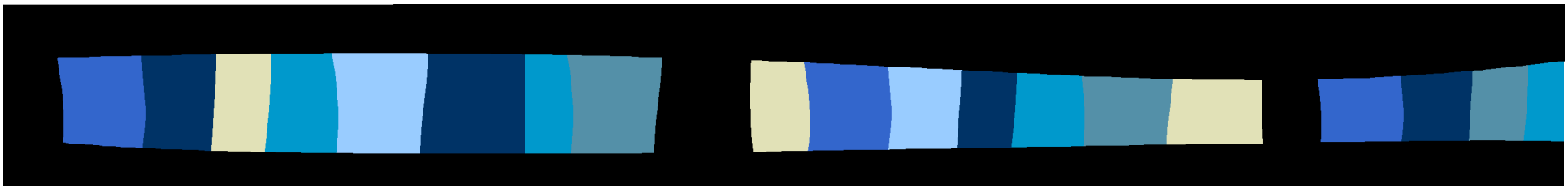
Language



Richard L. Lewis

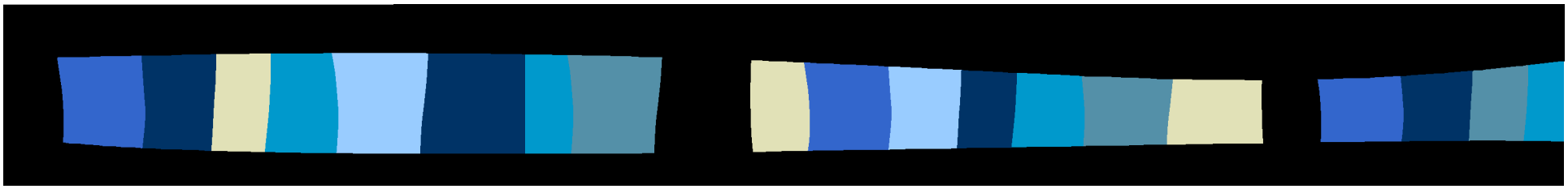
Psychology,

Linguistics, and Electrical Engineering & Computer Science

University of Michigan

July 24, 2001

ACT-R and Language Processing: *Opportunities, Challenges, and the Linguistic Killer Bees*



Richard L. Lewis

Psychology,

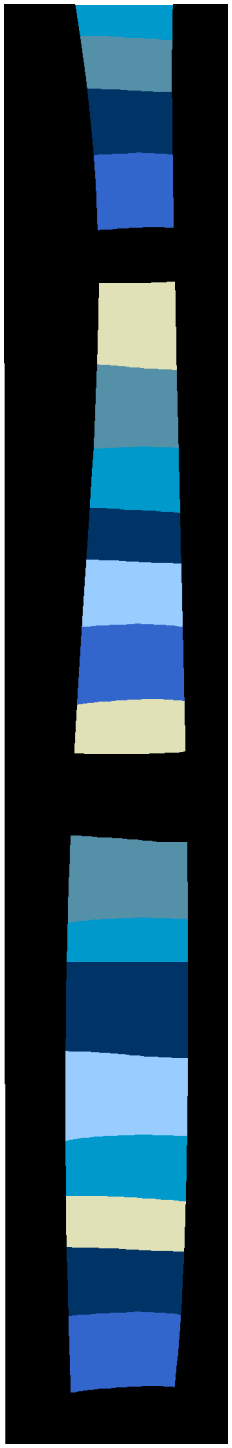
Linguistics, and Electrical Engineering & Computer Science

University of Michigan

July 24, 2001

Wexler's (1978) review of *Language, Memory, & Thought*

- Wexler's conclusions:
  - "*There is not much prospect of adding scientific knowledge by pursuing the methods … in LMT*"
  - *"There is remarkably little that a linguist (or even psychologist) could learn by reading LMT"*
- Some major gripes
  - Trying to achieve too much with a broad theory
    - Nature of explanation may be different in different domains (modularity)
  - Theory not "restrictive"—no explanatory principles
    - ACT can model *anything*, therefore explains *nothing*
  - Not dealing with sufficient range of complex linguistic phenomena



Wexler's diagnosis and cure

- Core problem: ***representational weakness***
- *"In studying language processing it seems obvious that the use of a strong representational theory would be very helpful...In particular, such a theory exists for syntax."*
- Work toward theory of "performability"
  - How can structures uncovered by linguistic theory be processed by a processor with human constraints?
  - Important that "entire representational theory be processed"

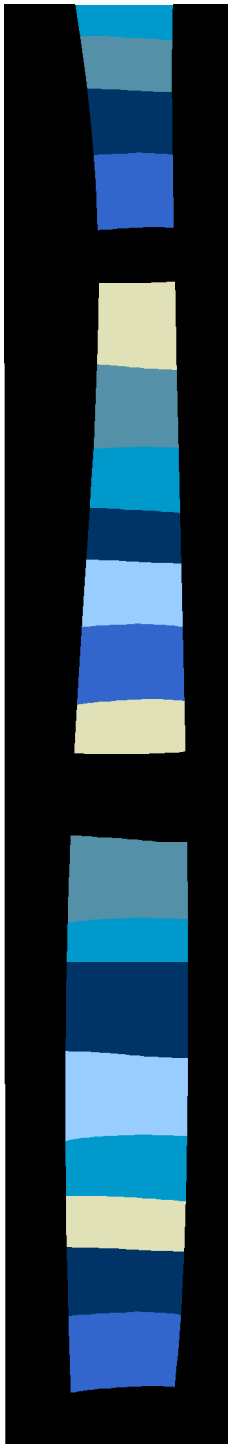
Good news, and a question

- On the prospects for a productive science of cognition grounded in cognitive architectures (& ACT), Wexler was ***simply wrong***
  - The present state-of-the-art in architectural theorizing (as represented in ACT-R) is evidence of significant progress over last 20+ years
- But for language, we can still ask:

What can ACT contribute to psycholinguistics and linguistics?

By the way, John is in good company

- From Bickerton's (1992) critique of Newell's (1990) UTC chapter on language:
  - "*To criticize the eighteen pages on language and language acquisition that follow would be like shooting fish in a barrel*"
- Part of Newell's response:
  - "*… I clearly faltered in dealing with the linguistic killer bees…. The lesson, of course, is never, never show fear.*"



Overview

(1) Preliminary: Major choices in developing ACT-R models of language

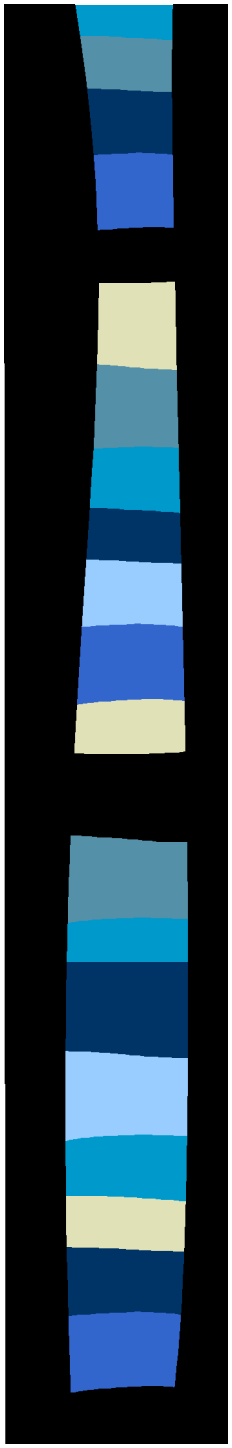
(2) An example domain: *sentence processing*

- Brief sketch of the structure of one model
- Some interesting—even unexpected—theoretical and empirical issues very closely tied to architecture

(3) What does ACT-R buy you?
Major opportunities for ACT-R in language

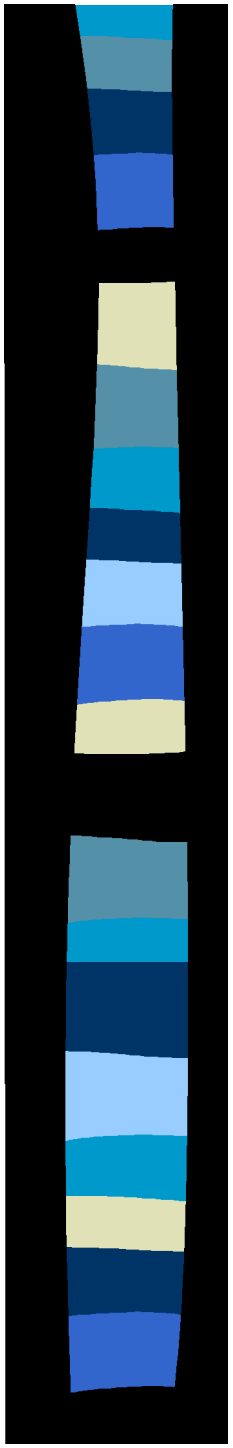
(4) Potentially serious challenges

(5) Revisiting the killer bees



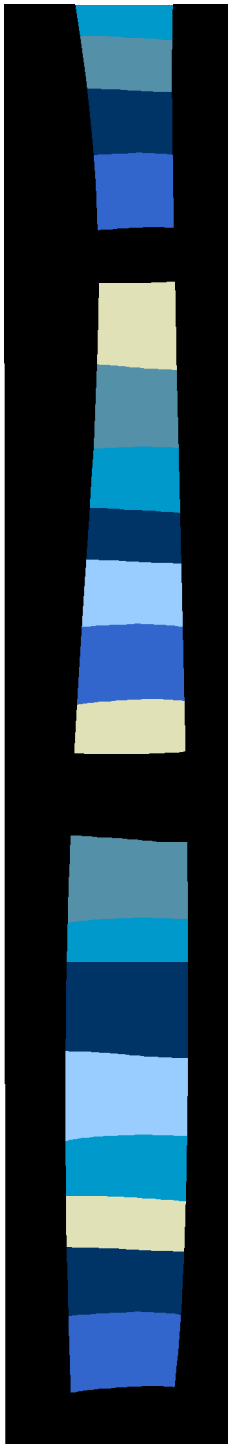
A major choice in developing language theories in ACT-R

- Should linguistic processes be realized *within* ACT-R, or should ACT-R be treated as central cognition, and a language module developed for it?
- Perhaps we should work on an *ACT-R/PML*
  - Where the “slash” corresponds to Jerry Fodor’s way of “carving the mind at the joints
- That’s not the approach I’ll discuss today—nor is it the approach traditionally taken in ACT research



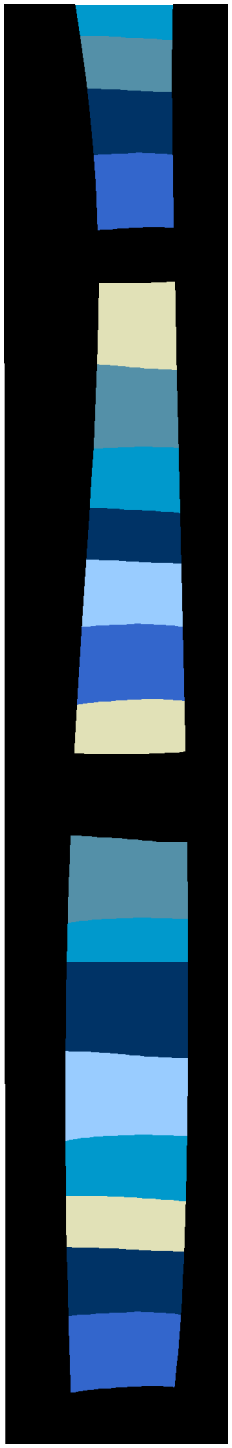
The alternative

- Treat language as a cognitive skill; embed linguistic processing in the architecture
- We know from Soar work that this can (surprisingly) yield processing models that are consistent in many respects with modular approaches
- And, even if it turns out to be wrong, we need to know why



Another major choice

- How to distribute ***lexical*** and ***grammatical*** knowledge across ***declarative*** and ***procedural*** memory
- Approach I'll assume:
  - Lexicon in declarative memory
  - Grammatical knowledge in procedural memory (for comprehension, in form of parsing productions)
  - Field typically doesn't phrase distinction in these terms, but one notable exception: Consistent with at least one neuropsychologically motivated model (the "declarative-procedural model", M. Ulmann)



Why sentence processing?

- Because that's mostly what I work on
- *Also*: A very interesting combination of symbolic and subsymbolic, and fast, real-time but complex processing
- Incredibly rich empirically and theoretically (perhaps too much so)

A classic processing limitation

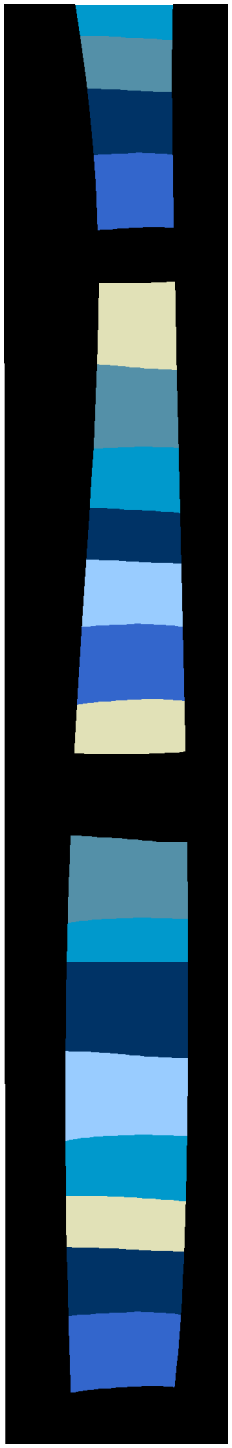
- Most people find one level of embedded clause comprehensible:

  The dog that the cat chased ran away.

- But double center embeddings are very difficult (Miller & Chomsky 1963):

  The salmon that the man that the dog chased smoked tasted bad.

WHY?

