The Advantage of Architectural Assumptions: Benefits for Baseball Batting

Christian Lebiere - Micro Analysis and Design

Robert Gray - Arizona State University

Dario Salvucci - Drexel University

Robert West - Carleton University

10th ACT-R Workshop July 27, 2003

A Weak Model?

"ACT-R is a useful theoretical framework, but as a computational model … a convincing case has yet to be made. ACT-R is best thought of as a "weak" model, in that the assumptions that are quantified do not highly constrain the model in any one testing situation. Rather, experiment-specific assumptions must be added to these for the model to be tested against experimental data. This is in part what gives ACT-R its flexibility to be applied in so many different domains, but it also makes it difficult to trace the model's performance back to those key foundational assumptions. For example, in the base-level learning equation, memory decay is expressed as a power function. Although I have not done the work to test it, I suspect it would not matter much if it were an exponential function. … What matters is that forgetting approximates human performance …. There is a trade-off in model specificity and generality. The strengths of ACT-R are in the latter. It suffers in the former. Precision is what models must achieve in simulating human performance. It is unclear whether the main properties of the model are actually responsible for its exceptional data-fitting abilities.

In some sense ACT-R is not a model to be disproved or falsified in a data-fitting exercise. It is a framework within which to study information processing. To reduce it to a computational model that fits data or simulates phenomena involves many hazards and such data should be interpreted very cautiously, even played down, for they can be misleading if one reads too much into them."

- *An anonymous reviewer*

Let’s Review the Case

- Typical models do require additional assumptions
- Some model parameters are adjusted to fit the data
- Some degree of that is to be expected given:
  - Lack of cognitive determinism
  - Individual differences in abilities, knowledge & strategies
- Cognitive architectures still do provide very significant qualitative and quantitative constraints
- Competing frameworks are even considerably looser
- But we need to resist the urge to over-engineer and over-fit and be clear what our constraints are

