “for the first time, and without separate conditioning, I have a drum”

p.120-1 “Other familiar units below the level of the word are the affixes used for inflectional, syntactical, or other purposes...These have their own histories....

The evidence is clearest when a speaker composes new forms of response with respect to new situations. Having developed a functional suffix -ed with respect to the subtle property of stimuli which we speak of as action-in-the-past, the suffix may be added for the first time to a word which has hitherto described action only in the present. The process is conspicuous..” (when children make incorrect regularizations). ‘He singed’ is obviously composed from separate elements, because the community reinforces the form ‘He sang’ ......

One kind of minimal unit is under control of the subtle properties of stimuli which distinguish with different 'parts of speech' - for example, the speaker may compose adverbs by adding -ly to adjectives. Suffixes such as -ness or -hood are usually readily manipulable as separate elements in composing new terms appropriate to “states of being.”

Innateness

p.462 “occasionally, through accidental circumstances, two or more children have grown up in partial isolation from established verbal communities and have developed fairly extensive idiosyncratic verbal systems, but the isolation has never been complete enough to prove that a verbal environment will arise spontaneously in the absence of prior verbal behaviour.”

Rumelhart, D.E. and McClelland, J.L. (1986b). 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* . MIT Press, London, 216-271.

Do we have **explicit inaccessible rules** ? Is there a Language Acquisition Device (LAD) to discover such rules?

Convention is Yes:
- the mechanism derives such rules
- hypotheses are rejected and replaced to account for evidence
- the LAD has **innate** knowledge of the possible range of languages and thus only considers hypotheses imposed by **linguistic universals** .

ALTERNATIVE: NO explicit rules (cf Honey comb cf Darwin)

EXAMPLE:
children's acquisition of regular and irregular English past-tense verbs.

regular = wiped and pulled
irr = came went (goed) gave got

there is a stage of goed and comed in Stage 3 both regular and irregular OK

– therefore "**U-shaped learning"**

acquistion is actually quite gradual

Goal was to simulate the stages with a simple connectionist model

---

There are at least 9 irregular types of the past-tense in English [p. 247 pf PDP2 – Rumelhart and McClelland, 1986]

1. **No change**:	*beat, cut, hit*
2. d to t:	*send/sent build/built*
3. Internal vowel change + d/t:	*feel/felt say/said*
4. Internal vowel change & delete consonant + d/t:	*bring/brought, catch/caught*
5. Internal Vowel Change change if d/t end:	*bite/bit find/found ride/rode*
6a Internal Vowel Change i to a:	*sing/sang drink/drank*
6b Internal Vowel Change i or a to u	*sting/stung hand/hung*
7 Other Internal Vowel Change:	*break/broke, give/gave*
8. Others with dipthong end:	*blow/blew fly/flew.*

and 3 Types of Regular English Verbs

1. end in a dental and add /d/	*start/started*
2. other end and add /t/	*look/looked ('t')*
3. other end and add /d/	*move moved*

