Psych229: Language Acquisition

Lecture 14
Past Tense & Symbols & Representation

Pinker & Ullman 2002: Past Tense Debate

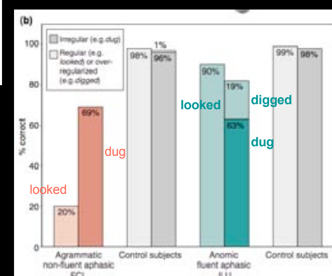
Results for brain-lesioned patients: predicted double dissociation

agrammatism

anomia

Agrammatism: more trouble inflecting regular than irregular

Anomia: more trouble irregular & overregularized



Pinker & Ullman 2002: Past Tense Debate

More results: Alzheimer's Disease, Parkinson's Disease, Huntington's Disease

1) Alzheimer's: impaired lexical knowledge & impaired irregular verbs
2) Parkinson's: impaired grammatical knowledge & impaired regular verbs
3) Huntington's: unsuppressed basal ganglion (~grammatical) & overuse of -ed rule (*dugged, walkeded*)

More results: lexical priming

1) Normal: regular & irregular forms prime stems (*walked~walk, found~find*)
2) Patients with left inferior frontal damage: priming only for irregulars & semantic priming (*goose~swan*)
3) Temporal-lobe damaged patient: priming only for regulars

Pinker & Ullman 2002: Past Tense Debate

More results: Electrophysiological Responses (Event Related Potentials - ERPs)

1) Regular suffix on irregular word (German *Muskels*) or left off of regular (Yesterday I *walk*): syntactic violation pattern (Left Anterior Negativity - LAN)
2) Irregular inflection illicitly applied (German *Karusellen*) or omitted (Yesterday I *dig*): semantic violation pattern (N400)

Point: reasonable evidence for double dissociations

Hard to get this is pattern associators - "lesioning" network (knocking out some chunk of it) tends to hurt irregulars more than regulars period…so how biologically plausible are these connectionist instantiations really?

They don't seem like they're meant to cover everything.

McClelland & Patterson (Rebuttal) 2002

But what about the quasi-regularity in the irregulars?

Though they don't view these as irregular rules the way Chomsky & Halle and Yang (2002) do, it's still something a pattern associator can capture without having to explicitly build in.

Quasi-regularity isn't just about English, either - find it many languages over the world. Words-And-Rules can't cover this with lexical-like memory for irregulars.

McClelland & Patterson 2002: Rules Schmules

McClelland & Patterson on rules

Bigger picture: rules = human cognition is symbolic, modular, innate, and domain-specific.

Pattern associators don't suppose any of this. Learning is just the gradual adjustment of simple processing units. Rules are about descriptions of language use, but there's no psychological reality to them.

The specific form of rule they're after here: rules as "discrete, categorical and symbolic objects used in a specialized, innate language module".

McClelland & Patterson 2002: Rules Schmules

McClelland & Patterson on rules

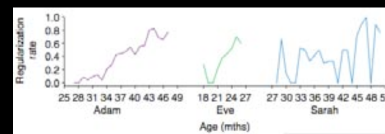
Predictions that symbolic rule models make

1) Acquisition of the symbolic rule is sudden
2) Rule is uniform in its applicability
3) Rule-based mechanism is separate from exceptions mechanism

Discussion: Are all these really true of the Words-And-Rules model? What about for any symbolic rule model?

McClelland & Patterson 2002: Rules Schmules

Acquisition pattern for regular past tense rule +ed: probabilistic & noisy

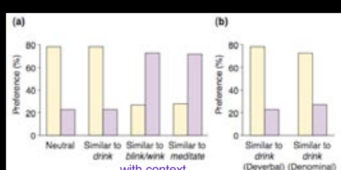


There is always a considerable period… in which production-when-required is probabilistic. This is a fact that does not accord well with the notion that the acquisition of grammar is a matter of the acquisition of rules, since the rules… either apply or do not apply. One would expect rule acquisition to be sudden. (Ref. [17], p. 257)

Discussion: Is it true that probabilistic performance does not accord well with the notion of acquisition of rules?

McClelland & Patterson 2002: Rules Schmules

Application of rule: subject to phonological & semantic influences



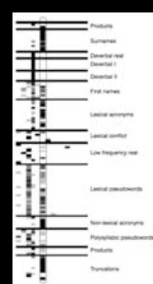
irregular
regular

Exocentric status didn't affect results

Point: regularization not just the Elsewhere rule (which we would expect novel verbs to fall under)

McClelland & Patterson 2002: Rules Schmules

Application of rule: against the Elsewhere application even for known words



German +s "default" plural usage - not so default

Fig. 3. Evidence that the German +s plural is not used uniformly across several situations supposedly calling for the use of a default as proposed by Marcus *et al.* [25]. Each row of the figure represents a different noun form, with the type of the form indicated; the horizontal bars separate the different types. Columns of the figure indicate alternative possible plural inflections, with the +s plural specifically highlighted. Grayscale darkness of the entry in each cell indicates the likelihood of using the particular plural for the given item, based on data from native German speaking adults. Reprinted with permission from Ref. [30].