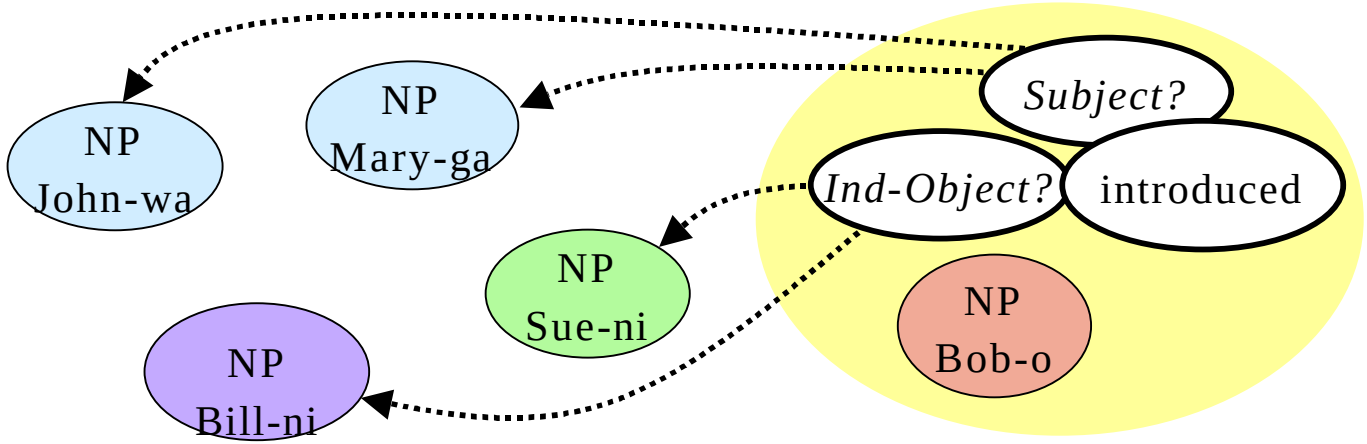


Examples of good/simple ideas that don’t quite work

- Kimball's *Principle of Two Sentences*: Can't parse more than two sentences at once
  - [$_s$ What [$_s$ the woman that [$_s$ John married] likes] is smoked salmon.] (Cowper 1976; Gibson 1991)
- Limited buffer for holding unattached NPs (say, two)
  - John-wa Bill-ni Mary-ga Sue-ni Bob-o syookai sita to it-ta. (Lewis, 1993, 1996)
  - *John Bill Mary Sue Bob introduced said*

Parsing as associative, cue-based retrieval from WM

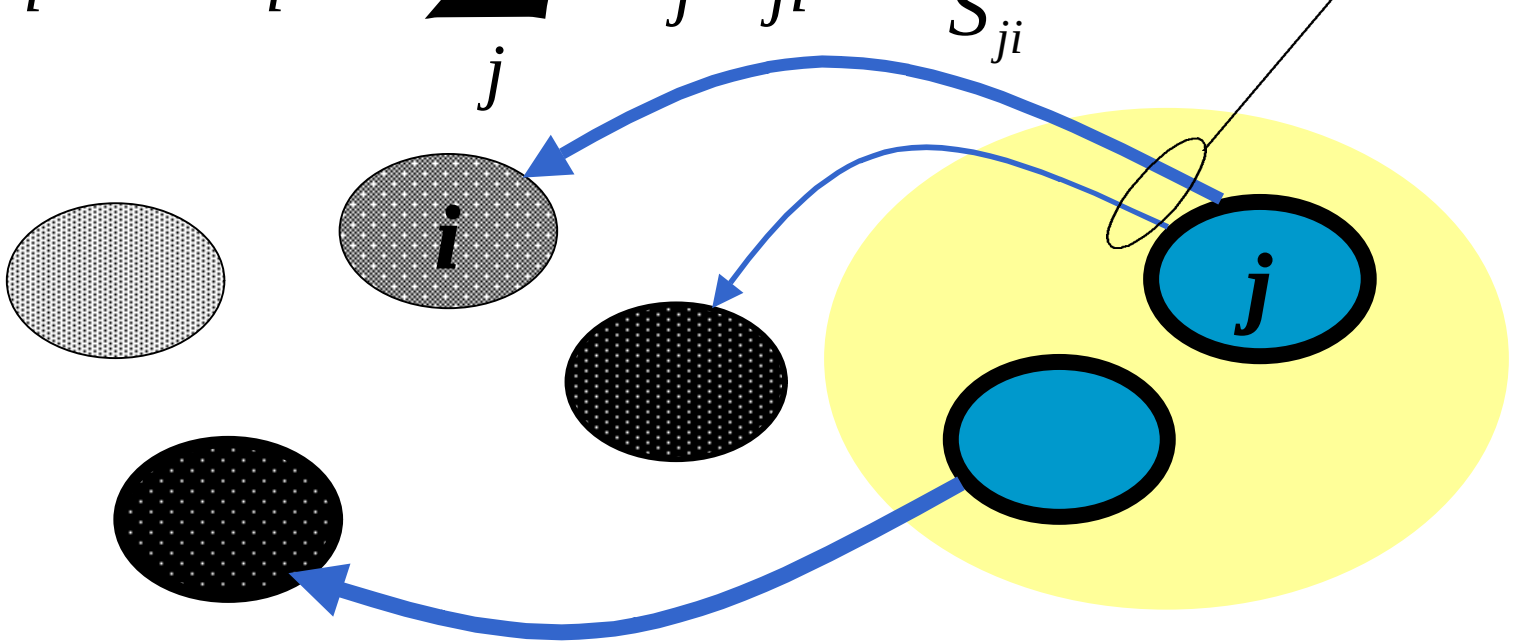
- Construes attachments as associative WM retrievals (Lewis 1998; McElree, 1998, 2000)
  - Interesting connections to MacWhinney's (1989) cue-based *competition model* (& other constraint-based approaches)
- Cues include (at least) *syntactic relations*
  - Retrieval interference arises from *cue-overlap* (a kind of similarity-based interference)



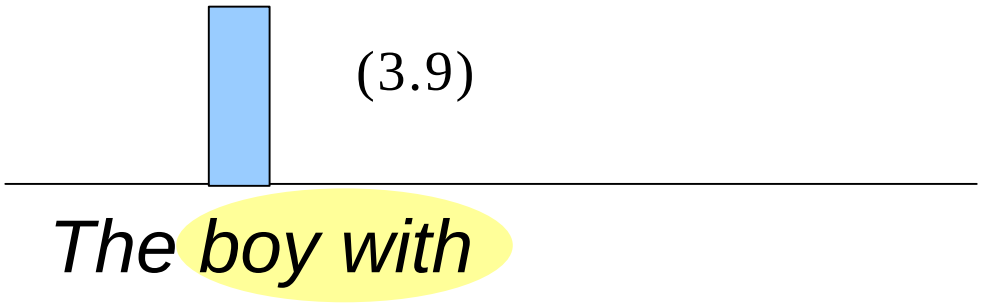
Decay, focus, & interference

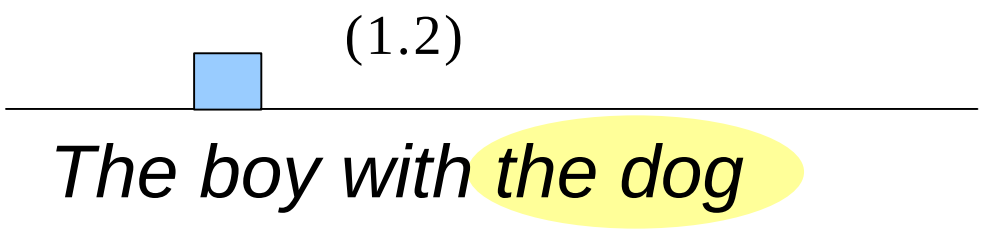
Focused elements serve as *retrieval cues.* Memory elements receive additional activation from associated focus elements.

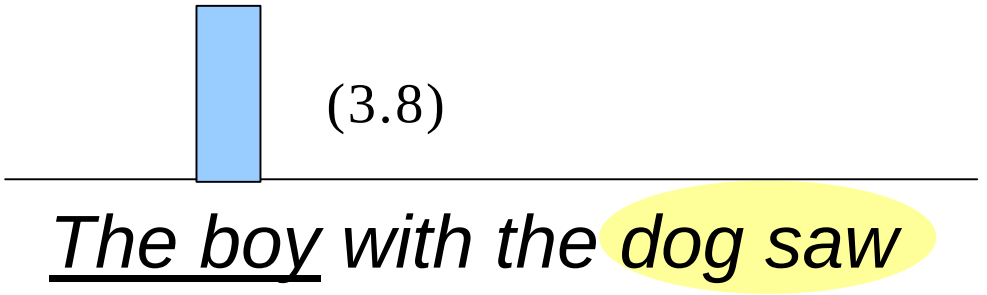
$$A_i = B_i + \sum_j W_j S_{ji}$$

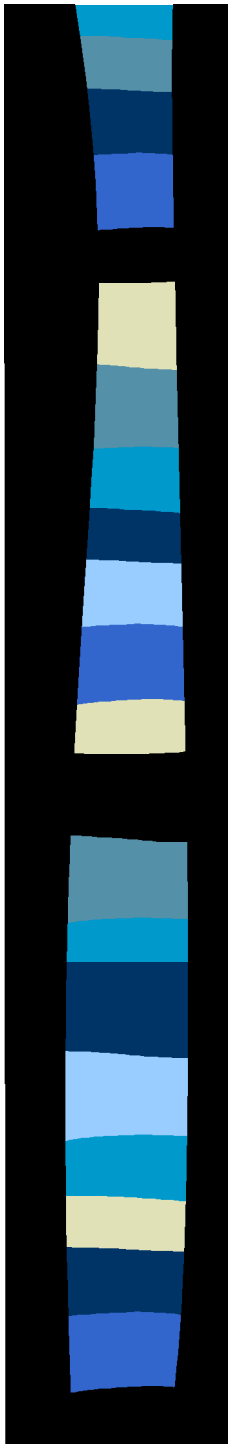


Example of activation dynamic in parsing









Four interesting theoretical issues

(1) How working memory limitations in parsing arise

(2) The representation of serial order information in sentence processing

(3) Modularity and control structure

(4) Decay vs. interference in processing ambiguities

Issue #1: Implications for WM limitations in parsing

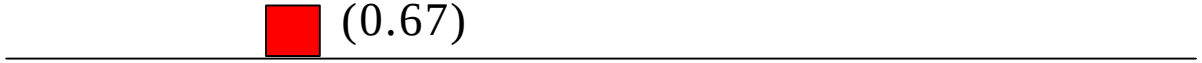
- Left alone ⇒ constituents decay
- More cues ⇒ less activation for each
- More constituents associated with a cue

  ⇒ less effective the cue is

Worst case: *multiple distal attachments with high retrieval interference*

Distance vs. interference

- Worst case for parser:
  - Multiple [limited focus], distal [decay] attachments, with multiple similar candidates [interference]

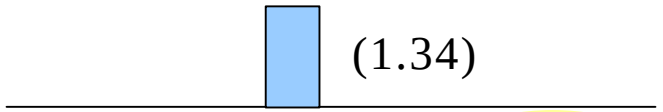


*The boy <u>who</u> the dog that the fish saw **ate** was*

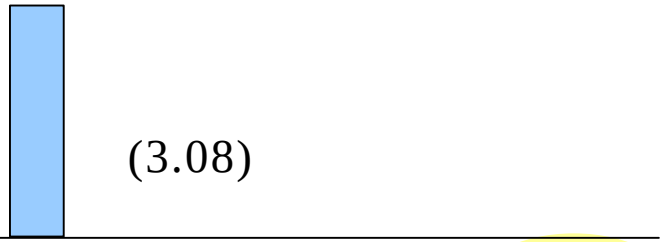
…..But long-distance attachments still possible

*<u>The girl</u> with the dog with the boy with the fish with the ticket **was**….*

Contrasts in center-embedding: SR vs. OR

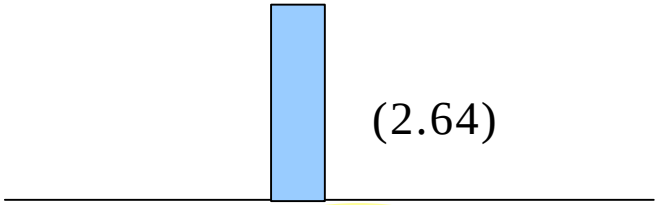


*The boy <u>who</u> the dog **bit***



*<u>The boy</u> who the dog bit **saw***

Object Relative

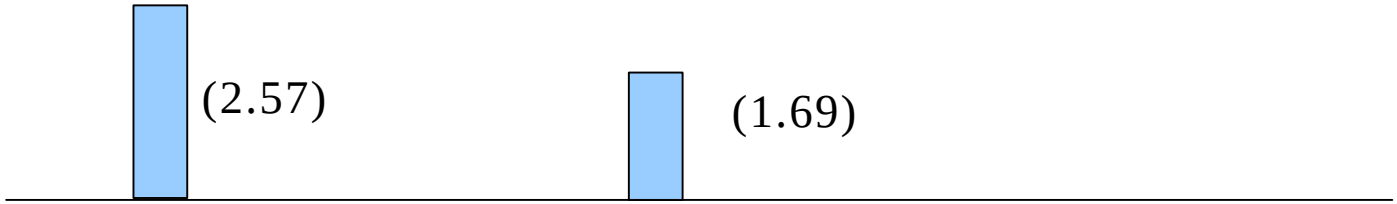


*The boy <u>who</u> **bit***

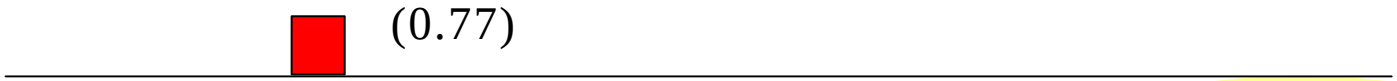
*<u>The boy</u> who bit the dog **saw***

Subject Relative

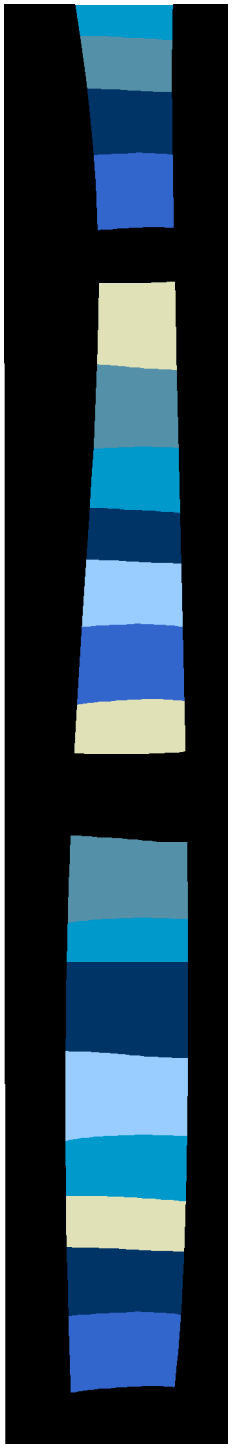
Effects of locality: RC/SC contrasts (Gibson, 1998)



*The claim that the boy who the dog bit **died upset** me*



*The man who the claim that the dog bit the boy **upset ...***



Issue #2: Serial order

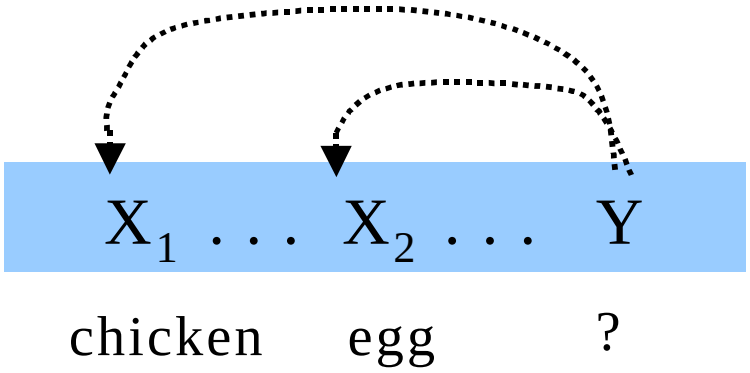
- Keeping track of serial order information is functionally required
- But very often, simply distinguishing *current item* from *preceding items* is sufficient
  - E.g., lets you distinguish the S from the O in SVO languages

Mary saw the dog.

