Evaluating “cost of insight” with rational inattention model

Pam Osborn Popp, Todd Gureckis, Daniel Bartels, Benjamin Newell

Overview

- *Background*: Caplin et al. (2020) present theory of rational inattention that can recover costs of attention from behavioral data
- *Goal*: Design a dataset to apply this theory to a task whose performance relies on higher order cognition
- *Approach*: Present category learning task from Shepard et al. (1961) in separate learning and test phases, varying incentive and problem type **between** subjects

Evaluating “cost of insight” with rational inattention model

Pam Osborn Popp, Todd Gureckis, Daniel Bartels, Benjamin Newell

Questions

- Can cognitive processes like “insight” be modulated by incentive?
- Do the costs of such learning competitively adapt to incentive analogously to basic microeconomics?
- What can we take away about how instructors should design cost functions in their classrooms?

Classic category learning task

- Six types of classification problems for the same 8 stimuli
- Performance decreases as type increases (approximately "difficulty")
- Stimuli are robust and results have been replicated many times

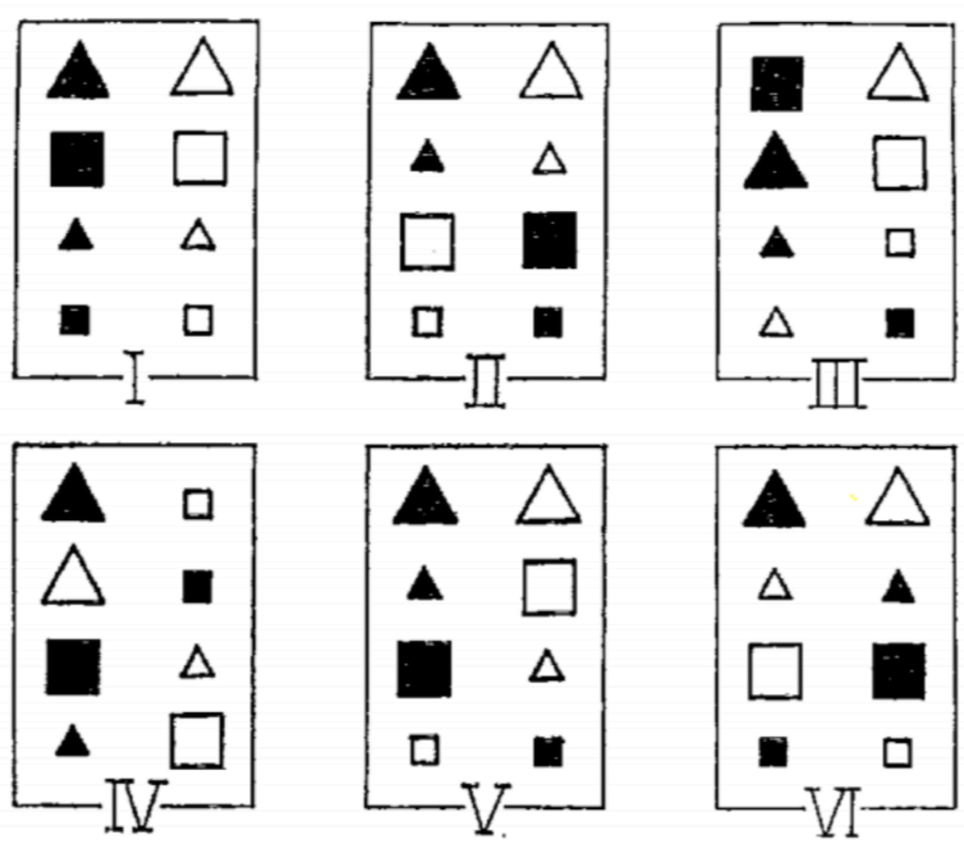


FIG. 1. Six different classifications of the same set of eight stimuli. (Within each box the four stimuli on the left belong in one class and the four stimuli on the right in the other class.)

Learning and Memorization of Classifications
Shepard, Hovland, and Jenkins
1961

Classic category learning task

- Six types of classification problems for the same 8 stimuli
- Performance decreases as type increases (approximately “difficulty”)
- Stimuli are robust and results have been replicated many times

