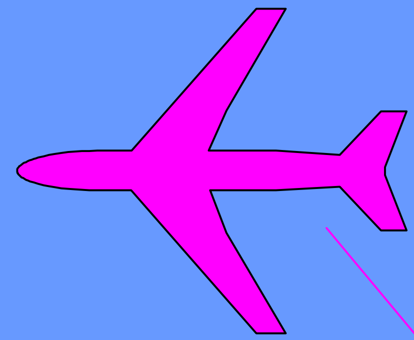
Quantitative Explorations of Category Learning with Symbolic Concept Acquisition

Robert Wray
Soar Technology

Ron Chong
George Mason University

Soar 23
27 Jun 2003



OLY 995
20 S 3

- Subjects instructed to learn to accept/reject from three two-valued features displayed to screen

Fuel %	Size	Turbulence
20 40	S L	1 3

- Learning features only displayed during learning trials
- 8 unique instances (4 positive/4 negative)

Symbolic Concept Acquisition: SCA

- Exemplar model
- Implemented within Soar architecture
  - Previously shown to qualitatively match Shepard et al problem type learning curves (Miller, 1993)
  - Critical constraint: Cumulation of results within an architectural approach mandates SCA (as extant model)
- Performs general-to-specific search over concept space when learning
- Performs specific-to-general search over feature space when making predictions
- This implementation: wholly symbolic (same learning results with Soar 8 distribution)

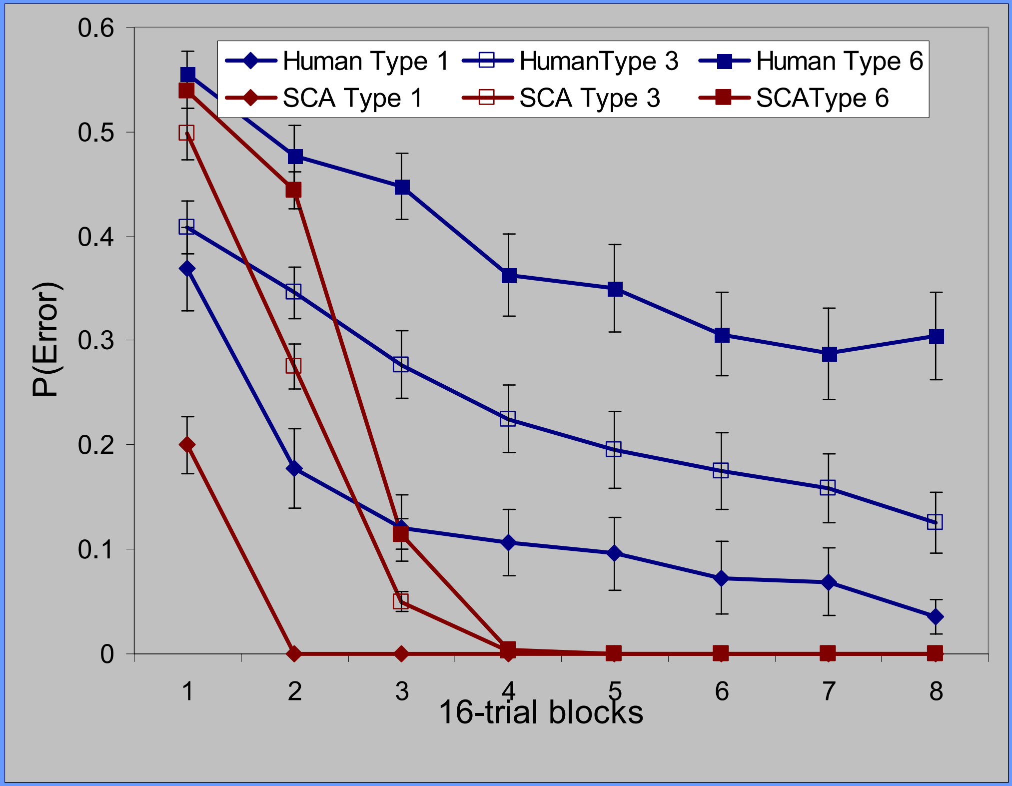
SCA & Human Behavior

Miller & Laird (*Cognitive Science,* 1996):

- Sensitive to complexity of problem type
- Generates responses in varying time
  - Practice effect
  - Typicality effect
- Learns linearly and non-linearly separable concepts at roughly equivalent rates
- Responds faster & with less error to basic level categories (basic level superiority)
- Improves accuracy with training on same data set (practice effect)