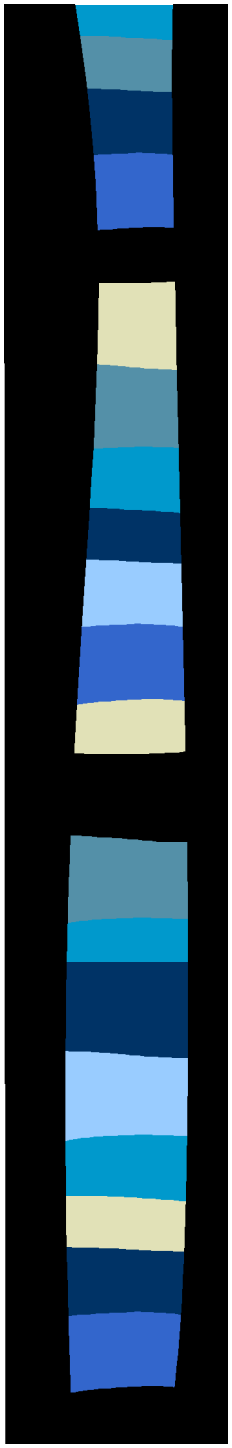
At "*saw*", attempt attachment of the subject (not the object).
No need to distinguish the relative order of *Mary* and *dog.*

When distinguishing *current* from *preceding* is not enough

- General case: *When two or more preceding items must be discriminated solely by their serial positions*



- Examples: Japanese sentential embeddings:
  Mary-ga Tom-ga butler-o killed knows
  Who killed the butler?



Why not use activation decay?

- Just attach to the most active (recent) candidate (simulate a stack)
- Two functional problems:

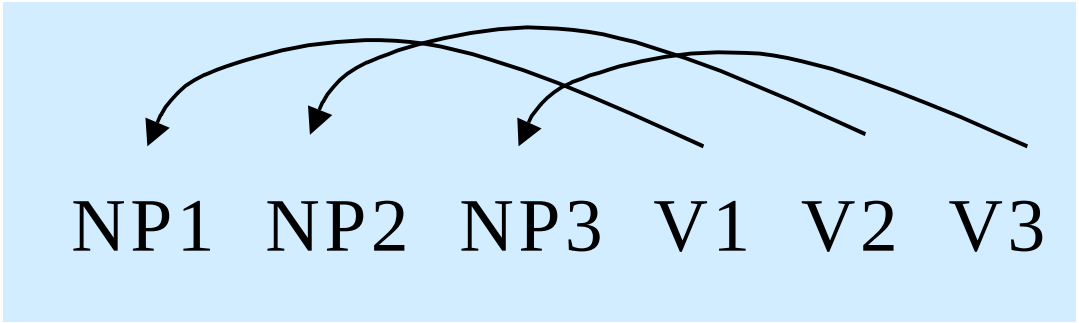
(1) The activation/strength of an item may not accurately reflect its serial position

  - Evidence from STM paradigms suggests item strength is not a good surrogate for position (e.g., McElree & Dosher, 1993)
  - Also, some items may be linguistically focused, or may have received additional processing
  - ***So this is a dicey theoretical path***

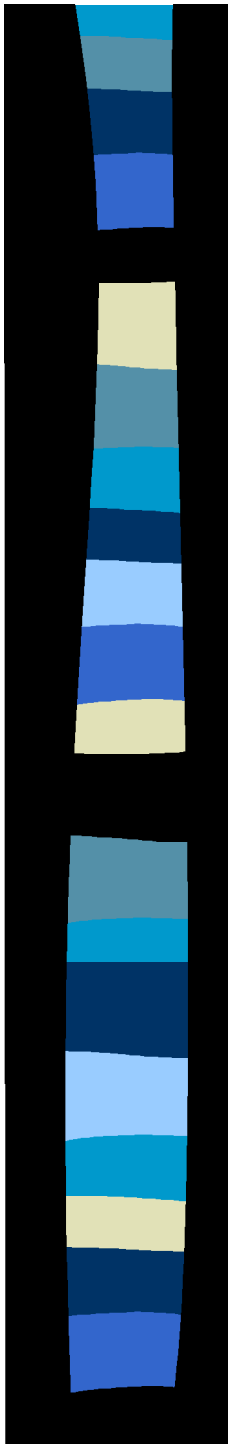
(2) Sometimes it is necessary to attach to the *first* item, not the most recent

Cross-serial dependencies

- In Dutch, a standard embedded construction requires crossed, not nested, dependencies

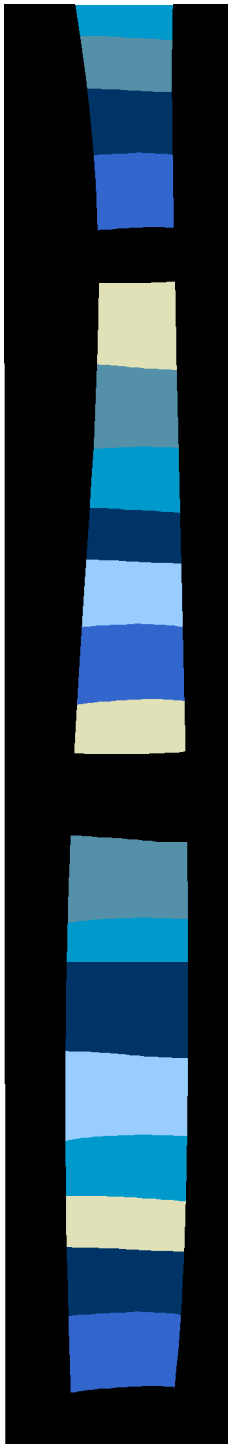


- …omdat ik Cecilia de nijlpaarden zag voeren
  because I Cecilia the hippopotamuses saw feed
  “because I saw Cecilia feed the hippopotamuses.
  (Steedman, 2000)



How should we represent serial order in such a parser?

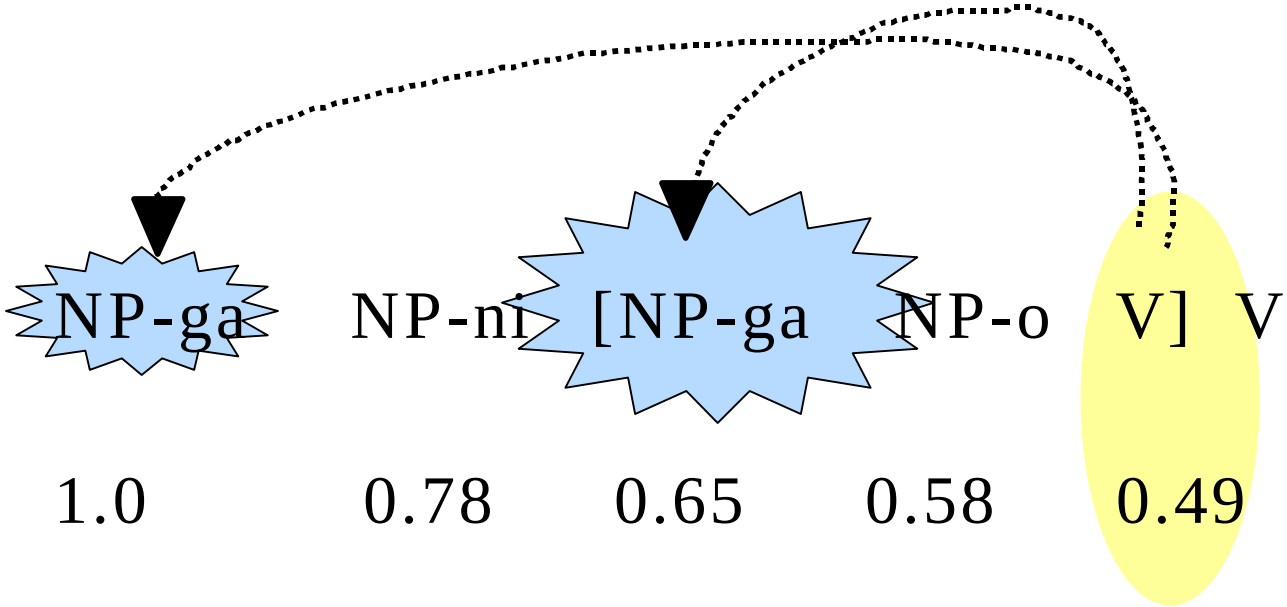
- Major choice: *position codes* vs. *associative chaining*
  - Relevant STM phenomenon: items in nearby positions tend to be confused (e.g., Estes (1972))
  - Confusable position codes adopted by many researchers (eg., Hensen (1998); Burgess & Hitch (1999), Anderson et al, (1998))
- Idea: associate each item with a position code that is a value along some gradient
  - There is a distinguished *START* anchor code (say,START = 1.0); other positions defined as some (probably non-linear) function



Using position codes as retrieval cues in parsing

- Only two codes are used as cues: START and END (the current position)
  - To effect a *recency/stack* discipline for *nested* dependencies, use the END code as a cue
  - To effect a *primacy/queue* discipline for *crossed* dependencies, use the START code as a cue
  - The best matching (closest) item will be retrieved (all other things being equal)
- No need to assume parser knows about any other position codes (e.g., the *2nd,* or *4th*)
  - I.e., these are not grammatically meaningful

Using position cue to retrieve most recent candidate



Retrieval cues:
END (= 0.49),
Subject

Prediction of positional similarity effects

- Given:
  - Confusable positional codes
  - And the functional requirement to distinguish items based on position
- Then it may be possible to make processing *easier* by *increasing distance*—increasing positional distinctiveness
- Data from Uehara (1997) bear this out:

- NP-ga NP-ga NP-ga NP-o V V V (4.31)
- NP-ga NP-ga **Adv** NP-ga NP-o V V V (3.61)

Testing positional similarity with single embeddings

- 2 x 2 design varying stacking (3 and 4 NPs) and positional similarity of subject NPs

(a) (ps=0, stack=3)

NP-ga NP-ni NP-ga V V

my brother teacher girl playing notified

(b) (ps=2, stack=3)

NP-ga NP-ga NP-o V V

dentist president interpreter called remembered

(c) (ps=0, stack=4)

NP-ga NP-ni NP-ga NP-o V V

professor president representative student examined promised

(d) (ps=2, stack=4)

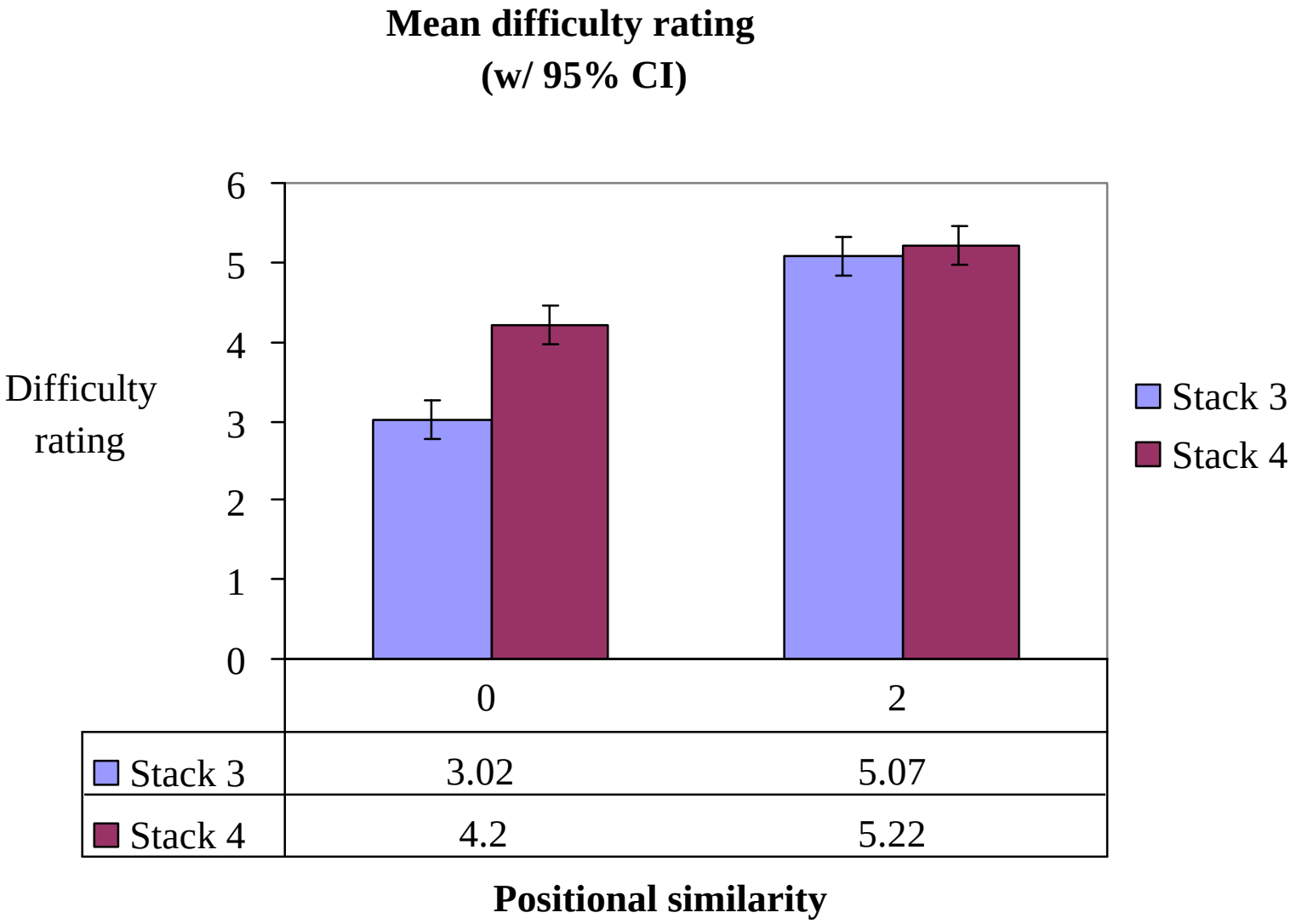
NP-ga NP-ga NP-ni NP-o V V

student lecturer reporter author introduced noticed

Difficulty rating study

- Participants rated difficulty on 7 point scale (1=easy, 7= difficult)
  - Each participant saw four versions of each experimental type (16 total experimental sentences interspersed with 34 fillers; 50 total)
  - Familiarity of lexical items controlled across conditions
- Participants were 60 female students from Kobe Shoin Women's University in Japan

