Developing a Domain-General Framework for Cognition: What is the Best Approach?

Jay McClelland
with
Dave Plaut, Steve Gotts, and Tiago Maia

CNBC
A Joint Project of Carnegie Mellon and the University of Pittsburgh

Overview

- A little bit of my own philosophy of science
- Why I became a connectionist
- The search for domain general principles
- How the search is carried out
- The current state of the PDP framework

- How we should think about
  'the symbolic level of thought'

- How to think about the relationship between different models and frameworks in Cognitive Science

Philosophy of Science

- There are no 'true' theories, just theories that are more or less useful.

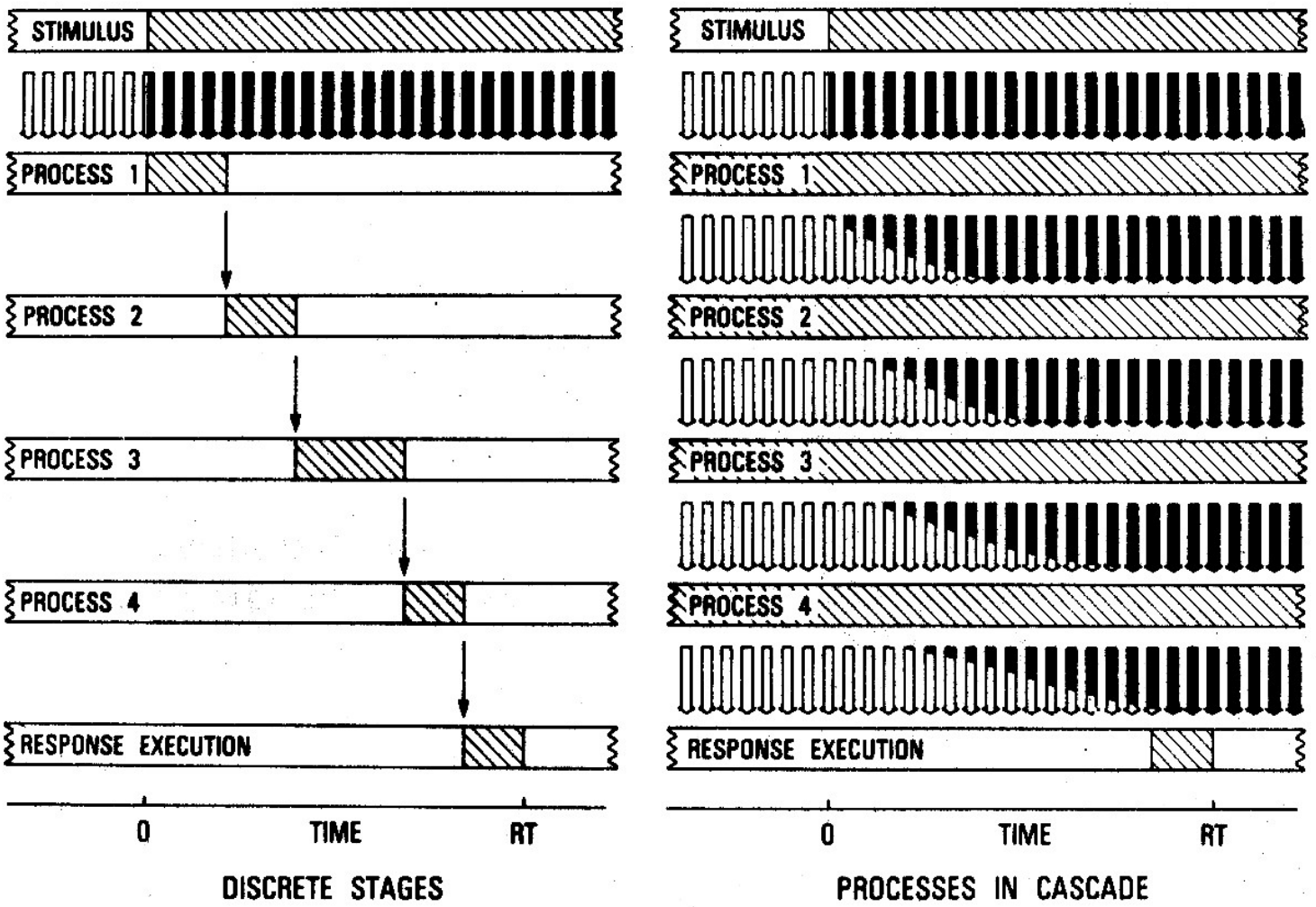
- Working within a particular model or framework doesn't require belief in the truth of a model, only a bet that it will prove useful.