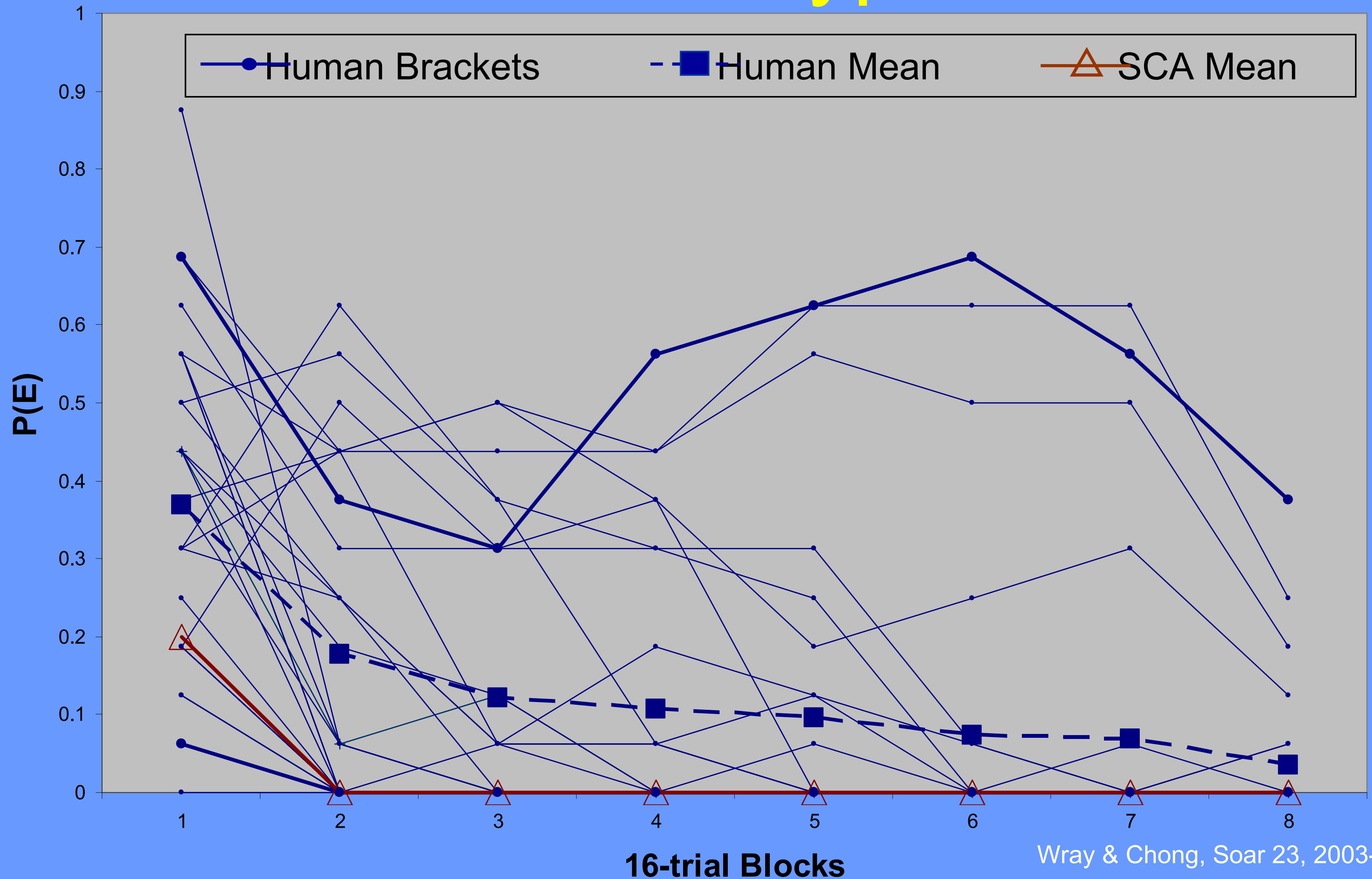
Initial Learning Results (Soar 22)

- Direct "out-of-box" application of SCA
  - Simple integration with ATC performance model
  - 3 features
  - 30 runs/problem
- \- Poor aggregate fit
- \- Poor qualitative fit
- \- Little variation from one model run to the next