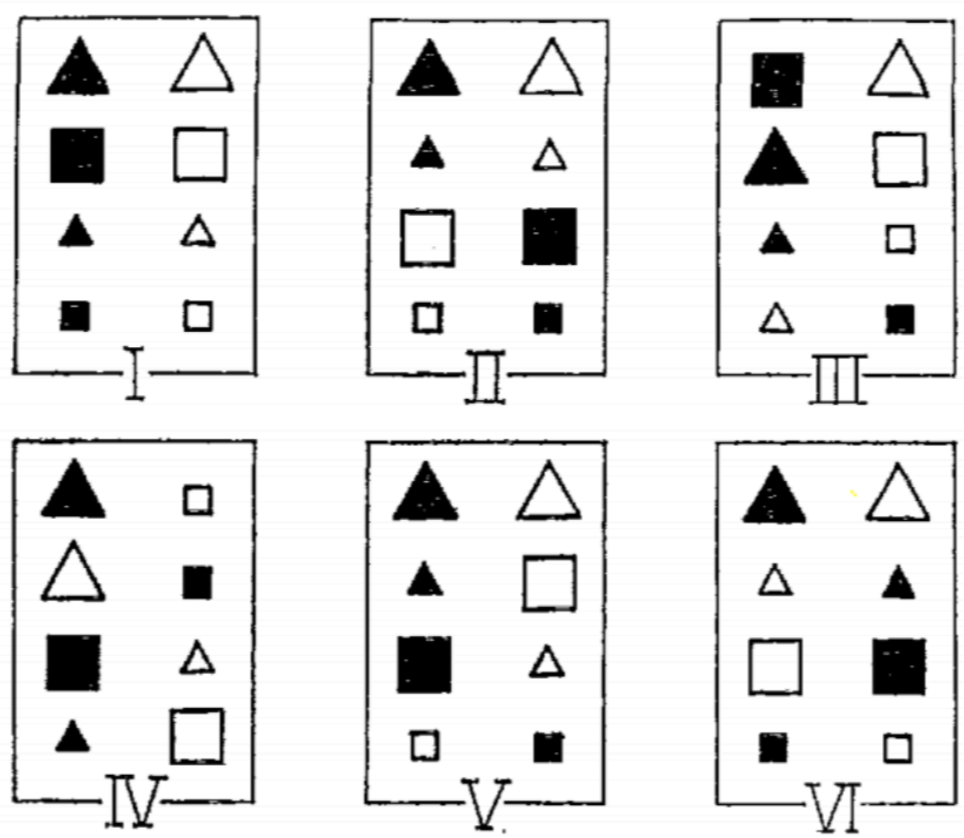
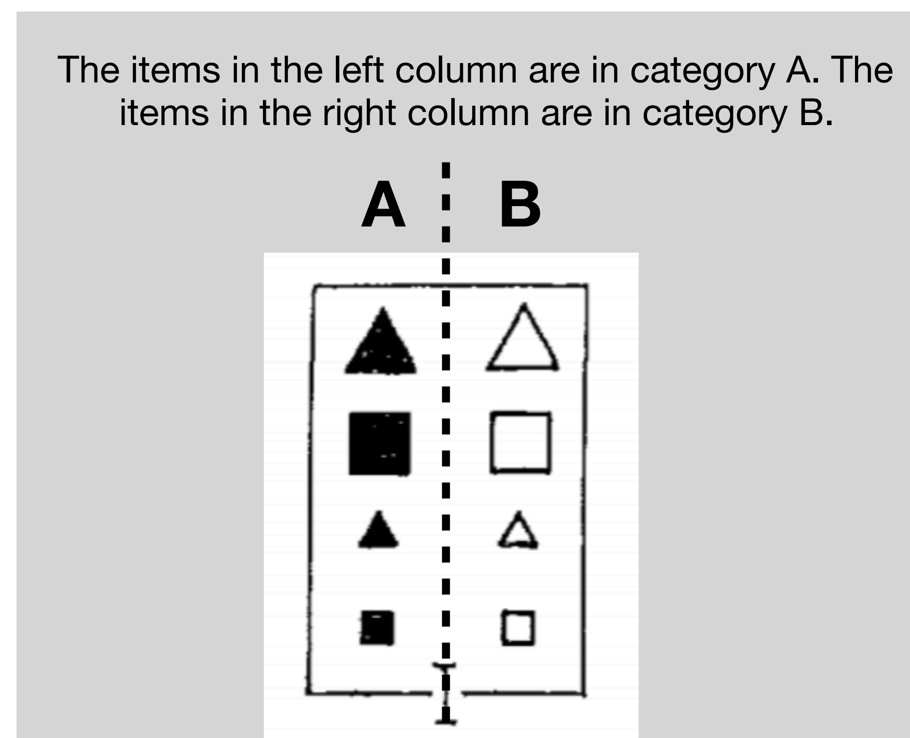
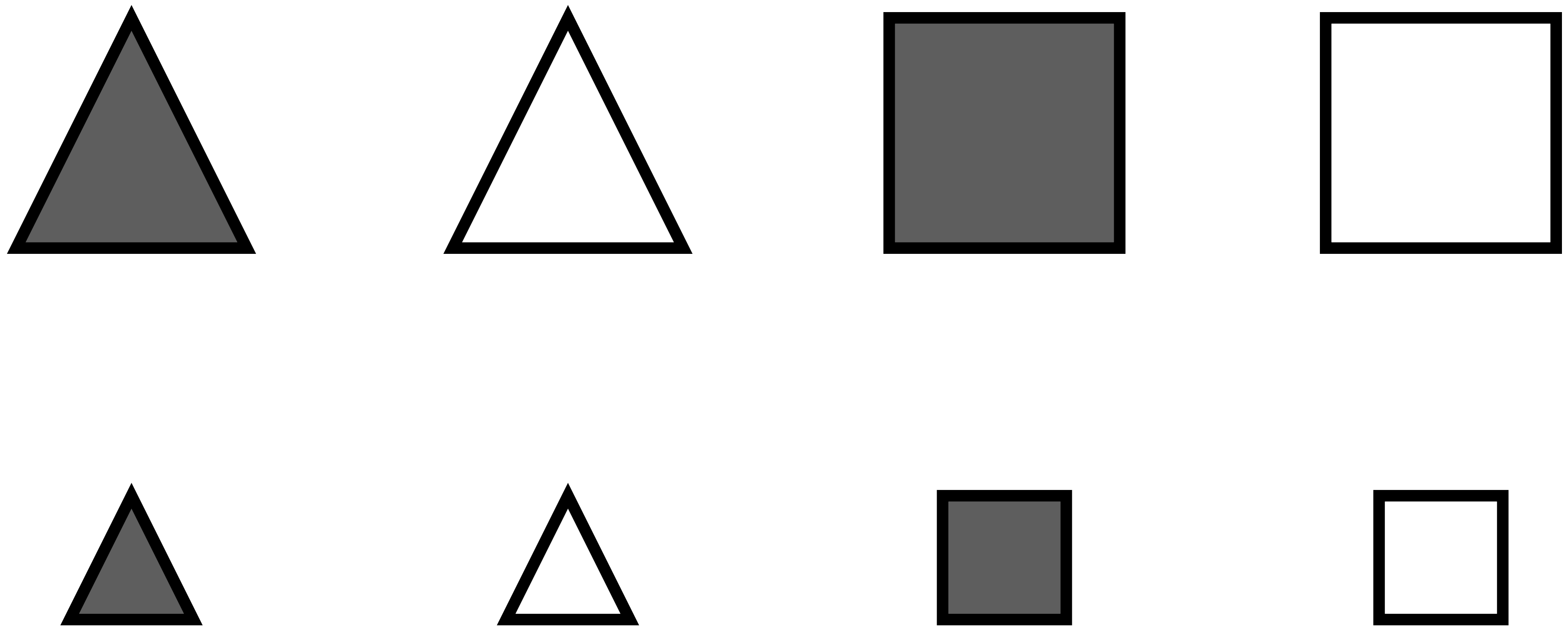


FIG. 1. Six different classifications of the same set of eight stimuli. (Within each box the four stimuli on the left belong in one class and the four stimuli on the right in the other class.)

