Computational modeling of analogy and similarity

Praveen Paritosh

CogSci 207

Fall 2004

Overview

- Examples
- Structure-mapping theory
- Simulating matching: SME
- Simulating generalizations: SEQL

An analogy

- Abusive husband (H), wife (W).
- H treates W badly
- H terriorises W, convinces everyone out there to hurt her, and only he can save her.
- When W goes out, H watches closely, to protect her.
- H alienates W from all her friends who try to warn her.
- But W doesnt quit, cant be alone, until a replacement lined up. In due course, another guy, G, comes along.
- Hmm.. why does W still stick with H? W is intrigued, but not sure. W's been scared, hurt, and cant help but see some of H in G. H didnt look all that abusive when we started dating, right?
- For W to pick G, he must prove himself to be a far superior choice, not slightly/maybe better.

Analogical reasoning in action

- To make prediction, use some similar event
- To argue for the prediction, focus on aspects that heighten similarity
  - The more similar the event, the more likely the prediction
- To argue against the prediction, focus on aspects that decrease similarity
  - Not just any differences will do.
  - Differences that are key to the prediction matter

Structure-Mapping Theory (Gentner, 1983)

- Analogy involves
  - correspondences between structured descriptions
  - candidate inferences fill in missing structure in target
- Constraints
  - *Identicality:* Match identical relations, attributes, functions. Map non-identical functions when suggested by higher-order matches
  - *1:1 mappings:* Each item can be matched with at most one other
  - *Systematicity:* Prefer mappings involving systems of relations, esp. including higher-order relations
- Also provides account of similarity, metaphor
- Growing body of evidence that same processes are used in perception, problem solving, conceptual change...

Representation assumptions in structure-mapping

- *Attributes* are unary predicates, representing category membership
  - (Bartender George)
  - Equivalently, (isa George Bartender)
- *Relations* describe connections between more than one thing
  - (worksAt George CineArts6)
- Logically equivalent not same as psychologically equivalent or computationally equivalent
- *Order* of an item in a description indicates height in nested structure
  - Order(entitties) = 0
  - Order(expression) = 1+max(order(arguments))
- Higher-order statement in this sense has deeper nesting
  - Contrast with higher-order in logic, meaning predicates can be used as terms.

Why Tiered identicality is important

- Analogy and similarity are not pure graph isomorphism; Content matters
- Postulating identical relations is a strong semantic constraint
  - (preferredPreyType Cats Mice)
  - (preferredPreyType ConArtist NaiveInvestors)
  - (typicalMaterialType AutomobileBody Metal)
- Identicality goes a long way, but sometimes want to weaken it slightly
  - (preferredFoodType ItalianPeople PastaDish)

Why non-identical functions can match

- Functions are ways of specifying terms.
- Psychologically, often correspond to dimensions or parts of some kind
  - e.g., (temperature water8) ↔ (pressure oil9)
  - e.g., (MilitaryFn Country18)
    ↔ (ImmuneSystemFn Animal6)
- Don't want to propose all possible function matches
  - Avoid combinatorial explosions
- Allow only if suggested by larger relational structure
  - Supports cross-dimensional comparisons

Why 1:1 is important

- Suppose we had
  - Titanic ↔ Enron
  - Captain of Titanic ↔ Ken Lay
  - Iceberg ↔ Ponzi scheme
  - Passengers ↔ Investors
  - Lookouts ↔ Whistle-blowers
- Then
  - "Lookouts warn captain but are ignored" ↔ "Whistle-blowers warn Ken Lay but are ignored"
- But if we also have
  - Lookouts ↔ Ken Lay
  - "Lookouts warn captain but are ignored" ↔ "Ken Lay warns Ken Lay but is ignored"