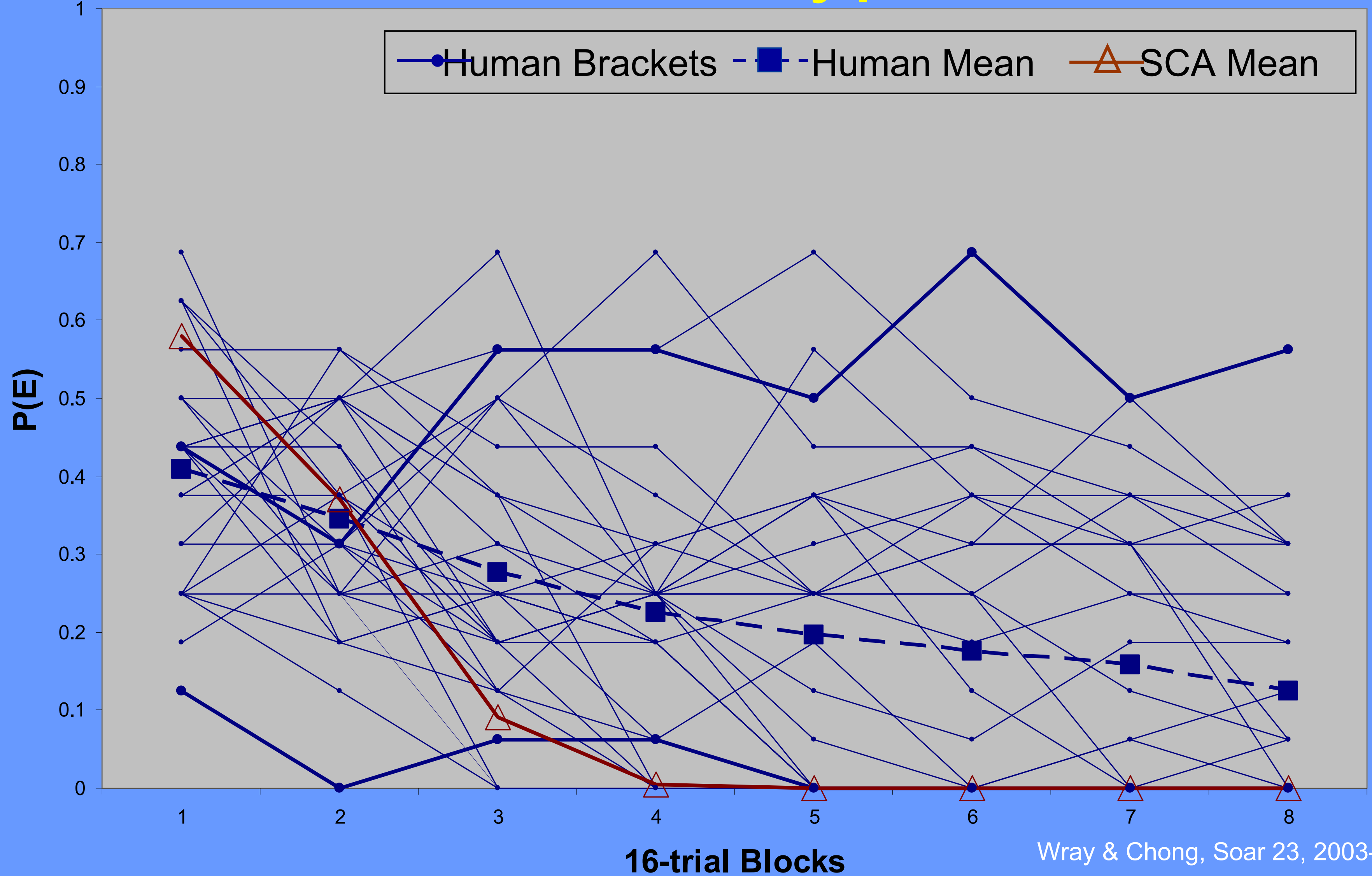


- \+ Qualitative match to human learning wrt problem type
- \+ A few individual SCA runs matched human P(Error) measures exactly

