

Perception, Thought and Language as Graded Constraint Satisfaction Processes: Models and Theory

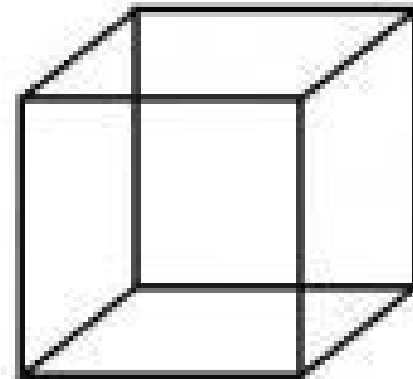
Jay McClelland

SymSys 100

April 12, 2011

Perception as Constraint Satisfaction

- For the first figure, one tends to see it as nothing until the 'solution' emerges.
 - The pieces and the whole simultaneously support each other
- For the second, there are two alternative solutions, each involving a re-interpretation of every part of the larger whole.
 - Again the parts and the whole are mutually consistent



Ubiquity of the Constraint Satisfaction Problem

- In sentence processing
 - I saw the grand canyon flying to New York
 - I saw the sheep grazing in the field
- In comprehension
 - Margie was sitting on the front steps when she heard the familiar jingle of the “Good Humor” truck. She remembered her birthday money and ran into the house.
- In reaching, grasping, typing...

Puzzles and Thought Problems that Can Involve Constraint Satisfaction

- Find a word that you can combine with each of the next three words to make a compound word:

Pine, crab, tree

- Make a single word by rearranging the letters in the phrase below:

A rope ends it

- You know how much gold weighs per cubic centimeter, and you want to test whether the king's golden crown is pure gold. But you don't know the volume in cc's of the crown. How can you find out?

How Geniuses Think

Feynman said to Dyson ... that Einstein's great work had sprung from physical intuition and when Einstein stopped creating it was because 'he stopped thinking in concrete physical images and became a manipulator of equations.' Intuition was not just visual but also auditory and kinesthetic. Those who watched Feynman in moments of intense concentration came away with a strong, even disturbing sense of the physicality of the process, as though his brain did not stop at the gray matter but extended through every muscle in his body.

Gleick, p. 244.

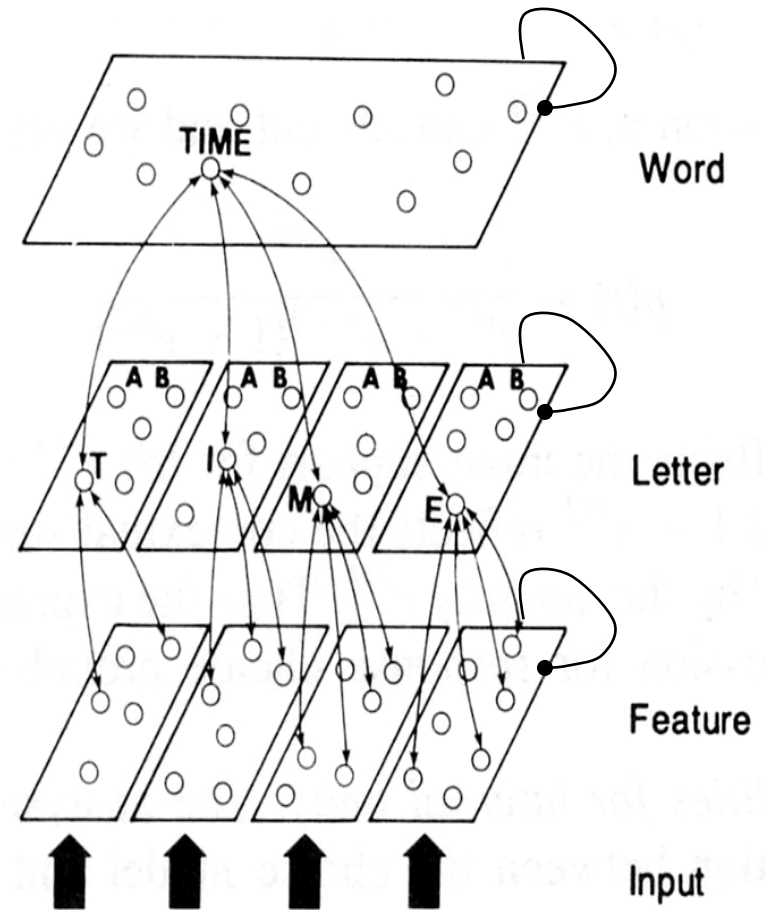
- Barwise and Etchemendy, two Stanford logicians, agree
 - They argue that nearly all great discoveries arise through intuition, not manipulation of equations
 - After mathematicians make a discovery, they then attempt to verify it formally
 - While the insight may have happened suddenly, the proof may take years to develop
 - Once a proof has been established, it is typically reported in a publication that says nothing about the intuition behind it