- 'Usefulness' can be with respect to many different goals. Here are two:
  - Building a better mousetrap
  - Exploring the implications of particular principles for understanding human cognition

Why I Became a Connectionist

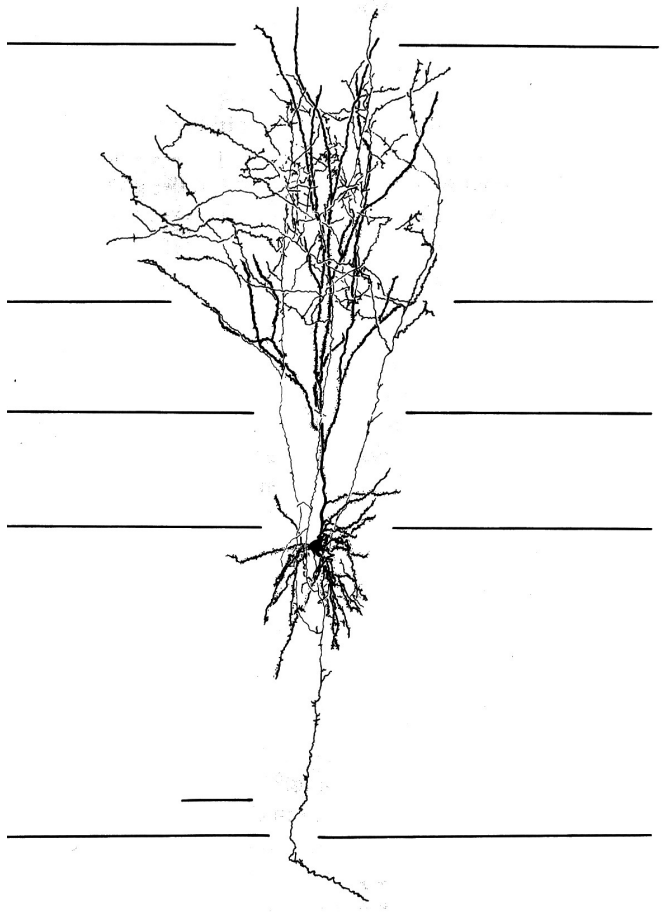
- People recognize a word more accurately than the letters that it contains, while at the same time depending on abstract knowledge of what the word looks like for its recognition.
  - That is, the word superiority effect can be obtained even when visual familiarity is disrupted using MiXeD UpPeR aNd LoWeR CaSe TyPe.
- How is this possible?
- Perhaps we use graded activations in a multilayer processing system.
- What are the implications of the use of such graded representations in a multi-layer hierarchy?

What are the consequences of assuming that mental processes involve a cascade of information through a series of continuous processing stages rather than a sequence of discrete information processing steps?