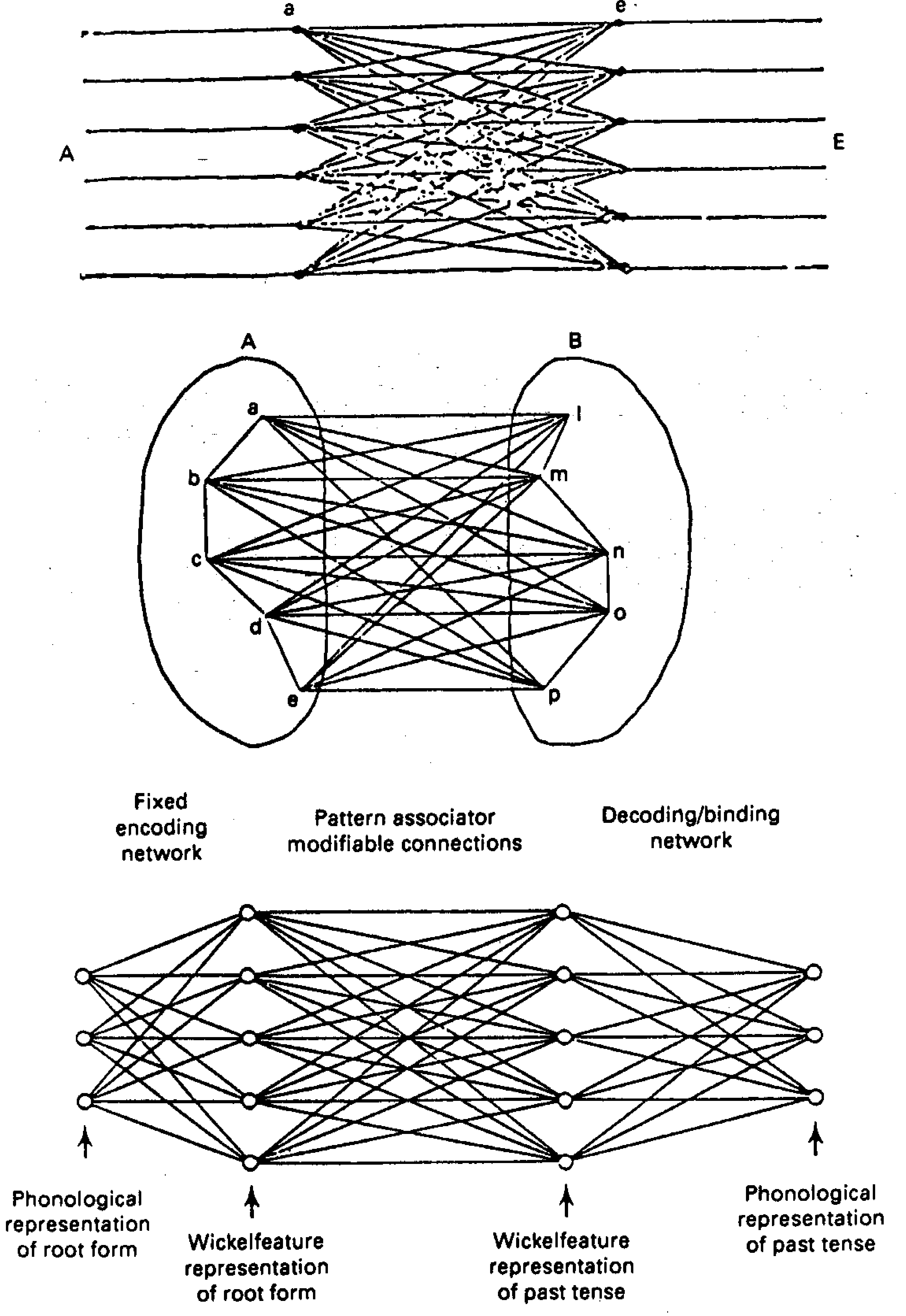


Fig. 8.1. Connectionist diagrams. *Top*: from Spencer (1855/1899 Vol. I, p. 527) in a section on "Genesis of simple nervous systems", used to show that even in a simple ganglion there may be many modifiable "lines of connection" between afferent (A) and efferent (E) fibres. *Middle*: from James (1890/1983, 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 or 1992)

Rules, representations, and the English past tense

William Marslen-Wilson and Lorraine K. Tyler (1998, see also 2003)

The significance of the English past tense in current cognitive science is that it offers a clear contrast between a potentially rule-based system – the procedures for forming the regular past tense – and an unpredictable and idiosyncratic set of irregular forms. This contrast has become a focus for a wide-ranging debate about whether mental computation requires the use of symbols. Highly regular combinatorial phenomena, such as the regular past tense, are prime candidates for rule-based symbolic computation. Earlier research concentrated on the evidence for this during language acquisition, looking at how children learned the English regular and irregular verb systems. Over the last five years attention has shifted towards the properties of the adult system, and we review here some recent research into the neural correlates of the two types of procedure. The evidence suggests that there are divergences in the neural systems underlying the generation and perception of regular and irregular forms. Regular inflected forms seem to involve primarily combinatorial processes, while irregular forms appear to have a hybrid status, sharing their semantic properties with the regular forms but diverging in the phonological domain, where their form representations are stored as complete units. This indicates that the regular and irregular past tenses may not, after all, provide a clean contrast in the types of mental computation they implicate.

William Marslen-Wilson is at the MRC Cognition and Brain Sciences Unit, 15 Chaucer Rd, Cambridge, UK CB2 2EF.

e-mail: w.marslen-wilson@mrc-apu.cam.ac.uk

Lorraine K. Tyler is at the Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge, UK CB3 3EB.

e-mail: l.tyler@psychology.bbk.ac.uk

A fundamental issue in the cognitive sciences is to determine the nature of mental computation. Over the last decade the focus of this debate has been the contrast between classical views of mental computation, seen as the rule-based manipulation of strings of symbols with a syntax, as opposed to more distributed systems, operating subsymbolically and without syntax. Despite the pervasive and crucial nature of this debate, it has nonetheless been hard to find specific domains where empirical evidence could be generated that might decide between these two broad classes of views. The English past tense, perhaps surprisingly, offers one of the few cases where this seems to be true. Protagonists on both sides of the debate generally agree that the mental representation of the regular and irregular past tense of the English verb is a crucial test case. In this brief essay, we begin by explaining why this should be so, and then go on to focus on the recent emphasis, in our own work as well as in the work of others, on the neural correlates of the cognitive systems supporting the English past tense, and on the resulting claims for the neurological as well as functional dissociability of these underlying systems in the brain.

The English verb

The significance of the English verb is that its procedures for forming the past tense offer an unusually sharp contrast, within the same cognitive domain, between a highly regular procedure and a highly irregular and idiosyncratic set of exceptions. The great majority of English verbs, numbering 10,000 or more, form their past tense by adding the regular [-d] affix to an otherwise unchanged stem. Depending on the final segment of the stem, this affix is realized as /d/, /t/ or /ed/, as in verbs like *jump/jumped*, *agree/agreed*, *state/stated*. This is an apparently paradigmatic example of a rule-based process, applying across the board to almost all the verbs in the language, and which functions as the default procedure for all new verb formations. The only exceptions are about 160 English verbs, many of them among the most common words in the language, which have irregular past tense forms, and which do not employ the regular affixing procedure. These are verbs like *give/gave*, *tell/told* and *buy/bought*, where the past tense form is idiosyncratic and phonologically unpredictable.