Finding Perceptual Solutions

- It appears that our brains can search for alternative solutions until one pops out.
- How are such solutions found?
 - One answer is that the process occurs through a noisy, interactive activation process.

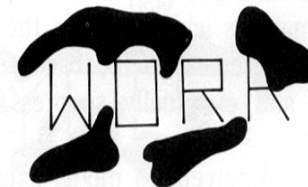
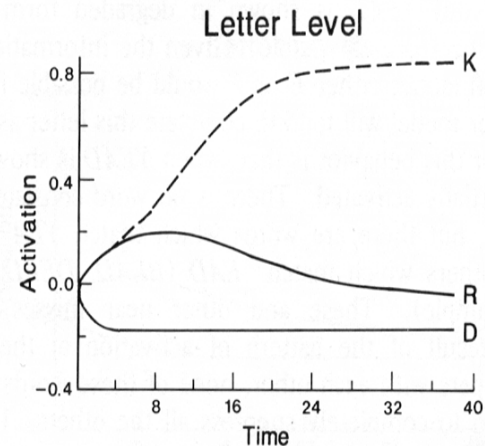
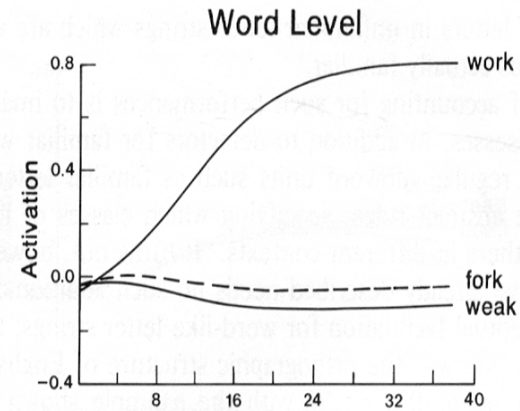
The interactive Activation Model: A Mutual Constraint Satisfaction Process

- Units represent hypotheses about the visual input at several levels and positions.
 - Features
 - Letters
 - Words
- Connections code contingent relations:
 - Excitatory connections for consistent relations
 - Inhibitory connections for inconsistent relations
 - Lateral inhibition for competition among mutually inconsistent possibilities within levels.
- Connections run in both directions
 - So that the network tends to evolve toward a state of activation in which everything is consistent.



Interactive Activation *Simultaneously* Identifies Words and Letters

- Stimulus input comes first to letter level, but as it builds up, it starts to influence the word level.
- Letter input from all four positions makes *work* the most active word unit (there is no word *worr*).
- Although the bottom up input to the letter level supports K and R equally in the fourth letter position, feedback from the word level supports K, causing it to become more active, and lateral inhibition then suppresses activation of R.



Findings Addressed by the IA Model

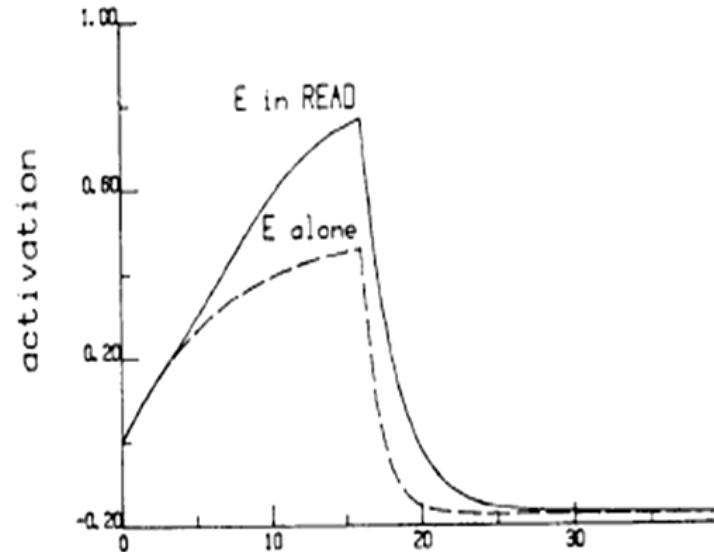
- The word superiority effect (Reicher, 1969).
 - Subjects identify letters in words better than letters in scrambled strings or single letters.
- The pseudoword advantage.
 - The advantage over single letters and scrambled strings extends to pronounceable non-words (e.g. LEAT / LOAT...)
- The contextual enhancement effect.
 - Increasing the duration of the context or of the target letter facilitates correct identification.
- Reicher's experiment:
 - Words are constructed in pairs differing by one letter.
 - E.g. READ vs ROAD.
 - The 'critical letter' is the letter that differs between members of the pair.
 - Critical letters occur in all four positions.
 - On a given trial, one member of the pair is shown, or a scrambled version of the word is shown, or just the critical letter is shown.
 - E.g., READ or EADR or E
 - The display is followed by a mask made of lines or letter fragments.
 - The critical letter and the alternative then appear as choice alternatives, with dashes indicating the tested position.
 - E.g. _O__ _E__