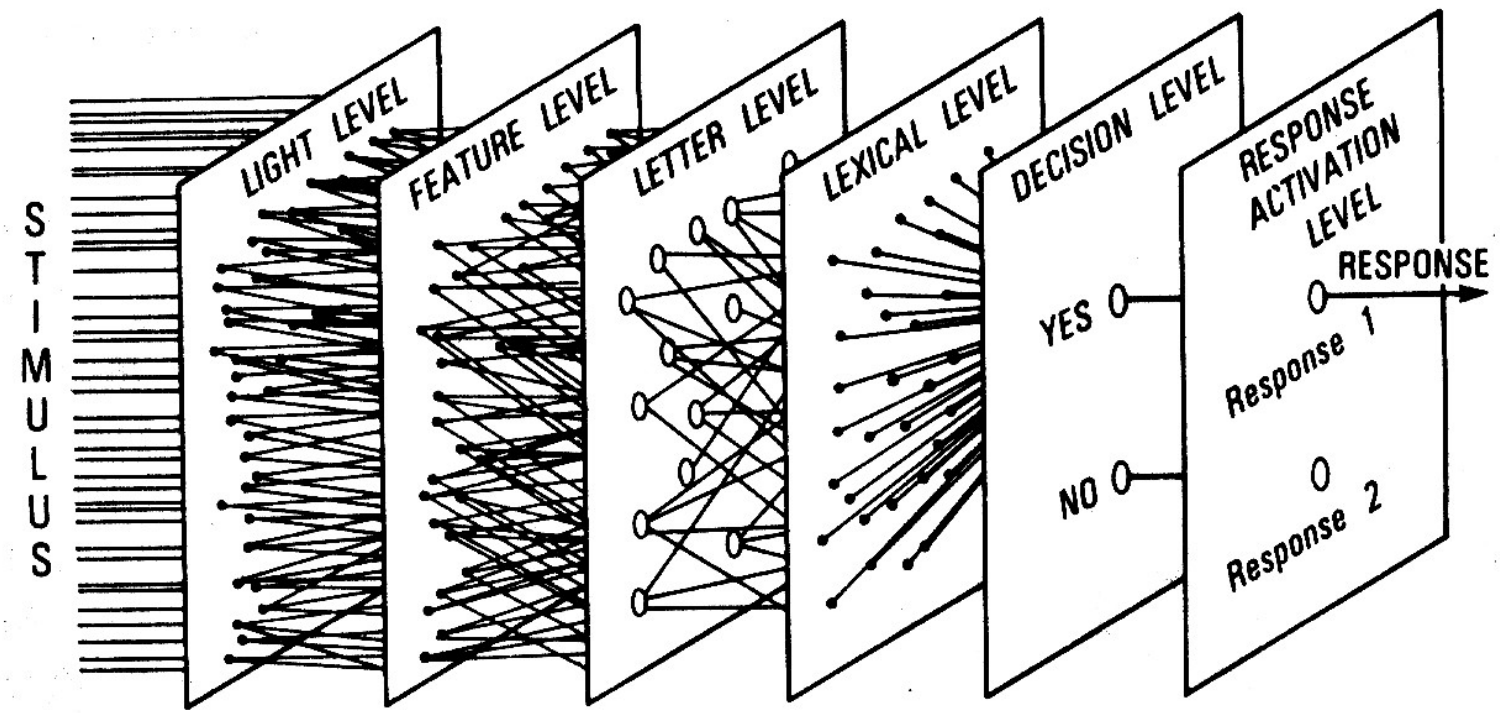


The Basis of Cognition I

- Cognitive processes occur via the propagation of excitatory and inhibitory signals among neurons via weighted synaptic connections.



It became clearer how exactly to think about this once I formulated the model as a neural network…



… and then added specific assumptions about how the units in the network propagate information via activation:

$$da_{nj}/dt = k_{nj}(i_{nj} - a_{nj}); \quad i_{nj} = \Sigma_i \, w_{ji} a_{(n-1)i}$$

- Adopting a neural network framework allowed us to address a number of specific questions:

  - What are consequences of assuming that mental processes involve a cascade of information through a series of continuous processing stages rather than a sequence of discrete information processing steps?

  - Does it really make sense to think that perception could involve a bi-directional propagation of information? Can we account for human behavior in psychological experiments through the interactive propagation of activation?

  - Is it possible for a single computational mechanism to simultaneously encompass both regularities across items and item-specific idiosyncrasies?

Table 2

Comparison of Inferences Derived From the Discrete Stage Model and the Cascade Model

Condition	Discrete stage model	Cascade model
If factors interact	They affect the duration of the same process.	They affect the rate of the same process, *or* they both affect relative asymptotic activation, *or* one affects the rate of the rate-limiting process and the other affects the relative asymptotic activation.
If effects are additive	They affect the durations of different processes.	They affect the rates of different processes, *or* one affects the rate of a fast process and the other affects the asymptote.

Note. Inferences from the cascade model assume a fixed criterion and very low error rates. The locus of an asymptote effect cannot be determined from the pattern of additivity and interaction.

Does it really make sense to think that perception could involve a bi-directional propagation of information? Can we account for context effects in perception through interactive propagation of activation?

- Fodor and others had argued that interactive processing is logically incoherent:

  How can you use word level information to help you process the letters if activation first propagates through the letter level to reach the word level?

- This doesn't seem to be a problem in a system that gradually builds up activation values via propagation of activation through excitatory and inhibitory links.

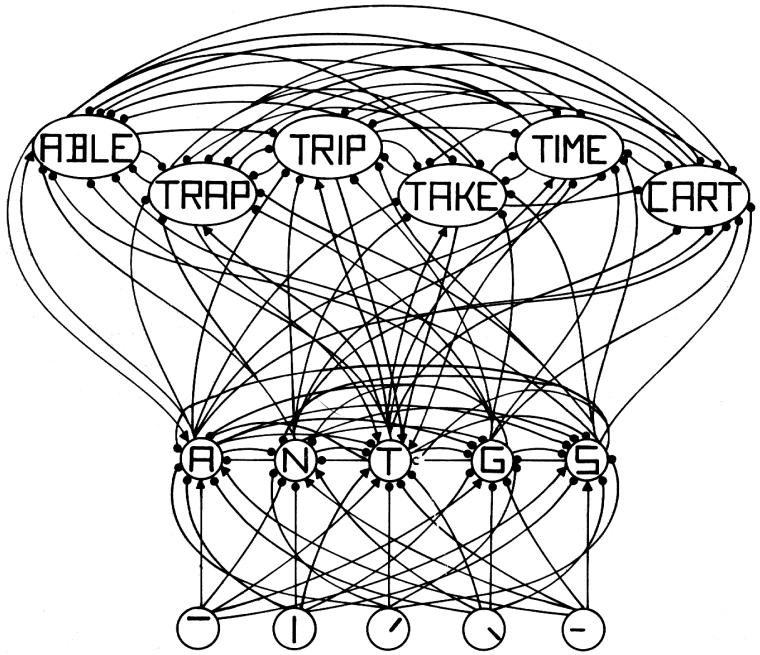


FIGURE 7. The unit for the letter *T* in the first position of a four-letter array and some of its neighbors. Note that the feature and letter units stand only for the first position; in a complete picture of the units needed from processing four-letter displays, there would be four full sets of feature detectors and four full sets of letter detectors. (From "An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings" by J. L. McClelland and D. E. Rumelhart, 1981, *Psychological Review, 88,* p. 380. Copyright 1981 by the American Psychological Association. Reprinted by permission.)