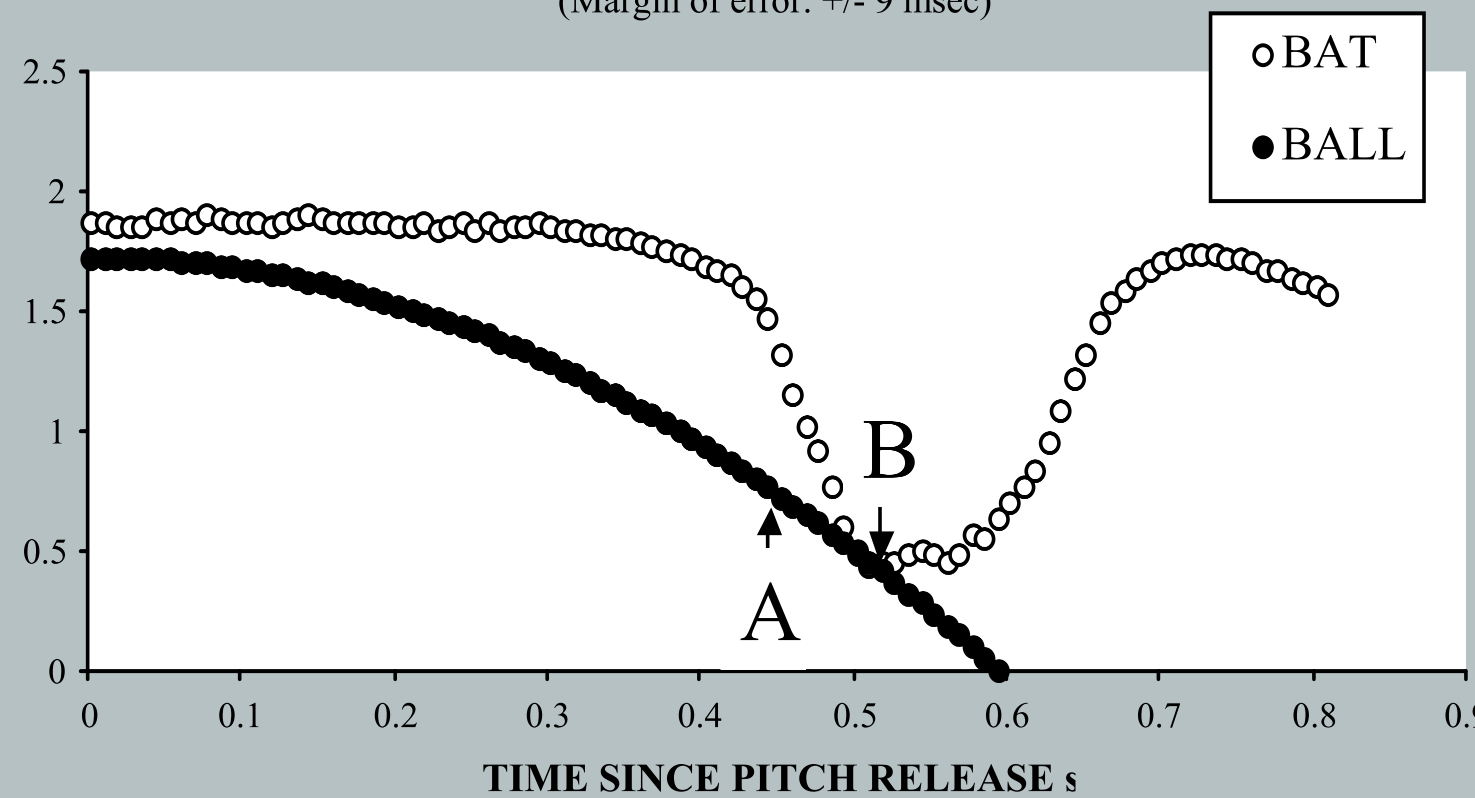
Baseball is Game of Expectations

- Everybody (in Americas & Japan) knows baseball
- Perceptual and mechanistic aspects have been studied but not (much) the cognitive aspects
- New domain to test ACT-R predictiveness
- Pitcher-batter duel is basically a game of expectations (paper-rock-scissors with ball and bat)
- Virtual reality experiment (Gray, 2001)
  - Ball projected on monitor
  - Tip of bat outfitted with position tracker
  - Subjects: 6 experienced college players

Defining Batting Performance

A: ball crosses the plate; B: minimum bat height; MTE=B-A

(Margin of error: +/- 9 msec)



Experiment 1: Unstructured

- Slow pitches (70 +/- 1.5 mph)
- Fast pitches (85 +/- 1.5 mph)
- Random sequence of 10 blocks of 25 pitches
- All pitches are strikes down the center of the plate
- Progress bar to indicate pitch release
- Visual feedback indicating trajectory and speed of ball if contact was made
- Large red “X” indicates strike if no contact made

Stupidity is an Asset?

… Hatteberg considers why everyone doesn't prepare for Jamie Moyer as he does - by watching tape, imagining what will happen, deciding what to look for, deciding what he will never swing at. "Some of the guys who are the best are the dumbest," he says. "I don't mean dumbest. I mean they don't have a thought. No system."

Stupidity is an asset?

"Absolutely. Guys can't set you up. You have no pattern. You can't even remember your last at-bat."