Problem Type 1

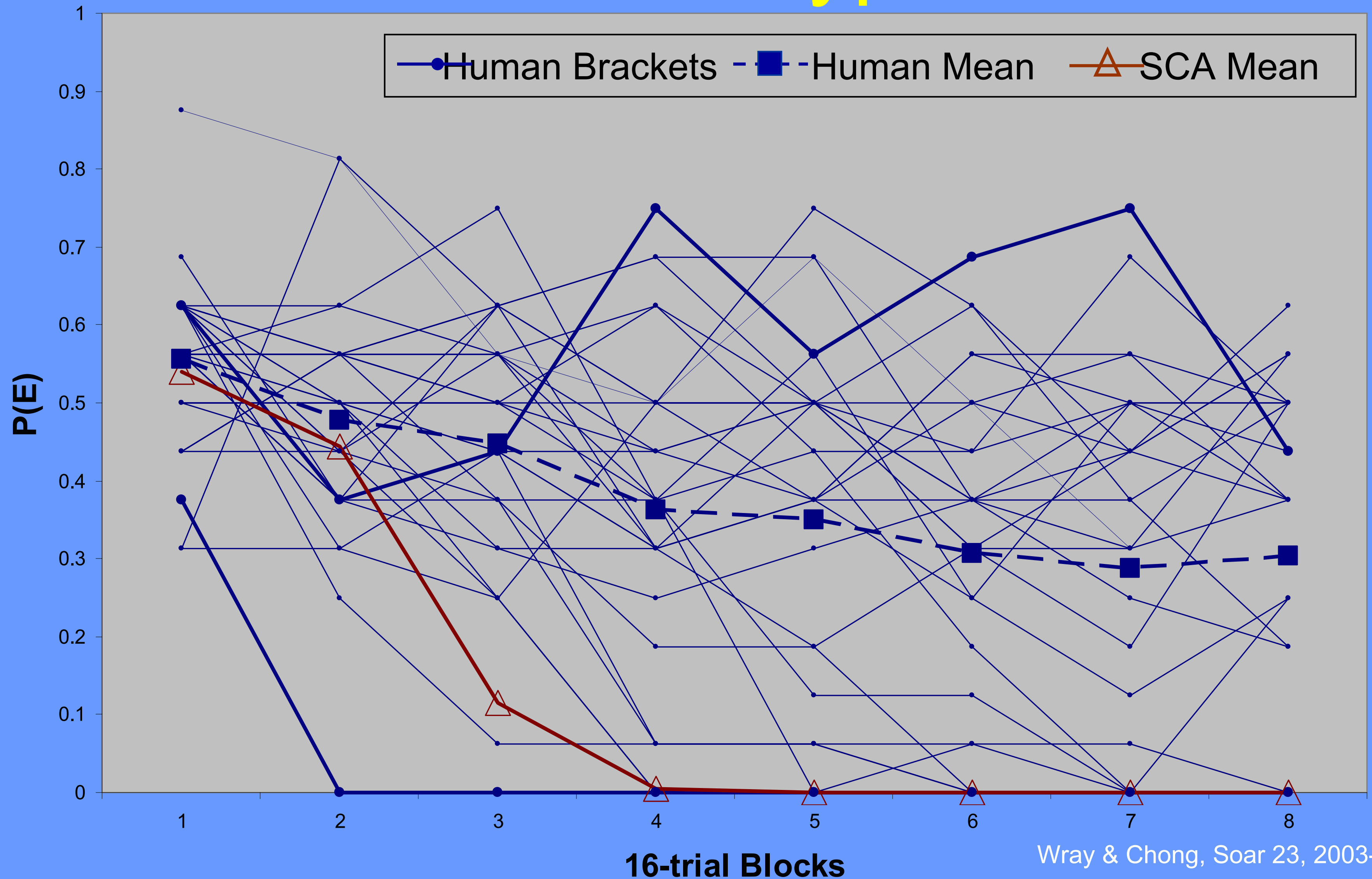


Problem Type 3



Wray & Chong, Soar 23, 2003-06-27

Problem Type 6



Lessons from the data

- Individual subjects showed remarkably different learning trajectories
  - Type 1: 4 learners above 20% error block8
  - Type 6: 1 learner P(E)=0 by second trial
  - ~1/3 of learners still at chance block8
- No individual subject's learning matched the aggregate curves
- Large differences between fastest and slowest human learners; large variation in human learning

FAST/LOW BRACKET = MIN $(\text{Total\_Error}_i)_{i=1 \ldots 30 \text{ subjects}}$

SLOW/HIGH BRACKET = MAX$(\text{Total_Error}_i)_{i=1 \ldots 30 \text{ subjects}}$

$$\text{Total\_Error} = \Sigma(P(\text{error})_b)_{b=2 \ldots 8 \text{ blocks}}$$

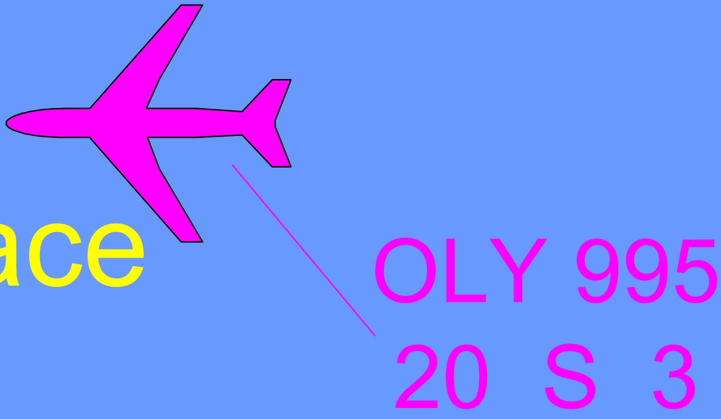
Reconsidering Initial Results

- Initial SCA model
  - Model of individual
    - Knowledge constant over blocks
    - Features constant over blocks
    - Little variation from one run to next (within type)
  - SCA mean within fastest/slowest brackets for all three types (and little variation)
    - Individual runs also within brackets
  - Some SCA individuals match some individual subjects qualitatively and quantitatively