Why parallel connectivity is important

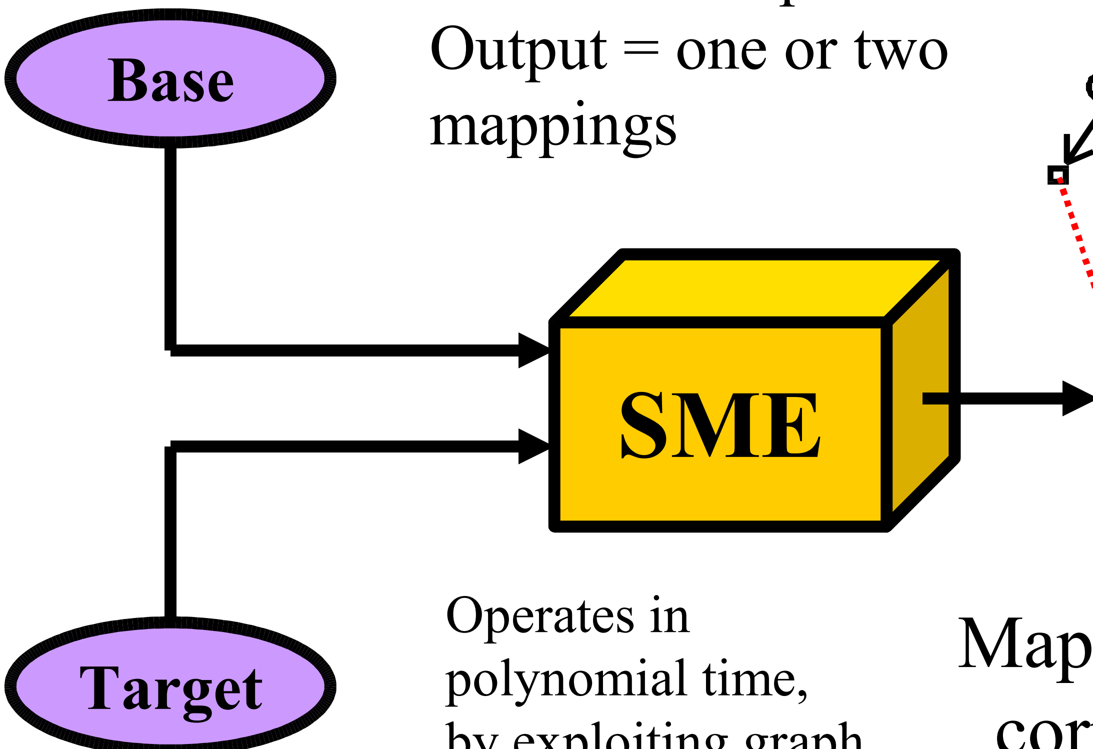
- Consider
  - ```
    (implies (and (LeakingFluidDevice BrakeCylinder2)
                  (partOf BrakeCylinder2 Car54))
             (DangerousDevice Car54))
    ```

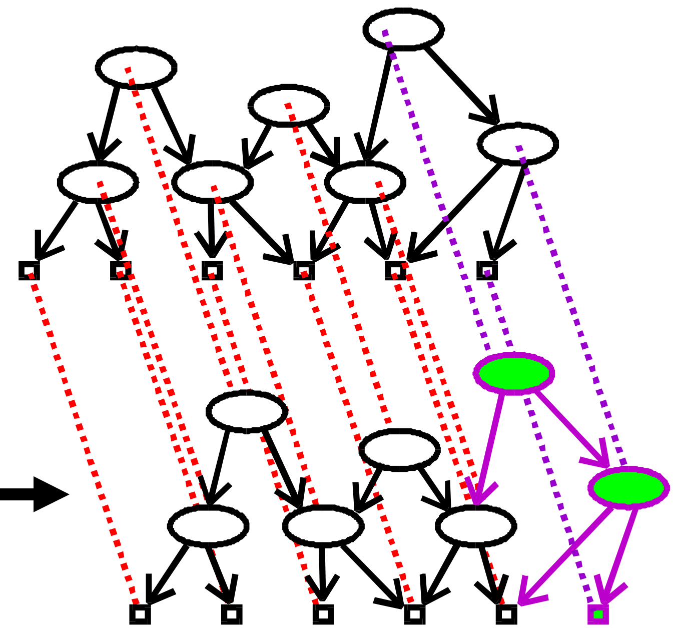
    ↔

    ```
    (implies (EjectsFlames BattleBot12)
             (DangerousDevice BattleBot12))
    ```
- What does BrakeCylinder2 go with?
- Best one can do in these cases is rerepresent
  - **(implies *<reason for being dangerous>* …)**
- But that operation lies outside the matcher, for tractability
  - Unwise to embed exponential processes in inner loop
    ```

Why systematicity is important

- Systems with higher-order relations structurally
  $\equiv$ well-justified arguments semantically
- Higher systematicity matching structures more likely to lead to stronger candidate inferences
- Example
  - ```
    (implies (applicable Theory8 Context12)
             (implies (and <fact1> <fact2>)
                      <conclusion>))
    ```
  - ```
    <factA> <factB>
    ```
    versus
  - ```
    <factA> <factB> (applicable Theory8 Context60)
    ```
    ```

SME: Structure-Mapping Engine

Inputs = propositional descriptions, w/ incremental updates
Output = one or two mappings

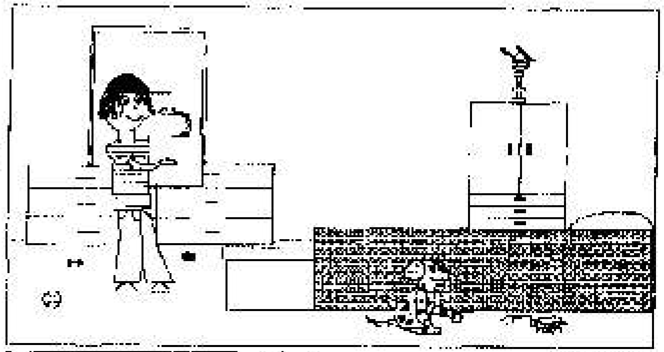




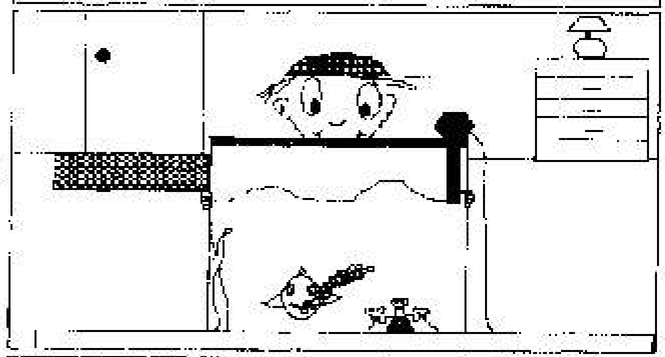
Operates in polynomial time, by exploiting graph labels & greedy algorithms

Mappings =
correspondences
+ structural evaluation
+ candidate inferences

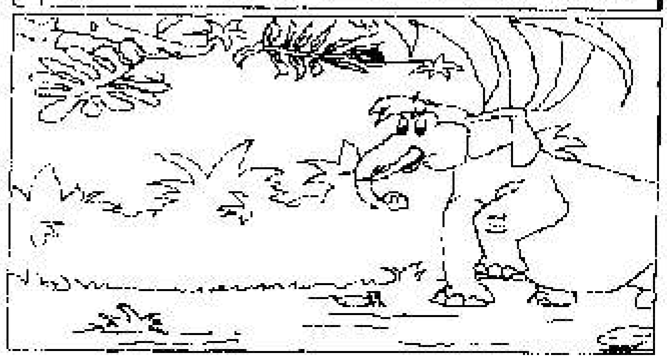
A)