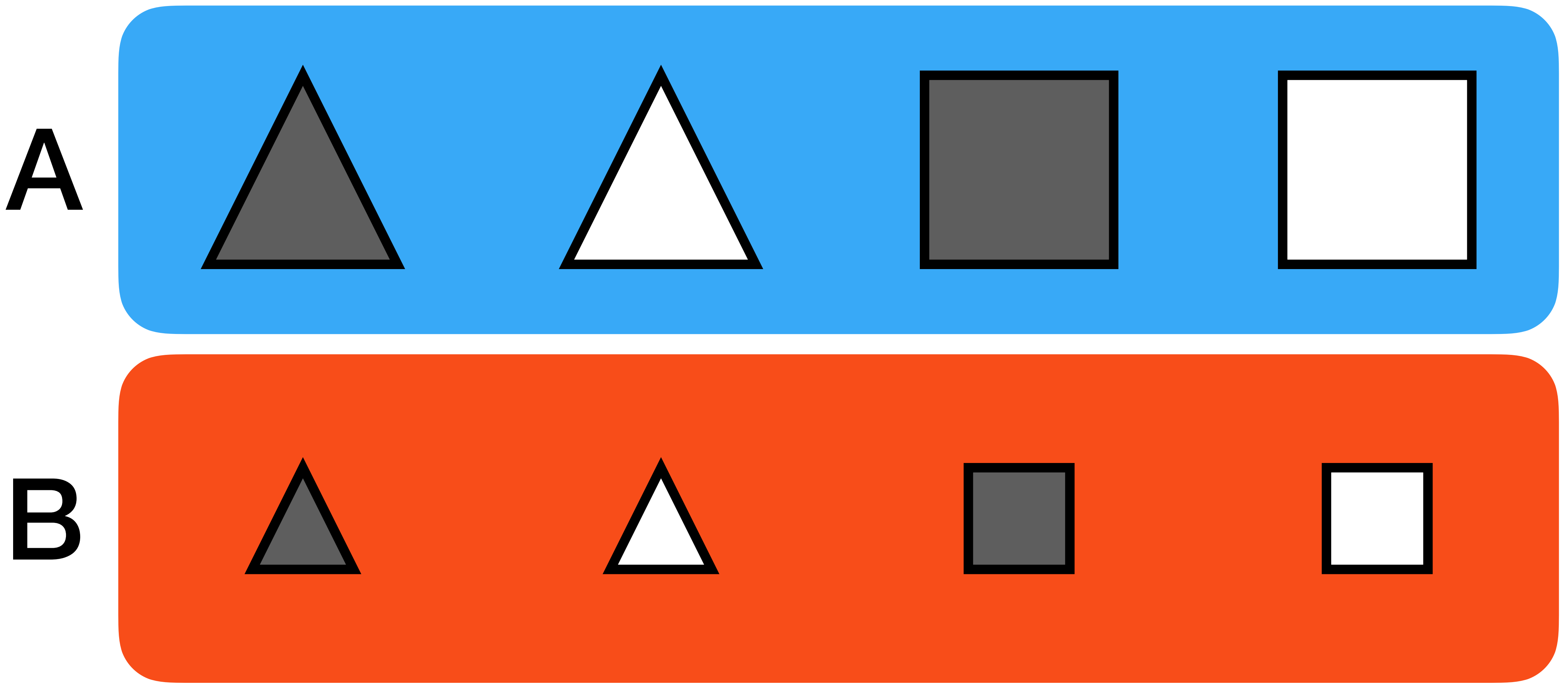


Stimuli



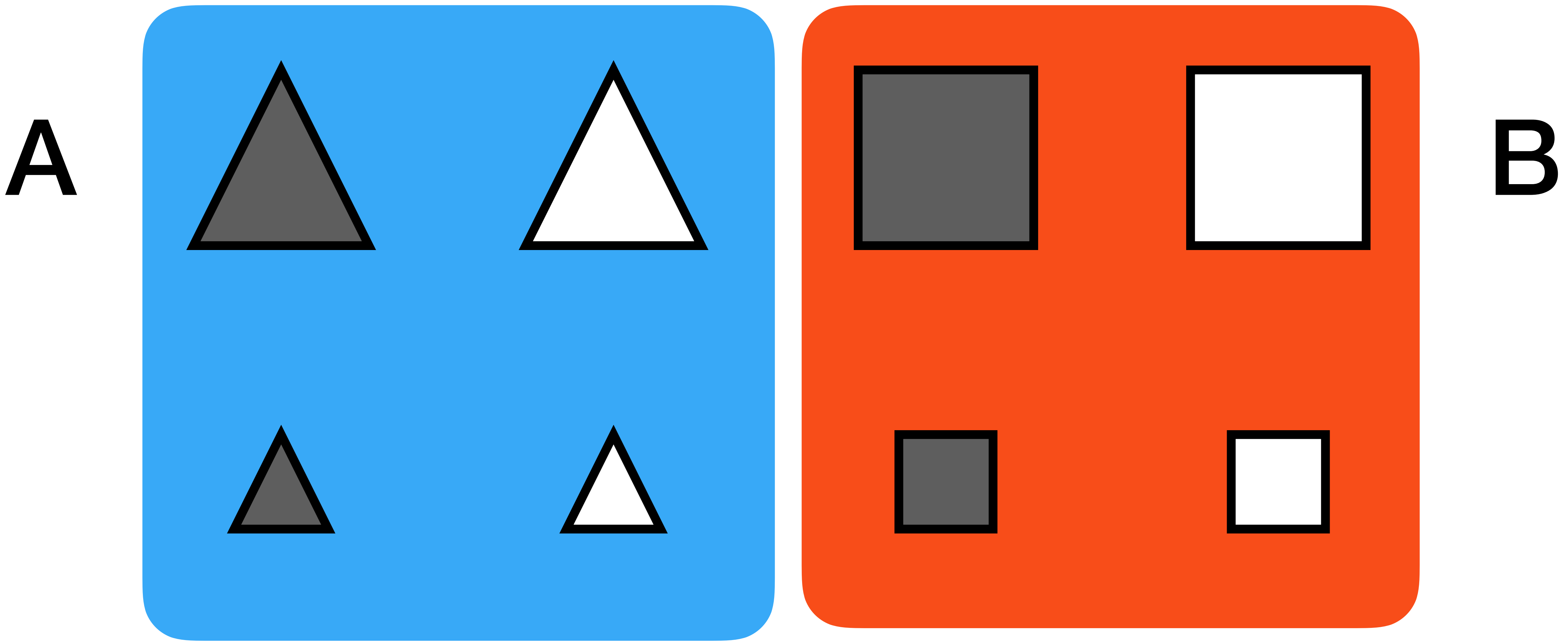
Type I Categorization

Grouped along single feature dimension
Large vs. Small



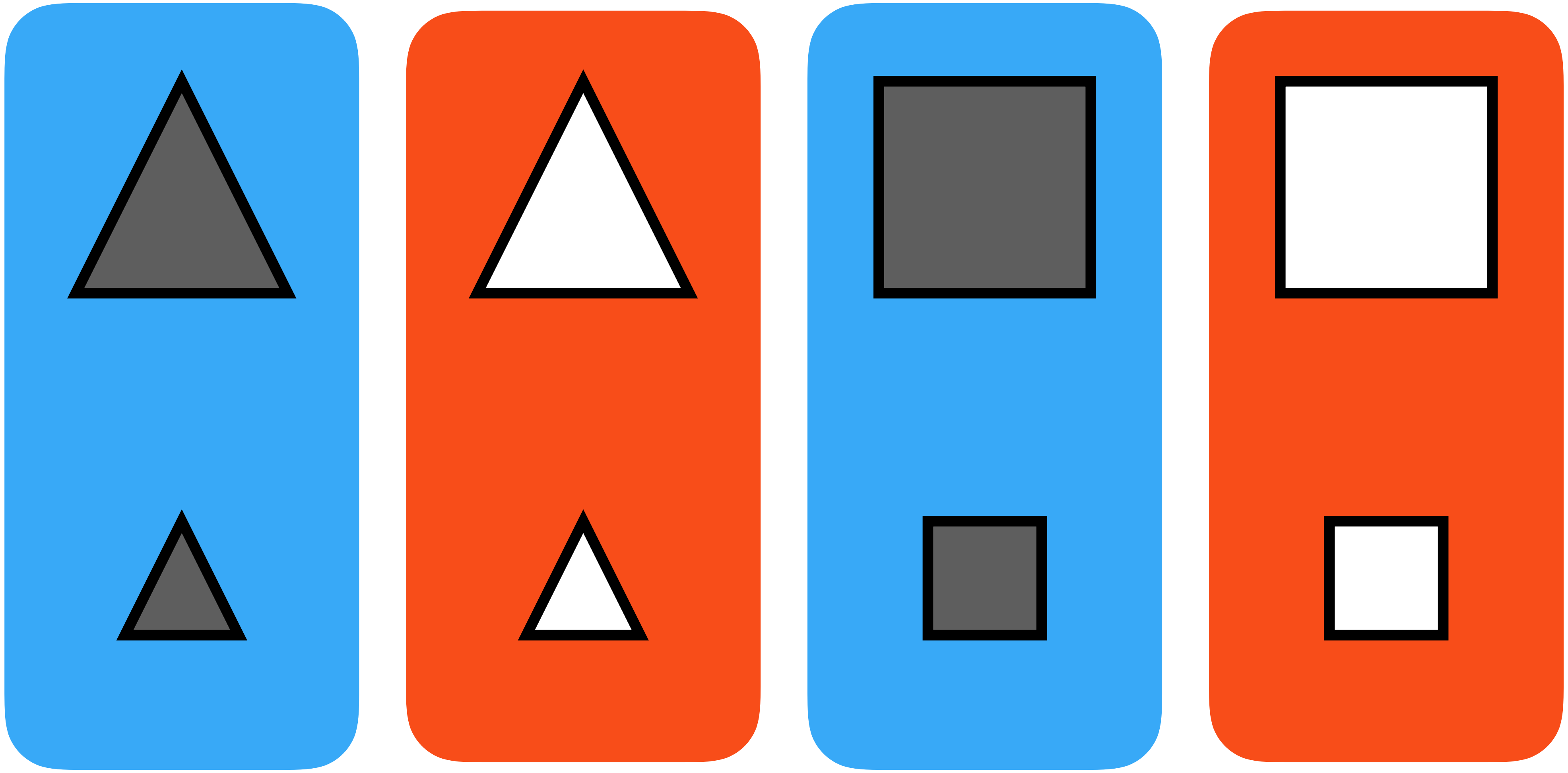