- *Michael Lewis, Moneyball, p. 185*

Between Smart and Stupid

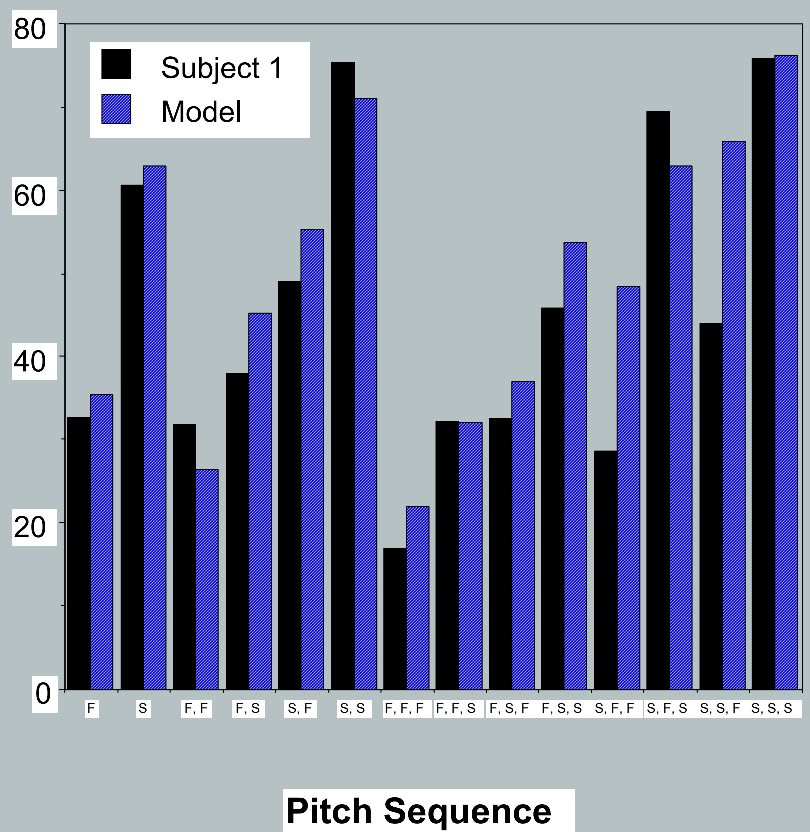
- Generating expectations is a natural human trait (sequence learning, PRS)
- It is essential in baseball (no time for preparation of motor command)
- On what basis, i.e. context, to anticipate? *Default assumption:* no context
- This maps to 1 simple production for activation-based memory retrieval
- Memory of 2 chunks, fast and slow, with activation determined by history

$$A_i = \ln \sum_j t_j^{-d} + N(0,s)$$ ***Activation Equation***

- Decay rate d and noise magnitude s *fixed* at 0.5 and 0.25 by prior models
- Need assumption to go from cognitive expectation to action timing
- *Default assumption*: no perceptual ability to adjust but perfect execution
- This is simply a linear mapping from probability scale to temporal scale