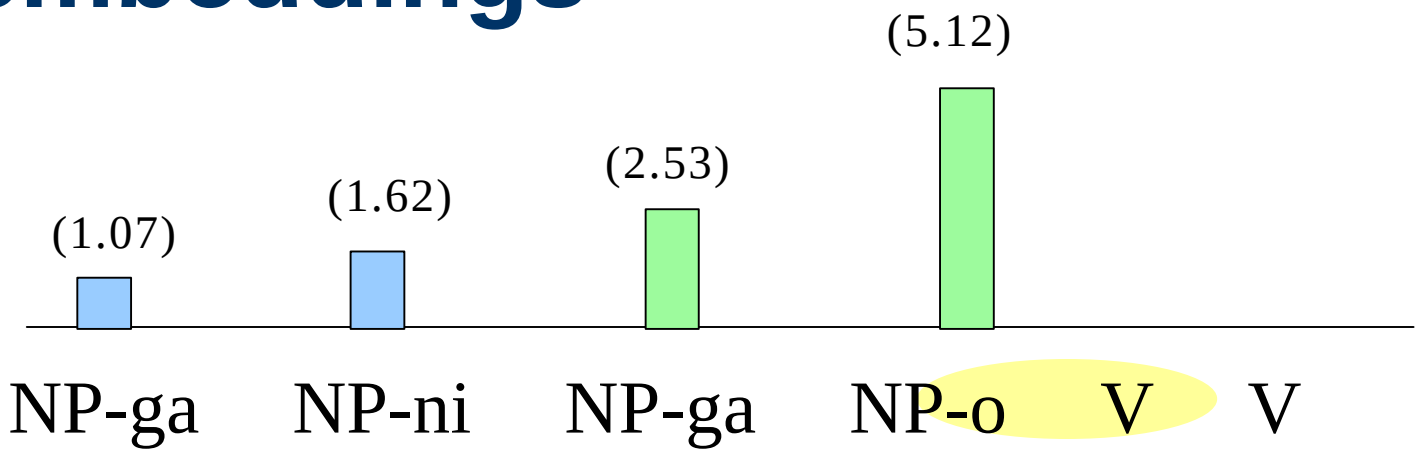
Results

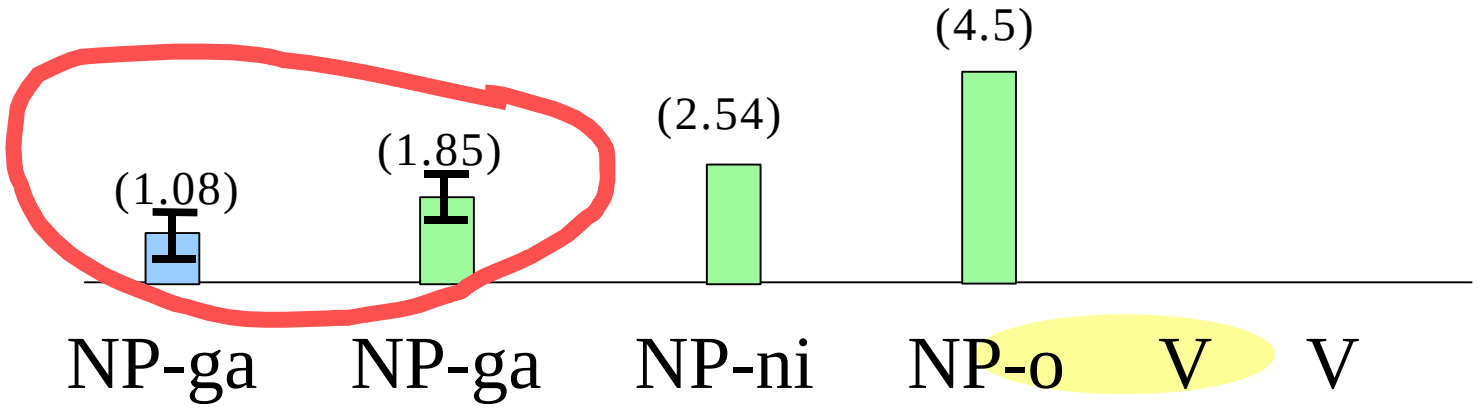

Mean difficulty rating
(w/ 95% CI)
Difficulty rating
6
5
4
3
2
1
0
0
2
Stack 3
Stack 4
Stack 3 3.02 5.07
Stack 4 4.2 5.22
Positional similarity

It’s robust

- Effect shows up in 2 rating paradigms: *1-7 fixed scale*, and *magnitude estimation*
- Effect shows up in 3 presentation paradigms
  - Paper & pencil questionnaire
  - Self-paced moving window
  - Self-paced central presentation

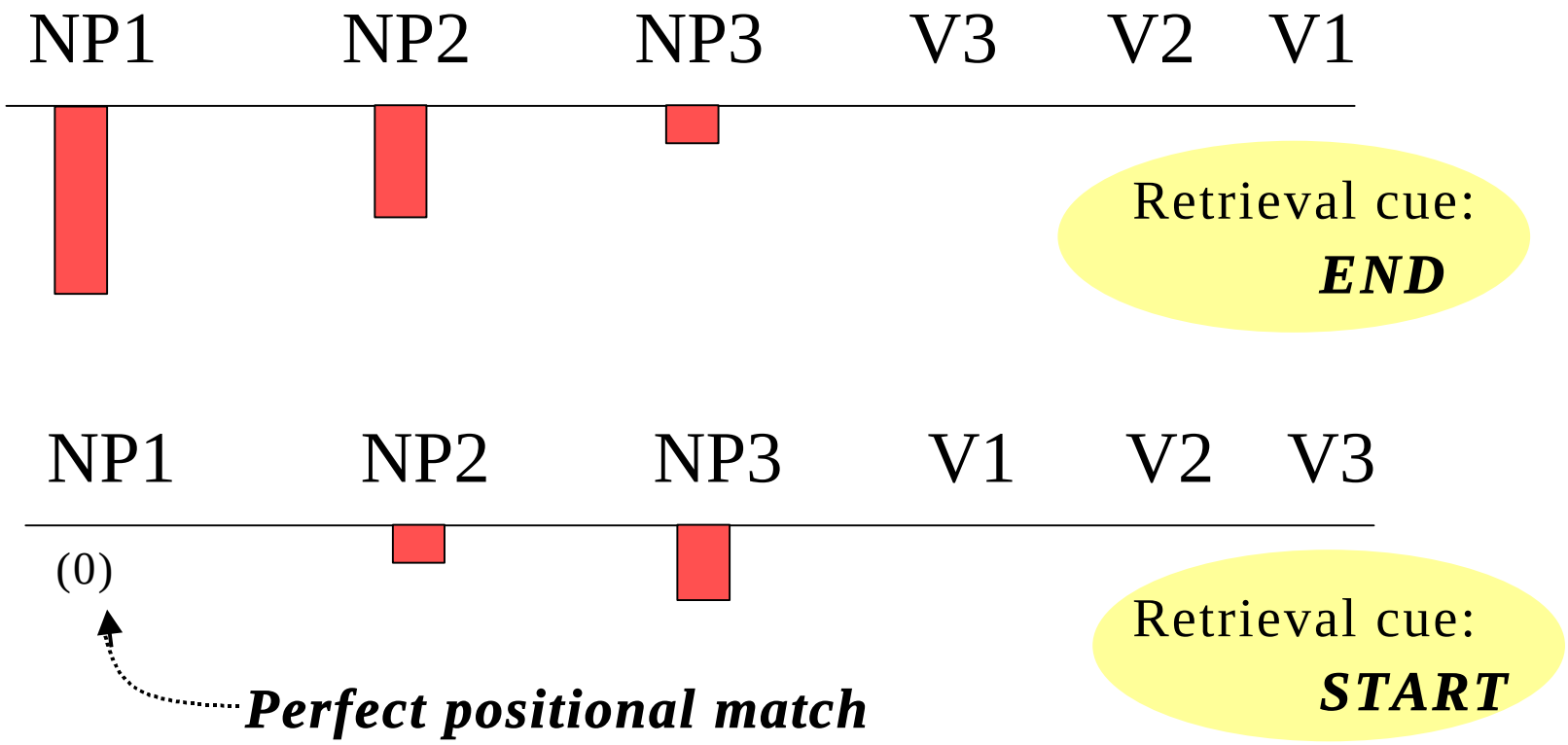
The ACT model on the single embeddings

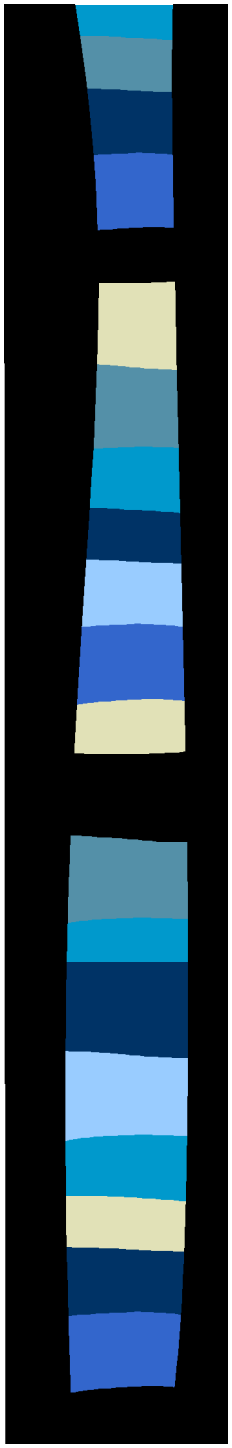




Why Dutch is easier than German

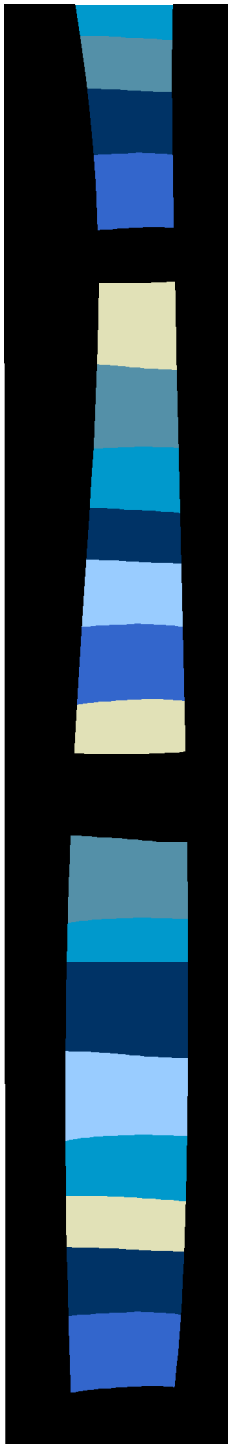
- Consider the *positional mismatch* when using START vs. END (current) position codes





Summary of qualitative coverage

- Classic difficult double relatives (French, Spanish, German)
- Subject vs. object relatives
- Subject sentences with relative clauses
- RC extrapositions (German)
- Many stacking contrasts (Japanese, Korean)
- Cross-serial vs. nested contrast (Dutch, German)
- RC/SC vs. SC/RC contrast
- Positional similarity contrasts (Japanese)
- Various pseudo-cleft and it-cleft w/relative contrasts
- Various Wh-movement w/ relative constructions

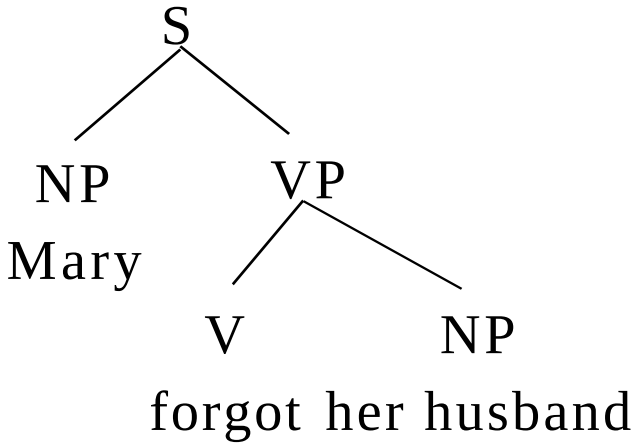


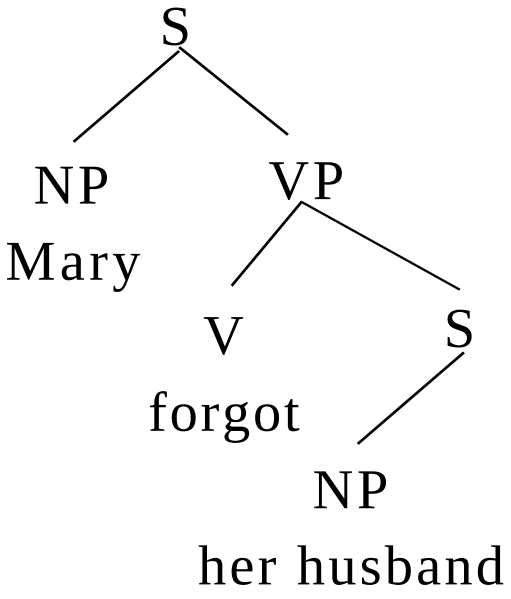
Issue #3: Modularity and control structure