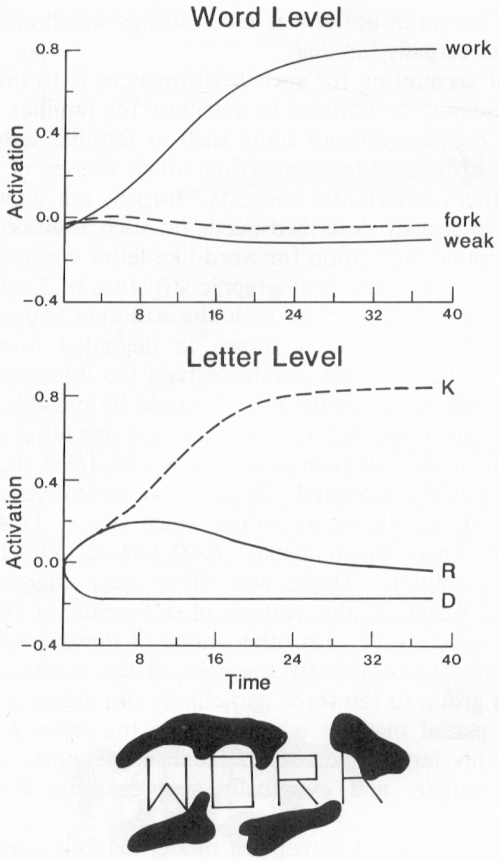


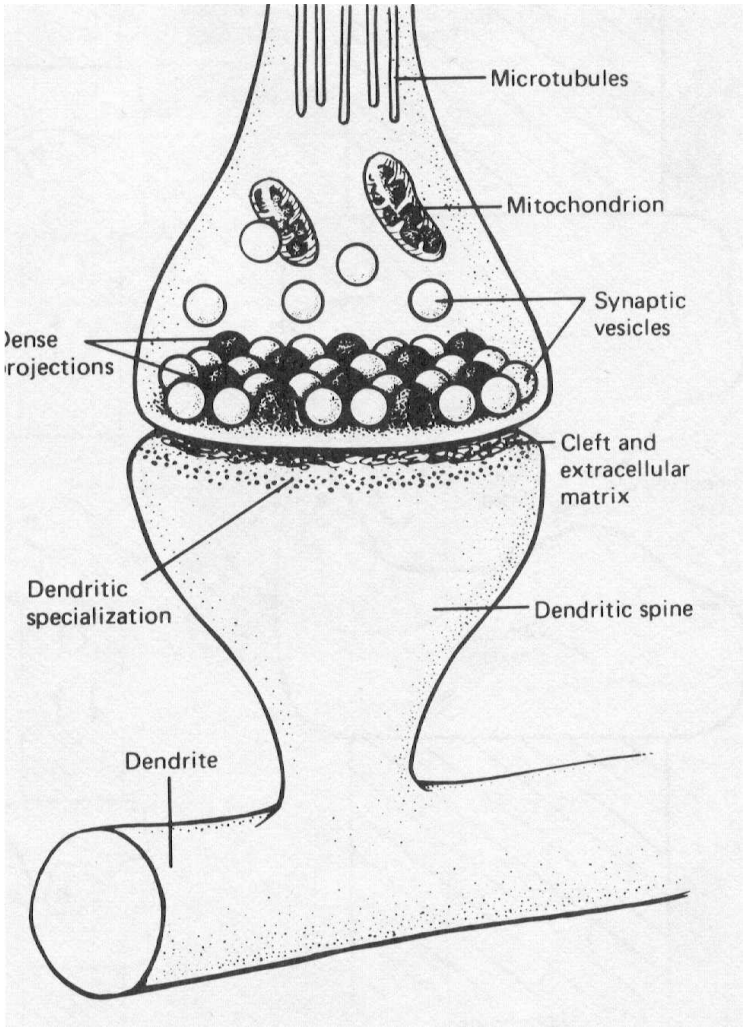
FIGURE 8. A possible display which might be presented to the interactive activation model of word recognition, and the resulting activations of selected letter and word units. The letter units are for the letters indicated in the fourth position of a four-letter display.

Comments

- I don’t claim that these ideas could only have been obtained by thinking of the problem in terms of a neural network.
- I simply claim that thinking of it in this way has helped guide me and others to explicit formulations that led to useful discoveries observations.
- I acknowledge that this thinking involves drastic simplification of many complex aspects of the propagation of activation among real neurons.
- The complexities can and do have implications for models at the cognitive level and it is part of my own agenda to understand what these implications are.
- Even so the simplifications are very useful and contribute to the ability to gain insight from the approach.