Type I Categorization

Grouped along single feature dimension
Triangle vs. Square



Type I Categorization

Grouped along single feature dimension
Black vs. White

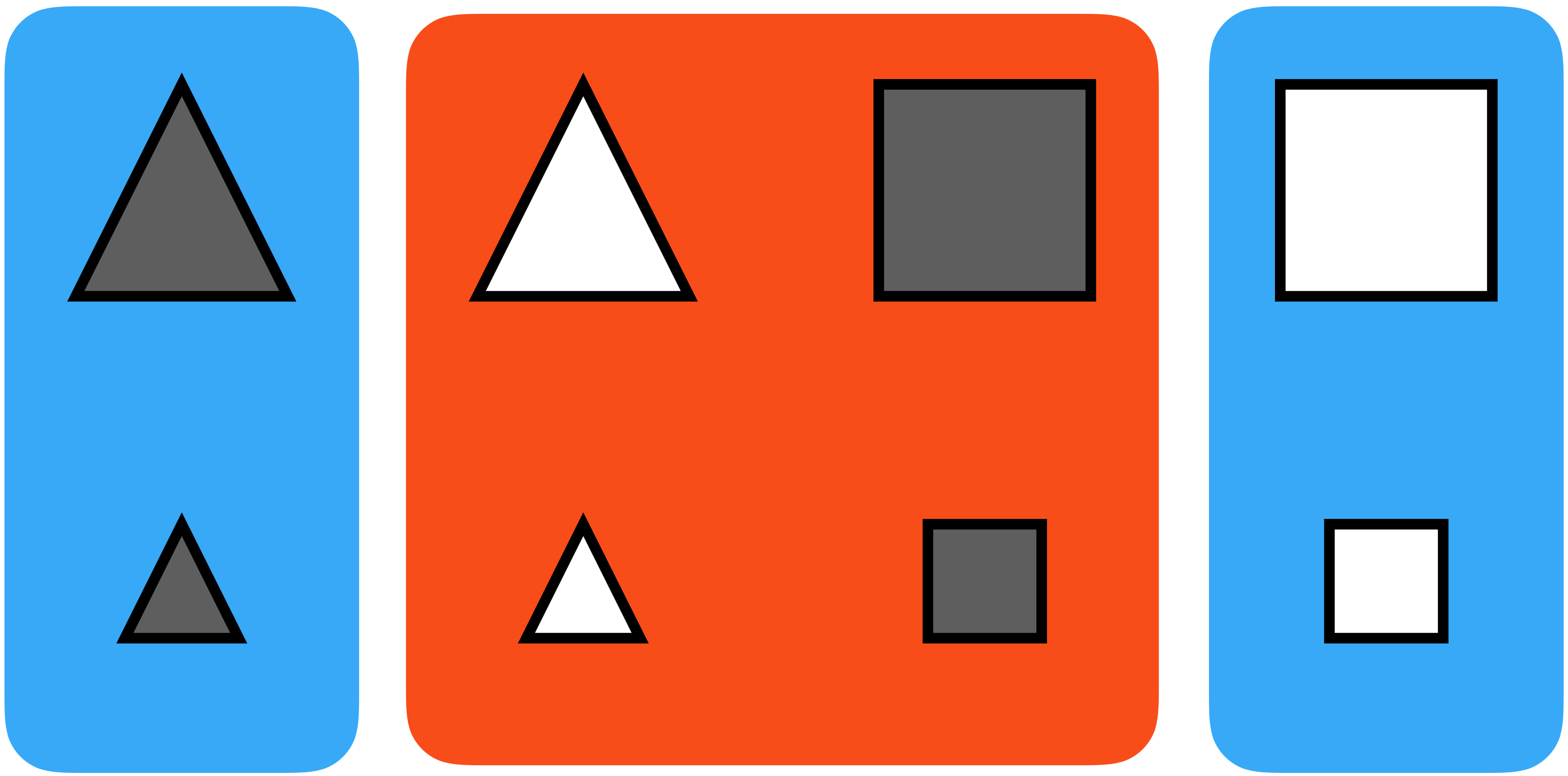