$$MTE = \frac{dist}{v_g} - \frac{dist}{v_a}$$ ***MTE Equation***

Model Meets Subject 1

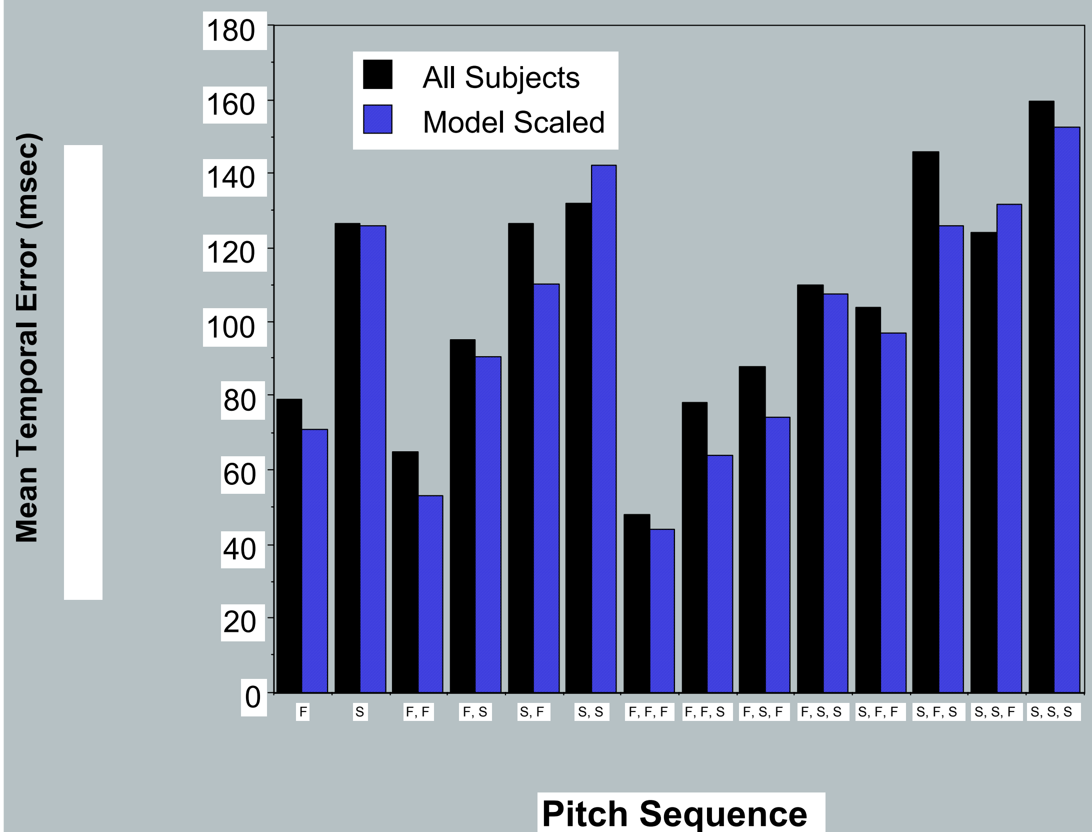


- MTE for fast pitches only (similar for slow)
- Effect 1: smaller error following F than S
- Effect 2: effect increases with history length
- Effect 3: order matters, e.g. F,S vs. S,F
- Qualitative and quantitative predictions

The Structure of the Environment

- The (base-level) activation equation reflects the structure of the environment (Anderson & Schooler, 1991) in the form of frequency and recency effects
- But this artificial domain violates that structure (it has none)
- However, both models and humans have internalized the general distribution and keep acting as if it still applies!
- The same qualitative effects are present in all subjects but different physical abilities produce different MTE levels
- To match that average level, we introduce *a parameter factor of 2* in the MTE equation to scale overall performance
- Despite the parameterization, it is still worth doing the quantitative fit because the effects are more reliable

Model Meets the Population



- Same effects as in Subject 1
- Effect 3 is more reliable (e.g. F,F,S vs F,S,F vs S, F, F)
- Frequency wins over recency (e.g. F,S,S vs S,F,F)
- Computational model is essential

Experiment 2: Real At-Bats

- From unstructured series of pitches to meaningful at-bats
- Horizontal location of pitches is varied to generate balls (12 +/- 1 inch on each side of the plate) & strikes (over the plate)
- Pitch location is randomly determined on each trial
- A swing and miss is considered a strike
- Standard rules: 4 balls is a walk, 3 strikes is a strike out
- Pitch speed still varies randomly but is now count-dependent
- On “ahead” counts (0-2 and 1-2), slow pitches were selected with 0.65 probability; on “behind” counts (2-0, 3-0, 3-1), fast pitches were selected with 0.65 probability

The Difference is Enormous

Each plate appearance they think of as a miniature game in itself, in which the odds shift constantly. … A first-pitch strike, for instance, lowered a hitter's batting average by about seventy-five points, and a first-pitch ball raised them by about as much. But it wasn't the first pitch that held the most drama for the cognoscenti; it was the third. "The difference between 1-2 and 2-1 in terms of expected outcome is just enormous," says Paul. "It's the largest variance of expected outcomes of any one pitch. On 2-1 most average major league hitters become all-stars, yet on 1-2 they become anemic nine-hole hitters."