The Basis of Cognition II

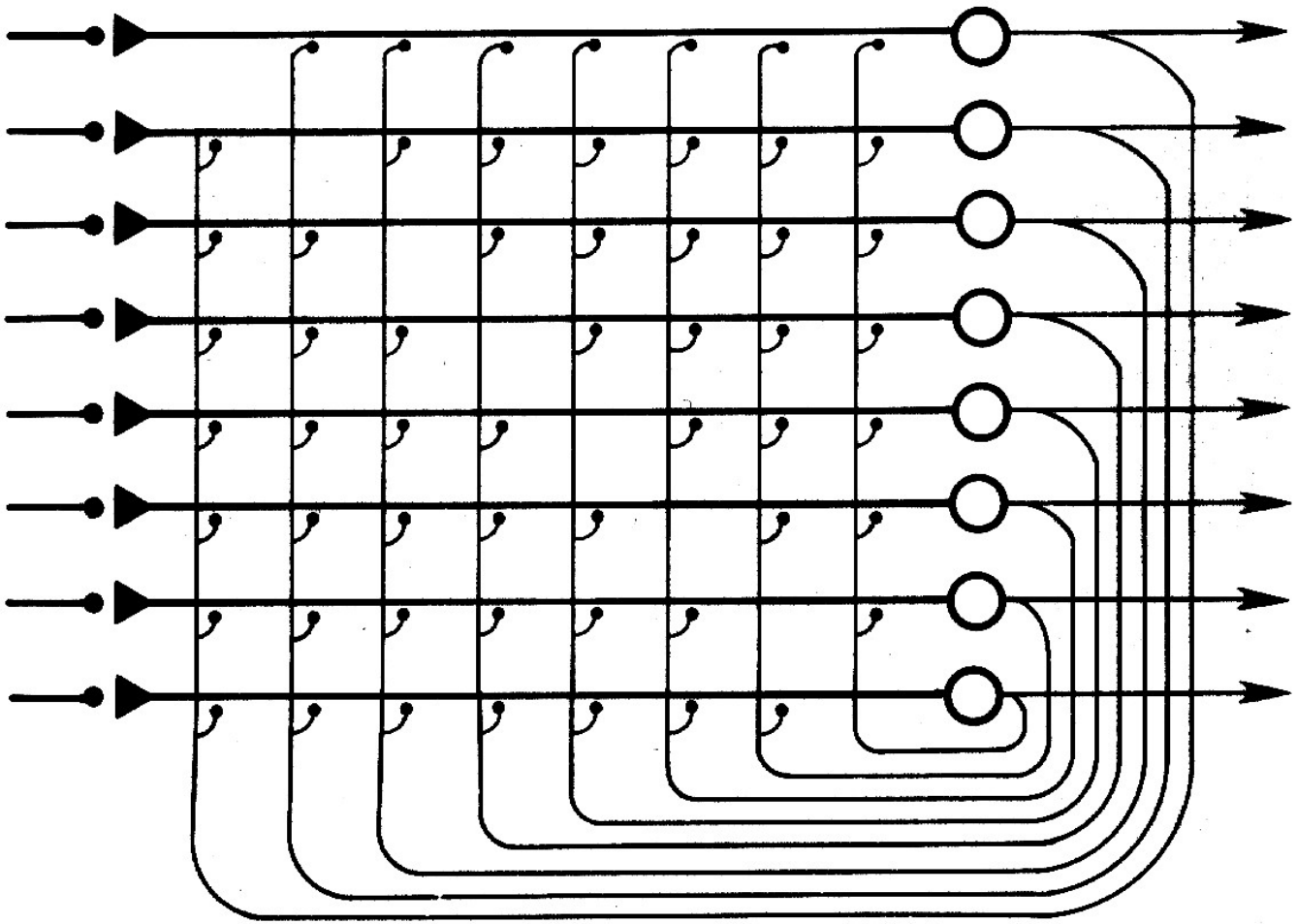
- Cognitive processes occur via the propagation of excitatory and inhibitory signals among neurons via weighted synaptic connections.
- The knowledge underlying cognition is stored in the strengths of the connections among the neurons.
- Learning occurs through the adjustment of the strengths of the connections.



Issues that arise when one considers learning in a neural network

- Biologically motivated learning rules are formulated in terms of equations specifying how pre- and post-synaptic signals modify the strengths of connections.
  - They are *not* formulated in terms of specifying under what circumstances a new cognitive entity such as a feature, letter or word unit should be added or inserted into a network (as in symbolic models or localist connectionist networks).
  - My attempts to force connectionist learning rules to assign units to represent cognitive entities resulted in failure…
  - But along the way it became clear that one might be able to do without such units, if learning lead the network to behave in a way that was consistent with the psychological data.

An Eliminative Connectionist Network (McClelland & Rumelhart, 1985)



Points Addressed with the MR85 Auto-Associator Model

- What is a 'memory trace', i.e. the trace left in the brain by an experience?
- It is not a copy of the mental representation formed as a result of processing the item, but…
  The ensemble of changes to the strengths of connections in the network.
- This changes influence processing of subsequent patterns, accounting for priming. Many later models have adopted a similar approach.
- The accumulated changes from a set of inputs shows sensitivity to the central tendency of an ensemble of patterns, and at the same time accounting for sensitivity to properties of individual items previously experienced.

➢ Perhaps we do not need explicit representations of items or of abstractions over items (categories, rules) to capture sensitivity to regularities in experience.

What has been eliminated?

- The need to postulate that each item ever encountered is stored separately in the brain (an idea that has always struck me as mechanistically untenable).

- The need to postulate the formation of explicit representations of prototypes (or other abstracted cognitive constructs like schemas, rules, and theories).