Listening to the Architecture

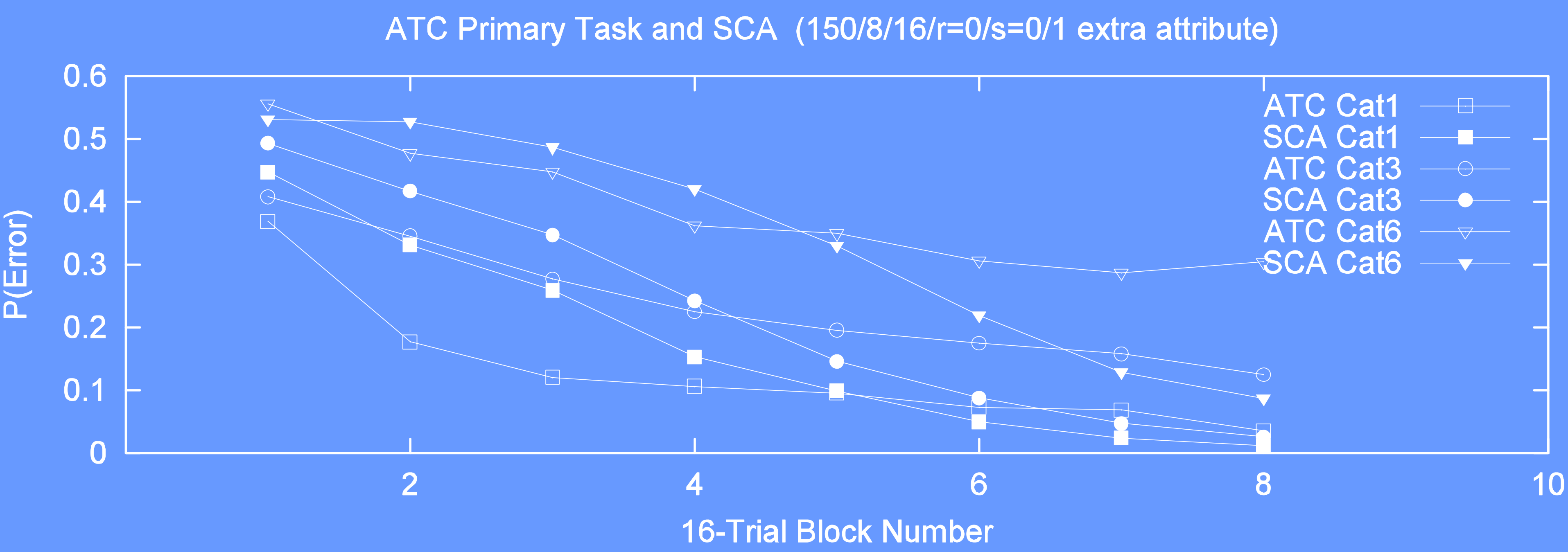
- Assume Soar & SCA are fixed
  - What variables can impact the results?
- SCA learning rate is sensitive to
  - Number of features
    - Size of the concept space is combinatorial in the number of features
    - Initial model assumed minimum 3 features
  - Abstraction order
    - The more systematic the abstraction, the less search of the concept space is needed before definition
      (Depth first search vs. Breadth first search)
    - Initial model assumed one strategy (relevant feature)

