Computational Psycholinguistics

Lecture 8: Constraint-based Models of Human Sentence Processing

Afra Alishahi

December 22, 2008

Probabilistic Models: Jurafsky, ICMM

- **Architecture**: Modular
  - Lexico-syntactic processor, no semantic knowledge
- **Mechanisms:** Parallel
  - Incremental, bounded parallel parsing, with reranking
- **Information resources:** Lexical and structural probabilities
- **Linking Hypothesis**:
  - Parse reranking causes increased RTs, if correct parse has been eliminated, predict a garden-path

Constraint-based Models

- Constraint-based models of sentence processing
    - focus on interactions of many probabilistic constraints to compute parallel competing interpretations
    - E.g., MacDonald et al. (1994), Spivey-Knowlton and sedivy (1995), McRae et al. (1998), Seidenberg and MacDonnald (1999), Kim et al. (2001)
    - Most of these models focus on the selection (and not the construction) of an interpretation for a sentence

Constraint-based Models

- **Architecture:** Non-modular
  - All levels are constructed and interact simultaneously
- **Mechanism:** Parallel
  - Ranking based on constraint activations
- **Information resources:** All
  - All relevant constraints are used immediately
- **Linking Hypothesis**:
  - Comprehension is easy when constraints support a common interpretation, difficult when they compete

Multiple constraints

"The doctor told the woman that ...

story

diet was unhealthy

he was in love with her husband

he was in love with to leave

story was about to leave

Prosody: intonation can assist disambiguation

Lexical preference: *that* = {Comp, Det, RelPro}

Subcat: *told* = { [_ NP NP] [_ NP S] [_ NP S'] [_ NP Inf] }

Semantics: Referential context, plausibility

- Reference may determine "argument attach" over "modifier attach"
- Plausibility of *story* versus *diet* as indirect object

The Interactive Activation Model

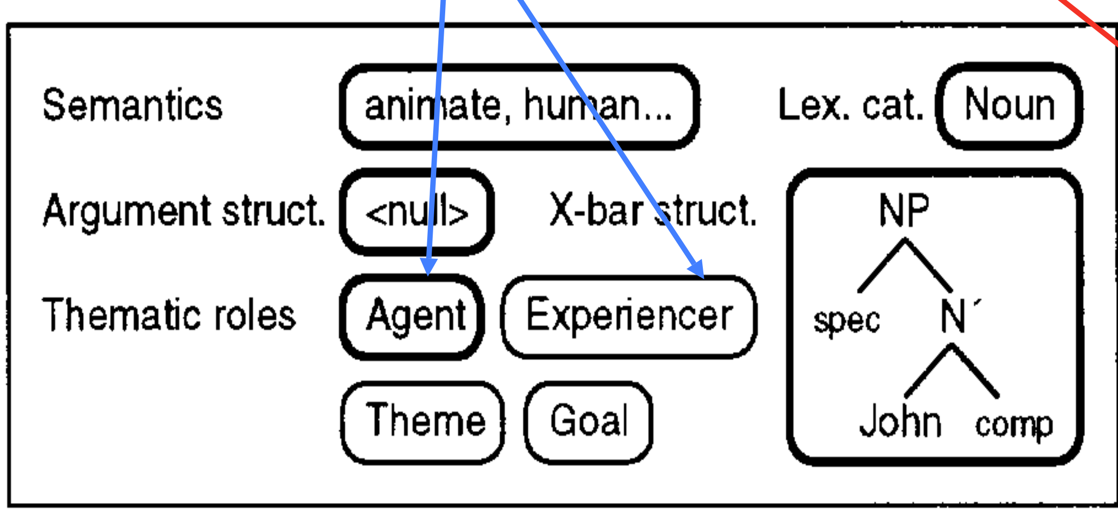
- The Interactive Activation Model (MacDonnald et al, 1994)
  - Simultaneous, multiple access is possible at all levels of representation, constrained by frequency/context
  - Detailed lexical entries are enriched with frequency information
  - Language processing is modeled as constraint satisfaction between lexical entries and across levels

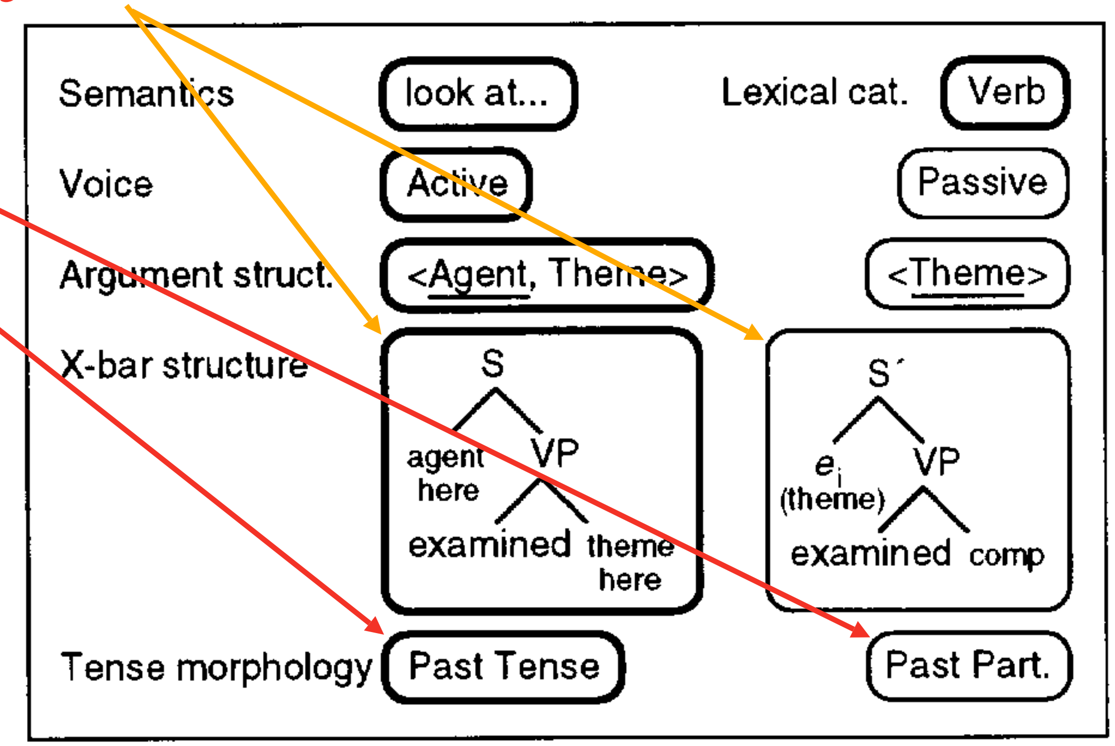
Interactive Activation

"John examined the evidence"

- "examined" is either a simple past or past participle
- Frequency determines 'activations'

➡ thematic fit, tense frequency, structural bias ...