- The need to identify individual units corresponding to a wide range of putative cognitive entities including
  - Feature detectors
  - Letter detectors
  - Logogens or word detectors

The Search for Domain General Principles

- One thing the PDP community shares with the ACT community (and the SOAR community) is a commitment to the search for Domain General Principles.
- Clearly there are divergent views on whether such a search can be worthwhile.
  - Chomsky, Marr, Fodor, and Keil have all rejected this approach.
  - Many others stress the existence of domain-specific evolutionary constrains as aptly point to evidence of domain specificity in humans and other animals.
- My own gut feeling is that there is a lot of mileage to be gained by seeking broad domain-general principles.
- There is domain-specificity but in my view most of it is either acquired or based on variations on general themes.
- In either case, it is likely that insight into each particular domain will be enhanced by bringing what has been learned from other domains into consideration.

How is the Search to be Carried Out?

- Breadth-first, approximative approach?
  - This is what I recall Newell advocated
- Depth-first approach, focusing on just one topic at a time?
- Both of these approaches seem inherently limiting.
  - It is important to take discrepancies seriously
  - It is important to bring ideas from one domain into the investigation of other domains
- This leads to the approach we have adopted in pursuing the PDP framework.

My own preferred approach:

Iteratively explore the adequacy of domain-general principles across a few target domains

a. Begin by formulating a putative set of principles.

b. Develop models based on these principles and apply them to a few carefully chosen target domains.

c. Assess the adequacy of the models so developed, and attempt to understand what really underlies both the successes and the failures of the models.

d. Use the results of the analysis to refine and elaborate the set of principles.

e. Return to step (b).

Advances in the Development of the Framework Arising From This Approach I

- Empirical shortcomings of the interactive activation model (McClelland and Rumelhart, 1981):
  - Failure to fit the quantitative pattern of trade-off between context and stimulus information
  - Did this reflect a fundamental shortcoming of the assumption of interactivity as Massaro claimed?
  - After careful analysis, we (McClelland, 1991; Movellan and McClelland, 2001) concluded that the answer was no.
  - Instead, the fact that the model was not intrinsically noisy was at fault.
- This work led to the principle that processing is inherently noisy or stochastic as well as graded, interactive, and nonlinear, embodied in much of our subsequent research (e.g., Usher and McClelland, 2001).

Advances in the Development of the Framework Arising From This Approach II

- Shortcomings of the Seidenberg and McClelland model of single word reading:
  - The model captured human performance in word reading, but its performance on non-words was not up to human levels.
  - This led two separate groups to argue that the problem lay fundamentally in the model's reliance on a single mechanism for processing both words and non-words.
  - However, Plaut, McClelland, Seidenberg and Patterson (1996) analyzed the SM model and found that the problem lay in its use of input and output representations that dispersed the regularities present in words across different units.
- The work led to the principle that the distributed representations used in connectionist models must be highly overlapping across items if they are to provide the basis for generalization.

The Current State of the PDP Framework

- We assume that the processing in networks of simple processing units is:
  - Graded, random, interactive, and non-linear
- We assume that representation is distributed, and recognize that variation of the extent of overlap may be an important strategy the brain uses to vary its sensitivity to shared structure vs. individual item characteristics.
- We are still exploring the key characteristics of the principles by which adjustments to the strengths of connections occur in the course of experience.
- We draw on the results of modeling studies as well as findings from neuroscience in our effort to further refine and develop the framework.

How should we think about 'the symbolic level of thought'?

- Many researchers appear to hold the view that human cognitive processes (or at least some modules within the human cognitive system) are strongly constrained by evolution and development so that they are effectively programmed or implemented in the brain at 'the symbolic level'.
  - Newell, Fodor, Anderson & Labiere, Pinker, and Marcus all appear to be among those adhering to this view.

How should we think about 'the symbolic level of thought'?

- My own belief is that approximate conformity to the characteristics of symbol processing machines, and even a tendency to promote such conformity, is a characteristic of the neural networks in certain parts of our brains
  - Especially those parts that exploit highly overlapping distributed representations.
- However, human performance is on close scrutiny more graded and interactive that one would naturally expect based on symbolic approaches.
- More importantly:
  - Directly modeling human cognition at the symbolic level leads to models that miss many of the ways in which our cognitive abilities exploit and promote regularity and systematicity.

How should we think about 'the symbolic level of thought'?

- By using models developed at the neural network level we have captured the systematicity in human performance more fully than has been achieved by existing models that have been formulated at the symbolic level.
- Thus, it is better to view conformity to principles of symbolic processing as an approximation or shorthand that may be useful for some purposes.
- The adequacy of the approximation that may be achievable by models formulated at this level may continue to improve as modelers working at the symbolic level continue the trend of incorporating principles of parallel-distributed processing.

Brief Elaboration of The Two Crucial Points in the Context of the Past Tense Debate

- Human performance is on close scrutiny more graded and interactive that one would naturally expect based on symbolic approaches.

- Directly modeling human cognition at the symbolic level leads to models that miss many of the ways in which our cognitive abilities exploit and promote regularity and systematicity.