Because of these unpredictabilities, it is unlikely that the acquisition of irregular forms involves the acquisition of rules of any sort, and it is widely agreed that they are learned and stored by some form of pattern-association process. The key theoretical issue, instead, is how to characterize the mechanisms underlying the regular past tense, and whether, in particular, the explanation of this classically

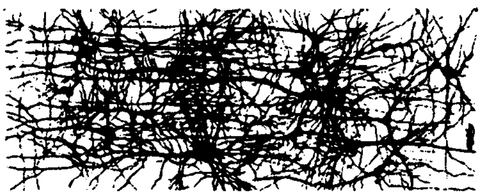
1364-6613/98/$ – see front matter PII: S1364-6613(98)01239-X

Trends in Cognitive Sciences – Vol. 2, No. 11, November 1998

12.4 The Neural Network Solution (to Association, Abstraction, and Everything . . .)

It appears that animals are capable of forming concepts, and that the methods they rely on for doing so are often similar to those used by humans. In both, associative processes play an important role, as subjects associate exemplars with a response (for example, a name; pecking) and then generalize this response to similar stimuli. In at least some situations, however, animals as well as humans combine information from exemplars to form a more abstract representation of the concept, whether in the form of a rule or a prototype.

Lieberman 2000 p517



A Neural Network Model

p 519

In outline, neural network models are surprisingly simple and rest on three basic assumptions:

1. *Neural network.* There is a network of neurons, with every neuron in the network connected to every other neuron.
2. *Transmission.* When one neuron in a network becomes active, this activity is transmitted to the other neurons in the network; the amount of excitation transmitted between any two neurons depends on the strength of the neural connection between them.
3. *Learning.* If two neurons within the network are active at the same time, the connection between them will be strengthened, so that future activity in one of these neurons will be more likely to produce activity in the other neuron.

In essence, these assumptions are virtually identical to those made by Pavlov almost 100 years ago: When two cortical centers are active simultaneously, the connection between them will be strengthened. Neural network models, however, incorporate two changes in Pavlov's ideas, which have far-reaching implications

p520

for these models' ability to predict behavior. First, they assume that the networks involved are quite massive, so that associations will be formed simultaneously among very large numbers of active neurons. Second, they provide a mathematical formula that allows us to calculate exactly how much each of these connections will be strengthened. Together, these assumptions allow us to make predictions about the brain's functioning in a way that goes far beyond anything Pavlov ever attempted or could have attempted—because of the large number of connections, the model's predictions can be calculated only with the aid of computers.

The predictions of neural network models depend critically on the formula used for calculating how connections are strengthened. A number of formulas or rules have been suggested, but one of the most influential has been the *delta rule*.