\- *Michael Lewis, Moneyball, p. 147*

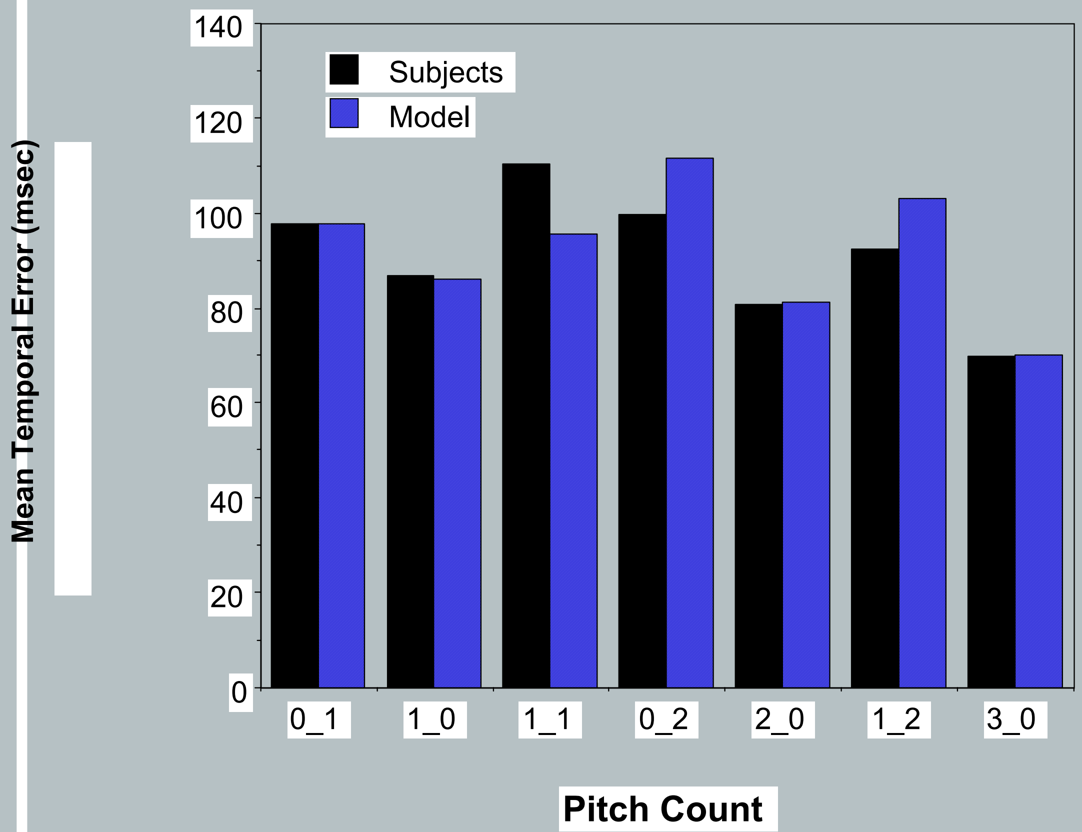
The Count is the Context

- The context of each pitch needs to be represented explicitly
- The *default assumption* is to represent the ball-strike count
- Each pitch is now recorded as a ball-strike-speed triplet chunk
- There are now 4x3x2=24 chunks instead of simply 2 chunks
- The same activation learning and MTE equations apply
- The retrieval production is similar but also matches context
- Since balls and strikes are numbers, partial matching applies:

$$M_i = A_i + \sum_{context} MP \cdot Sim_{vd}$$ ***Partial Matching Equation***

- Similarities Sim_{vd} taken from other models; *MP* at default 1.5
- Rest of model & parameters left unchanged (incl. MTE factor)