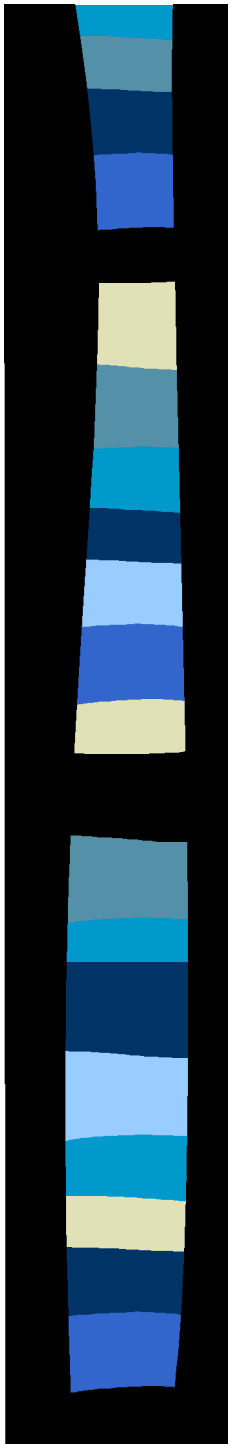
- Does ACT-R yield a modular or interactive account of sentence processing?
  - The answer may surprise you
  - [Answers were surprising in Soar as well; see Lewis, 1998; Newell, 1990]
- Where to look: *Factors affecting on-line structural ambiguity resolution*

Quick review

- Example:
  - *Mary forgot her husband would .....*
- Structural accounts like *Minimal Attachment* prefer the direct object structure over the sentential complement
  - Because the direct object structure is less complex (fewer nodes); hence should be computed faster

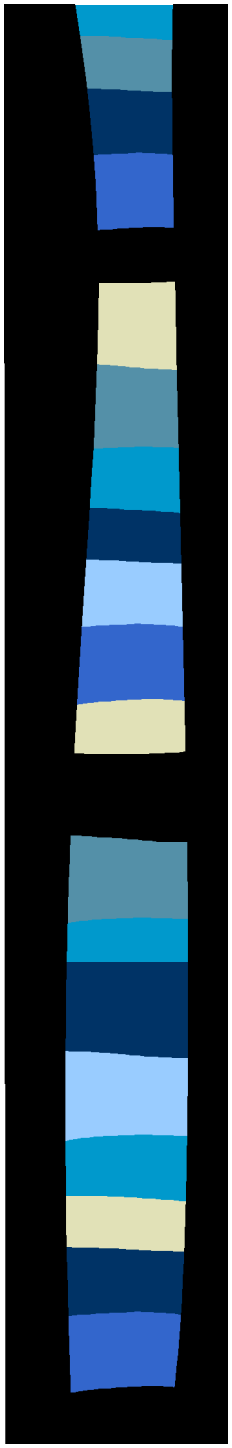






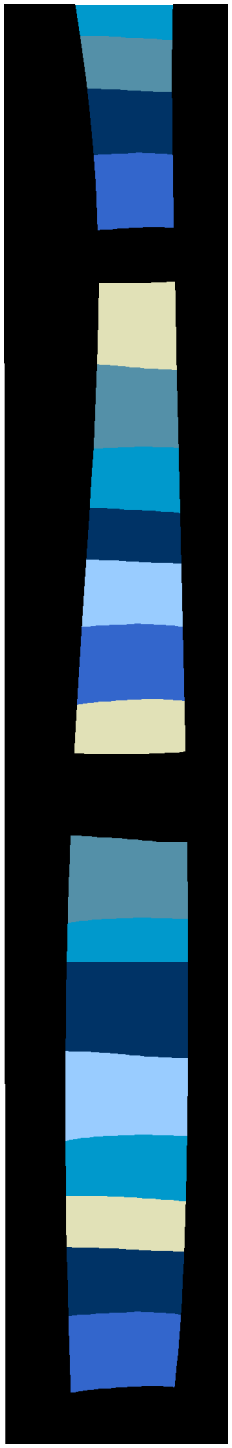
Structural ambiguity resolution in the ACT-R model

- Choice between competing parsing productions corresponds to selection of a path to pursue in the parsing search space
- Thus, structural ambiguity resolution happens via conflict resolution
- *I.e., the theory of ambiguity resolution is ACT-R's theory of conflict resolution*



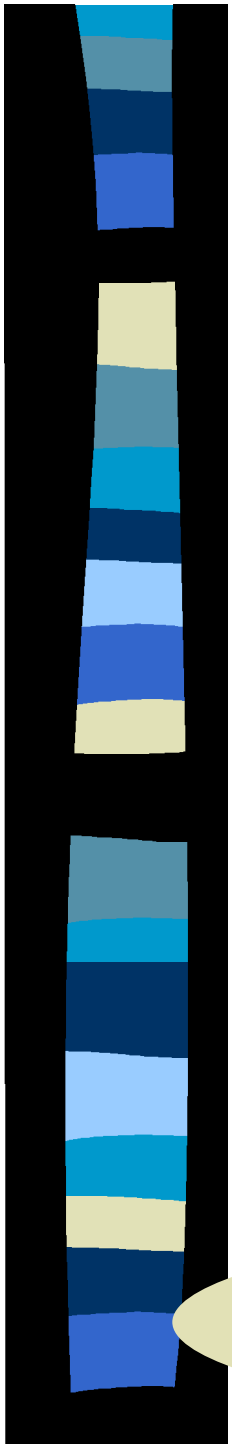
Modularity and conflict resolution in ACT-R 3.0 vs. 4.0

- Key issue in modularity: *are there architectural boundaries that prevent certain kinds of information from being brought to bear on some processing decision?*
- In ACT-R 4.0, there is a clear move in the direction of *limiting* the information flow, compared to 3.0, 2.0
  - Fewer factors (fewer knowledge sources) affect initial production choice in ACT-R 4.0
- This has a fairly dramatic effect on the nature of the resulting sentence processing theory



ACT-R 3.0 vs. ACT-R 4.0

- ACT-R 3.0: *Particular attachments (production instantiations) compete*
  - Activations of declarative elements are a factor in the conflict resolution
  - Provides way to integrate decay/recency, and frequency effects into ambiguity resolution
- ACT-R 4.0: ***First**, attachment types compete*
  - Based on their expected gain
  - ***Then**,* given an attachment type, different instantiations of that attachment compete
    - *I.e.,* different chunks "compete" for retrieval

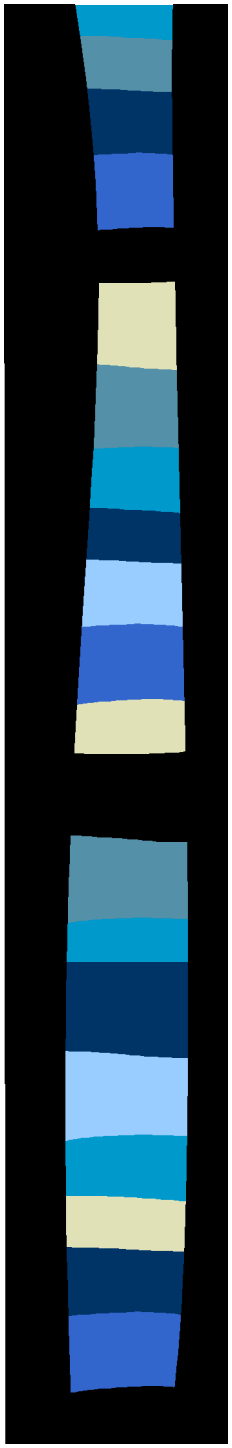


An old favorite...

The horse raced past the barn fell.

- At ***raced***, main verb attachment production will win the *initial* competition
  - A much more successful construction; and predicted cost of reduced relative is higher
- But considerable evidence now that GP effect can be reduced (even eliminated?) by various lexical, contextual, semantic factors

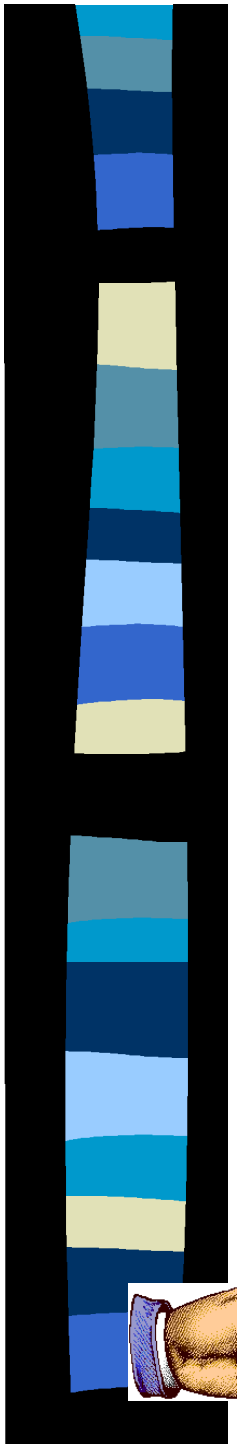
The students taught by the Berlitz method failed miserably.



An interesting asymmetry

- The evidence in favor of on-line semantic/contextual effects always shows how various factors make the *dispreferred* structure *easier*
- But Frazier (1995, 1998) has argued that it is *"completely accidental in the constraint-satisfaction model that garden paths have not been demonstrated for the simpler (preferred) structure"*

The children pushed quickly by the armed guards.

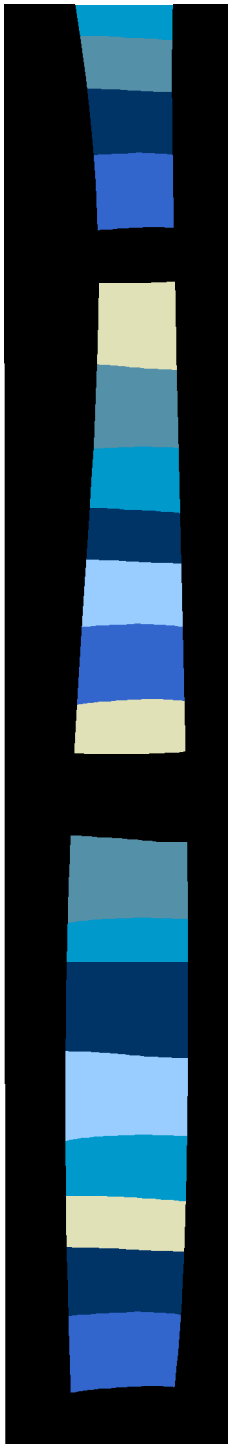


The predictions

- Constraint-satisfaction models predict:
  - Effects of semantic fit,context on processing *dispreferred* structure (could reduce garden path)
  - Effects of semantic fit, context on processing *preferred* structure
    - because evidence *for* the dispreferred structure is evidence *against* the preferred structure
    - So could actually cause a garden path if dispreferred structure becomes preferred
- Classic structural models predict:
  - Effects of semantic fit, context on processing dispreferred structure due to *easier reanalysis* (second pass processing)
  - But *NO effect on preferred structure, because it is <u>always</u> pursued first*

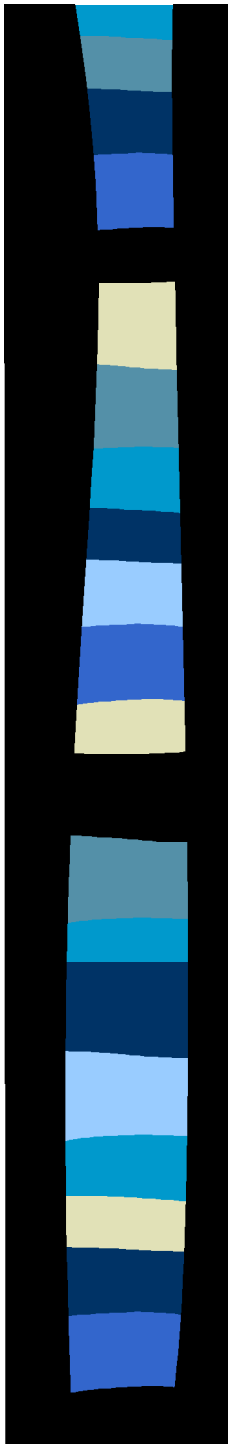
The ACT-R 4.0/5.0 prediction

The ACT-R predictions should pattern with the classic structural/ Minimal Attachment-style theories



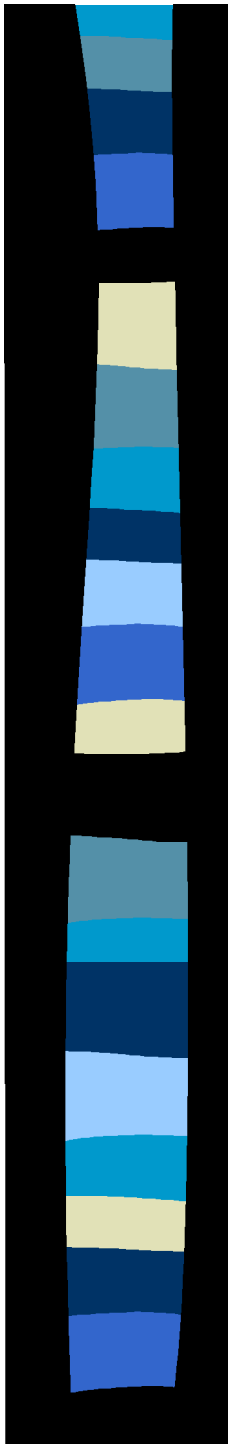
Binder, Duffy & Rayner (2001)

- This is *exactly* what Binder et al found, in a carefully done eye-movement study
- They found *no hint* of garden path effects in first-pass measures for the main verb construction, even when both semantic fit and referential context conspired *against* the main verb reading, and *for* the reduced relative reading
  - Importantly, they used materials (and even improved on them) that have been demonstrated to show clear effects of GP reduction for the relative clause, and showed in off-line norming that they were equi-biased



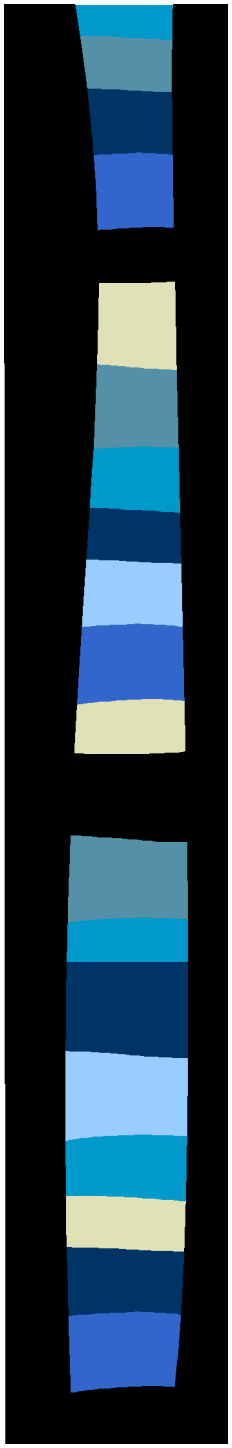
Another unexpected asymmetry

- Competition between structure *types* should *not* show effects of recency (decay)
  - E.g., VP-PP attachment vs. NP-PP attachment
  - Because the conflict resolution won't consider the activation of the attachment site
- But competition between sites for the *same* type of structural attachment *should* show effects of site activation (perhaps recency)
  - E.g., VP1-PP vs. VP2-PP