Feature Space

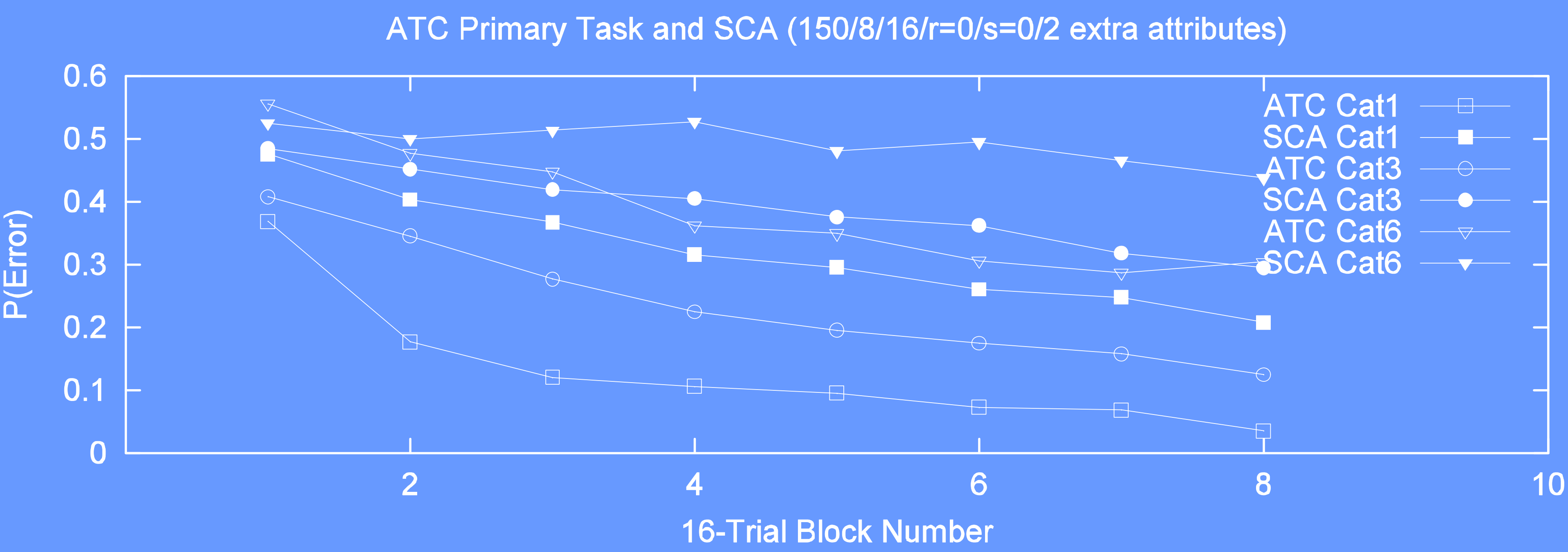


- 18% of human subjects reported being influenced by extraneous features during learning (even though instructed to ignore all but fuel, turbulence, & size):
  - *It might be the direction, the area it was in, location of nearby planes,…*
  - *I noticed a pattern between numbers & letters at the bottom of the plane*
  - *how close they were to the intersection with other planes …*
  - *The planes that had all the lowest descriptions together or the highest descriptions all together were accepted*
  - *a small plane experiencing heavy turbulence and light on fuel would definitely need make some adjustments*