McClelland & Patterson 2002: Rules Schmules

Neural basis for rules vs. words

1) Non-fluent aphasics (agrammatism): effects of regular vs. irregular difficulties disappear once test words are controlled more thoroughly for phonological properties
2) Parkinson's Disease (extra rule application - *dugged*, *walkeded*): could be due to phonological complexity of test words not being controlled

McClelland & Patterson 2002: Rules Schmules

Well, maybe rules aren't all bad….

Albright & Hayes (2003) is an example of a rule-based model that has good properties: graded rule activation, probabilistic outcomes, allow rules to strengthen gradually with experience, incorporate semantic and phonological constraints, and use rules within a mechanism that incorporates word-specific information.

But then is this empirically indistinguishable from a connectionist account? (M&P think not - "rules" are just higher-level descriptions of regularities in pattern associator.)

Pinker & Ullman (rebuttal) 2002: Combination and Structure!

Sure, there's quasi-regularity….but that's not the big deal
(which doesn't mean it's not important…)

Big deal: Does human cognition use mechanisms that are combinatorial and sensitive to grammatical structure and categories?

Rule = combinatorial operation (ex: +ed)

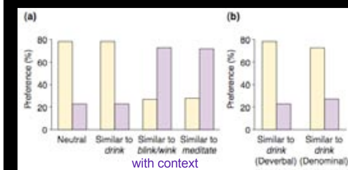
Of course they can be acquired and used probabilistically.

More important:

1) Do they apply when memory fails to retrieve exception?
2) Do they apply to heterogeneous situations with only grammatical category as the common denominator?
3) Does it disassociate neurophysiologically with memory lookup and associate with combinatorial processing?

Pinker & Ullman (rebuttal) 2002: Combination and Structure!

About when the rule applies (novel items): bad experiment



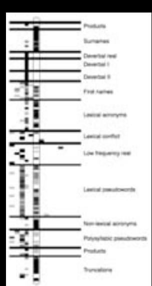
irregular
regular

Relationships aren't like any real verb, very short familiarization - how do we know subjects took restriction seriously?

Does experimenter want me to treat "frink" as a distorted form of "blink" or "drink"?

Pinker & Ullman (rebuttal) 2002: Combination and Structure!

Application of rule: German +s plural is messy but…



Pattern associator story is no better - German speakers learn to connect +s with "each arbitrary property that must be associated with a specific use of an item in context", ex: surnames.

Coincidence in this story: circumstances eliciting -s (names, unusual-sounding words, acronyms) have nothing in common except failure to access irregular root for grammatical category noun

Pinker & Ullman (rebuttal) 2002: Combination and Structure!

Double dissociation critique:

Non-fluent aphasics (agrammatism): effects of regular vs. irregular difficulties disappear once test words are controlled more thoroughly for phonological properties

…but reappeared in other tasks that were also controlled!

Also, later manipulations included stems rhyming with irregulars, so not so perfectly controlled after all.

Pinker & Ullman & McClelland & Patterson (2002): Caricatures

Pinker & Ullman:

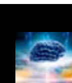
Rules are what produce the regularities in human language. They are part of the human mind. Human cognition uses combinatorial processing that is more than simply a strong connection strength for certain regularities that appear.

symbolic structure

McClelland & Patterson:

No, human cognition doesn't. You can get everything you need without recourse to a separate rule structure.

Marcus (2003): Symbols



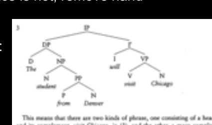
The Mind & Symbols

Big question: Is the mind something that manipulates symbols? Or is the basic unit of cognition something that has nothing to do with the "sentences and propositions" of symbol-manipulation (Churchland, 1995)?

Symbol-manipulating models: typically described in terms of production rules & hierarchical trees

Production rule: If precondition 1 is true, do action 1
"If surface is hot, remove hand"

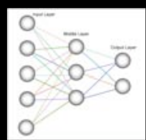
Hierarchical tree:

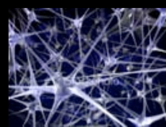


Marcus (2003): Symbols

The Mind & Symbols

Connectionist models: tend to be “neurally-inspired”, described in terms of neuron-like units and synapse-like connections





Important point: Just because something is connectionist doesn’t mean it can’t also manipulate symbols (connectionist = implementational level, symbols = computational level)

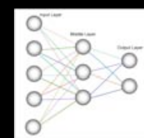
Marcus (2003): Symbols

The Symbolicists vs. the Non-Symbolic Connectionists

Symbolicist assumption: circuits in the brain correspond in some way to the basic devices necessary for symbol manipulation (e.g. some circuit supports representation of a rule)

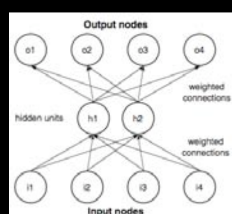
Non-Symbolic Connectionist assumption: there will not be any brain circuits like this (rules are epiphenomena of regularity in patterns of activation)

Non-symbolic connectionists tend to focus on multilayer perceptrons as a model of cognition, and this is the model in general that’s brought up whenever symbols (or no symbols) are. (This is because it’s an explicitly-formed model.)