Interactive Activation: Limitations

- Complex interaction behaviors are difficult to predict
  - Conflicting constraints should cause difficulty. Do they?
  - Difficult to actually implement, and estimate frequencies
- No distinct parser is modeled

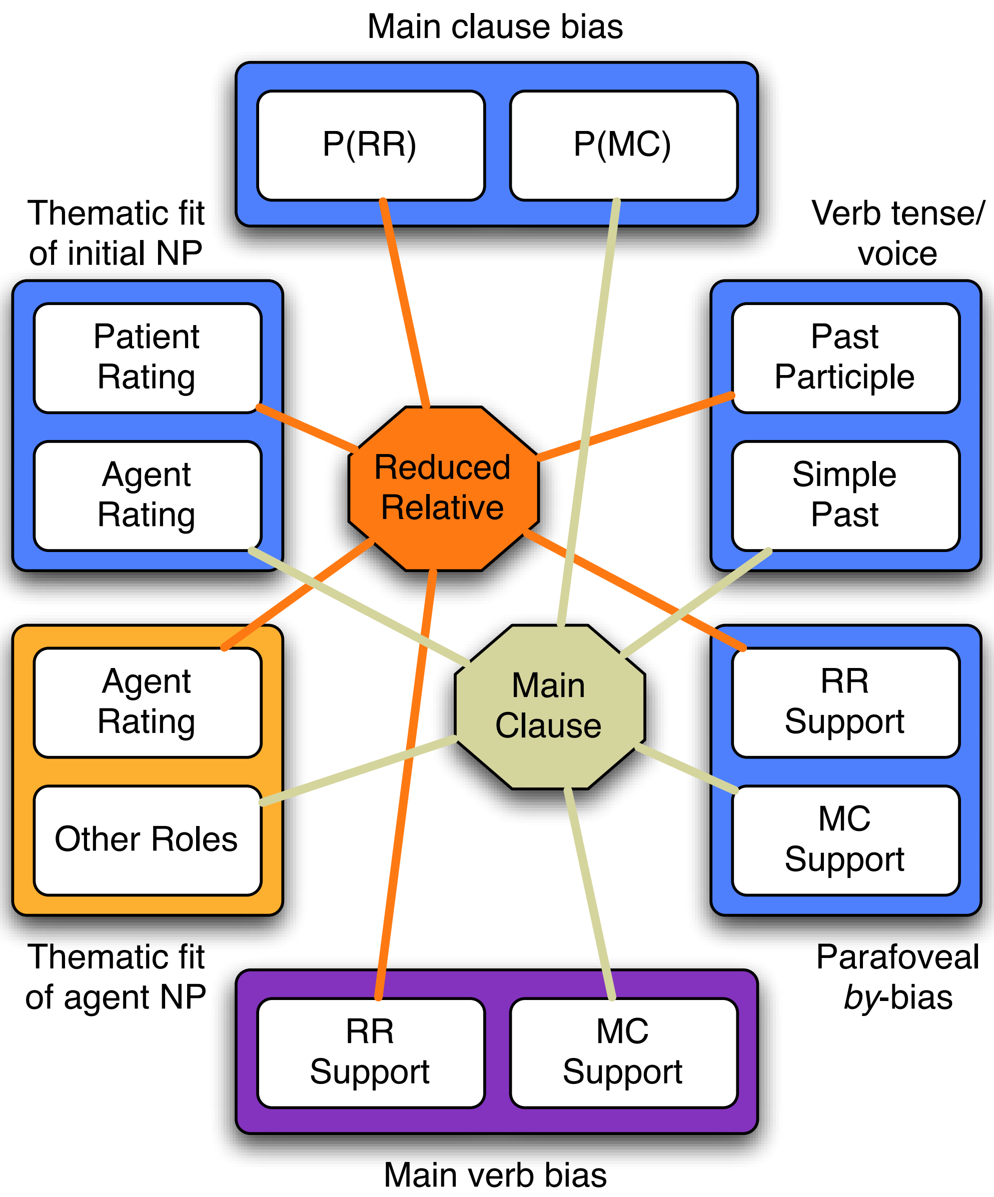
Competitive-Integration Model

- The Competitive-Integration Model (Spivey-Knowlton & Sedivy, 1995)
  - Diverse constraints (linguistic and conceptual) are applied simultaneously in ambiguity resolution.
- Assumption: all analyses are constructed
  - Weighted and normalized constraints provide probabilistic support for analyses
- Goal: Simulate reading times
  - RTs are claimed to correlate with the competition time required to settle on one of the alternatives