How the Model Works: Words vs. Single Letters

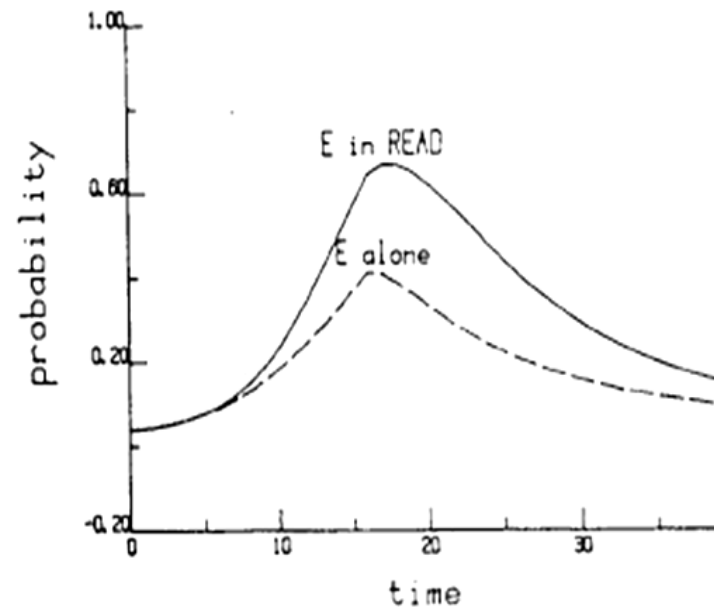
Top panel shows activations, driven by the letter display and then the mask

Bottom panel shows choice probabilities, based on a moving average of activations.