Marcus (2003): Symbols

Multilayer Perceptrons

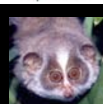


Nodes: have activation values (0.5, 1.0)
- input/output: have meaning associated with them (+ed, *walk*, …)
- meaning affects what things are considered alike (c onset (cat ~ cab) vs. +animal (cat ~ dog))

Activation values: numbers assigned to nodes, based on input

Ex: +furriness is set to 1.0 is input is furry, 0.0 otherwise

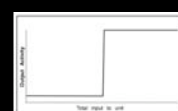
Furriness node = 1.0



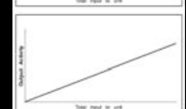
Marcus (2003): Symbols

Activation of nodes - based on total input fed in (weighted sum of values)

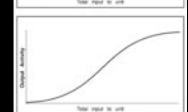
Step or binary threshold function - either on or off, based on threshold



Linear function - activation scales linearly with input

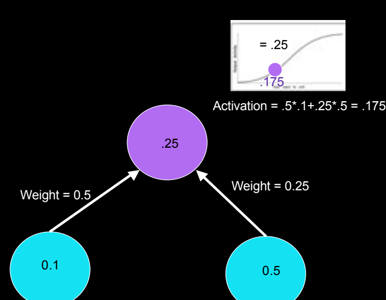


Sigmoid function - activation scales curvily with input (models with hidden units tend to use this kind)



Marcus (2003): Symbols

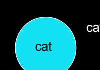
Activation of nodes - based on total input fed in (weighted sum of values)



Marcus (2003): Symbols

Localist vs. Distributed Models

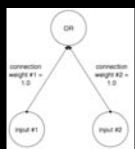
Localist: each input and output corresponds to a particular word or concept (*cat, furry*)



Distributed: each input and output is encoded by the simultaneous activation of a number of nodes (combine features to get meaning: *furry, 4 legs, meows* = cat)

Marcus (2003): Symbols

Implementing functions

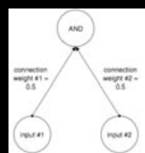


Input1	Input2	Input1 OR Input2
0	0	0
0	1	1
1	0	1
1	1	1

OR node activated if total input >= 1.0

Input1	Input2	Input1 AND Input2
0	0	0
0	1	0
1	0	0
1	1	1

AND node activated if total input >= 1.0



Marcus (2003): Symbols

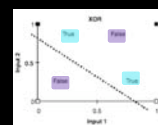
Implementing functions: The need for another layer

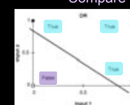
Input1	Input2	Input1 XOR Input2
0	0	0
0	1	1
1	0	1
1	1	0

Not linearly separable = no straight line can separate true from false outputs = can't be done with just 2 layers (input and output)

Exclusive-or: Only true if one or the other, but not both, are activated

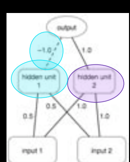
Compare to OR and AND





Marcus (2003): Symbols

Implementing functions: The need for another layer



Input1	Input2	Input1 XOR Input2
0	0	0
0	1	1
1	0	1
1	1	0

Exclusive-or: Only true if one or the other, but not both, are activated

OR and not AND

Input 1	Input 2	Input to hidden unit 1	Output from hidden unit 1	Input to hidden unit 2	Output from hidden unit 2	Input to output unit	Output
F = 0	F = 0	0	0	0	0	0	0
F = 0	T = 1	0.5	0	1	1	1	1
T = 1	F = 0	0.5	0	1	1	1	1
T = 1	T = 1	1	1	2	1	0	0

Marcus (2003): Symbols

About hidden layers

Sometimes thought of as recoding the input (ex: XOR hidden layer has OR and AND in it) - similar to internal representations of input

About learning with multiple layers: initially, connection weights are random and need to be adjusted

One way: Hebbian learning
"Cells that fire together wire together" - strengthen connection weight between input node and output node every time they are active simultaneously

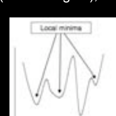
Another way: Delta ("difference") rule learning
Change weight of connection between input and output node, based on activation of input node multiplied by difference between what output node should have done and what output node actually did (involves parameter = learning rate = how much adjustment)
For hidden layers, use back propagation variant that estimates what hidden layer input and output activations should be.

Marcus (2003): Symbols

The nice thing about back propagation

If learning rate is small, back propagation is a gradient descent algorithm - gradually getting closer and closer to a right answer (set of weights), which is at a metaphorical "valley" on the answer "landscape"

One pitfall: local minima



Bonus: Small learning rate = gradual learning (which is what children seem to do)

But these algorithms require supervision - need to know what the right output activation should have been. Where does this come from? One answer: The data to the learner. (Need to verify this for each learning problem, though.)

Example: past tense model
Scenario: past tense of *run*
Model predicts: *run+ed*
Data = *ran*

…therefore, adjust weights