The Computational Model

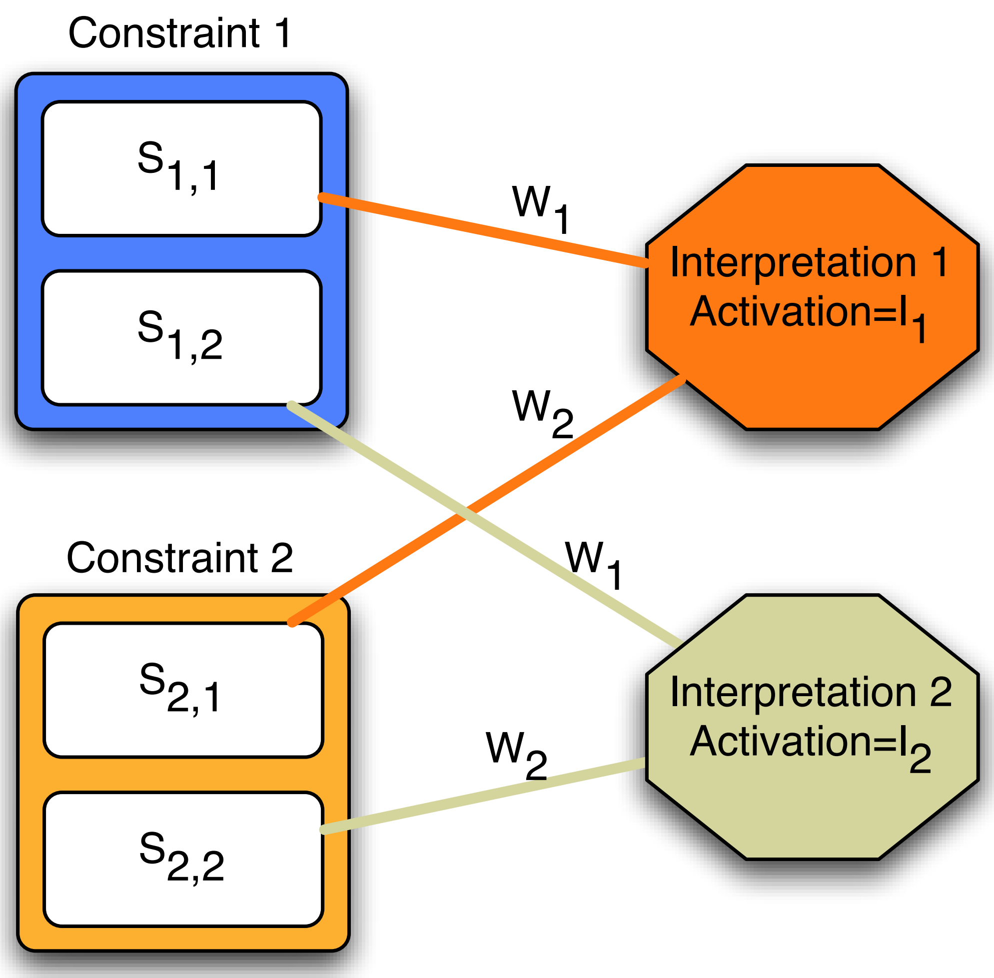
"The crook arrested by the detective was guilty of taking bribes"

1. Combines constraints as they become available in the input
2. Input determines the probabilistic activation of each constraint
3. Constraints are weighted according to their strength
4. Alternative interpretations compete to a criterion
5. Cycles of competition mapped to reading times



The Recurrence Mechanism

- $S_{c,a}$ is the raw activation of the node for the c^{th} constraint, supporting the a^{th} interpretation,
- w_c is the weight of the c^{th} constraint
- I_a is the activation of the a^{th} interpretation
- 3-step normalized recurrence mechanism:
  - Normalize $$S_{c,a}(norm) = \frac{S_{c,a}}{\sum_a S_{c,a}}$$
  - Integrate $$I_a = \sum_c \left[w_c \cdot S_{c,a}(norm) \right]$$
  - Feedback $$S_{c,a} = S_{c,a}(norm) + I_a \cdot w_c \cdot S_{c,a}(norm)$$



$$\sum_i w_i = 1$$

Implementing the Model

- Constraints contribute to the activation of competing analyses, over time

1. Identifying the relevant constraints
2. Estimate bias of each constraint towards each interpretation
3. Set the weight of each constraint so that behaviour of the model matches with previous studies
4. Make predictions for reading times

Evaluation (McRae et al., 1998)

- Constraint bias can be established using experience-based measures
  - corpus frequencies (e.g., for structural and lexical constraints)
  - completion norms (e.g., thematic constraint)
- Evaluation against reading time studies
  - Make predictions for reading times by the constraint-based model and the Garden-path model
  - Compare actual reading times with the predictions

Constraint Parameters

"The crook/cop arrested by the detective was guilty of taking bribes"

Verb tense/voice constraint: is the verb preferentially a past tense (i.e. main clause) or past participle (reduced relative)

⇒ Relative log frequency is estimated from corpora: **RR=.67, MC=.33**

Main clause bias: general bias for structure for "NP verb+ed …"

⇒ Corpus: **P(RR | NP + verb-ed) = .08, P(MC | NP + verb-ed) = .92**

by-Constraint: extent to which 'by' supports the passive construction

⇒ Estimated for the 40 verbs from WSJ/Brown: **RR= .8, MC= .2**

Thematic fit: the plausibility of crook/cop as an agent or patient

⇒ Estimated using a rating study

by-Agent thematic fit: good Agent is further support for the RR vs. MC

⇒ Estimated using a rating study

Thematic Fit Parameters

"The crook/cop arrested by the detective was guilty of taking bribes"

- Estimating thematic fit with an off-line rating (1-7) study

```
How common is it for a
    crook   _____
    cop     _____
    guard   _____
    police  _____
    suspect_____
To arrest someone?
To be arrested by someone?
```