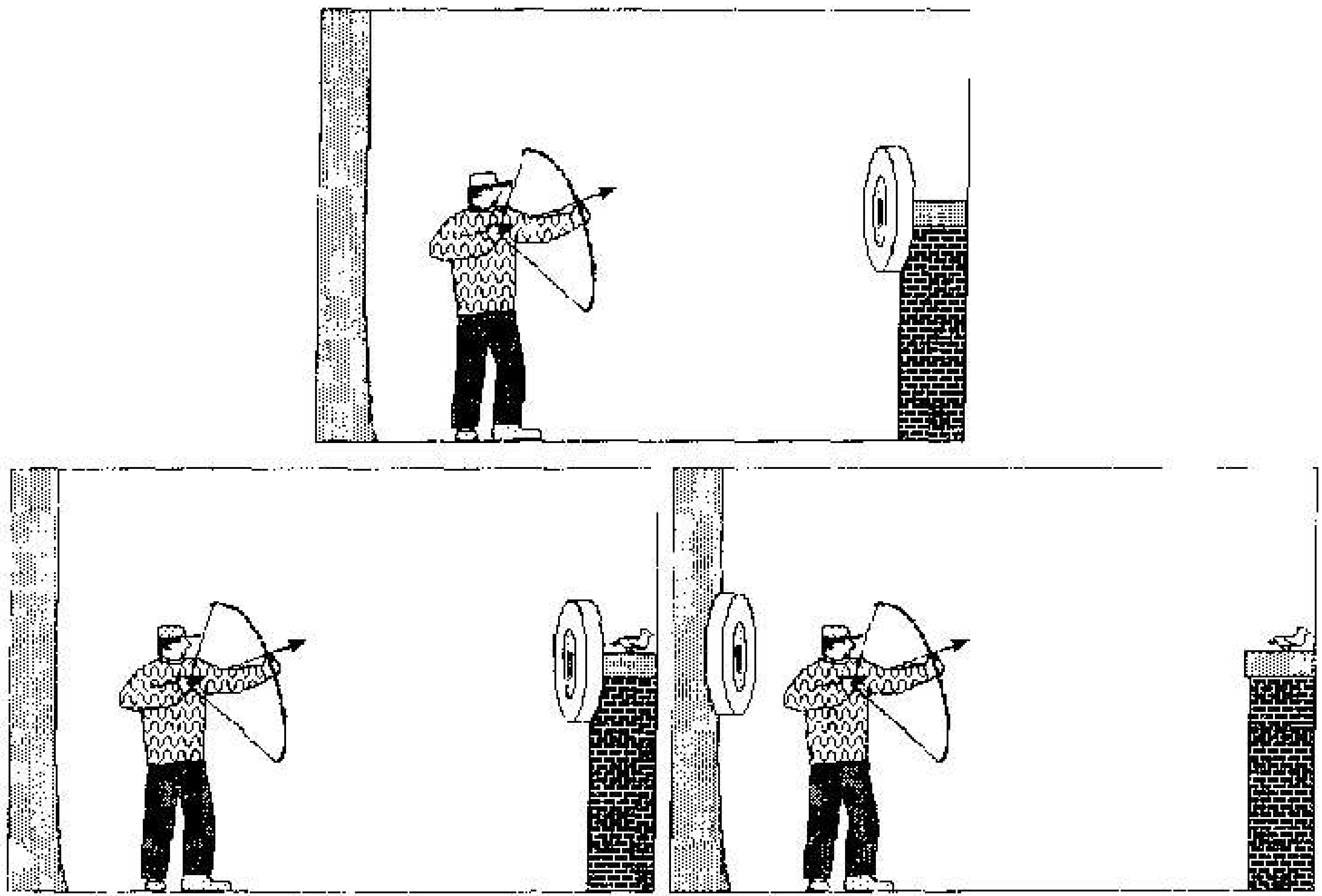


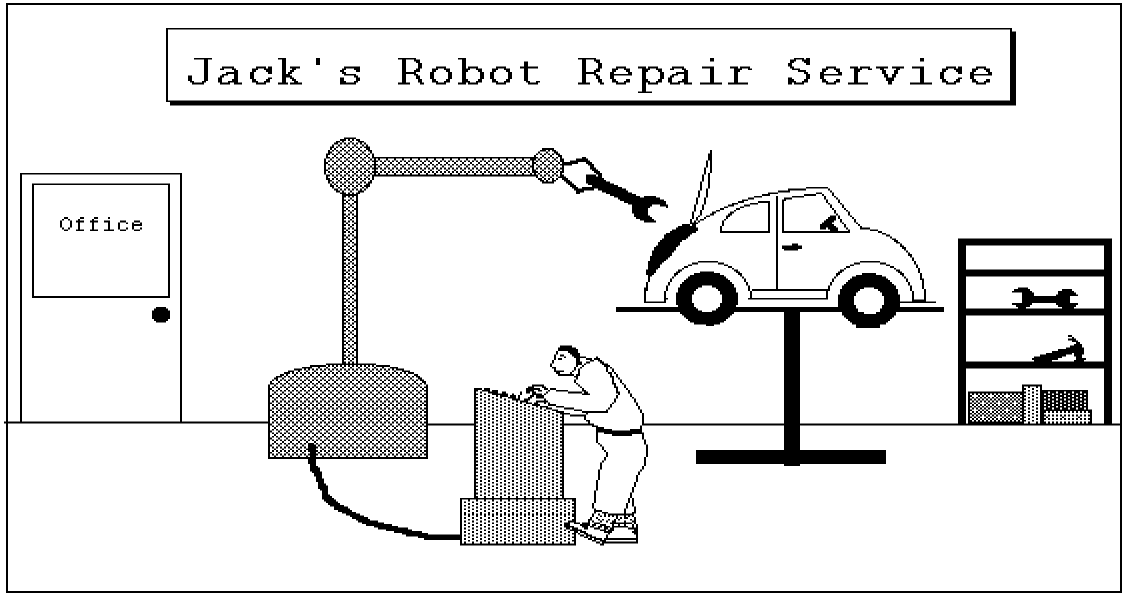
B)