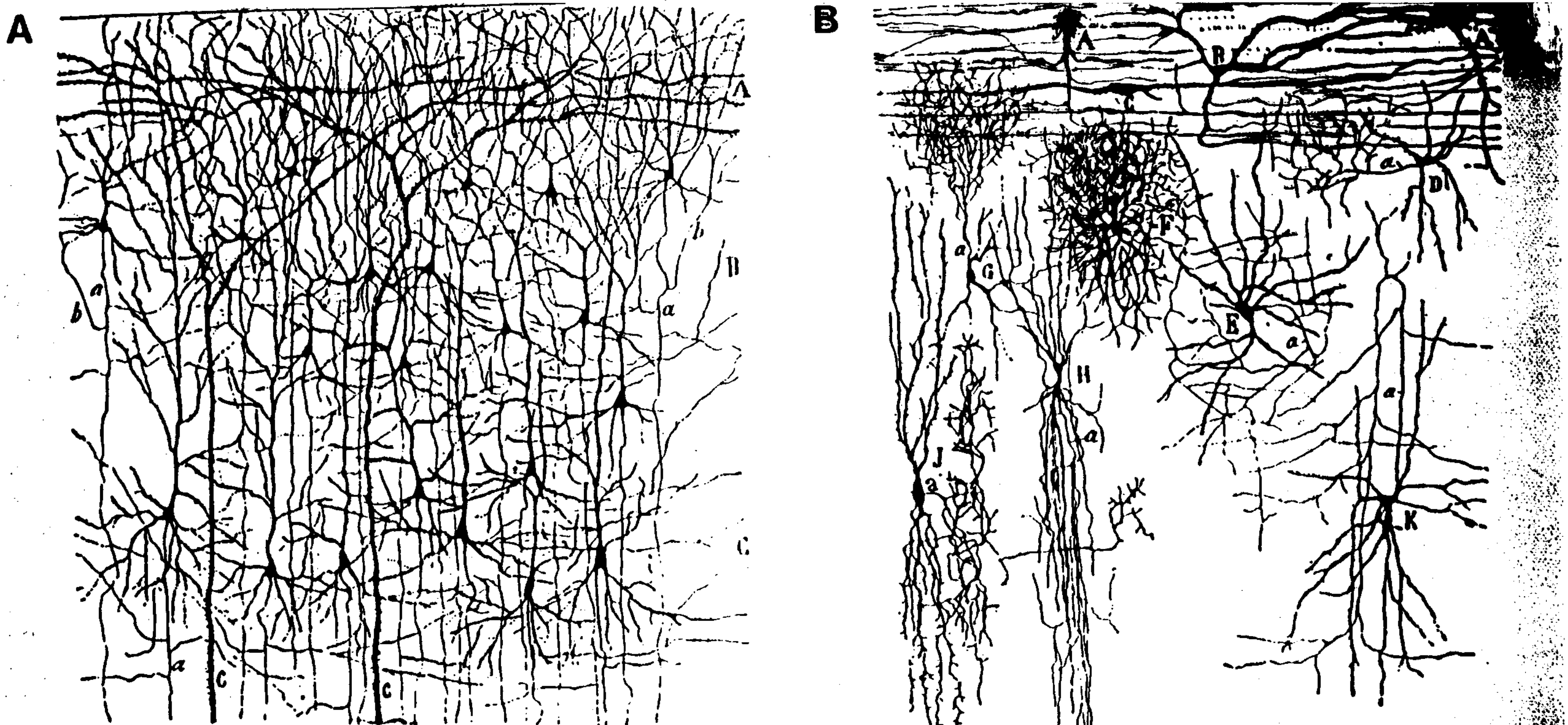


FIGURE 14. Drawings of Golgi-stained neurons of the cerebral cortex. *A*: The most characteristic neurons of the cerebral cortex are the pyramidal cells. They are characterized by a pyramid shaped cell body, a thick apical dendrite, and a number of basal dendrites. There are a variety of different types of pyramidal cells. *B*: The remainder of neurons in the cerebral cortex can be broadly referred to as nonpyramidal cells. There are a variety of different types of nonpyramidal cells. See text for further details. (From *Histologie du Systeme Nerveux de L'homme et des Vertébrés, T. 2. [Histology of the Nervous System of Man and of Vertebrates, Vol. 2]*, by S. Ramón y Cajal, 1911, Paris: Maloine. Copyright 1911 by Librairie Maloine. Reprinted by permission.)

Crick and
Asanuma (1986)

Thorndike
(1905/1919)

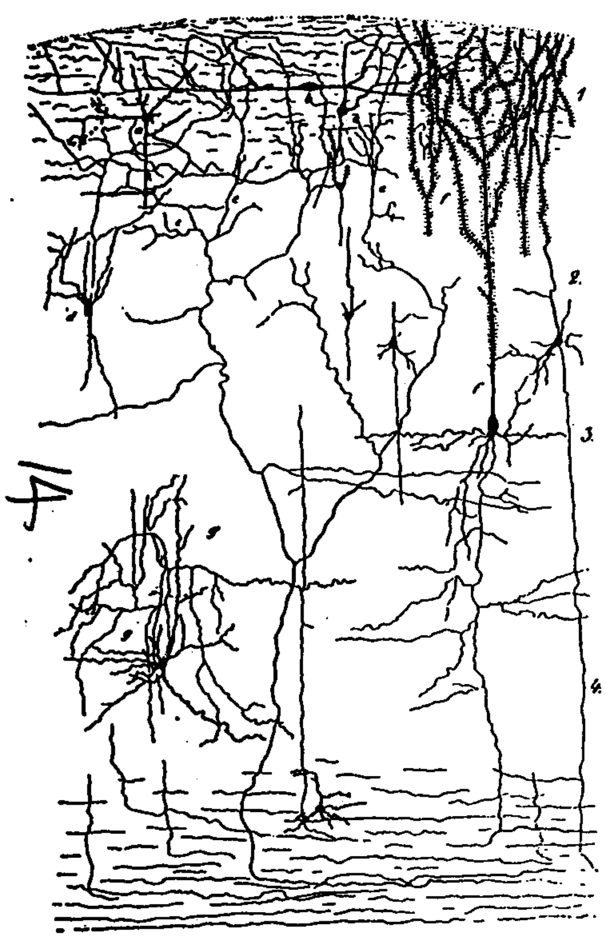


FIG. 6. A sketch showing elements of the structure of the brain cortex in mammals. These are drawn from actual specimens. Greatly magnified. After Edinger, 221, 152.