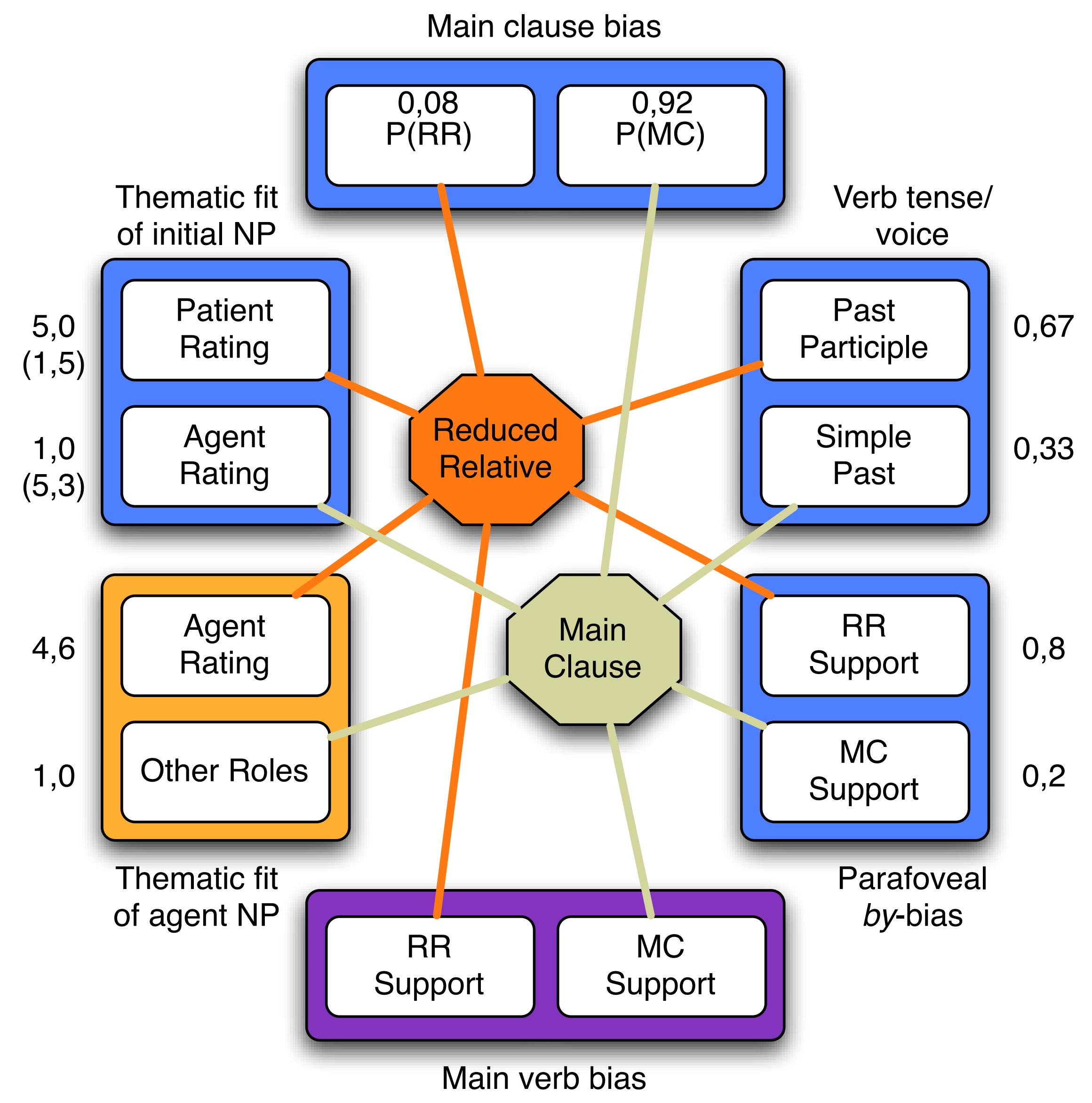
NP 1	Rel	Main
Agent	1,5	5,3
Patient	5,0	1,0

by NP	Rel	Main
Agent	4,6	1,0

The Computational Model

"The crook arrested by the detective was guilty of taking bribes"

1. Combines constraints as they become available in the input
2. **Input determines the probabilistic activation of each constraint**
3. Constraints are weighted according to their strength
4. Alternative interpretations compete to a criterion
5. Cycles of competition mapped to reading times



A Gated Completion Study

Gated sentence completion study:

The cop/crook arrested ...

The crook arrested by ...

The crook arrested by the ...

The crook arrested by the detective...

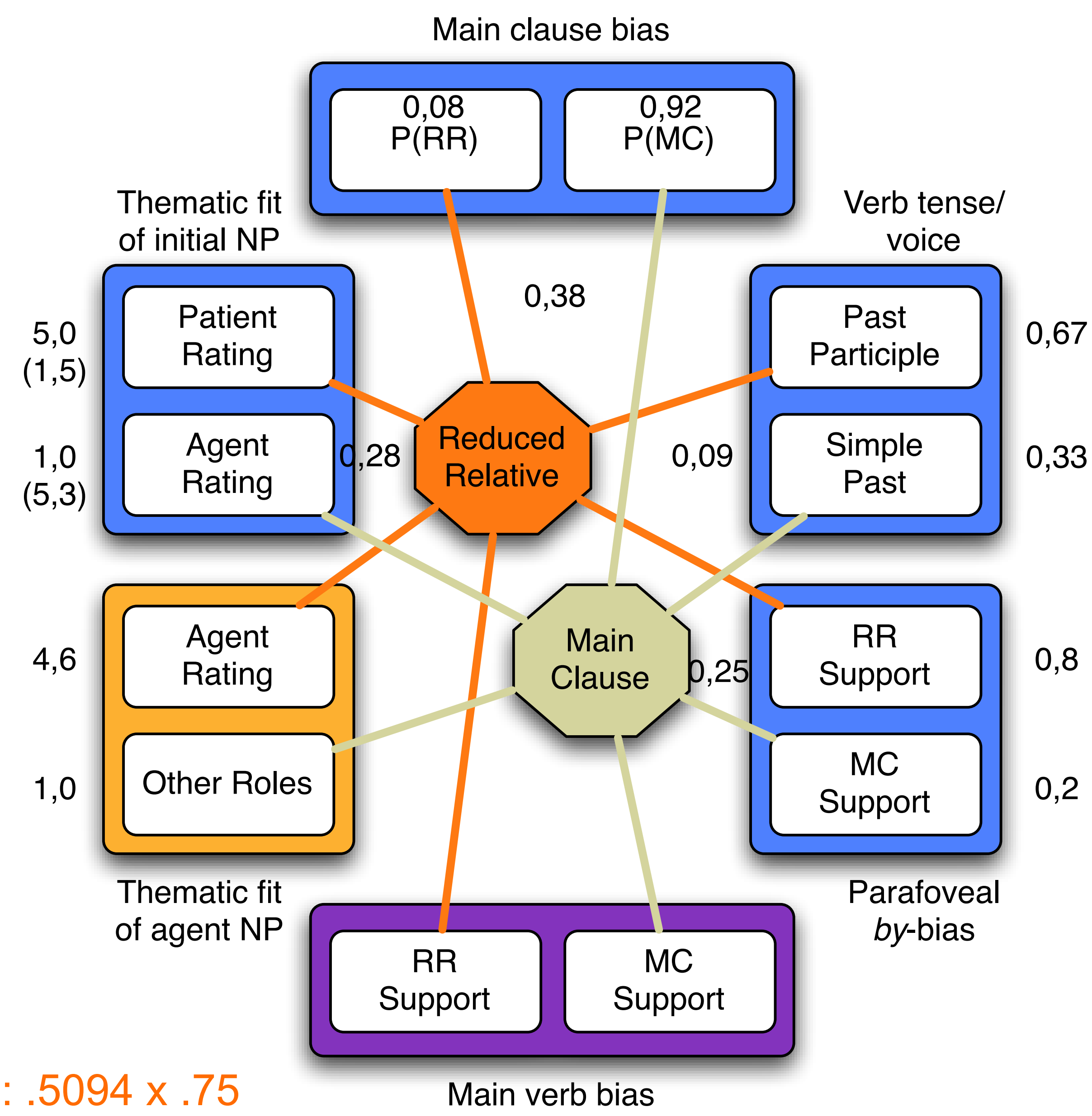
- Establish that thematic fit does influence off-line completion, and use that to adjust the weights
- Manipulate the fit of NP1:
  - Good agents (and atypical patients)
  - Good patients (and atypical agents)
- Hypotheses: Effect of fit at verb
  - Additional effect at ‘by’