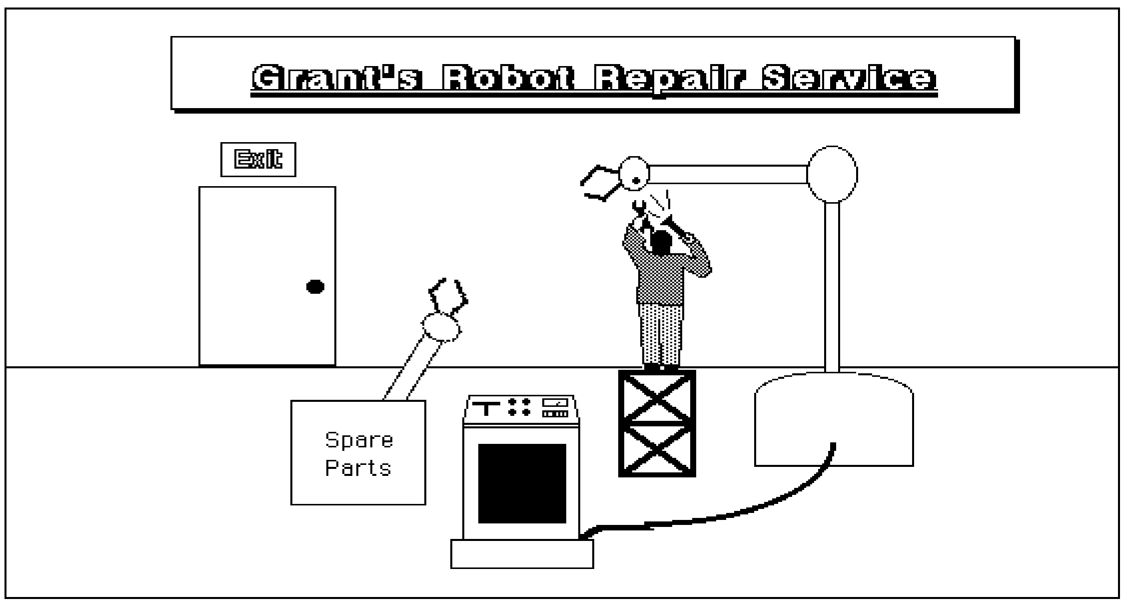


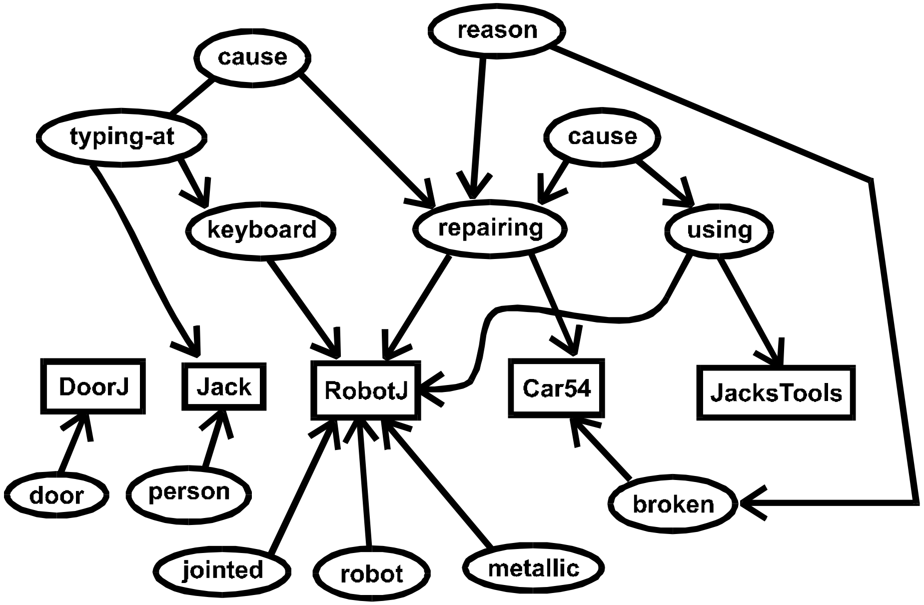
C)





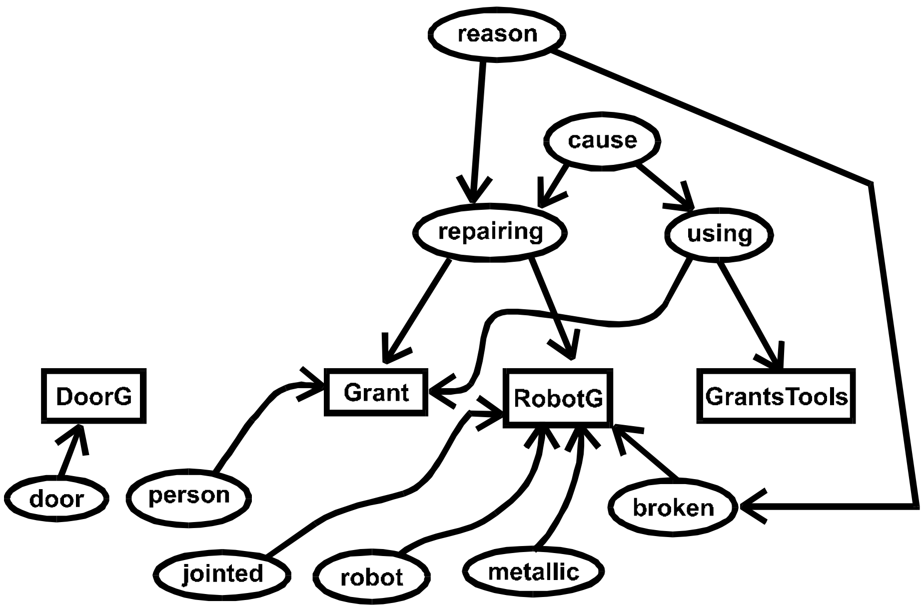

Jack's Robot Repair Service
Office


Grant's Robot Repair Service
Exit
Spare
Parts



Jack's Robot Repair description:

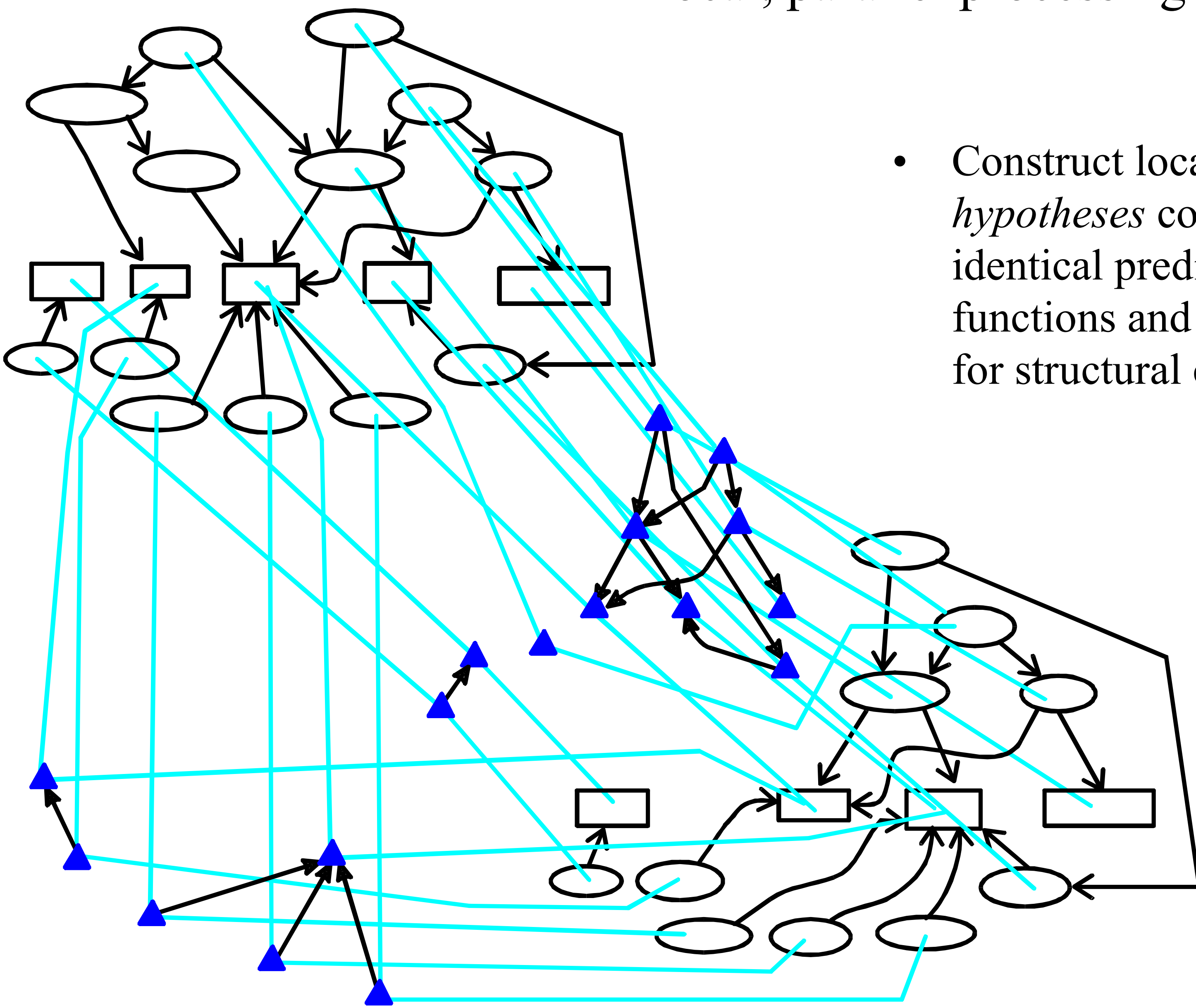
(REASON (REPAIRING ROBOTJ CAR54)
(BROKEN CAR54))
(CAUSE (TYPING-AT JACK
(KEYBOARD ROBOTJ))
(REPAIRING ROBOTJ CAR54))
(CAUSE (REPAIRING ROBOTJ CAR54)
(USING ROBOTJ HANDTOOLSJ))
(DOOR DOORJ)
(JOINTED ROBOTJ)
(METALLIC ROBOTJ)
(ROBOT ROBOTJ)
(PERSON JACK)



Grant's Robot Repair description:

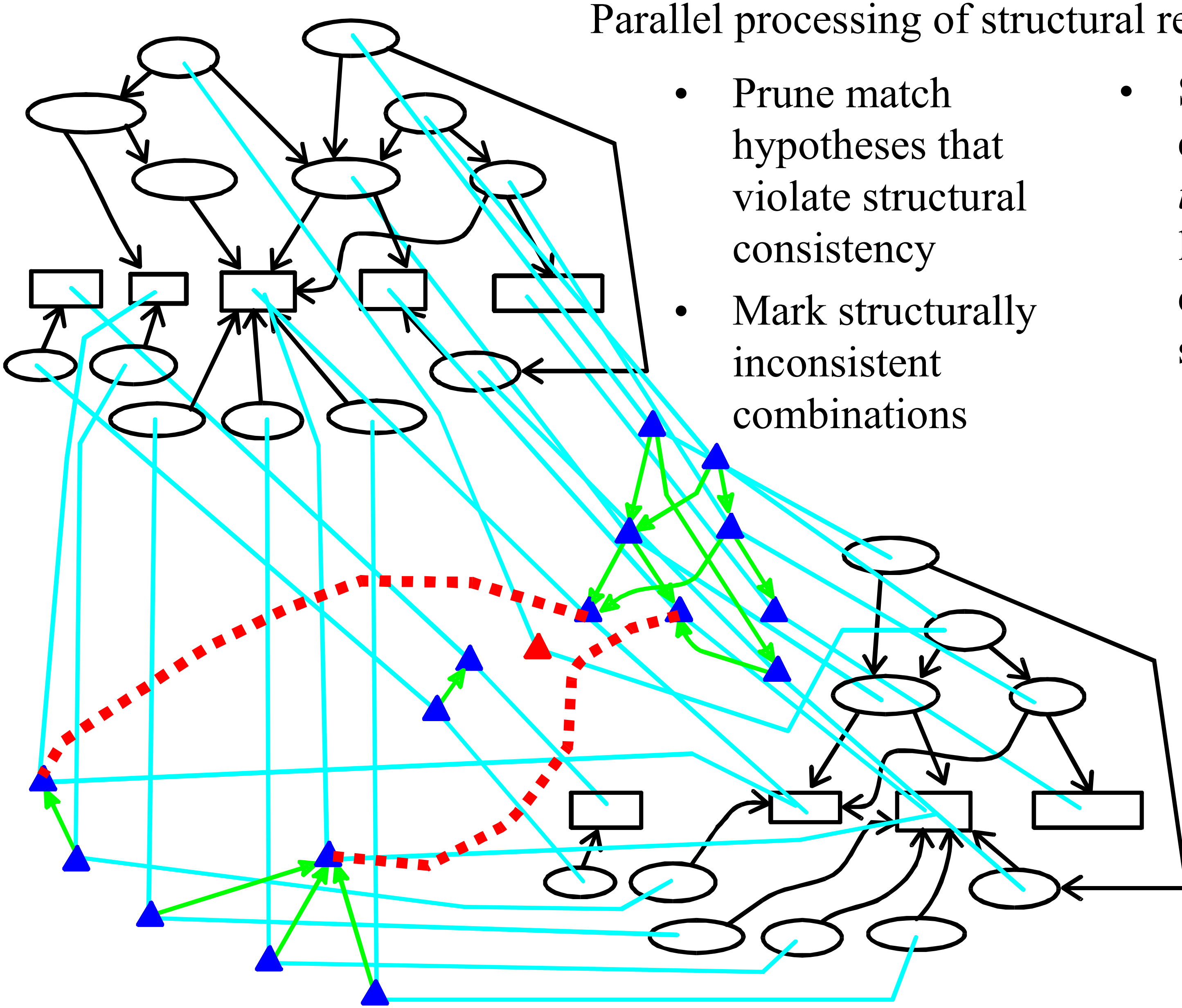
(REASON (REPAIRING GRANT ROBOTG)
(BROKEN ROBOTG))
(CAUSE (REPAIRING GRANT ROBOTG)
(USING GRANT HANDTOOLSG))
(TOOL-SET HANDTOOLSG)
(DOOR DOORG)
(JOINTED ROBOTG)
(METALLIC ROBOTG)
(ROBOT ROBOTG)
(PERSON GRANT)