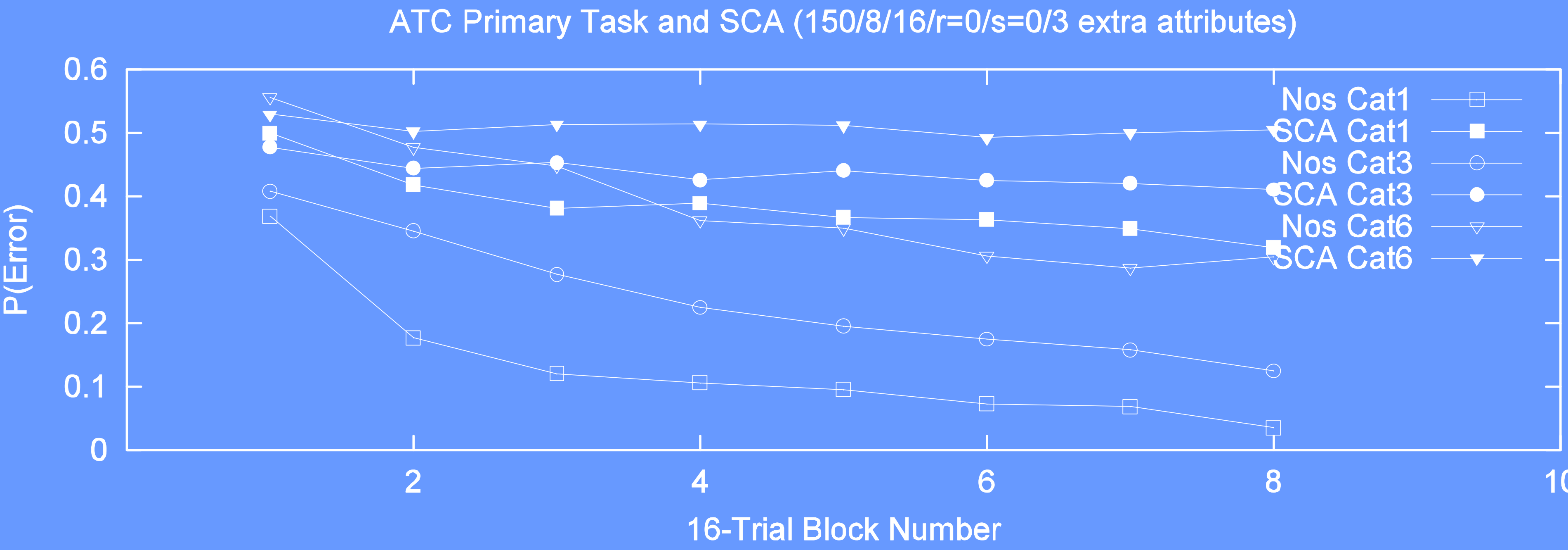
SCA Sensitivity to Features
4 Features



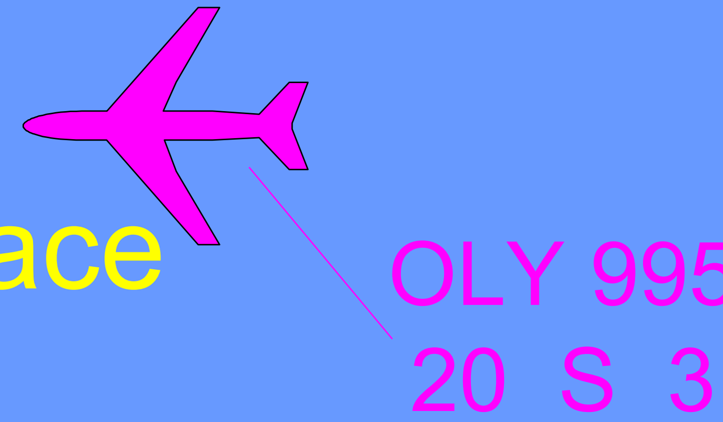
SCA Sensitivity to Features
5 Features



SCA Sensitivity to Features
6 Features

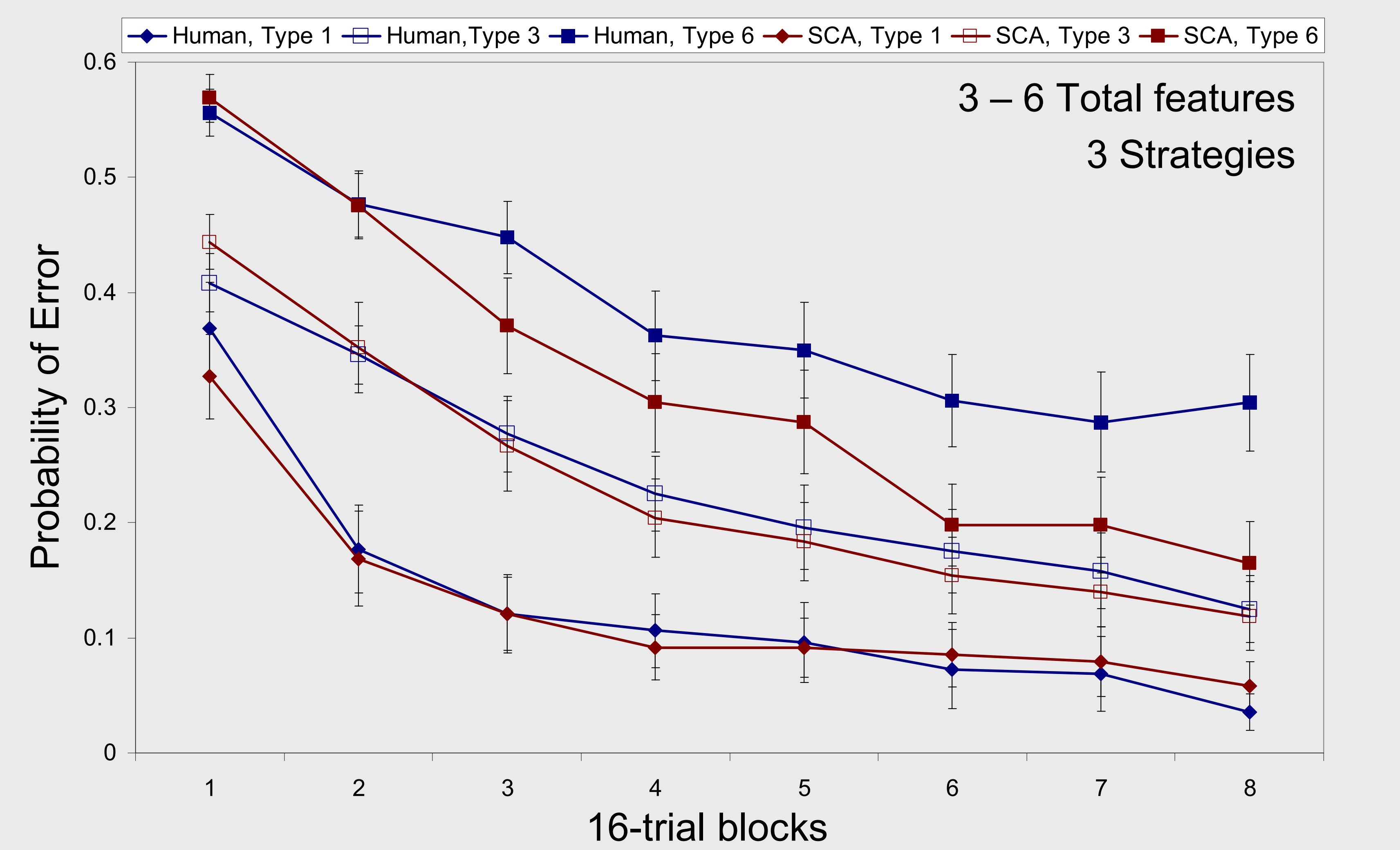


Feature Space



- 18% of human subjects reported being influenced by extraneous features during learning (even though instructed to ignore all but fuel, turbulence, & size):
  - *It might be the direction, the area it was in, location of nearby planes,…*
  - *I noticed a pattern between numbers & letters at the bottom of the plane*
  - *how close they were to the intersection with other planes …*
  - *The planes that had all the lowest descriptions together or the highest descriptions all together were accepted*
  - *a small plane experiencing heavy turbulence and light on fuel would definitely need make some adjustments*
- Estimated feature space via sensitivity analysis of SCA
  - 3 additional random binary features:
    - Nearly flat learning over 8 blocks (3 or less add'l features necessary)
  - Result: uniform distribution of 0-3 additional features in addition to 3 instructed features