Pitch Count Matters

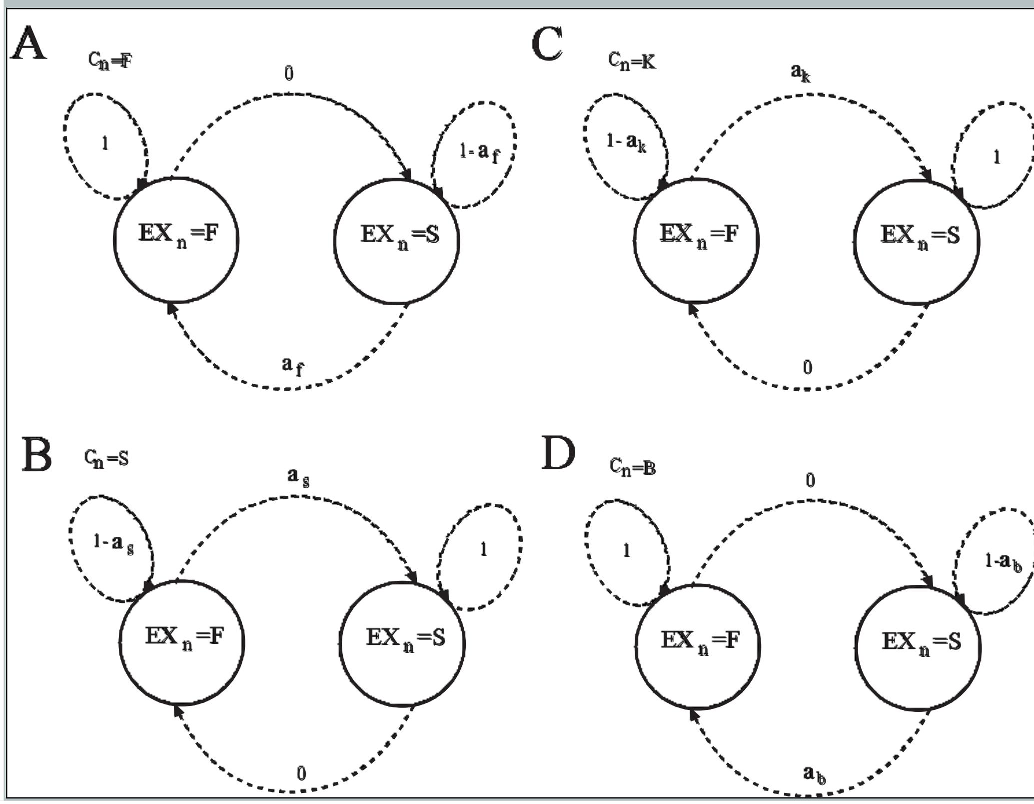


- Model sensitive to pitch distribution (ahead vs behind)
- Effect generalizes to nearby "even" counts (e.g. 0-1 vs 1-0)
- Effect also reflects pitch count depth (e.g 1-0 vs 2-0 vs 3-0)
- Model overreacts slightly to behind counts (0-2 and 1-2)?

Markov Model (Gray, 2001)

Basic Markov assumption:

Current state determines future



- 2 states: expecting fast or slow pitch
- Probabilities of switching state a_s, a_f and temporal errors when expecting fast and slow pitch T_f, T_s need to be estimated
- 2 more transition rules and associated parameters (a_k, a_b) to handle pitch count

Markov vs. ACT-R

- State representation
  - Markov has discrete states that represent decisions
  - ACT-R has graded states that reflect the state of memory
- Transition probabilities
  - Markov needs to estimate state transition probabilities
  - ACT-R predicts state change based on theory of memory
- Pitch count
  - Markov has to adopt additional rules and parameters
  - ACT-R generalizes using previously constrained methods

The Short Story

- Cognitive architectures are not a panacea for modeling
- Modeling still requires specifying representation and processing assumptions and maybe even fitting parameters
- However, constraints from the architecture and from other models buy much predictiveness (need to make that clear)
- Frameworks like Markov modeling (and neural networks) depend considerably more on parameter estimation
- ACT-R also makes detailed trial-by-trial predictions
- This raises possible applications to training, games, etc
- Architectural implication: no rehearsal on retrieval!