Type II Categorization

Grouped along two feature dimensions

Type II Categorization

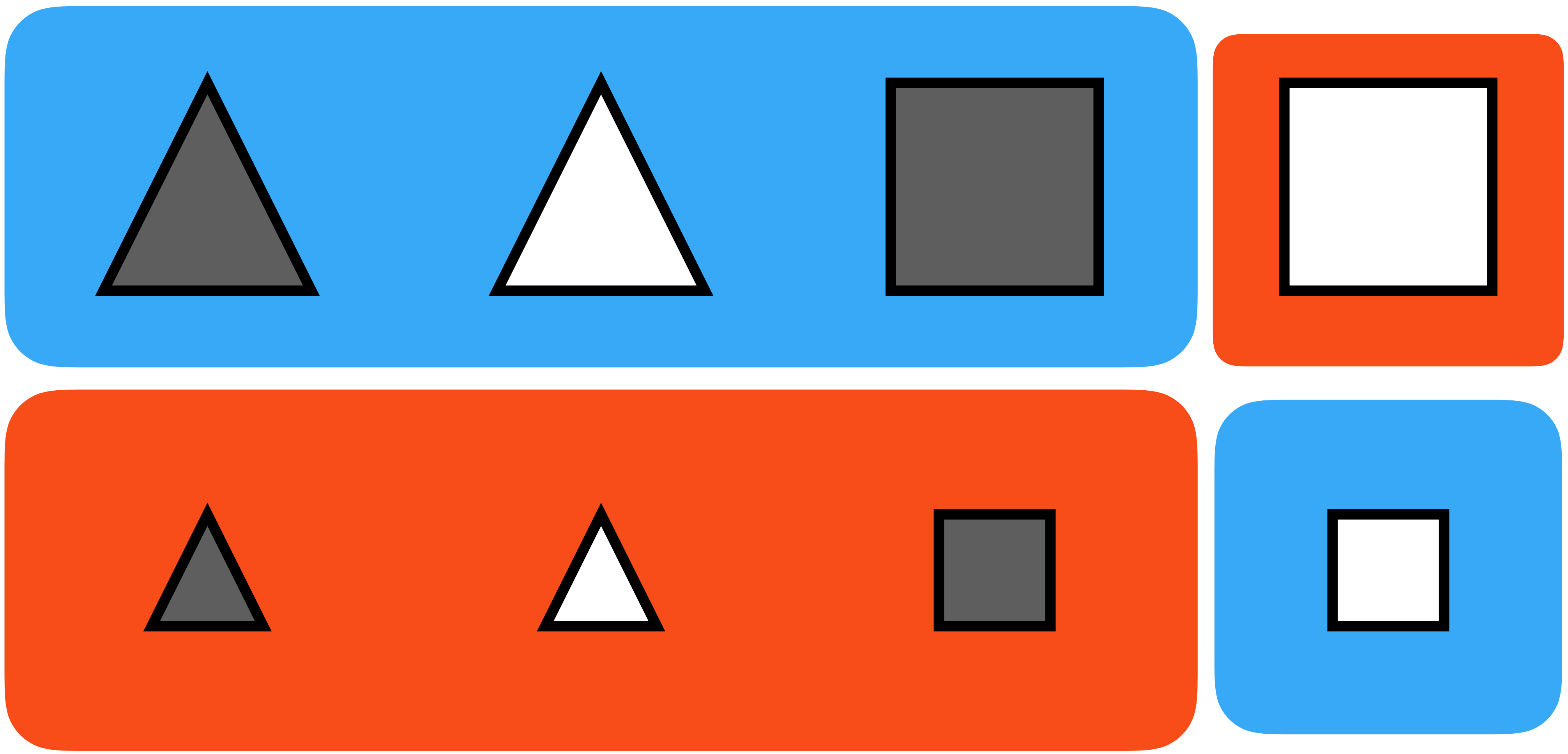
Grouped along two feature dimensions
“Black triangles and white squares are in group A”