Local, parallel processing first



- Construct local *match hypotheses* connecting identical predicates, functions and entities needed for structural consistency

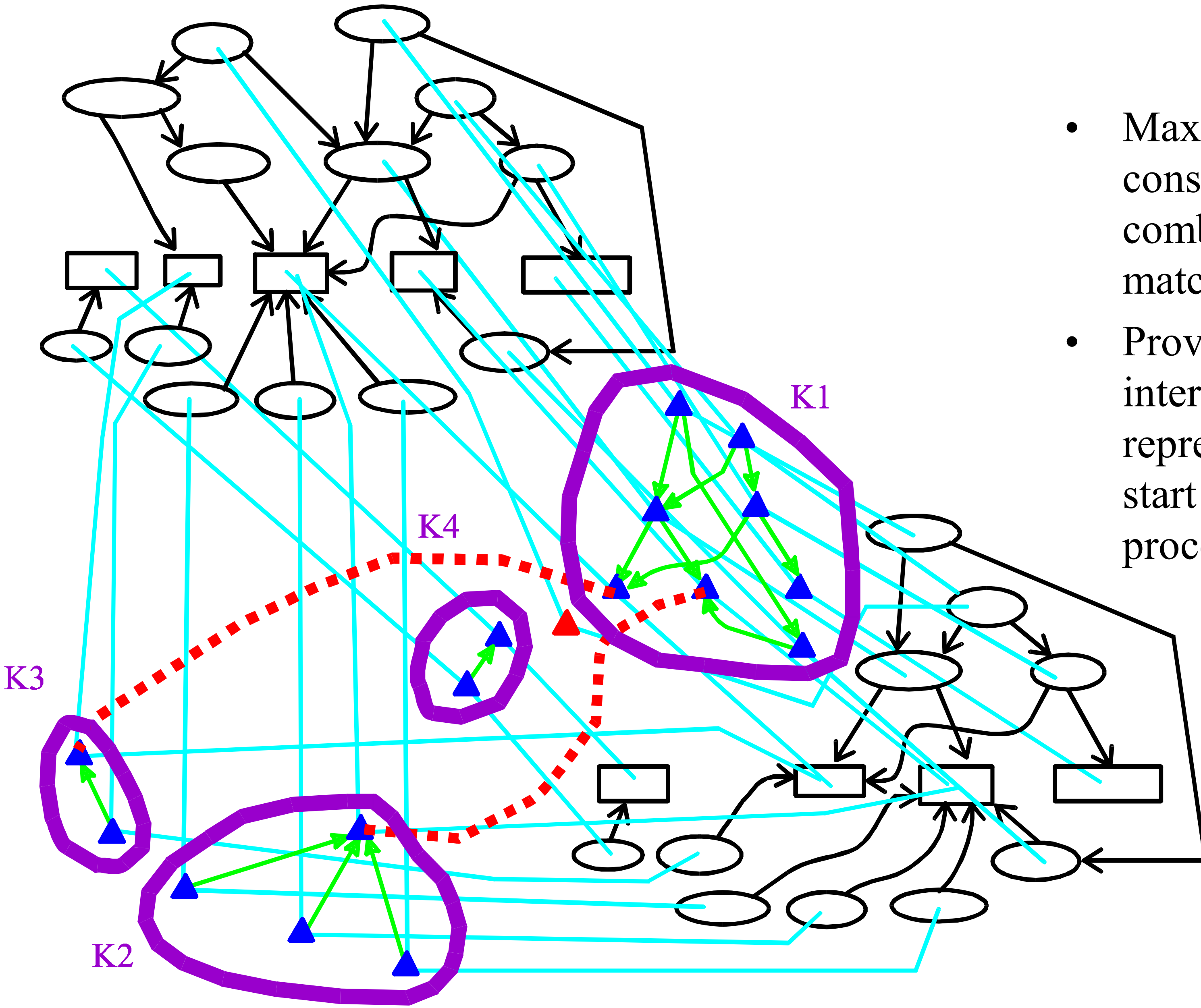
Parallel processing of structural relationships



- Prune match hypotheses that violate structural consistency
- Mark structurally inconsistent combinations