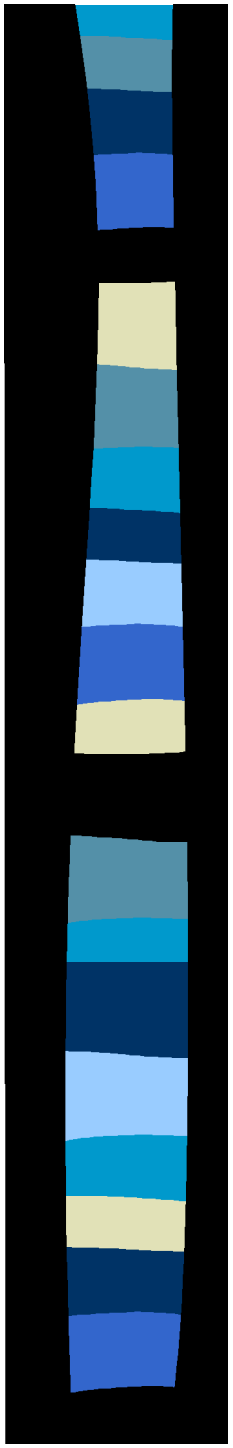
Example of the asymmetry in PP attachment

- Between structure-types: VP vs. NP
  - ***Mary painted the wall with cracks.***
  - Competing sites are the VP *painted* and the NP *wall*
- Within structure-type: NP1 vs. NP2
  - ***The father of the queen with the beard.***
- This actually maps roughly onto distinction between two major preferences: *Minimal Attachment* and *Late Closure* (Frazier)
  - BUT: It has always been *stipulated* that when the two factors *conflict*, MA wins



Bottom line theoretical implications

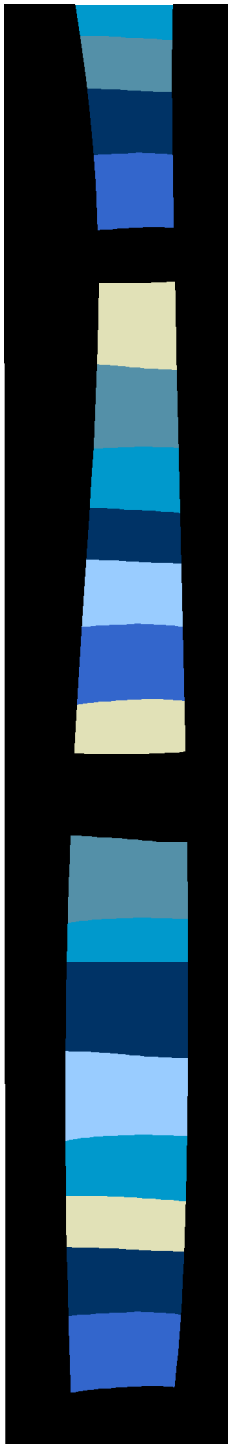
- ACT-R 3.0 yields something a bit closer to *lexical-constraint-based approach* to ambiguity resolution (e.g., Tanenhaus, MacDonald…)
- ACT-R 4.0/5.0 yields something a bit closer to a *modular structure-first* approach to ambiguity resolution (e.g., Frazier, Clifton)
  - Actually, closest to statistical tuning models (e.g, Mitchell et al, Crocker et al.)
- **WARNING: I'm oversimplifying the issues here considerably**



Example of possible problems: Major category frequency

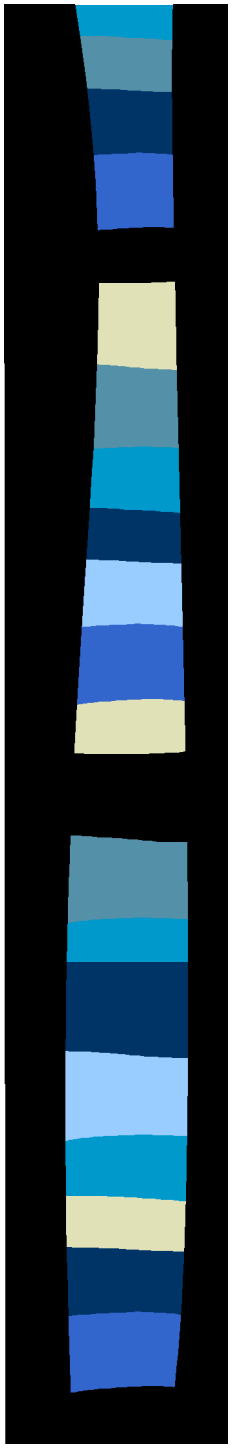
Boland 1998; Corley & Crocker 1998

- Base activation of lexical entries reflects frequency; determines retrieval latencies
- Ambiguous: bias affects resolution
  - *the German **makes** the beer/are cheaper…*
    *the warehouse **prices** the beer /are cheaper*
  - All things being equal, base-level activations will determine which lexical entry is attached first
- Unambiguous: bias affects processing times
  - Lower base-levels = slower times for subordinate
  - *…the German **make** is cheaper than ..*



Factoring into multiple productions *may* save the day

- A production could make an initial retrieval of the dominant lexical entry, followed by the attachment productions
- More consistent with smaller-grain-sized productions anyway
- Something like this happens already in the ACT-R 3.0 model
  - But these productions are category-specific; this solution will only work if productions are general
- Critical issue, then, may be *TIME*



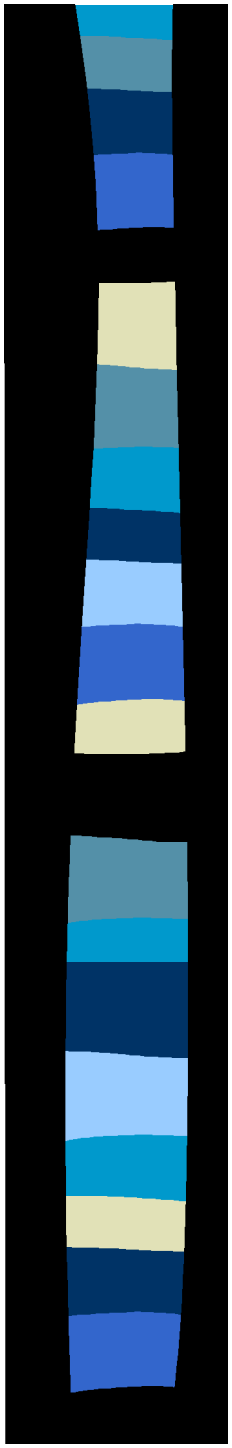
Issue #4: Decay vs. interference in reanalysis

- Any serial model (such as the ACT-R model) of sentence processing must be capable of *reanalysis* when the wrong path is pursued
  - Because not all “garden paths” are difficult:

The boy understood the man was paranoid.

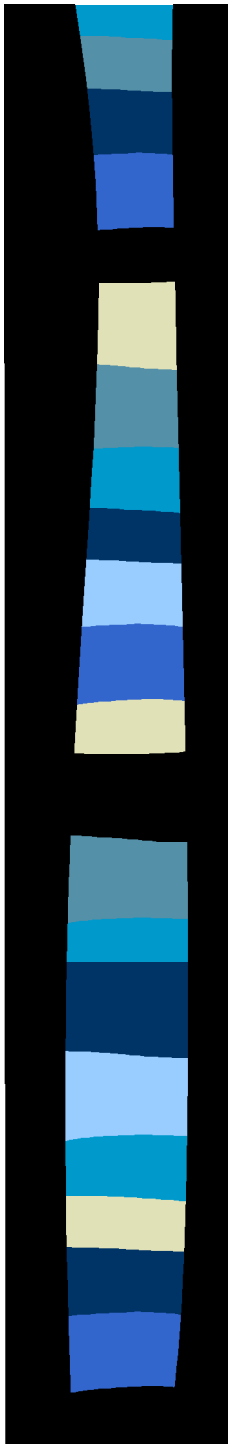
- What factors affect reanalysis difficulty?
  - One common assumption is length; but even long ambiguous regions can be pretty easy:

The boy understood the man who was swimming near the dock was paranoid.



Another interesting asymmetry prediction

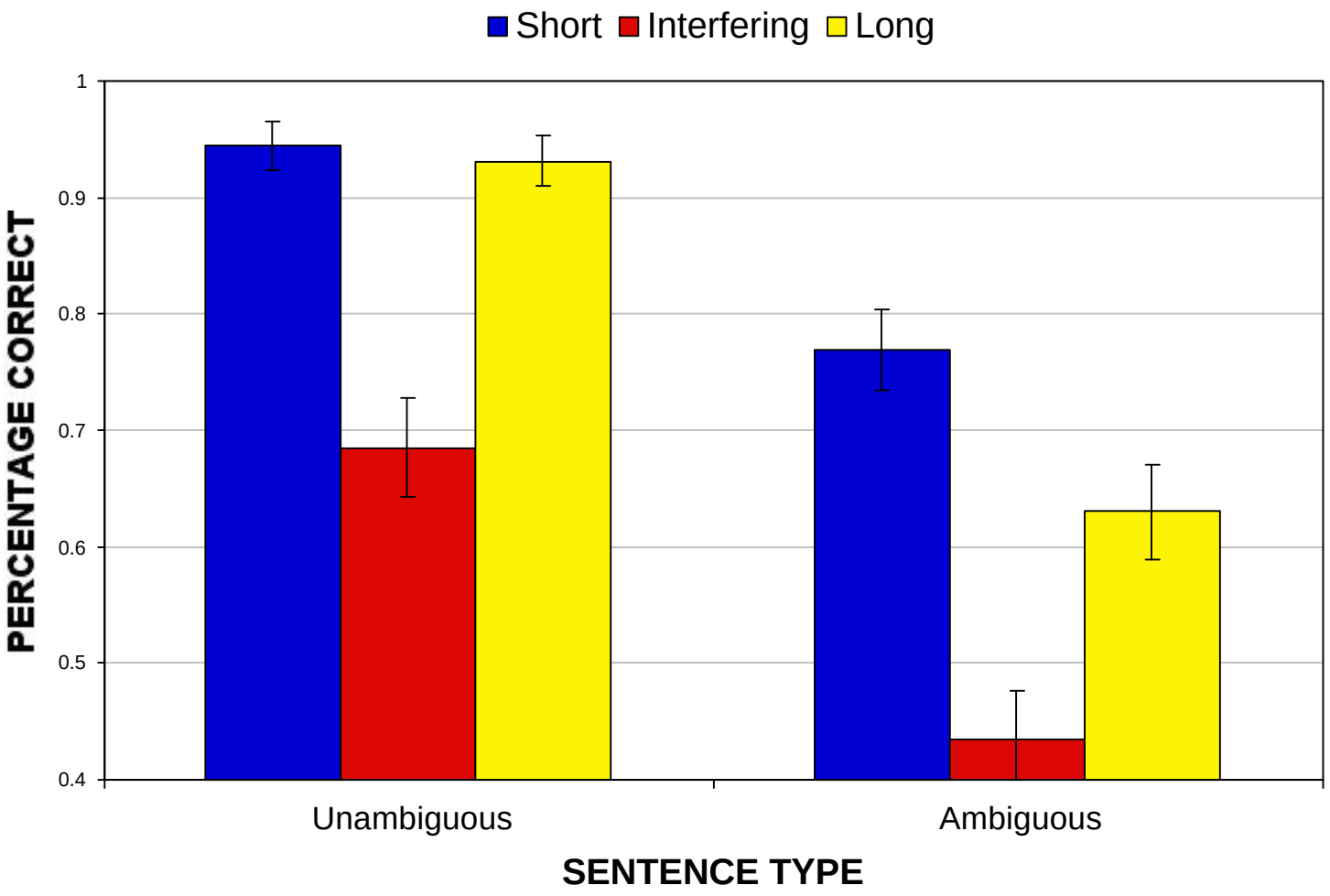
- Both structural *interference* (due to associative retrieval interference) and *decay* should affect syntactic attachments
  - Hence, should also affect reanalysis; because reanalysis requires attachment to the correct (dispreferred, discarded) structure
- But a discarded structure will suffer MORE *decay* than the chosen structure, because the chosen structure receives activation boosts from being used
- But it will NOT suffer more *interference*



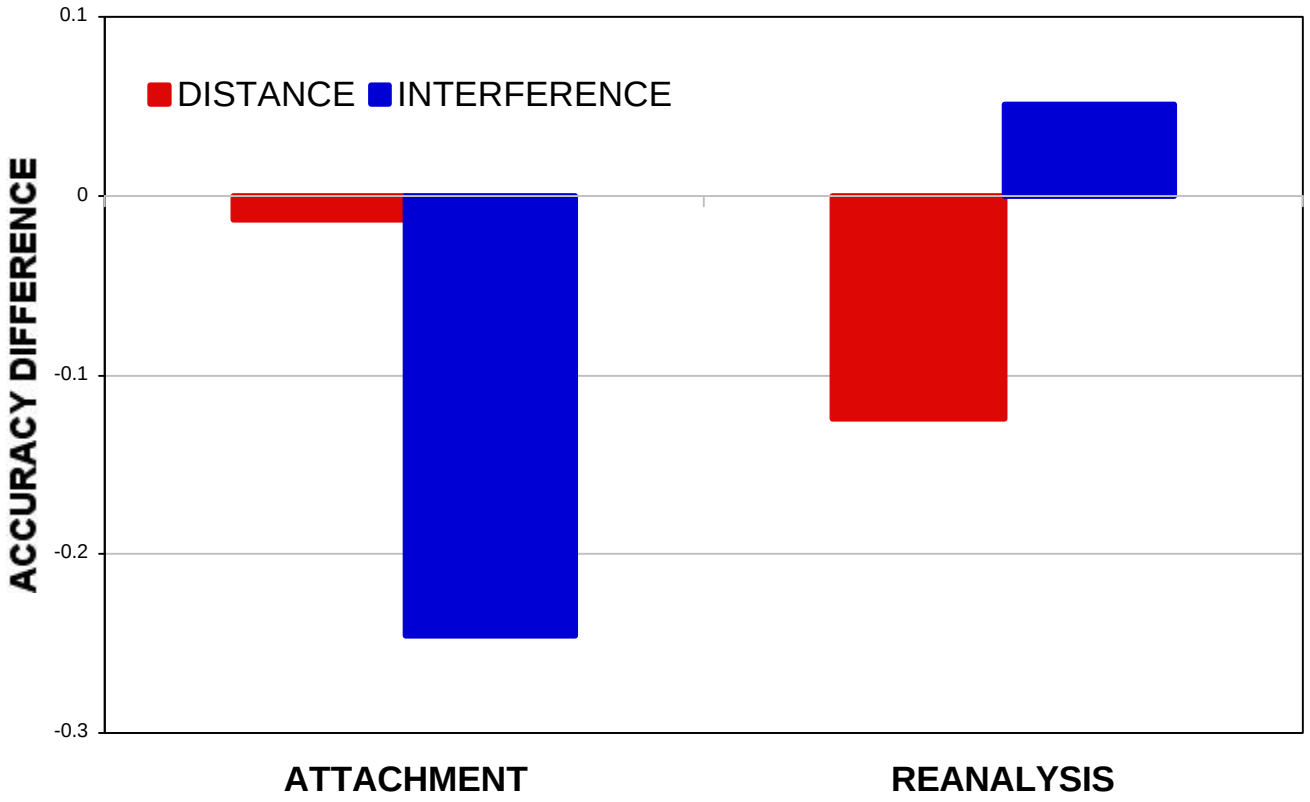
A reanalysis study
(w/ Julie Van Dyke, Pitt/UM)