letter level activations



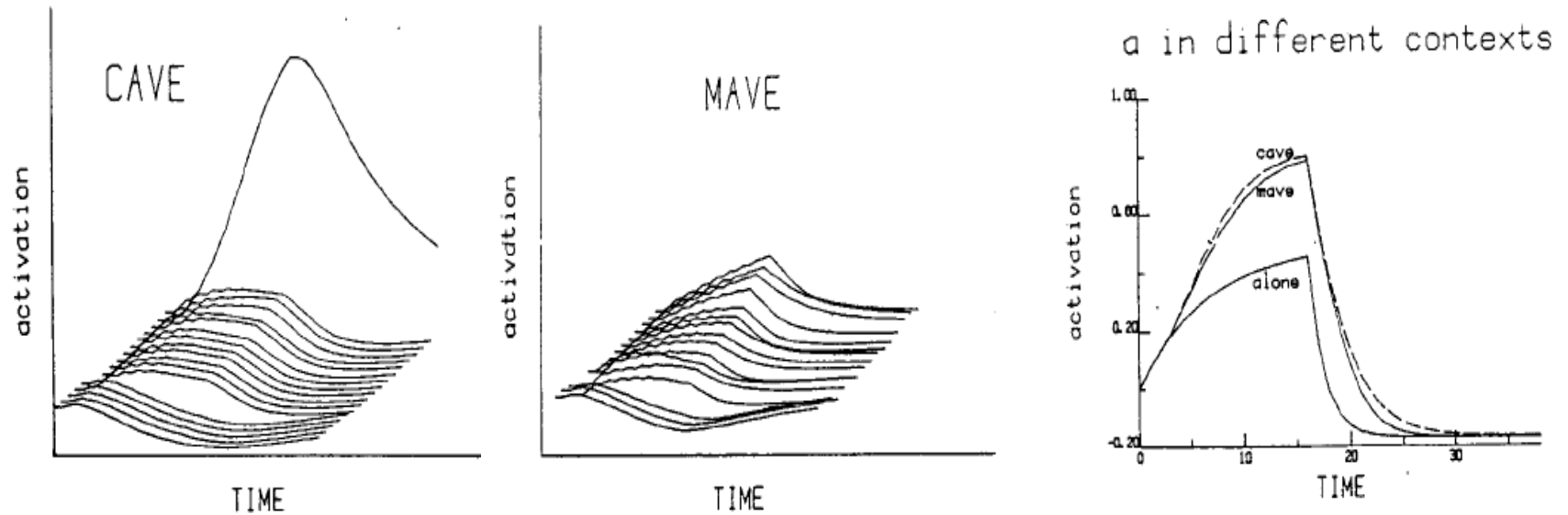
output values



Explaining the Pseudoword Advantage as a Conspiracy of Mental Agents

- The fact that people show facilitation for items in nonwords (LEAT, MAVE) as well as words could mean they 'know the rules' by which words are formed.
- The model offers an alternative explanation, based on the idea that a 'conspiracy of mental agents' each partially consistent with the input, produces the facilitation.
- Models based on such conspiracies are used to explain how we categorize novel examples, infer missing properties of things, and bias our inferences based on other similar things that we know.
- This idea is a precursor of the idea that rules of language arise in models using distributed representations, as discussed in the *Past Tense Debate* reading.

Word and Letter Level Activations for Words and Pseudowords



Left and Center:

Each curve indicates the activation of one of the word units in the model after presentation of the indicated input.

Right:

Activation of unit for 'a' in the second position, for CAVE, MAVE, and A alone.

A Computational Theory of Graded Constraint Satisfaction

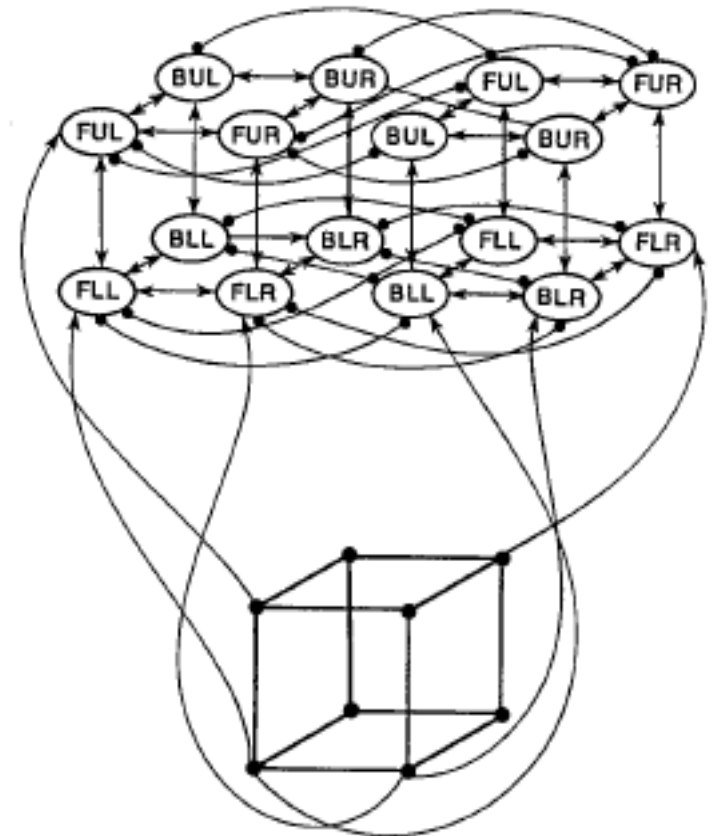
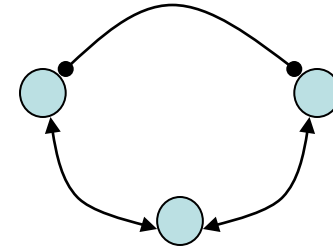
- Consider a network with symmetric connections, i.e. for all pairs of units i, j :

$$W_{ij} = W_{ji}$$

- Provide external input e_i to some of the units.
- Define the Goodness of a state of the network as:

$$G(s) = \sum_{i>j} w_{ij} a_i a_j + \sum_i e_i a_i$$

- As the network settles it tends toward states of higher goodness
- Examples:
 - Three-unit network
 - Necker Cube Network
 - Interactive activation network
- Noise in the settling process allows networks to jump out of local Goodness maxima.
- If we gradually reduce the noise, we can guarantee finding the best solution.