14

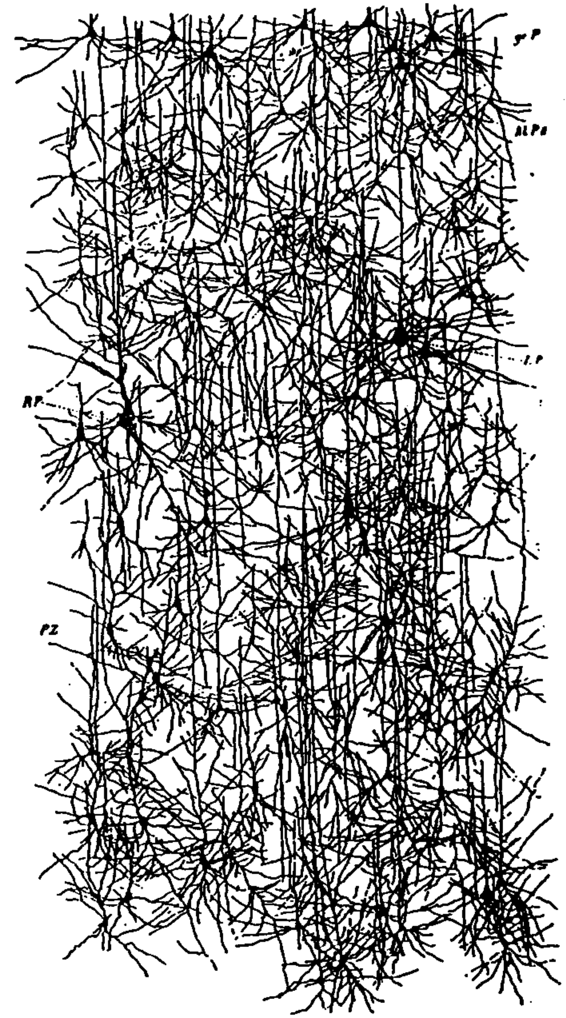


FIG. 7. A section through the brain cortex. Greatly magnified. After Kölliker, 652, 732.

9-a

denote this relationship of close proximity or contact, though it should be understood that there may be, and commonly is, no connection in the sense of one neurone growing into the other, fusing with it, making a structural connection. Every neurone, then, is in connection[1]

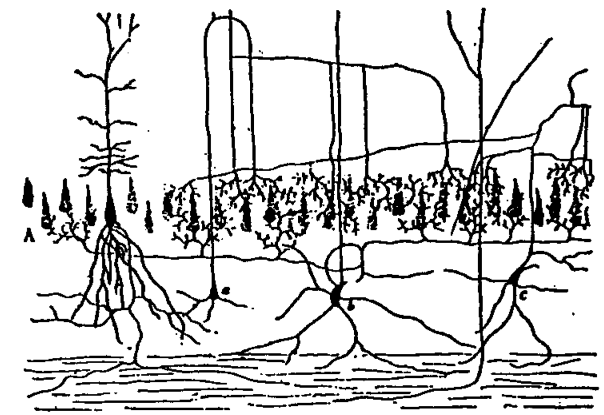


FIG. 37. Association cells (a, b and c). Greatly magnified. After Edinger, 28, 9.

with some other or others. Fig. 41 shows the general plan of such connections. Figs. 42 and 43 are drawings of the actual connections in two cases where they can be clearly inferred from what the microscope reveals.

[1] The word *synapsis*, meaning a clasping together, has been suggested as a useful descriptive term for the peculiar connections that exist between neurone and neurone.

Christiansen, M.H. and Chater, N. (1999) Connectionist natural language processing: The state of the art. *Cognitive Science, 23,* 417-437.
This **Special Issue on Connectionist Models of Human Language Processing** provides an opportunity for an appraisal both of specific connectionist models and of the status and utility of connectionist models of language in general. ………We argue that **connectionist psycholinguistics has already had a significant impact** on the psychology of language, and that connectionist models are likely to have an important influence on future research.

Coleman, S. L., Brown, V. R., Levine, D. S., & Mellgren, R. L. (2005). A neural network model of foraging decisions made under predation risk. *Cognitive Affective & Behavioral Neuroscience, 5*(4), 434-451.
This article develops the cognitive-emotional forager (CEF) model, a novel application of a neural network to dynamical processes in foraging behavior. The CEF is based on a neural network known as the gated dipole, introduced by Grossberg, which is capable of representing short-term affective reactions in a manner similar to Solomon and Corbit's (1974) opponent process theory. The model incorporates a trade-off between approach toward food and avoidance of predation tinder varying levels of motivation induced by hunger. The results of simulations in a simple patch selection paradigm, using a lifetime fitness criterion for comparison, indicate that the CEF model is capable of nearly optimal foraging and outperforms a run-of-luck rule-of-thumb model. Models such as the one presented here can illuminate the underlying cognitive and motivational components of animal decision making.

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review, 112*(2), 347-382.
In the novel noun generalization task, 2 1/2-year-old children display generalized expectations about how solid and nonsolid things are named, extending names for never-before-encountered solids by shape and for never-before-encountered nonsolids by material. This distinction between solids and nonsolids has been interpreted in terms of an ontological distinction between objects and substances. Nine simulations and behavioral experiments tested the hypothesis that these expectations arise from the correlations characterizing early learned noun categories. In the simulation studies, connectionist networks were trained on noun vocabularies modeled after those of children. These networks formed generalized expectations about solids and nonsolids that match children's performances in the novel noun generalization task in the very different languages of English and Japanese. The simulations also generate new predictions supported by new experiments with children. Implications are discussed in terms of children's development of distinctions between kinds of categories and in terms of the nature of this knowledge.