The complete model

The Computational Model

"The crook arrested by the detective was guilty of taking bribes"

1. Combines constraints as they become available in the input
2. Input determines the probabilistic activation of each constraint
3. **Constraints are weighted according to their strength**
4. Alternative interpretations compete to a criterion
5. Cycles of competition mapped to reading times



MC bias : .5094 x .75
Thematic Fit : .3684 x .75
Verb tense : .1222 x .75
by-bias : .25

On-line study

- Two-word, self-paced presentation (similar to completion studies):

The crook / arrested by / the detecti

The cop / arrested by / the detective

The crook / that was / arrested by /

The cop / that was / arrested by / th

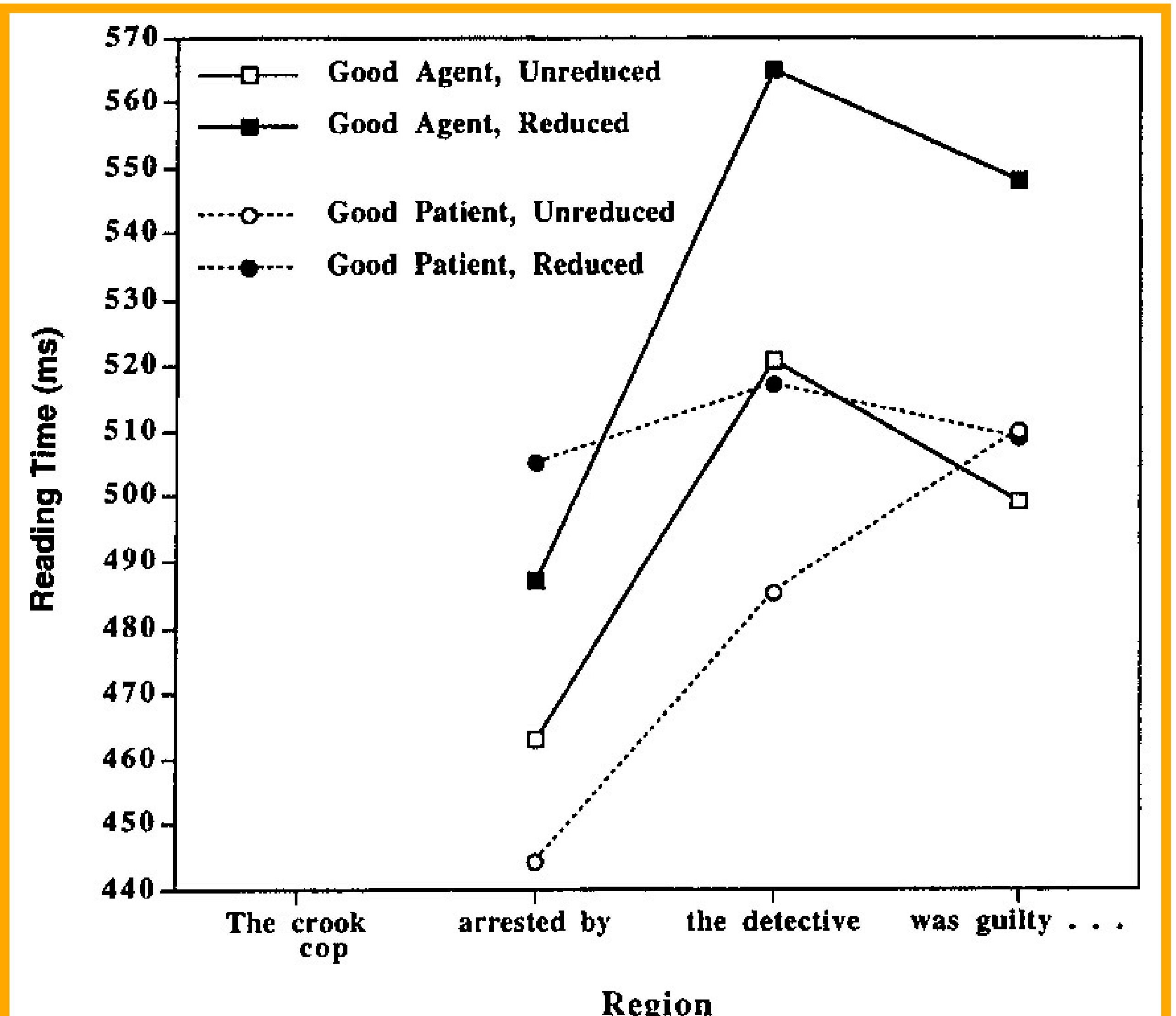


FIG. 5. Self-paced reading times for the Experiment.

Model Predictions

- Two “versions” of the model are implemented:
  - Constraint-Based: constraints apply immediately for each region
  - Garden-Path: MC and main-verb biases apply immediately, other constraints are delayed
- Predict per-region reading times for each model
  - Each region is processed until it reaches a threshold
  - As more cycles are computed, threshold is relaxed

CB vs. GP predictions

Constraint Based (CB) Model
MC bias: .5094 x .75
Thematic Fit: .3684 x .75
Verb tense: .1222 x .75
by-bias: .25