The Relationship Between Goodness and Probability in the Boltzmann Machine

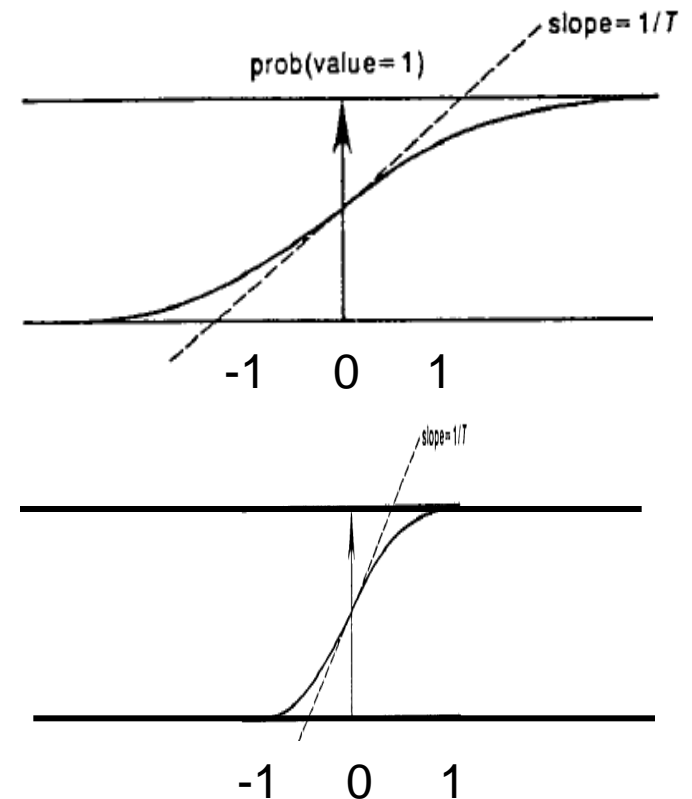
- A 'Boltzmann Machine' is a stochastic neural network in which units' activations are set to 0 or 1 with probability:

$$P(a_i = 1) = e^{\text{net}_i/T} / (1 + e^{\text{net}_i/T})$$

- When T is large, the Boltzmann machine can jump from state to state easily.
- If one gradually reduces T , one can reach an *equilibrium* in which the probability of being in a particular state is given by:

$$P(s) = e^{G(s)/T} / (\sum_{s'} e^{G(s')/T})$$

- Here s refers to one state and s' ranges over all possible states.



Whole Network States as Perceptual Inferences

- It can be shown for some kinds of cases that if the weights, biases, and external inputs are set appropriately, then the quantity

$$e^{G(s)/T} / (\sum_{s'} e^{G(s')/T})$$

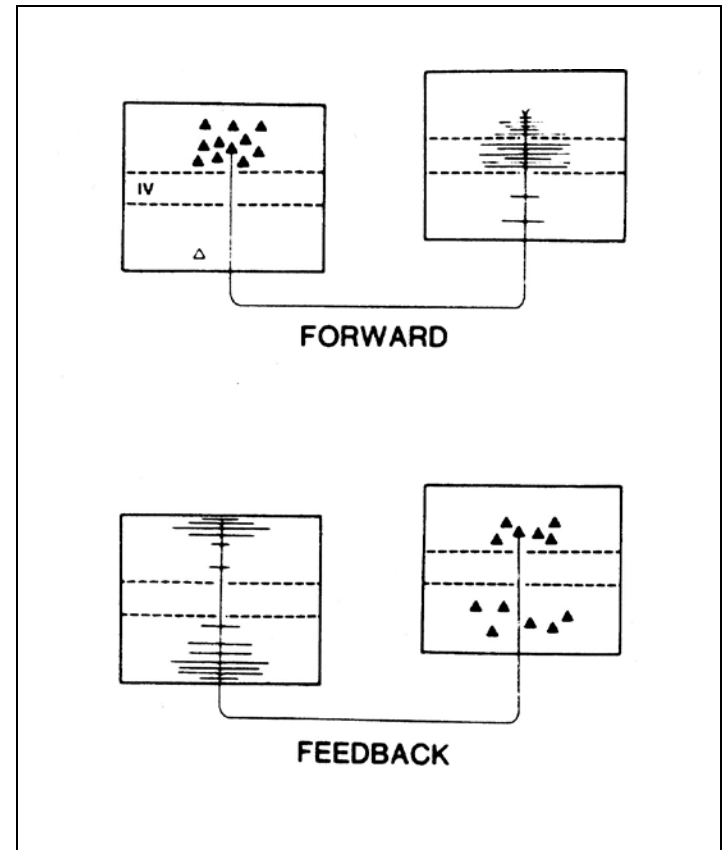
represents the posterior probability that the state corresponds to the correct interpretation of the input.