Joanisse, M. F., & Seidenberg, M. S. (2005). Imaging the past: Neural activation in frontal and temporal regions during regular and irregular past-tense processing. *Cognitive Affective & Behavioral Neuroscience, 5*(3), 282-296.
This article presents fMRI evidence bearing on dual-mechanism versus connectionist theories of inflectional morphology. Ten participants were scanned at 4 Tesla as they covertly generated the past tenses of real and nonce (nonword) verbs presented auditorily. Regular past tenses (e.g., walked, wagged) and irregular past tenses (e.g., took, slept) produced similar patterns of activation in the posterior temporal lobe in both hemispheres. In contrast, there was greater activation in left and right inferior frontal gyrus for regular past tenses than for irregular past tenses. Similar previous results have been taken as evidence for the dual-mechanism theory of the past tense (Pinker & Ullman, 2002). However, additional analyses indicated that irregulars that were phonologically similar to regulars (e.g., slept, fled, sold) produced the same level of activation as did regulars, and significantly more activation than did irregulars that were not phonologically similar to regulars (e.g., took, gave). Thus, activation patterns were predicted by phonological characteristics of the past tense rather than by the rule-governed versus exception distinction that is central to the dual-mechanism framework. The results are consistent with a constraint satisfaction model in which phonological, semantic, and other probabilistic constraints jointly determine the past tense, with different degrees of involvement for different verbs.

Joanisse, M.F. and Seidenberg, M.S. (1999) Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences of the United States of America, 96*, 7592-7597.
The formation of the past tense of verbs in English has been the focus of the debate concerning connectionist vs. symbolic accounts of language. Brain-injured patients differ with respect to whether they are more impaired in generating irregular past tenses (TAKE-TOOK) or past tenses for nonce verbs (WUG-WUGGED). Such dissociations have been taken as evidence for distinct "rule" and "associative" memory systems in morphology and against the connectionist approach in which a single system is used for all forms. We describe a simulation model in which these impairments arise from damage to phonological or semantic information, which have different effects on generalization and irregular forms, respectively. The results provide an account of the bases of impairments in verb morphology and show that these **impairments can be explained within connectionist models that do not use rules** or a separate mechanism for exceptions.

Maia, T. V., & Cleeremans, A. (2005). Consciousness: converging insights from connectionist modeling and neuroscience. *Trends in Cognitive Sciences, 9*(8), 397-404.

Over the past decade, many findings in cognitive neuroscience have resulted in the view that selective attention, working memory and cognitive control involve competition between widely distributed representations. This competition is biased by top-down projections (notably from prefrontal cortex), which can selectively enhance some representations over others. This view has now been implemented in several connectionist models. In this review, we emphasize the relevance of these models to understanding consciousness. Interestingly, the models we review have striking similarities to others directly aimed at implementing 'global workspace theory'. All of these models embody a fundamental principle that has been used in many connectionist models over the past twenty years: global constraint satisfaction.

Mareschal, D., & Johnson, S. P. (2002). Learning to perceive object unity: a connectionist account. *Developmental Science, 5*(2), 151-172.

To explore questions of how human infants begin to perceive partly occluded objects, we devised two connectionist models of perceptual development. The models were endowed with an existing ability to detect several kinds of visual information that have been found important in infants' and adults' perception of object unity (motion, co-motion, common motion, relatability, parallelism, texture and T-junctions). They were then presented with stimuli consisting of either one or two objects and an occluding screen. The models' task was to determine whether the objector objects were joined when such a percept was ambiguous, after specified amounts of training with events in which a subset of possible visual information was provided. The model that was trained in an enriched environment achieved superior levels of performance and was able to generalize veridical percepts to a wide range of novel stimuli. Implications for perceptual development in humans, current theories of development and origins of knowledge are discussed.

Plunkett, K. and Juola, P. (1999) A **connectionist model of English past tense and plural** morphology. *Cognitive Science, 23*, 463-490.

The acquisition of English noun and verb morphology is modeled using a single-system connectionist network. The network is trained to produce the plurals and past tense forms of a large corpus of monosyllabic English nouns and verbs. The developmental trajectory of network performance is analyzed in detail and is shown to mimic a number of important features of the acquisition of English noun and verb morphology in young children. These include an initial error-free period of performance on both nouns and verbs followed by a period of intermittent over-regularization of irregular nouns and verbs. Errors in the model show evidence of phonological conditioning and frequency effects. Furthermore, the network demonstrates a strong tendency to regularize denominal verbs and deverbal nouns and masters the principles of voicing assimilation. Despite their incorporation into a single-system network, nouns and verbs exhibit some important differences in their profiles of acquisition. Most importantly, noun inflections are acquired earlier than verb inflections. The simulations generate several empirical predictions that can be used to evaluate further the suitability of this type of cognitive architecture in the domain of inflectional morphology.

Queller, S., & Smith, E. R. (2002). Subtyping versus bookkeeping in stereotype learning and change: Connectionist simulations and empirical findings. *Journal of Personality and Social Psychology, 82*(3), 300-313.