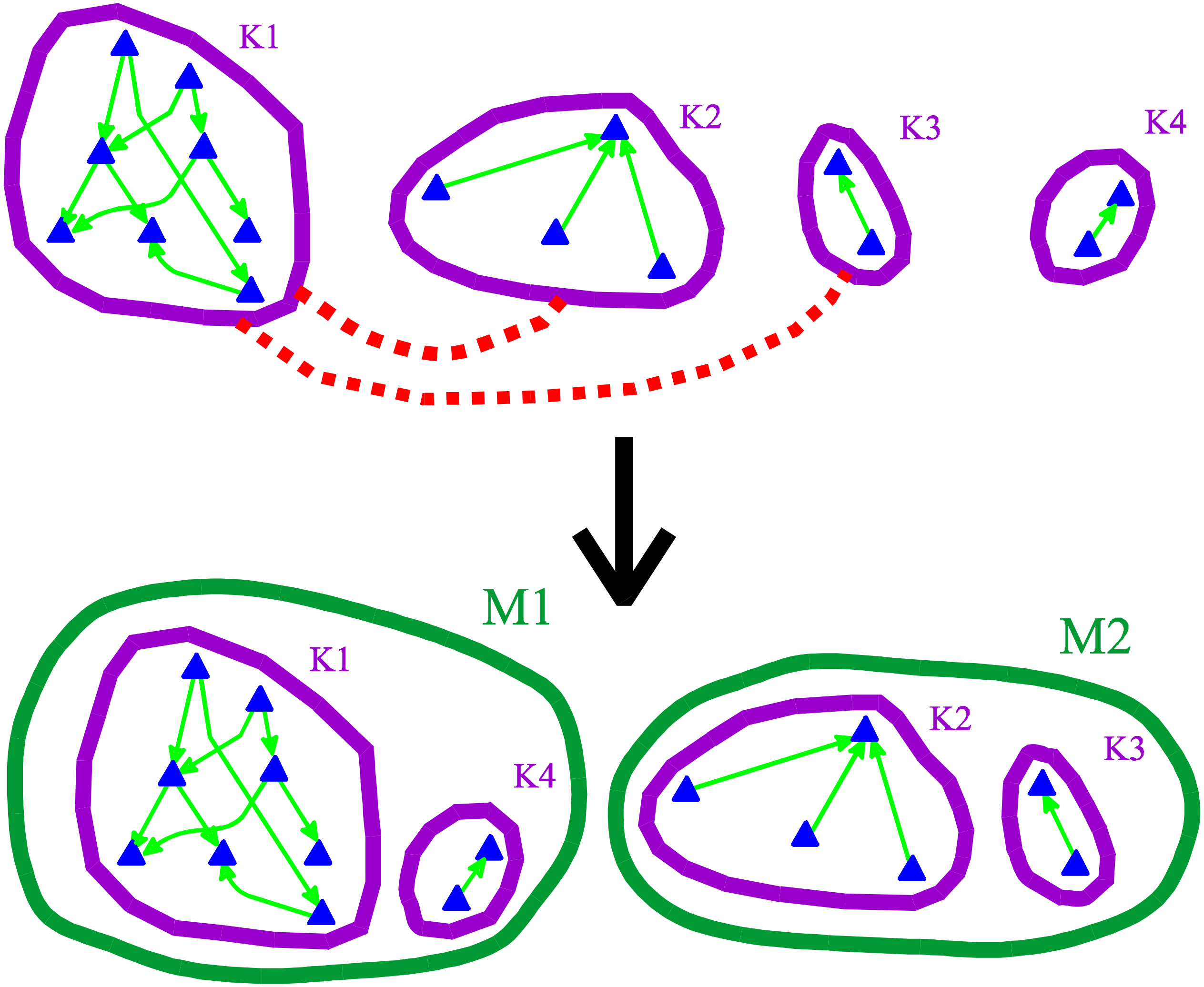
- Start structural evaluation via *trickle-down*, local method of enforcing systematicity

Identify Kernel mappings



- Maximal locally consistent combinations of match hypotheses
- Provides intermediate-level representation for start of serial processing

Construct global mappings



- Use greedy merge method
- Construct candidate inferences from leftover base structure
- Generate until less than *dropoff* of max, up to *n* mappings

Complexity of the SME Algorithm

1. Local Match construction *$O(n^2)$ worst case*
2. Structural consistency, local evaluation
   $O(n)$ worst case
3. Kernel construction *$O(n)$ worst case*
4. Structural evaluation *$O(n)$*
5. Merge *$O(k)$ worst case*

Adding incrementality does not affect these bounds

SME consistent with psychological findings

- Systematicity and structural consistency influence interpretation of analogies
  - Clement & Gentner, 1991
- Structural consistency influences inference in analogical reasoning
  - Clement & Gentner, 1991; Keane, 1996; Spellman & Holyoak, 1992, 1996; Markman, 1997)
- Structural consistency influences inference in category-based induction
  - Lassaline, 1996; Wu & Gentner, 1998
- Systematicity influences inference in analogical reasoning and category-based induction
  - Clement & Gentner, 1991, Wu & Gentner, 1998
- Ordinary similarity comparisons utilize structural alignment and mapping
  - Gentner, 1989; Gentner & Markman, 1995, 1997; Markman & Gentner, 1993, in press; Medin, Goldstone, & Gentner, 1993
- Similarity-based retrieval is surface-driven, but similarity-based reasoning is structurally driven
  - Gentner, Rattermann, & Forbus, 1993; Holyoak & Koh, 1987; Ross, 1989

Some SME psychological predictions

- Online processing of similarity and analogy is influenced both by object richness and by relational depth
  - Gentner & Rattermann, 1991; Markman & Gentner, 1993; Rattermann & Gentner, 1998
- *Relational shift:* Early in development object matches win over relational matches (because of inadequate relational knowledge)
  - Gentner & Toupin, 1986; *Halford, 1987; Gentner, 1988; Gentner & Rattermann, 1991; Rattermann & Gentner, in press
- Learning higher-order domain relations enables children to perform relational mappings
  - Gentner & Rattermann, 1991; Kotovsky & Gentner, 1996; Rattermann & Gentner, 1998; Goswami & Brown, 1989

Marcus et al experiments

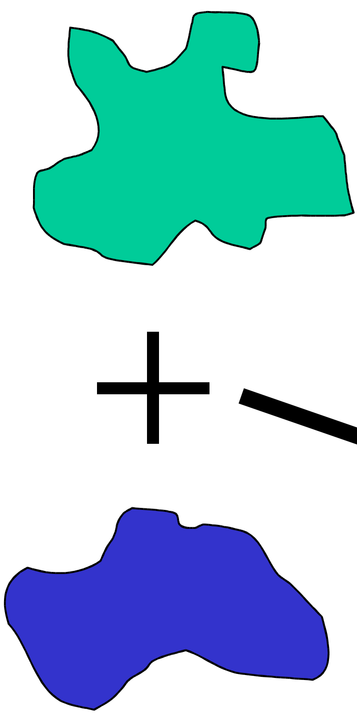
- 7 month old infants
- Exposed to sequence of artificially generated spoken input, with grammatical regularities
  - ABA pattern: “pa ti pa”, “po wa po”
  - ABB pattern: “pa pa ti”, “po po wa”
  - Training: 3 repetitions of 16 three-word sentences with single underlying pattern
- Test: 12 sentences with entirely new words (e.g., “wo fe wo”), ½ from familiar pattern, ½ from other pattern
- Result: 15 out of 16 infants distinguished inconsistent sentences.
- Variations:
  - Controlled for voiced/unvoiced
  - Used AAB versus ABB, to eliminate detection of duplication as criterion
- Claim: Statistical learning methods by themselves insufficient to model this task; variables needed
  - Simple recurrent network simulation attempt failed
  - Transition probabilities between specific words insufficient
  - Other learning mechanisms needed in addition