The beauty of this is that now whole states – constellations of hypotheses – play the role of individual hypothesis.

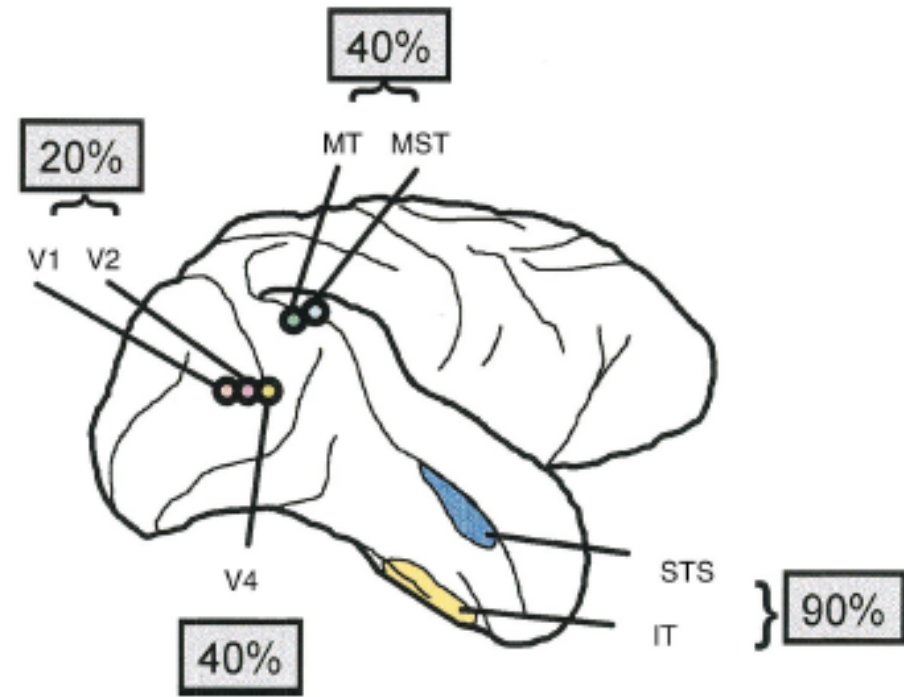
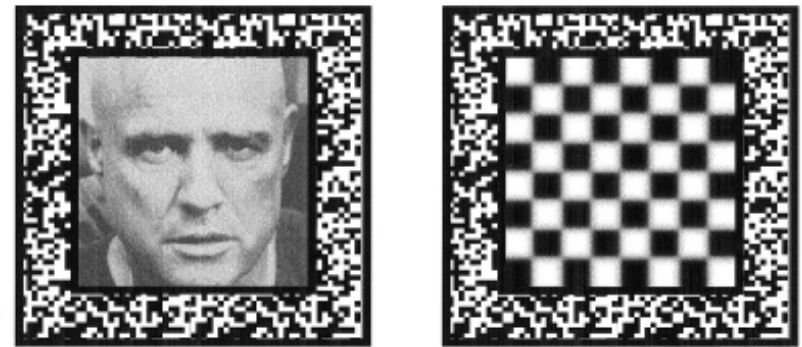
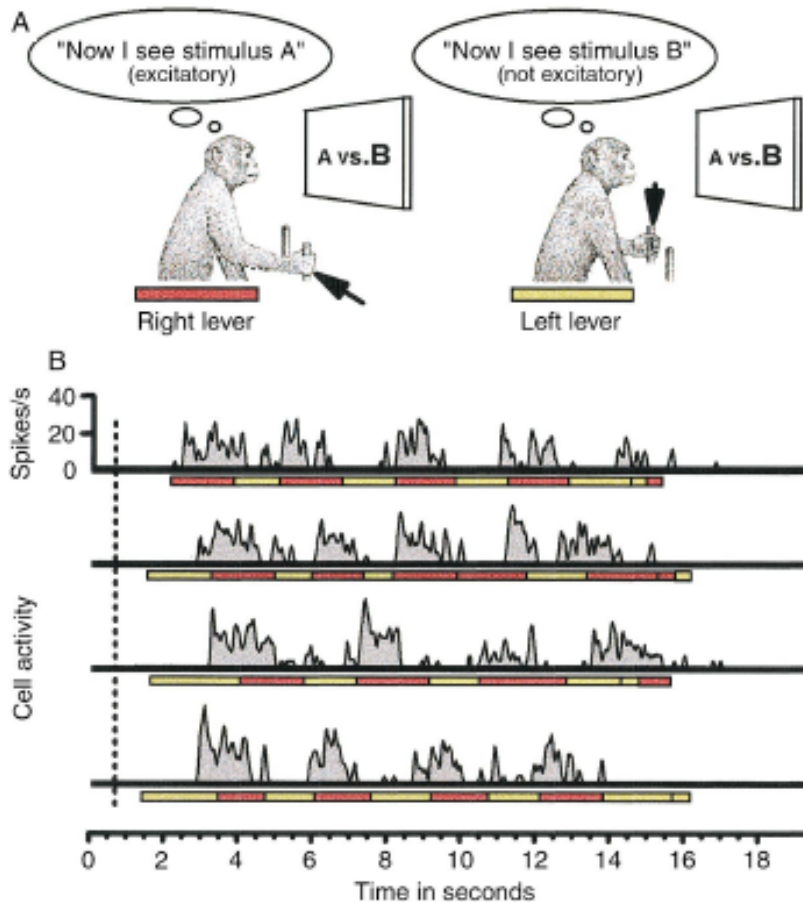
- When $T = 1$ the probability of being in a state matches the probability that the state is the correct one.
 - When $T = 0$, only the best state (or each of the N equally good states) are possible
- In short, neural networks are capable of achieving optimal perceptual inference over entire *ensembles of hypotheses*, if the weights, biases, and external inputs are set appropriately.