A distributed connectionist network can account for both bookkeeping (M. Rothbatt. 1981) and subtyping (NI, B. Brewer, V. Dull, & L. Lui, 19 1 S. E. Taylor, 1981) effects. The finding traditionally regarded as demonstrating subtyping is that exposure to moderate (compared with extreme) disconfirmers leads to subsequent ratings of the group that are less stereotypic. Despite learning that is incremental and analogous to bookkeeping, the simulations replicate this finding and suggest that the "subtyping" pattern of, results will be drastically reduced if disconfirmers are encountered before the stereotype is well-established. This novel prediction holds with human participants and offers, a tantalizing suggestion: Although moderate disconfirmers may produce more stereotype change, stereotype development might be discouraged by exposure to either extreme or moderate disconfirmers.

17

Quinlan
(1991)

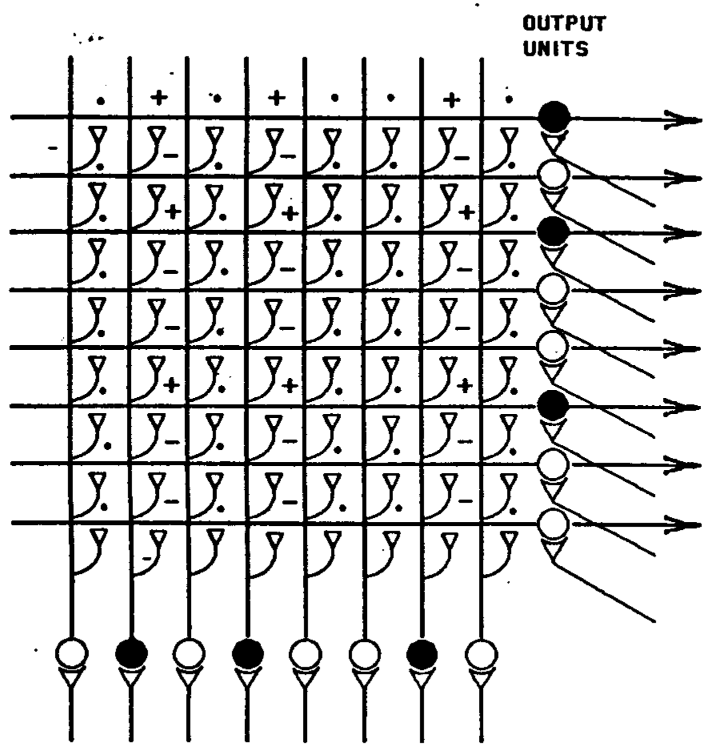


Figure 2.5. Pattern associator comprising eight input units and eight output units. Every input unit is connected to every output unit. All connections are weighted. A '+' signifies an excitatory link and a '−' signifies an inhibitory link. (From McClelland and Rumelhart, 1988, adapted with permission, *see* p. xv.)

Kaplan *et al*
(1992)

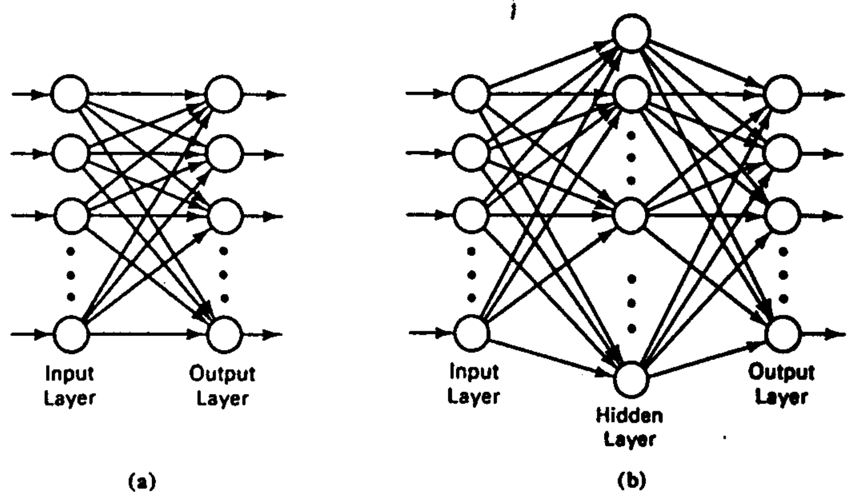


Fig. 6.1. Feed-forward, pattern-associator models: (a) a simple, two-layer pattern associator; (b) a multi-layer, back-propagation network.

that the connectionist models themselves have unavoidable behaviourist implications which their proponents did not intend. We will argue that this, in fact, is the case, but only for a limited (although popular) class of models. We will discuss a class of connectionist models which are not at all vulnerable to the behaviourism characterization.

The argument surrounding "connectionism equals associationism" is more complex. Association *is* a fundamental property of connectionist models, but we will argue that this is not the disadvantage that it is sometimes made out to be.