Fitting SCA to Aggregate Data



Resulting Predictions

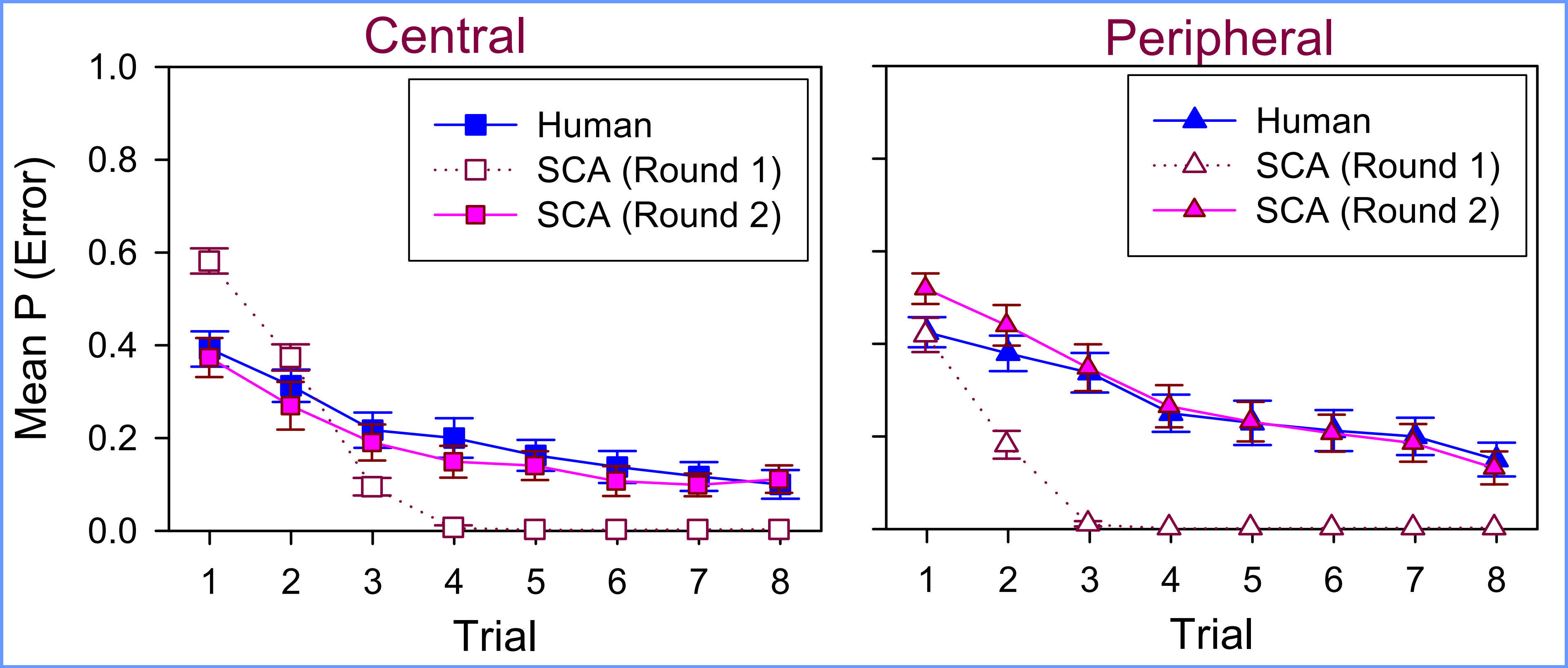
Predictions after fitting to aggregate (no additional modifications):

- Good quantitative fits for both central & peripheral Type 3
- Poor quantitative but good qualitative match for prediction response time
  - Need more refined task switching knowledge? (current model switches immediately)
- Comparable number of perfect learners for Types 1 & 3

	Human	**SCA**	**Initial SCA**
Type 1	**24**	**23**	**30**
Type 3	**15**	**15**	**30**
Type 6	**6**	**12**	**30**

- Increased number of exact correspondences between SCA runs and individual human data

Central vs. Peripheral Type 3



Limitations & Future Work

- Subject feature space is dynamic/SCA's is static
- SCA learns something every trial
- Deliberate abstraction
- Exemplar recall insensitive to time & interference
  - ⇨ SCA-inspired model based on attention & episodic indexing (Altmann & John, 1999)
- Exemplar & hypothesis testing hybrids
  - ⇨ RULEX/SCA hybrid

Conclusions

- \+ SCA able to fit aggregate data reasonably well
  - Good predictions post fitting; fitting tied to observed data
  - Match to individual as well as aggregate data (Estes, 2002)
  - Model reuse (including code reuse!)
- \+ Architectural constraint led to unique views of the data
  - SCA (+ UTC) effectively predicted additional features played a role
  - Instances of model are instances of simulated individuals
- \- No single model/approach likely to account for human data
  - \- Individual data is highly variable
  - \- Some subjects learn measurably little
- \- Need *a priori* feature identification
  - \- Future experiments need to control or estimate features

SCA (Soar 8) available: http://www.speakeasy.org/~wrayre/soar/sca/html/
(Fully documented with SoarDocs)

Book on AMBR project in preparation