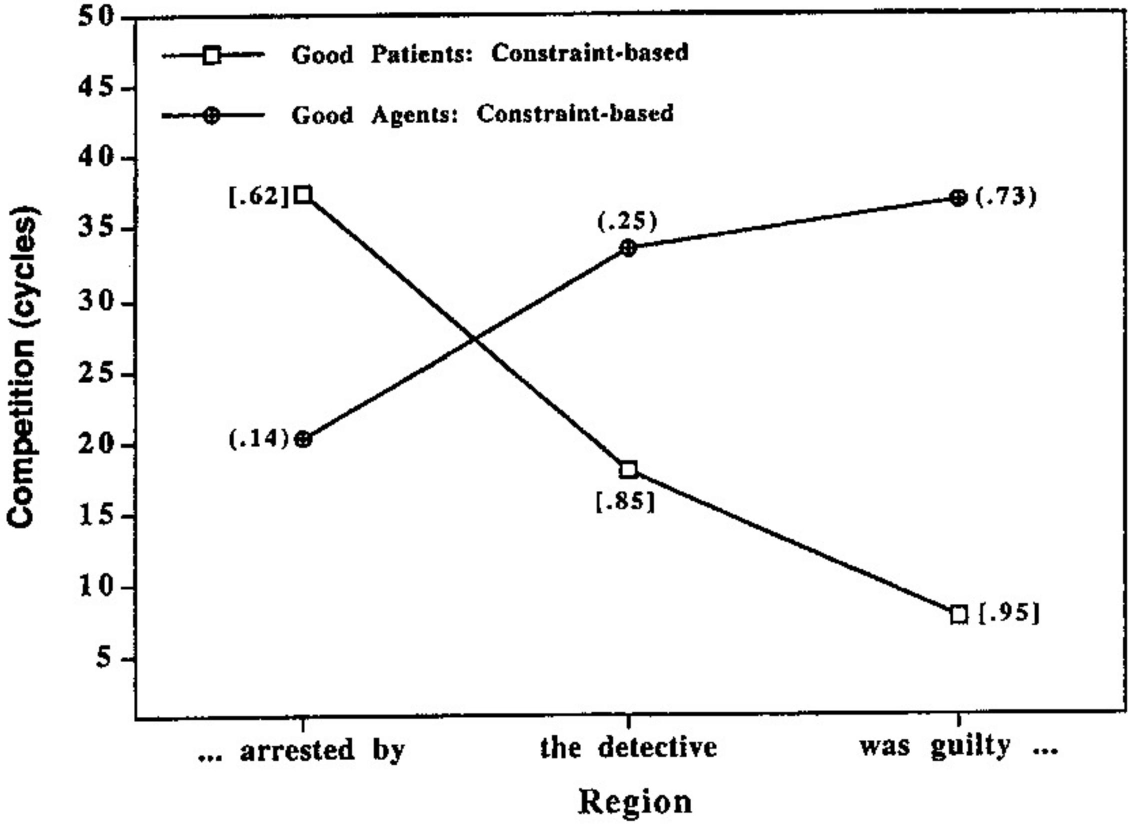


FIG. 3. Self-paced predictions derived from the constraint-based competition model. In this and all following model figures, the number beside each model datum is the mean activation of the reduced relative node after competition in that region for either (good agents) or [good patients].

Garden Path (GP) Model:
MC bias: 1

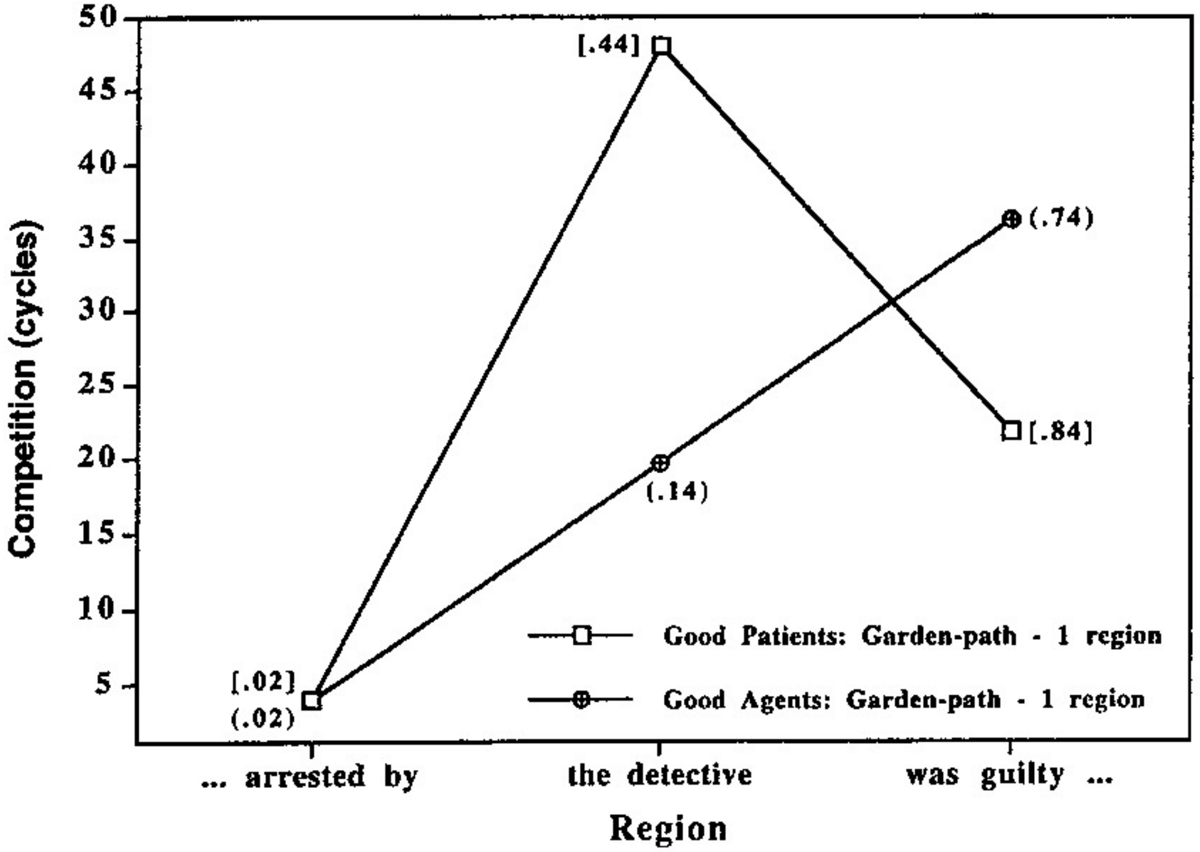


FIG. 4. Self-paced predictions as derived from the garden-path model when constraints other than the nain clause and main verb biases were delayed by a region.