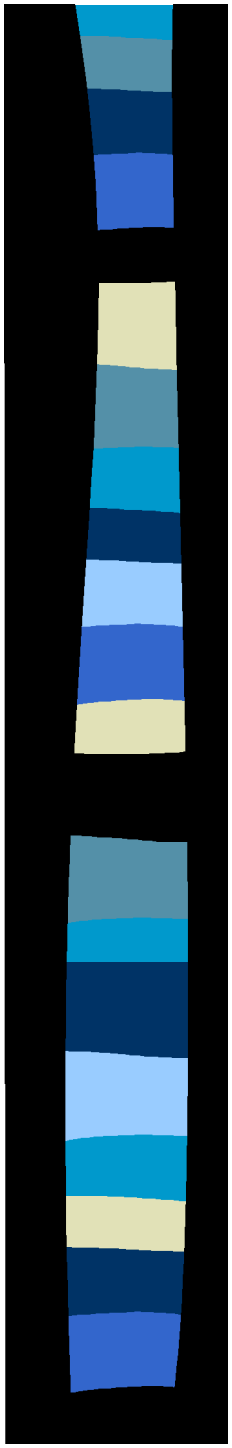
- Compare ambiguous and unambiguous versions of short, long, and interfering structures
  - We compute cost of reanalysis per se by comparing ambiguous and unambiguous conditions
- Example of interfering condition:
  - ***The boy understood the man who said the townspeople were dangerous was paranoid.***
- Task: rapid grammaticality judgment (Ferreira & Henderson, 1991)

The results


Short
Interfering
Long
PERCENTAGE CORRECT
1
0.9
0.8
0.7
0.6
0.5
0.4
Unambiguous
Ambiguous
SENTENCE TYPE

Results replotted

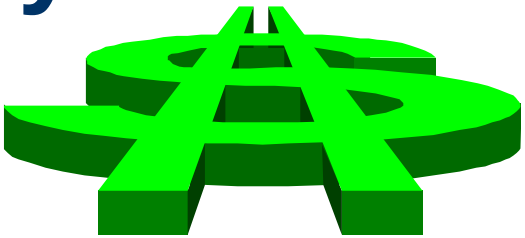




Some distinctive features of the ACT-R approach

- Explicit theory of **retrieval in WM**
  - Fits well with emerging modern views of STM/WM (McElree, Tehan & Humphreys, …)
  - And effects of decay, interference
- Explicit theory of **serial order representation**
  - Tackles long-neglected functional problem
  - Potential unification with verbal STM theory
- New perspective on **modularity**, grounded in rational analysis and computational concerns
  - “Independently motivated” theory (ACT-R conflict resolution) providing great constraint
- Generally, contact with cog psych theory

What does ACT-R buy you?



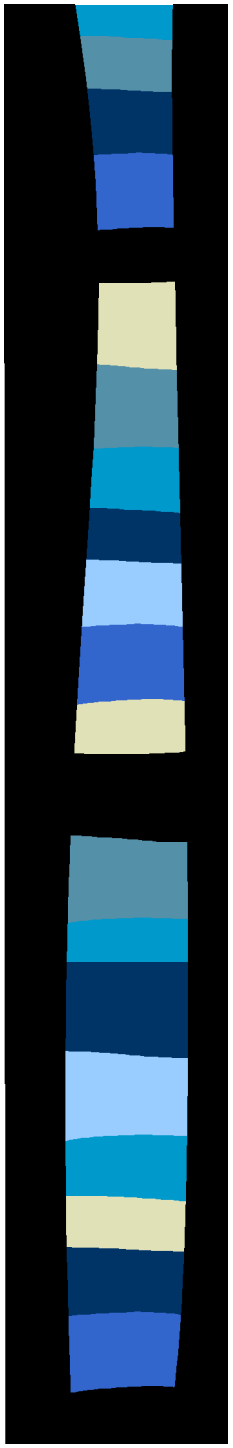
(1) ACT-R is a vehicle for making *psycholinguistics* come into contact with the *theoretical vocabulary of cognitive psychology*

(2) Evidence in some areas that the *details* of ACT-R may be pushing the theory in just the right direction

(3) Provides *unification* with cognitive theory; hence greater explanatory power

(4) Provides framework for building detailed *quantitative* processing models

- Hence, permits bringing to bear quantitative data on theory construction



A new kind of psycholinguistics?

- ACT-R modeling, and the empirical work it motivates, could help lead to a psycholinguistics characterized by
  - Complete, detailed processing theories; models of the fine structure of sentence processing
  - Quantitative, parameter-free(?) models
  - Models that make explicit predictions about the dependent measures used in the experiment (eye-movements, button presses, judgments)
  - Highly *constrained*, hence explanatory and predictive models

Major opportunities and challenges

(1) Incorporating independently motivated ACT language models in all models involving verbal material.

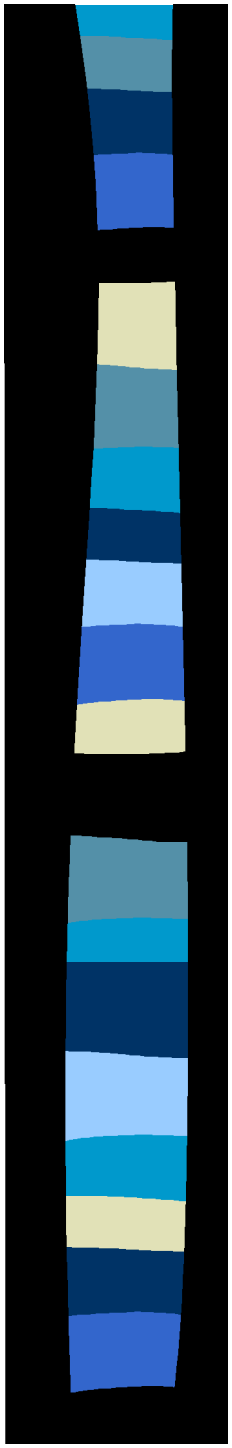
(2) Instruction taking.

(3) Functional NLP concerns: Scaling up

(4) Linguistic task operators

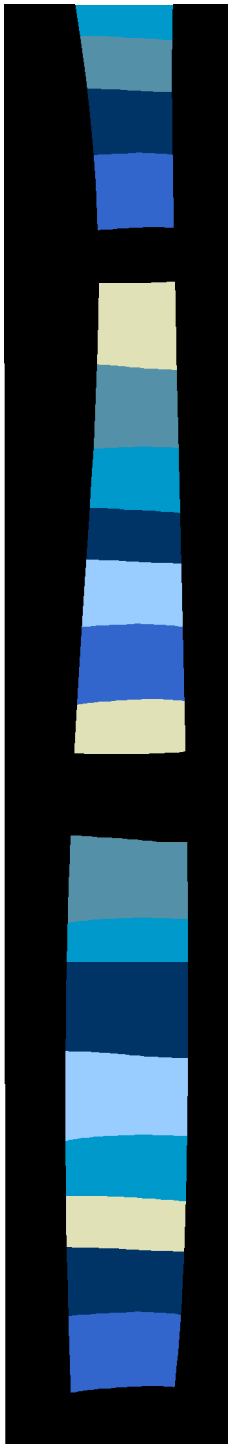
(5) Closing the perception-motor loop via ACT-R/PM

All are unique to cognitive architectures



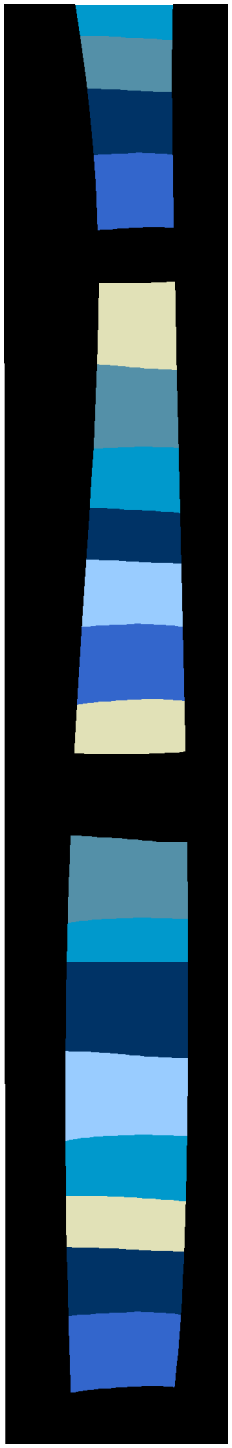
Opportunity/challenge #1

- Routinely incorporating independently motivated ACT language model(s) in *all* models of experiments with verbal materials
  - Closing the loop so that the linguistic processing is completely constrained; no theoretical degrees of freedom on the language side (cf. Kintsch models of comprehension)
  - Some examples moving in this direction: Anderson, Budiu, & Reder (2001), and Altmann & Davidson (2000)
  - Verbal rehearsal



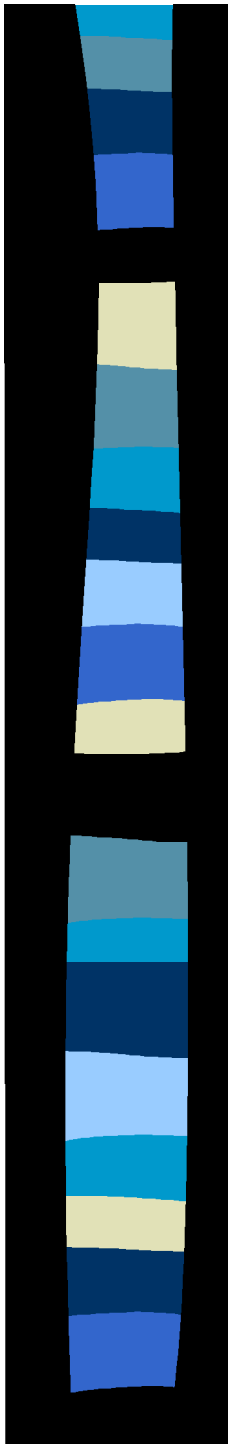
Opportunity/challenge #2

- Instruction taking (one of Newell's dreams for a UTC)
  - Finally "close the loop": rather than posit productions and chunks that encode knowledge of a task, have models that read instructions and carry out the task (Lewis, Polk, Newell, 1989)
  - *Considerably* reduces theoretical degrees of freedom
  - Build systems that accomplish variants on some experimental paradigm



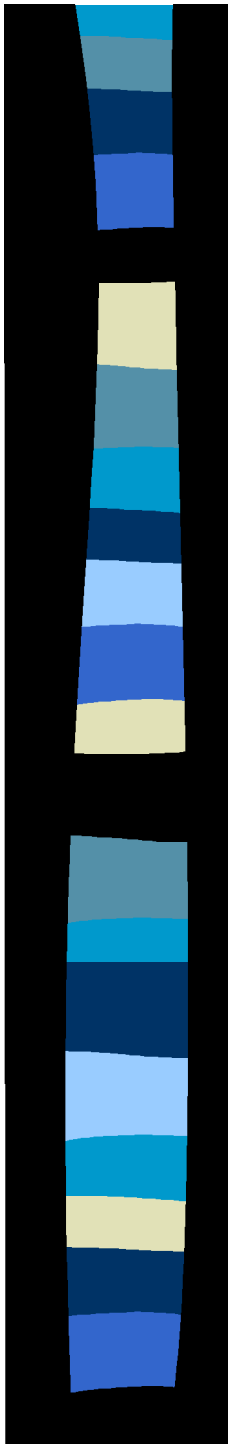
Opportunity/challenge #3

- Linguistic task operators (Lehman, Polk, Newell, Lewis, 1991)
  - Build models in which language is used to perform cognitive tasks (thinking by talking to oneself)
  - Uses language comprehension operators themselves as the *interpretive process* that yields "behavior"
- Turns standard instruction-taking process on its head
  - Uses NL itself as the language for representing behaviors
  - Newell had produced a set of LTOs that accomplished the blocks world



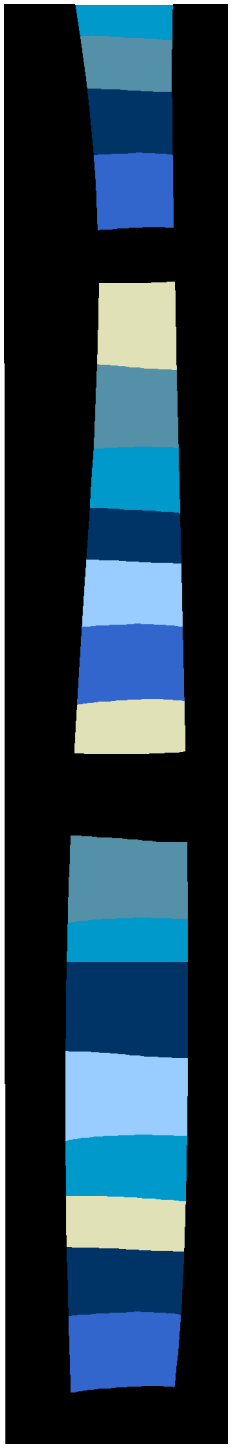
Some speculations

- Could this offer solution to John's ugly interpretive code?
- Depends on an NL semantics very closely grounded in perceptual-motor representations
  - So that one's understanding of "push the button" is quite close to the motor program that will be set up to actually push the button
  - Then "interpretive execution" is more like "releasing" the motor program rather than "interpreting a declarative representation"
- Need to be careful: could lead to a procedural semantics to NL
  - Learn from Miller & Johnson-Laird's program