Types III, IV, and V

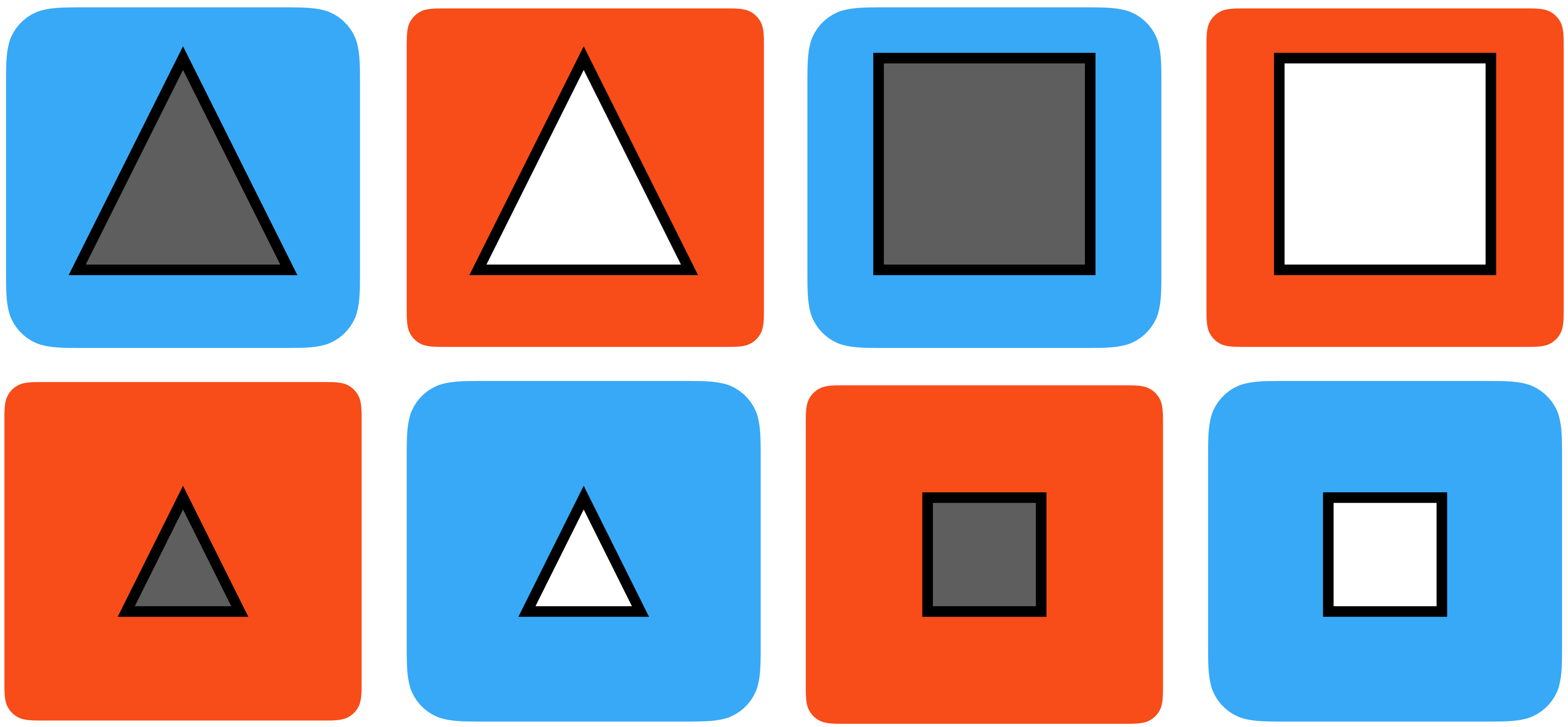
Rule relies on all three features

“Large shapes are in category A, **except for** the white square”



Type VI Categorization

No simple feature-based rule that can describe grouping
Requires memorization of group membership



Six types of classification problems

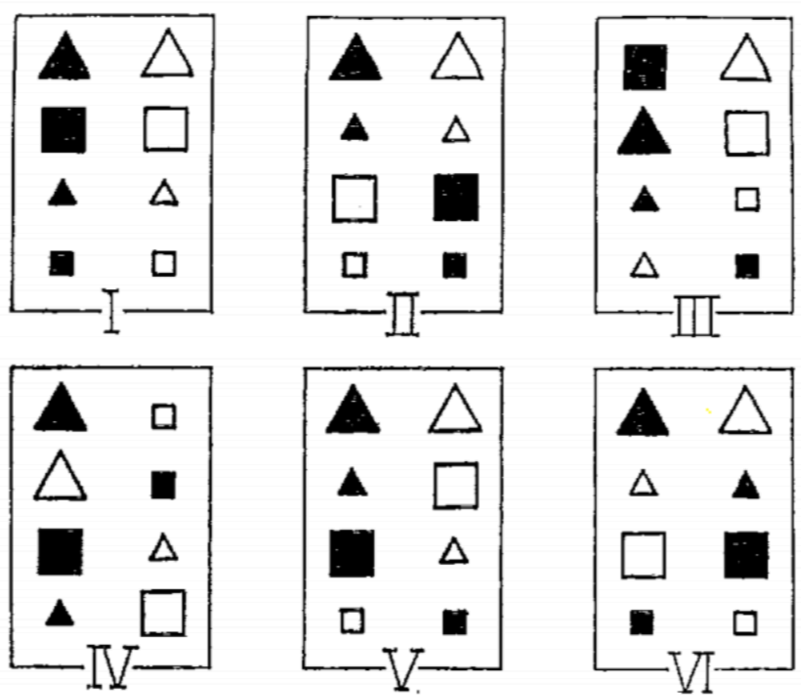


FIG. 1. Six different classifications of the same set of eight stimuli. (Within each box the four stimuli on the left belong in one class and the four stimuli on the right in the other class.)

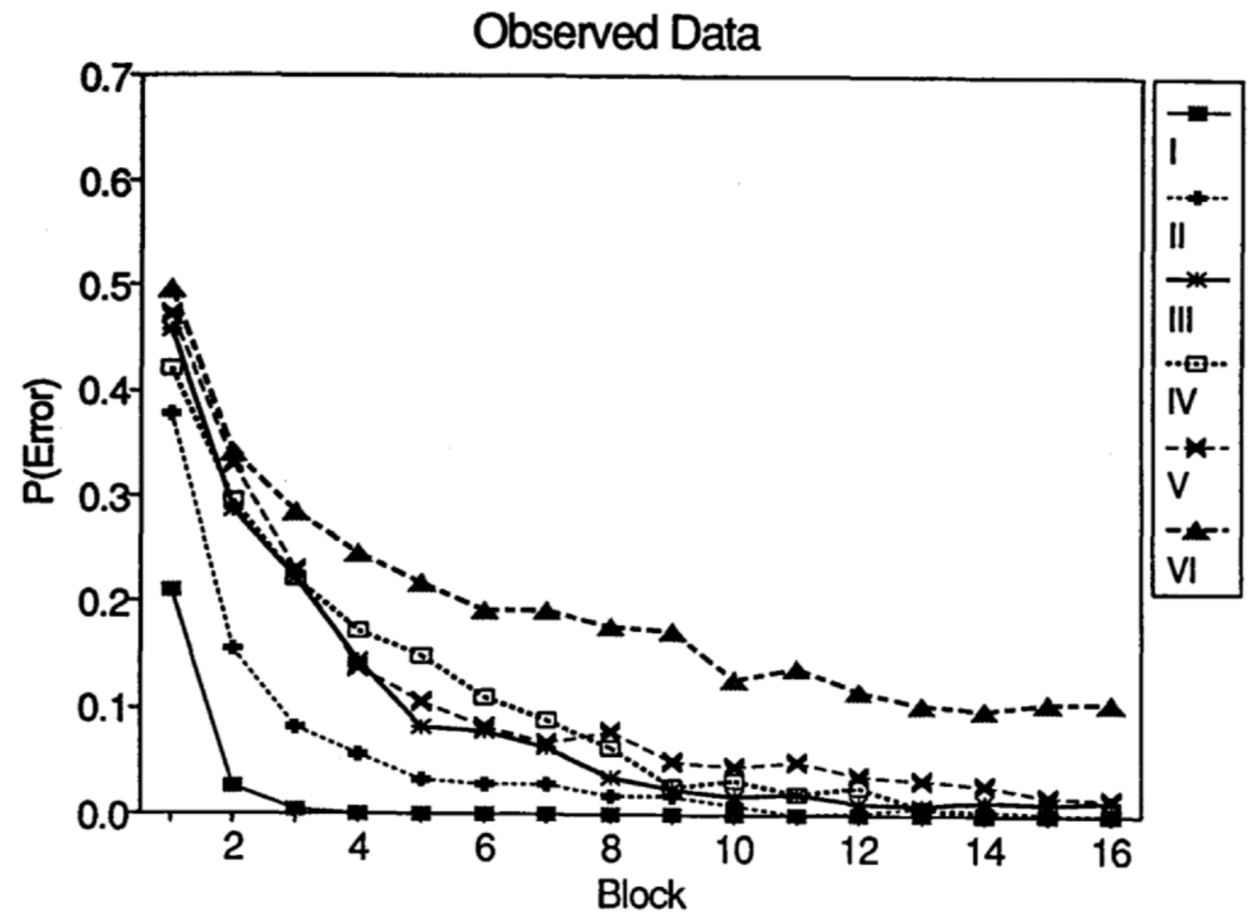


Figure 2. Average probabilities of errors for each problem in each block of 16 trials. The data from only Blocks 1–16 are shown. There were essentially zero errors during Blocks 17–25 for all problems except Type VI (see Table 1).