GP vs CB Modelling of the Reading

Human reading times:

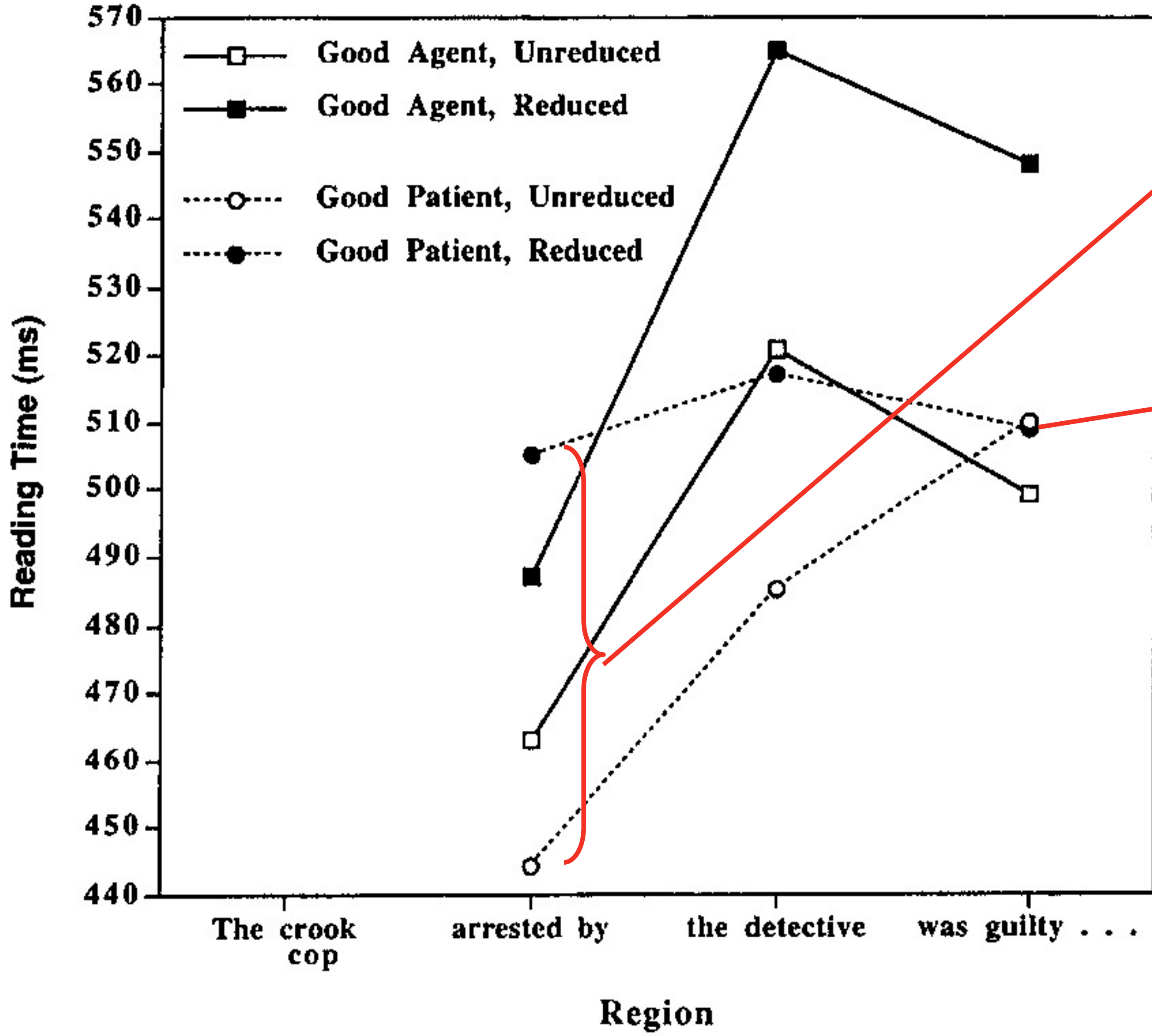


FIG. 5. Self-paced reading times for the Experiment.