Interactivity in the Brain

- Bidirectional Connectivity
- Temporary inactivation of 'higher' areas degrades representations in 'lower' visual areas.
- Constraint Satisfaction in Binocular Rivalry



Distributed Alternation of Brain Activity in Binocular Rivalry



Models of Binocular Rivalry

- The tendency to perceive only one of the two whole percepts, and not a blend of the two, is consistent with a stochastic constraint satisfaction model like that in the Boltzmann machine, in which the neurons representing parts of each interpretation mutually support each other.
- Why do the interpretations alternate?
 - ‘Fatigue’ could play a role
 - Models in which fatigue occurs on two different time scales can account for many features of the data (new work since the Leopold and Logothetis article).
 - Signals that increase noise from time to time, or temporarily ‘turn down’ the input to allow the settling process to restart, could play a role as well.
 - Clearly, as Leopold and Logothetis argue, the phenomenon involves participation of many interacting brain areas.

Neural Network Models of Learning, including Language Learning

- Neural network models address how we learn to process inputs, as well as the processing of the inputs once learning has occurred.
 - Changes in connection strengths based on pre-and post-synaptic activity can allow networks to learn weights that capture appropriate conditional relations between hypotheses.
- When applied to topics such as language learning, they provide alternatives to classic symbolic accounts in which language acquisition is a matter of:
 - Discovering rules
 - Memorizing lists of exceptions
- A fierce debate grew up around these issues in the late 1980's.
- The second reading for today (*Past Tense Debate* papers) presents both sides of this argument.

The Essence of The Past Tense Debate (from my perspective)

- Are patterns of language use and language learning well-characterized by simple, categorical rules, such as:

To form the past tense of a verb, just add 'ed'

- Or do they exhibit sensitivity to frequency of occurrence, influence of semantic and phonological similarity between items, and gradual rather than sudden acquisition?
- If the second alternative is correct, an approach based on similarity-based distributed representations, gradual learning of connection weights, and graded constraint-satisfaction appears to provide a promising approach for further developing a detailed account.
- But this approach does still face challenges, as noted by Pinker and Ullman, and the debate continues.
- Perhaps we will find a time to revisit these points later in the course when we consider other aspects of language.