Nosofsky et al. 1994

Approach

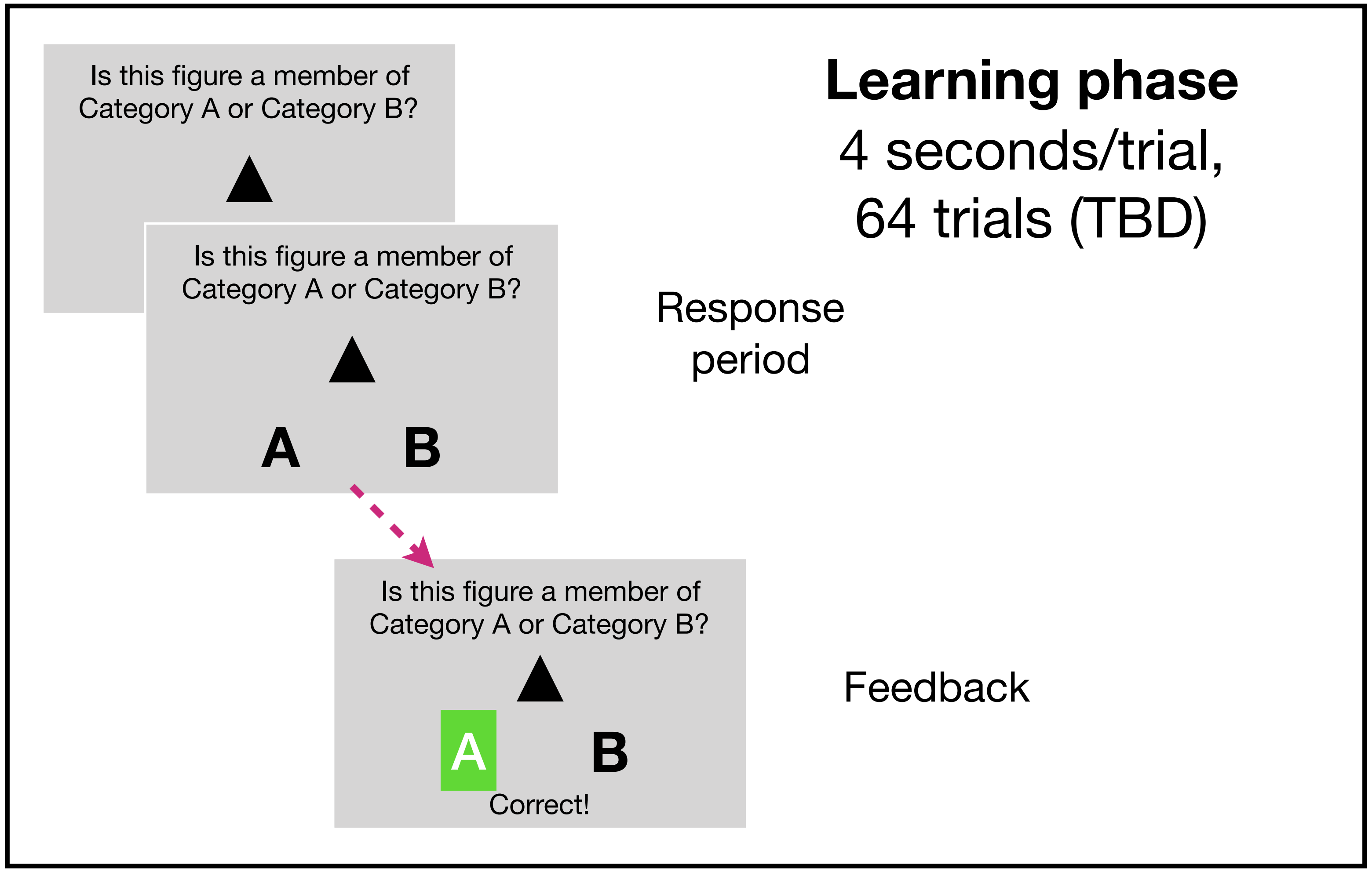
- In our task, **both** incentive and difficulty (Type) will be varied **between** subject to avoid learning effects across blocks
- Each subject will encounter one fixed length Learning Phase and one Test Phase of one problem type at one incentive level

	I	II	III	IV	V	VI
0 pts	N	N	…			
1 pt	N	N				
2 pts	…		…			
4 pts						
8 pts						
16 pts						
32 pts						

- 6 problem types, 7 incentive levels
- N=30 subjects per condition
- 1260 total subjects