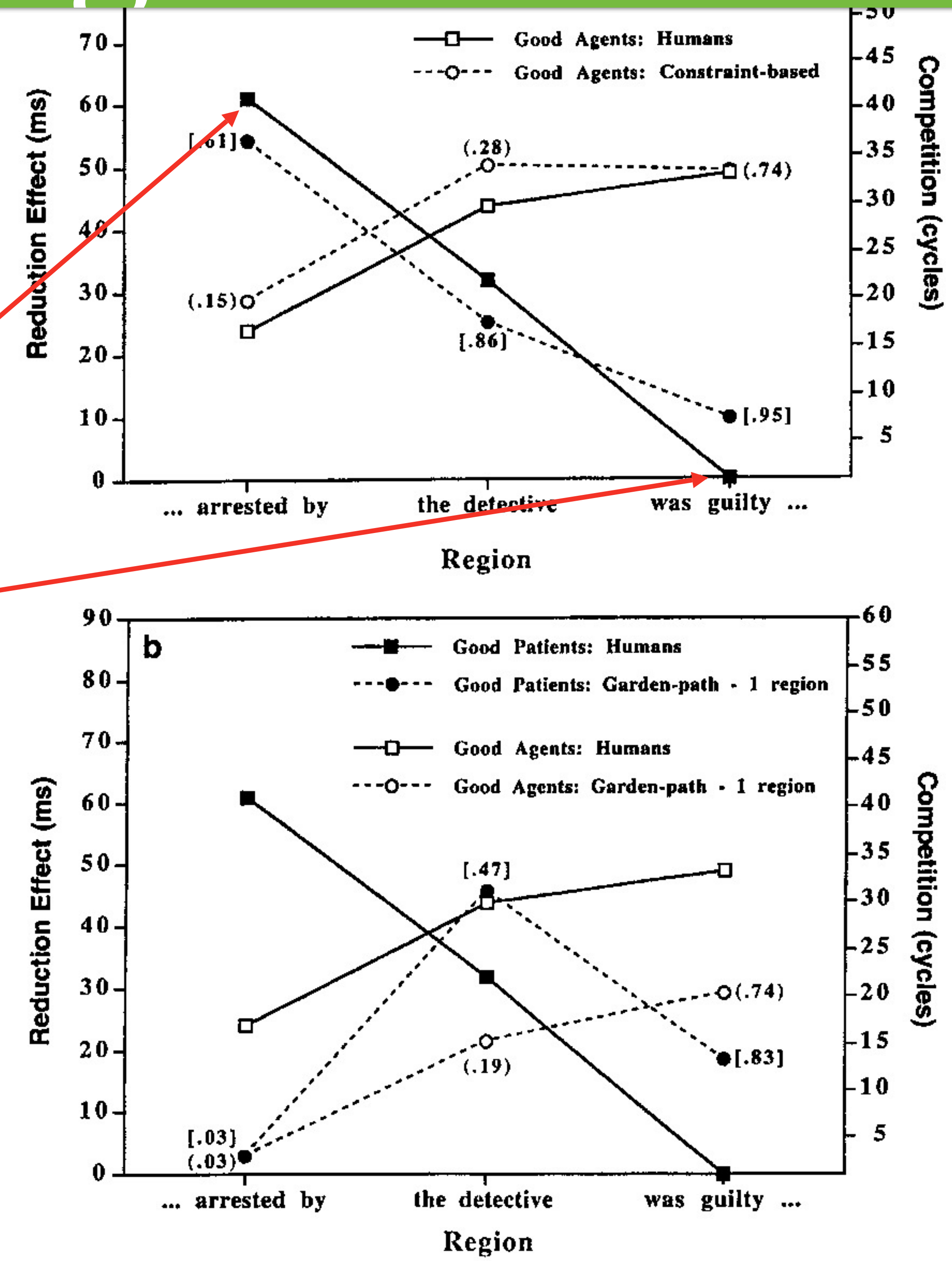


FIG. 6. Simulations of self-paced reading by (a) the constraint-based model, and (b) the one-region delay garden-path model.

3rd Model: Short Delay GP Theory

- The GP-model, has a 1-2 word delay in use of information, what if this delay is reduced?

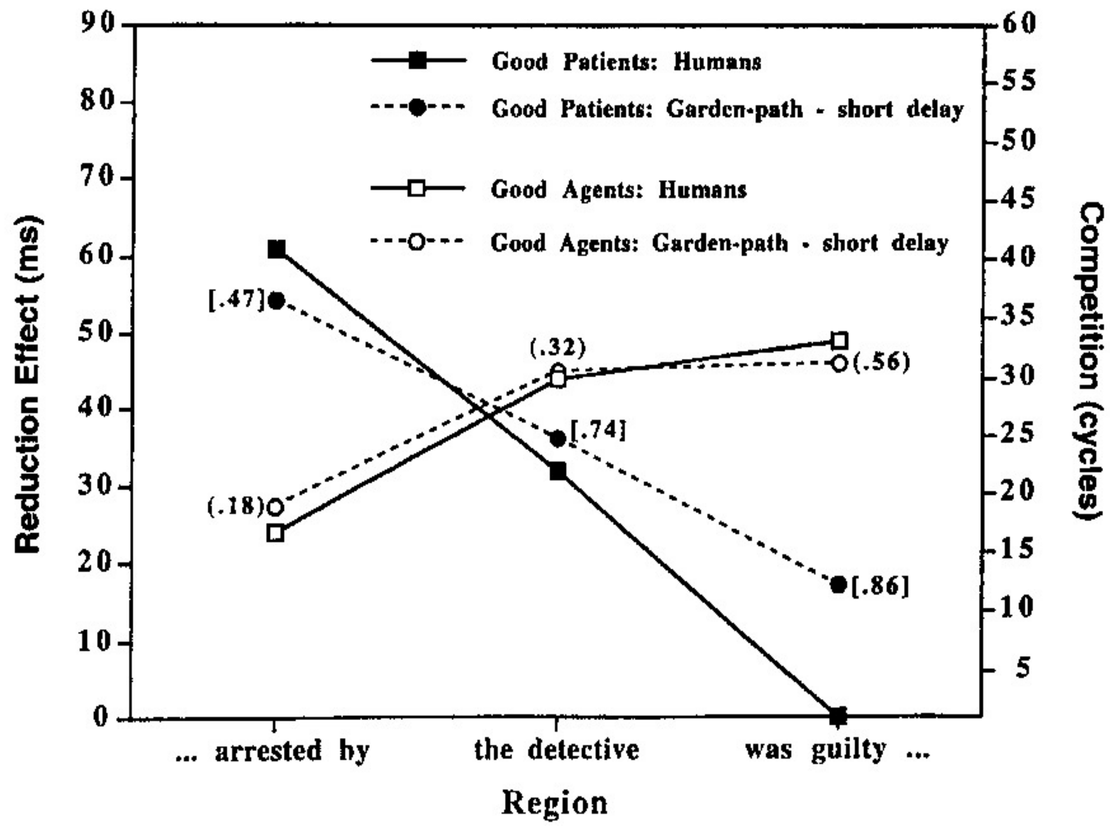


FIG. 7. Garden-path simulations of self-paced reading when constraints other than the main clause and main verb biases were delayed by 4 cycles of competition, or approximately 10–25 ms. Note the large predicted reduction effect at the main verb for the good patients.

Issues and Criticisms

- What constraints to include/exclude:
  - Exclude a constraint that correlates with other constraints, or does not vary among exp. materials
- Constraint integration independent of parsing?
  - What is *really* being modeled?
- Is the implementation of the GP model a fair comparison?
- The model is not easily scalable

Constraint-based vs. Probabilistic

- Similarities:
  - Weighting of different constraints
  - Simultaneous integration of constraints
- Differences:
  - Probabilistic models scale more easily, and typically are not "handcrafted"
  - Constraint-based models directly predict processing difficulty (competition among constraints), whereas probabilistic models do not

Other Recent Approaches

- Narayanan & Jurafsky (1998, submitted):
    - Use bayesian belief networks to combine SCFG like probabilities with other semantic and thematic probabilities.
- Pado (2006):
    - A wide-coverage model of role-assignment and thematic fit (plausibility), which can be integrated with a syntactic parser.

- Expectation-based approaches: Hale (2001), Levy (2006):
  - Based on the probability distribution of all parses, processing difficulty is associated with its *surprisal*: a words conditional probability based on context.
- Stochastic models: Kempen & Vosse (2000), Tabor (2004)
  - Argue for mechanisms which emphasize local coherence, rather than "perfect" incremental parsing.

Readings so far ...

- Matthew Crocker. Mechanisms for Sentence Processing. In: Garrod & Pickering (eds), Language Processing, Psychology Press, London, UK, 1999.
- Dan Jurafsky. Probabilistic Modeling in Psycholinguistics. In Bod et al (eds.). Probabilistic Linguistics. The MIT Press, 2003.
- Ken McRae, Michael Spivey-Knowlton, Michael Tanenhaus. Modeling the Influence of Thematic Fit (and Other Constraints) in On-line Sentence Comprehension. Journal of Memory and Language, 38, 283–312 (1998).