Connectionism Equals Behaviourism

It is our contention that the behaviourism characterization applies only to a limited class of models which we will refer to as feed-forward pattern associator (FFPA) models. FFPA models are those which are trained to associate arbitrary sets of state vectors, and we include in this class both those models which lack layers of "hidden" units and employ the standard perceptron learning rule (or delta rule) as well as models which do possess hidden units and which employ the more sophisticated "back-propagation" learning rule (Fig. 6.1). The focus on FFPA models is appropriate for several reasons. These models are the subject of ongoing and widespread study within the connectionist community. FFPA models have been used to produce some of the most striking and best-known results of modern connectionism in such areas as speech synthesis (Sejnowski and Rosenberg 1986) and past-tense verb learning (Rumelhart and McClelland 1987). As such, they have been the basis of some of the most revolutionary claims made by the connectionists. Finally, there are even wide-spread efforts to find com-

Rumelhart & McClelland (1986a)

The Three-Stage Learning Curve

The results described in this and the following sections were obtained from a single (long) simulation run. The run was intended to capture

approximately the experience with past tenses of a young child picking up English from everyday conversation. Our conception of the nature of this experience is simply that the child learns first about the present and past tenses of the highest frequency verbs; later on, learning occurs for a much larger ensemble of verbs, including a much larger proportion of regular forms. Although the child would be hearing present and past tenses of all kinds of verbs throughout development, we assume that he is only able to learn past tenses for verbs that he has already mastered fairly well in the present tense.

To simulate the earliest phase of past-tense learning, the model was first trained on the 10 high-frequency verbs, receiving 10 cycles of training presentations through the set of 10 verbs. This was enough to produce quite good performance on these verbs. We take the performance of the model at this point to correspond to the performance of a child in Phase 1 of acquisition. To simulate later phases of learning, the 410 medium-frequency verbs were added to the first 10 verbs, and the system was given 190 more learning trials, with each trial consisting of one presentation of each of the 420 verbs. The responses of the model early on in this phase of training correspond to Phase 2 of the acquisition process; its ultimate performance at the end of 190 exposures to each of the 420 verbs corresponds to Phase 3. At this point, the model exhibits almost errorless performance on the basic 420 verbs. Finally, the set of 86 lower-frequency verbs were presented to the system and the transfer responses to these were recorded. During this phase, connection strengths were not adjusted. Performance of the model on these transfer verbs is considered in a later section.

We do not claim, of course, that this training experience exactly captures the learning experience of the young child. It should be perfectly clear that this training experience exaggerates the difference between early phases of learning and later phases, as well as the abruptness of the transition to a larger corpus of verbs. However, it is generally observed that the early, rather limited vocabulary of young children undergoes an explosive growth at some point in development (Brown, 1973). Thus, the actual transition in a child's vocabulary of verbs would appear quite abrupt on a time-scale of years so that our assumptions about abruptness of onset may not be too far off the mark.

Figure 4 shows the basic results for the high frequency verbs. What we see is that during the first 10 trials there is no difference between regular and irregular verbs. However, beginning on Trial 11 when the 410 midfrequency verbs were introduced, the regular verbs show better performance. It is important to notice that there is no interfering effect on the regular verbs as the midfrequency verbs are being learned. There is, however, substantial interference on the irregular verbs. This interference leads to a dip in performance on the irregular verbs.

18

Rumelhart & McClelland (1986a)

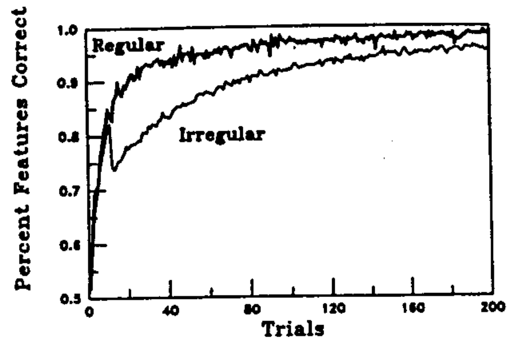


FIGURE 4. The percentage of correct features for regular and irregular high-frequency verbs as a function of trials.

Equality of performance between regular and irregular verbs is never again attained during the training period. This is the so-called U-shaped learning curve for the learning of the irregular past tense. Performance is high when only a few high-frequency, largely irregular verbs are learned, but then drops as the bulk of lower-frequency regular verbs are being learned.

while the strengths of regularized alternatives jump up. From about Trials 11 through 30, the regularized alternatives together are stronger than the correct response. After about Trial 30, the strength of the correct response again exceeds the regularized alternatives and continues to grow throughout the 200-trial learning phase. By the end, the correct response is much the strongest with all other alternatives below .1.

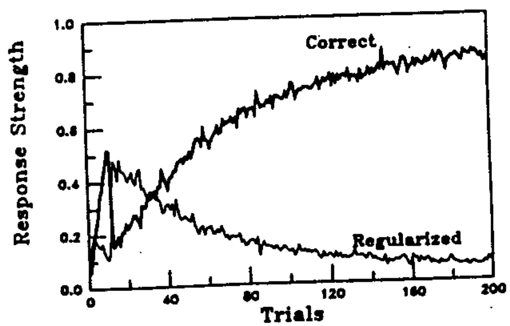


FIGURE 5. Response strengths for the high-frequency irregular verbs. The response strengths for the correct responses are compared with those for the regularized alternatives as a function of trials.

[4] Unless otherwise indicated, the regularized alternatives are considered the base+ed and past+ed alternatives. In a later section of the paper we shall discuss the pattern of differences between these alternatives. In most cases the base+ed alternative is much stronger than the past+ed alternative.