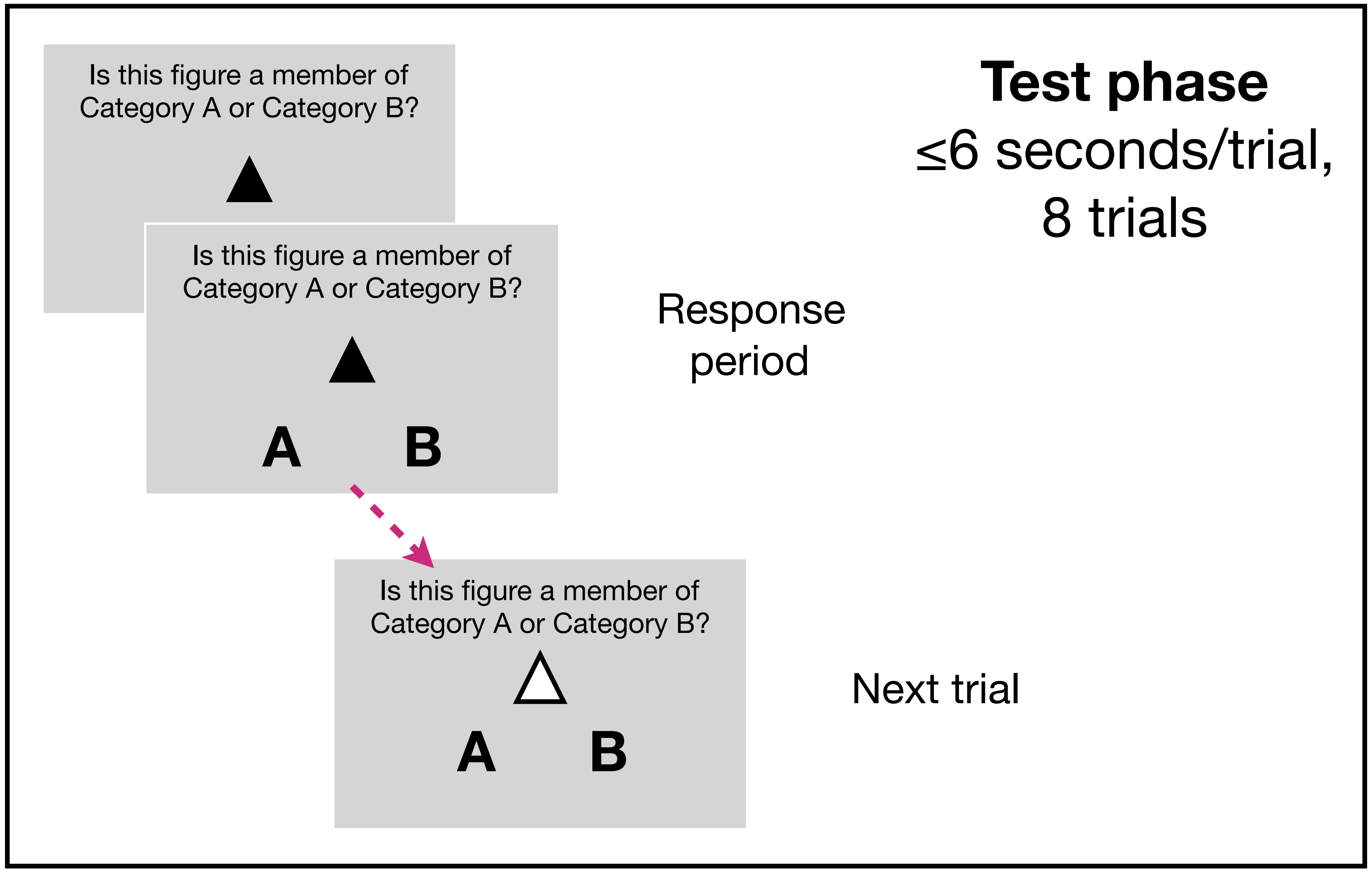
Approach

- Subjects read instruction and complete a comprehension check
- Learning phase
  - 64 trials of equal temporal length
  - Response period allows subject to guess category membership
  - Feedback provided after response
  - No reward or punishment for correct/incorrect answers during this period



Approach

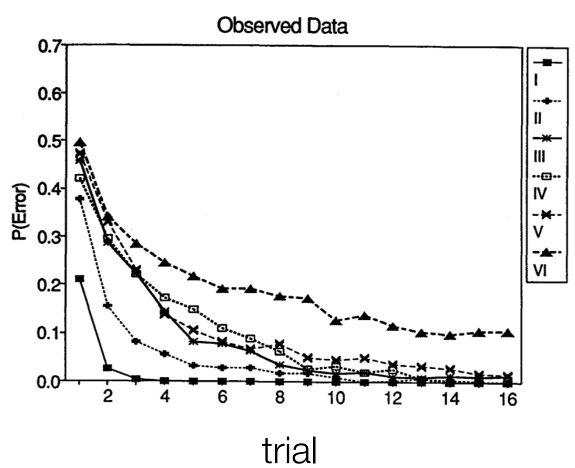
- Test phase
  - 8 trials testing category membership belief of each stimulus (one-shot response)
  - No feedback
  - Performance above chance determines bonus calculation based on incentive condition



Approach

- This experiment will generate data to fill the following plot types:

Learning Phase Data



- For x incentive levels, we will have x plots of this nature
- (Can also present data from a single problem type with x separate lines as incentive levels)

Test Phase Data

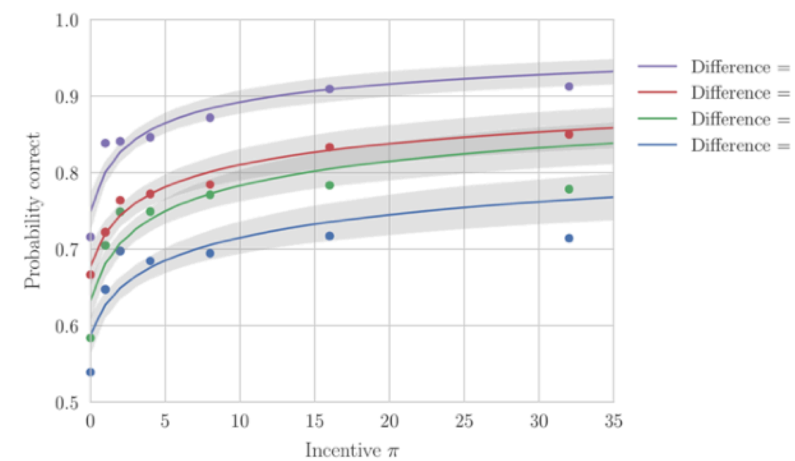


Figure 9: IPCs for all difficulty levels

- Instead of Difference, separate lines will indicate Problem Type
- Average percent correct across N subjects with standard error bars

Incentives and Bonus Payment

- 8 trials in test phase, and chance performance is 50% correct (4 trials)
- Option A: Subjects' maximal probability of winning a fixed bonus changes with condition. Condition ∈ {0, 1, 2, 4, 8, 16, 32}. Prob[Bonus] = (actual_score - chance_score) / 128

condition (points per correct answer)	chance score	**Perfect Score**	probability of bonus	bonus amount	expected value
0	0	0	0	$10	$0.00
1	4	8	0.03125	$10	$0.3125
2	8	16	0.0625	$10	$0.625
4	16	32	0.125	$10	$1.25
8	32	64	0.25	$10	$2.50
16	64	128	0.5	$10	$5.00
32	128	256	1	$10	$10.00

- Option B: Subject performance above chance corresponds to **equivalent probability** of **variable bonus**. Prob[Bonus] = %correct above chance = {0, 0.25, 0.5, 0.75, 1.00}. Bonus is calculated from expected values of table above, such that Condition ∈ {$0.00, $0.31, $0.63, $1.25, $2.50, $5.00, $10.00}