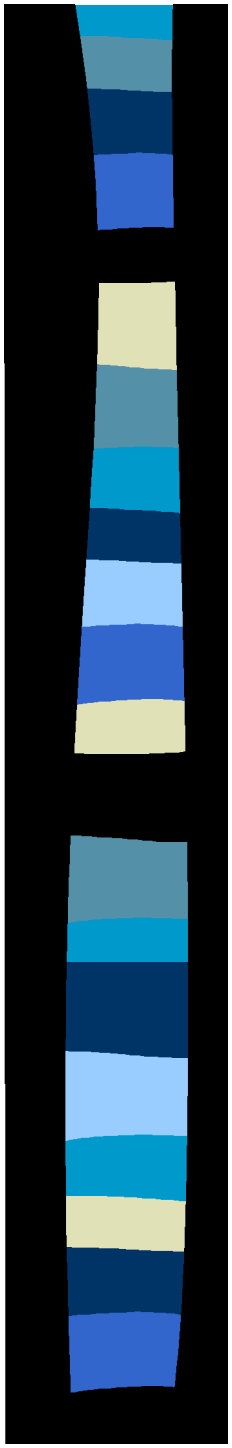
Opportunity/challenge #4

- Functional NLP challenges
  - **Scaling** NL systems in ACT-R: Can ACT handle a lexicon of 30,000 words? A grammar base of 1,000s of production rules?
  - "Train" ACT-R on large corpora of text to set production/declarative memory parameters
  - Not just a technical engineering question; of critical *theoretical* importance psycholinguistically
  - ***"Scale counts in cognition"***



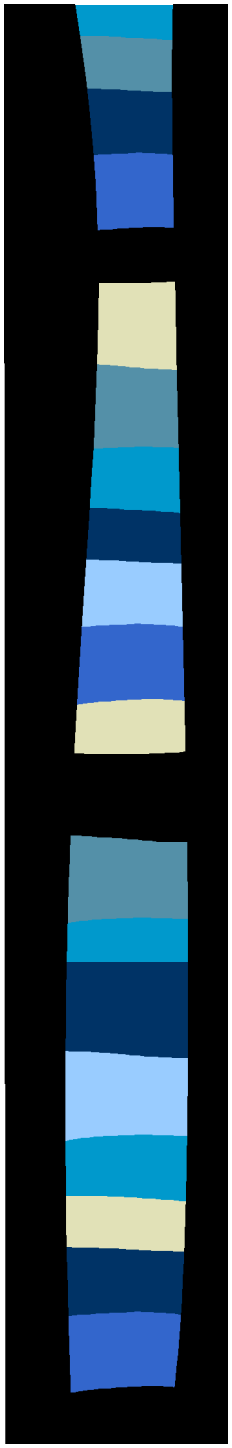
Opportunity/challenge #5

- Use ACT-R/PM/EMMA to develop explicit models of eye-movements in reading
  - Good for the psycholinguistic theory
  - Good for ACT-R/PM—enormous literature on eye movements in reading
- Develop models of eye-movements in context
- Develop models of button-pressing paradigms
  - Word-by-word reading
- Develop explicit theories of *global judgments* (grammaticality, difficulty, acceptability)
  - Binary, 1-7 scale, magnitude estimation



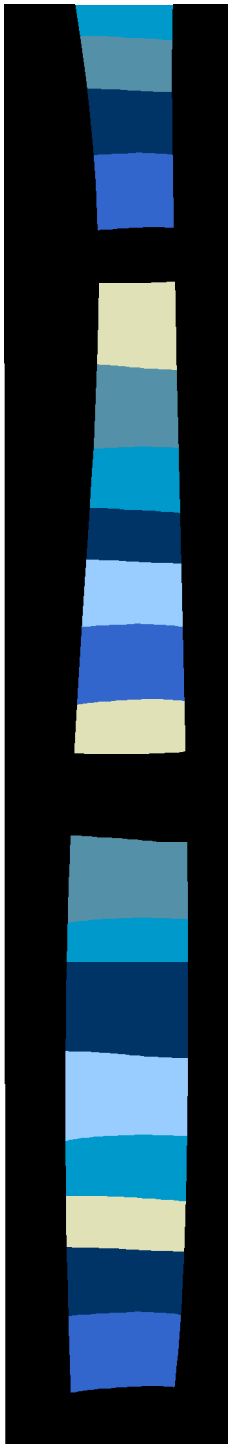
What does it take to meet these challenges?

- Common requirements for many:
  - Incorporating a *(broad coverage) theory of semantic representations*
    - Possibilities include: Sowa's Conceptual Graphs; Jackendoff's Conceptual Structures; Miller et al's WordNet
  - Incorporating a *(fairly broad coverage) theory of syntactic and lexical representations*
    - Possibilities include: HPSG, combinatory categorial grammar
- Technically, this incorporation will involve bringing some existing large database into ACT-R



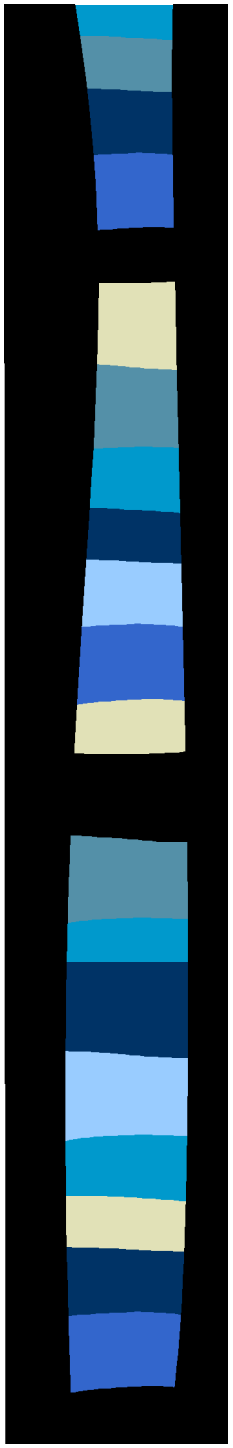
Potentially serious architectural challenges

- Time—is there enough?
  - Probably not.
  - Only have 200-300ms/word (on average)—Time for a couple of productions and retrievals
  - Anderson et al (2001) met constraint by combining considerable amount of syntactic, semantic structure building into single productions
  - But I have separated these in my model (and in earlier NL-Soar model, in which there was *just barely* enough time)
  - Also referential processing happening on-line
- I don't taking timing seriously in current model



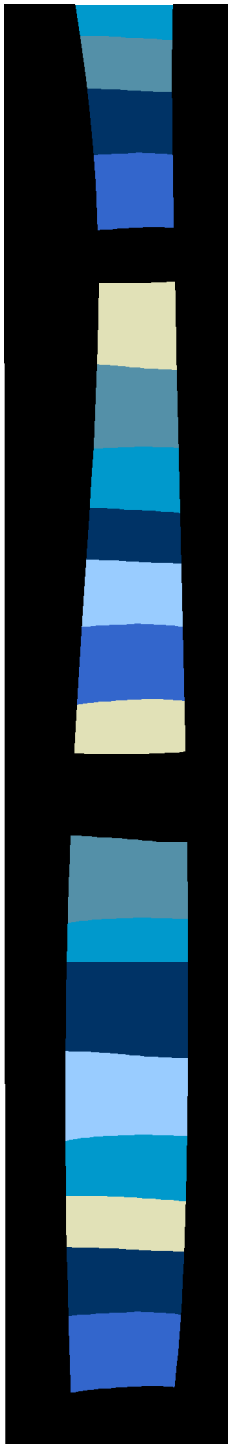
Potentially serious architectural challenges

- Is ACT-R too hopelessly symbolic and serial for language processing?
  - Many think of “lexical entries” and “lexical access/retrieval” as old-fashioned
- Right approach: Go after signature phenomena addressed by connectionists
  - *Good candidate*: Tanenhaus et al work on eye-movements in context that track time-course of lexical access and sentence processing
  - Can see neighborhood effects on-line, extremely rapid match to referential context
  - Can ACT-R work fast enough to do this?



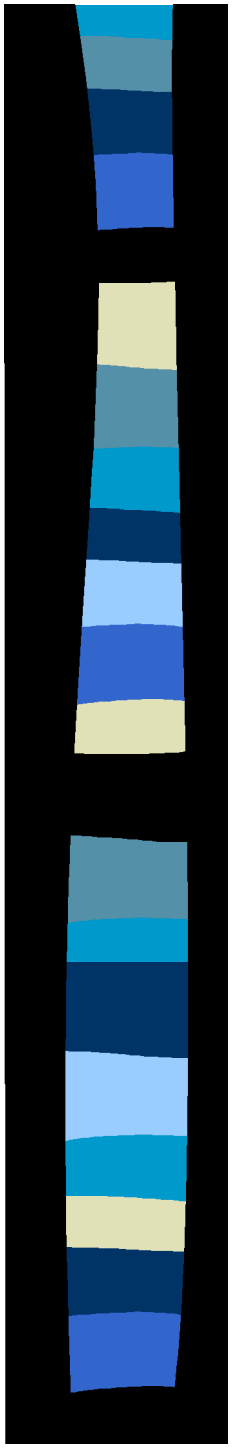
Potentially serious architectural challenges

- Acquisition
  - Hard to work on this problem without a stable production learning mechanism
  - Perhaps compilation will be a reasonable base
- Control structure
  - Can ACT-R's conflict resolution handle interactive/lexical effects in ambiguity resolution?
  - (Perhaps, but in appropriate time limits?)



What about ACT-R 5.0?

- By gosh, it's the best thing since ACT-R 4.0!!
- A big potential win as I see it now: *competitive declarative memory retrievals*
  - Could provide natural account of differential pattern of reading times on *lexical ambiguities (slow-down)* and *syntactic ambiguities (no-effect or even faster)*
  - Could provide less heavy-handed approach to getting associative interference effects**;** e.g., may not have to worry about dynamically resetting cues so fan doesn't build up too much
  - Potentially cleaner account of similarity-based interference
    - Probably incorporating Raluca's representational similarity

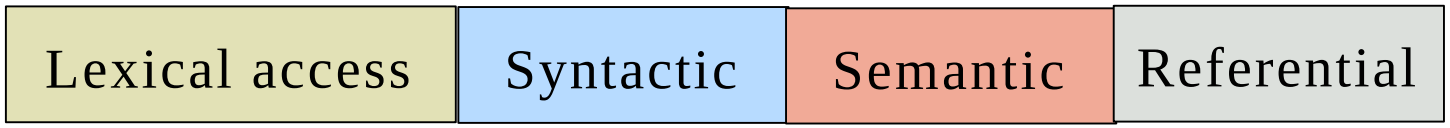


Other potential wins in 5.0

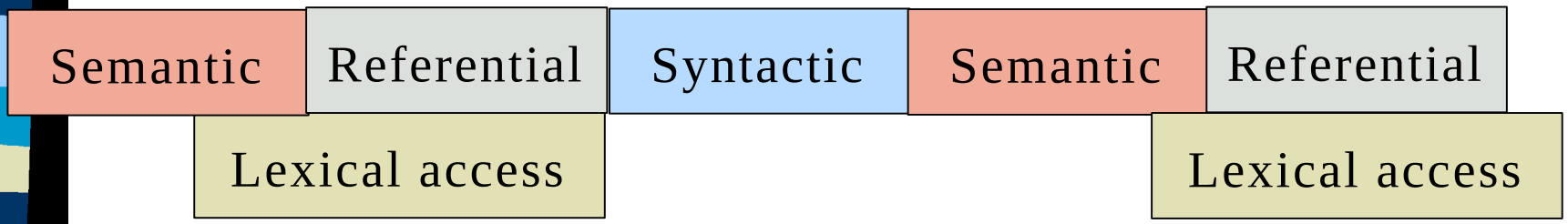
- Another potential win: *parallel retrieval and production firing*
  - Issue: may not be enough time for firing syntactic, semantic, referential processing productions *AND* perform lexical access
  - But there *MIGHT BE* enough time if lexical access for the *next* word can be initiated while finishing up processing of the last
  - Predicts “spill-over” effects in reading, for which there is *ample* evidence, in both self-paced reading and eye-tracking

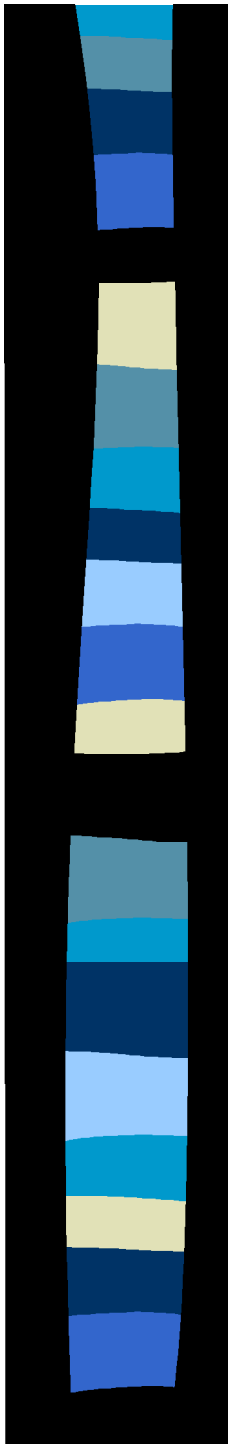
Timing in ACT 3/4 vs. 5.0

- ACT 3.0/4.0



- ACT 5.0





Yet another possibility

- Use production compilation to compile out the lexical access
  - Produces word-specific comprehension production
    - Similar to original idea of “comprehension operators” in Soar
  - Would then shift burden of lexical frequency effects to procedural memory
- Has interesting effect of distributing (redundantly in quite specific ways) grammatical knowledge across the “lexicon”
  - This might be exactly right
- Actually, may be impossible to avoid

Revisiting the killer bees

- Let's reconsider some of Wexler's gripes
  - Tackling broad range of complex linguistic (and psycholinguistic) phenomena? YES
  - Constrained, explanatory principles? YES
  - Asking for too much from a broad theory (because modularity is right)? NO
- And his suggestion of taking advantage of existing representation theories
  - This is basically right on target and exactly what was done in NL-Soar, and is right path for ACT-R
  - ***Linguistic theory provides the ontology of representational features, ACT-R the processing architecture***

Minimal Attachment review

- Dominant theory in the 80's, early 90's: Frazier's *Garden Path Model*
  - Serial (single structure pursued)
  - Decision principle: *Minimal Attachment:* Structural ambiguities resolved by pursuing the simplest structure (determined by counting number of syntactic nodes); simplest structure assumed to be computed most quickly
  - (Other principles involved; e.g., *Minimal Chain Principle, Right Association; Construal)*

Opportunities & challenges, cont.

- Connectionism
  - Will the symbolic side of ACT-R language models be *hopelessly symbolic*?
  - E.g., many researchers reject the idea of "retrieving entries" from a lexicon—stored lexical entries are old-fashioned