Pinker's statements about the past-tense rule

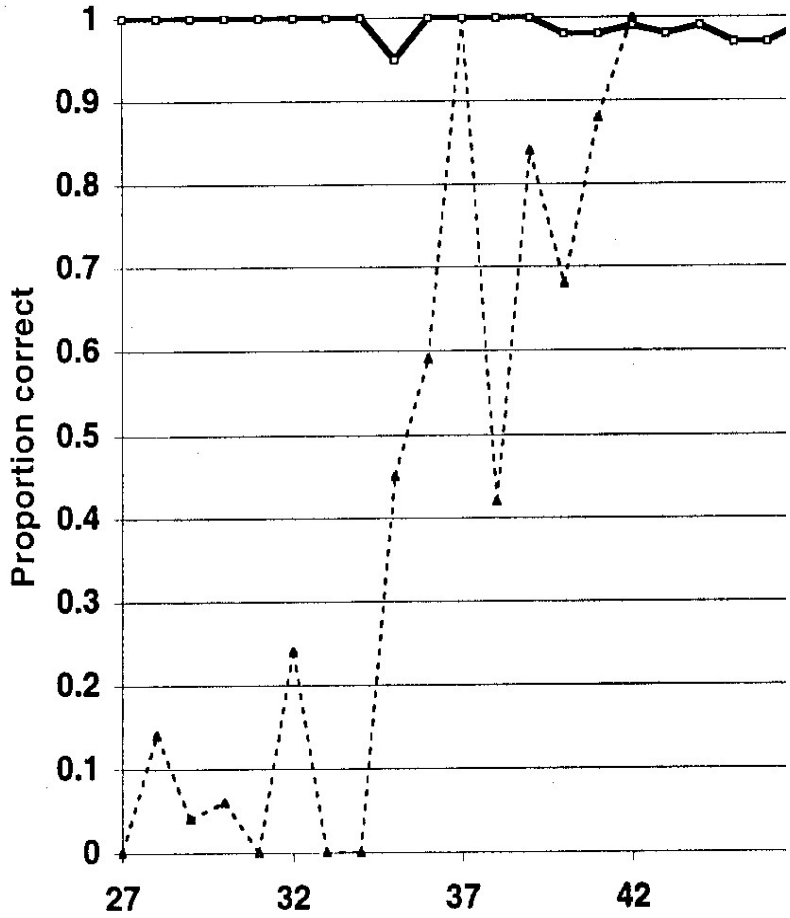
- It is a symbolic rule that applies to items regardless of their phonological or semantic characteristics.
- It is something one either has or does not have.
- Its discovery occurs as an an 'epiphany'; or in a 'Eureka moment' (Pinker, *Words and Rules*, 2001).

Marcus et al's portrayal of the acquisition of the regular past tense.

- "Adam's first over-regularization occurred during a three-month period in which regular marking increased from 0 to 100%"

Did Marcus et al see the data through rule tinted glasses?

- Actually Adam's use of the regular gradually increased over a period of about 1 year.
- The particular data points cited by Marcus et al were based on very small samples and would be expected to be intrinsically noisy if use of the rule is variable.
- The data also indicate that initially the rule applies only to certain subtypes of verbs and gradually spreads to others over the course of development (Shirai and Anderson, 1995).



"It is true of all of the grammatical morphemes in all three children that performance does not abruptly pass from total absence to reliable presence. This is always a considerable period … in which production where required is probabilistic. This is a fact that does not accord well with the notion that the acquisition of grammar is a matter of the acquisition of rules, since rules in a generative grammar either apply or do not apply. One would expect rule acquisition to be sudden."

Roger Brown, A First Language, 1973

An Alternative Characterization of the Data (McClelland and Patterson, 2002)

- There is no sudden onset in the acquisition of the regular past tense.
  - It is acquired gradually over the period of about one year.
  - It is initially applied to a subset of words and only gradually extends across the full range of regular words.
- Other key points:
  - There are frequency effects and semantic and phonological similarity/neighborhood effects in regular as well as exception words.
    - The effects are weaker in the regular words, consistent with a robust characteristic of the relevant connectionist models.
  - There is no double dissociation in past-tense inflection.
    - Although a deficit in semantics differentially impacts exceptions, as expected under connectionist models that incorporate semantics, there is no reverse dissociation.
- Thus, it appears that the rule-based characterization is at best an approximation even for the regular inflections.

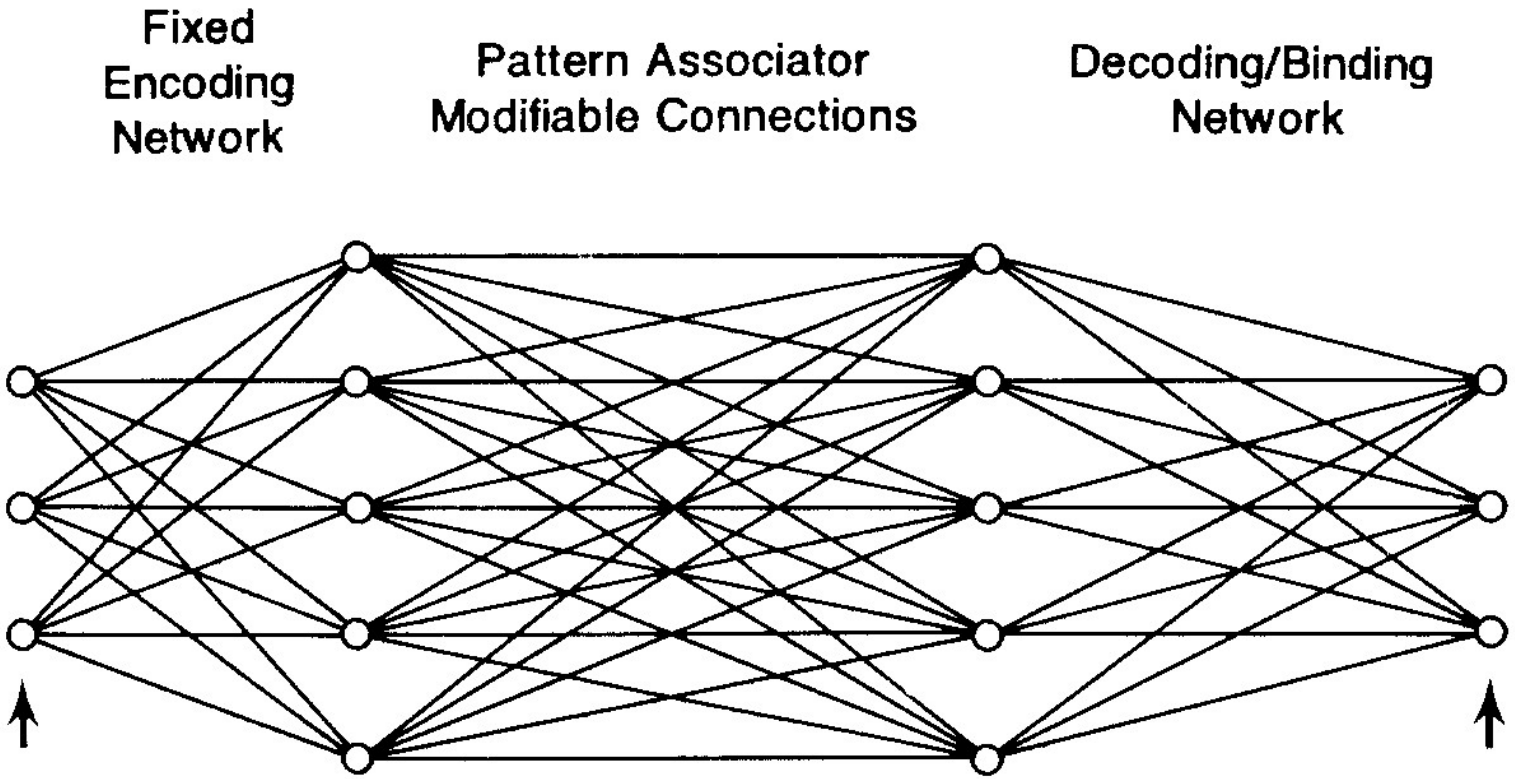
Quasi-Regularity

- There is systematicity/regularity in almost all exceptions as well as in regular items.
- We call this 'quasi-regularity' and in our view it is ubiquitous.
- In word pronunciation:
  - PINT, BREAD, and even AISLE and HYMN share may aspects of regular items
- In past tense inflection:
  - DID, MADE, HAD, SAID, KEPT, WEPT, MEANT, DREAMT, THOUGHT, BOUGHT, CAUGHT, SOUGHT, and many other exceptions add /d/ or /t/ as other regular items do.
- In idiomatic language:
  - 'Their goose is really gonna get cooked.'
- In semantics:
  - While chickens, penguins and ostriches differ from 'regular' birds in important ways they still share many properties with robins, sparrows, and eagles

Handling of Quasi-Regularity in Symbolic and PDP Models

- Symbolic models such as those of Pinker and of Taatgen and Anderson fail to capture this 'quasi-regularity' in exceptions.
- Distributed models such as the ones used by Rumelhart and McClelland and subsequently by many others are intrinsically structured in such a way as to make capturing such structure automatic.
- Models of this type have been successfully applied to comprehension (St. John and McClelland, 1989; Rohde, 2001) as well as single-item processing, and their future is very promising.
- The connectionist models therefore capture more structure than existing symbolic approaches. They capture the regularity in the regular items and the quasi-regularity in exceptions.

The Rumelhart & McClelland Past Tense Model



"keep"

"kept"

How should we think about the relationship between different models or frameworks in Cognitive Science?

- The debate between connectionist models and other models has often been construed as a fundamental clash of paradigms.
- Very often, however, there clash appears to be one between different levels of description.
- Models cast at more abstract levels can be extremely useful.
- So can models cast at more detailed levels.
- It's time we stopped asking which level is the 'right' level.
- It seems more likely that insight will come from the parallel exploration of a range of approaches, together with an effort to understand the relationships between them.
- This is an idea that Allan Newell promoted, and that John Anderson also appears to endorse.
- An understanding of the importance of multiple approaches is one of the great strengths of the Carnegie Mellon environment.