Connectionist simulation attempts

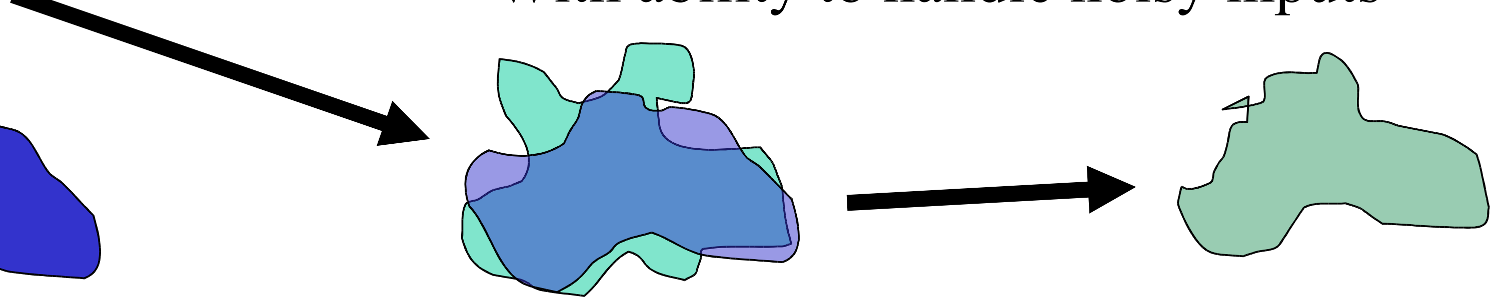
- Elman
  - Simple recurrent network
  - Added pretraining phase to teach words (50,000 out of a set of 120, 6 passes), supervised same/different
  - Supervised training of ABA versus ABB patterns. Full set of habituation sentences shown 347 times to train.
- Dienes, Altmann & Gao
  - Simple recurrent network, with extra layer of hidden units
  - 10 iterations for each test item (3+ epochs), unclear whether supervised or not
  - Smaller Euclidean distance between prediction and target for consistent versus inconsistent test stimuli
- Shastri & Chang
  - Localist connectionist simulation
  - Unsupervised training of ABA or ABB patterns. 34 epochs for ABB, 30 epochs for ABA (epoch = full set of habituation sentences)
- Problems with these attempts
  - Don't learn within span of stimuli given to infants
  - No demonstrated ability to handle input noise
- Current debate
  - Eliminative connectionism versus algebraic rules

A Third Alternative: Symbolic structural alignment

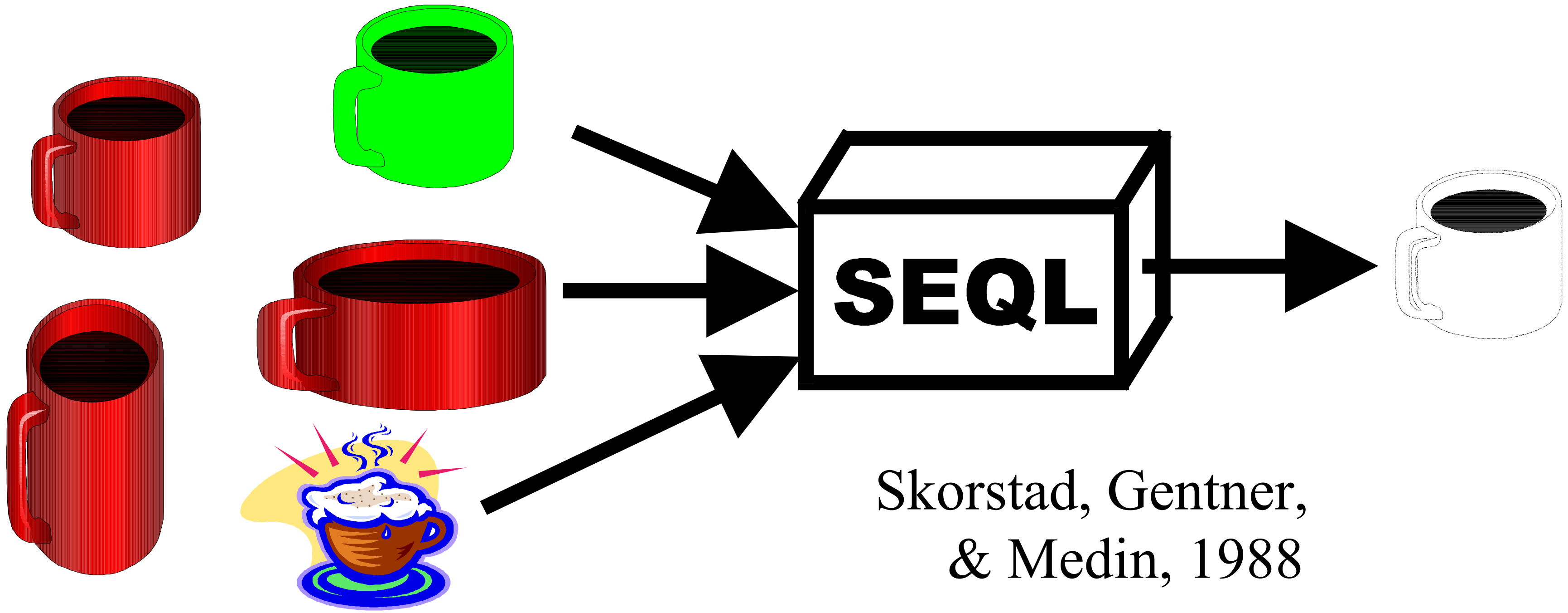
- Claim: Symbolic, structured representations are crucial
- Claim: Rules are not essential
  - Generalizations constructed via comparison sufficient



- Claim: Incremental abstraction via structural alignment over symbolic descriptions can capture the phenomena
  - With no more data than infants had
  - With ability to handle noisy inputs



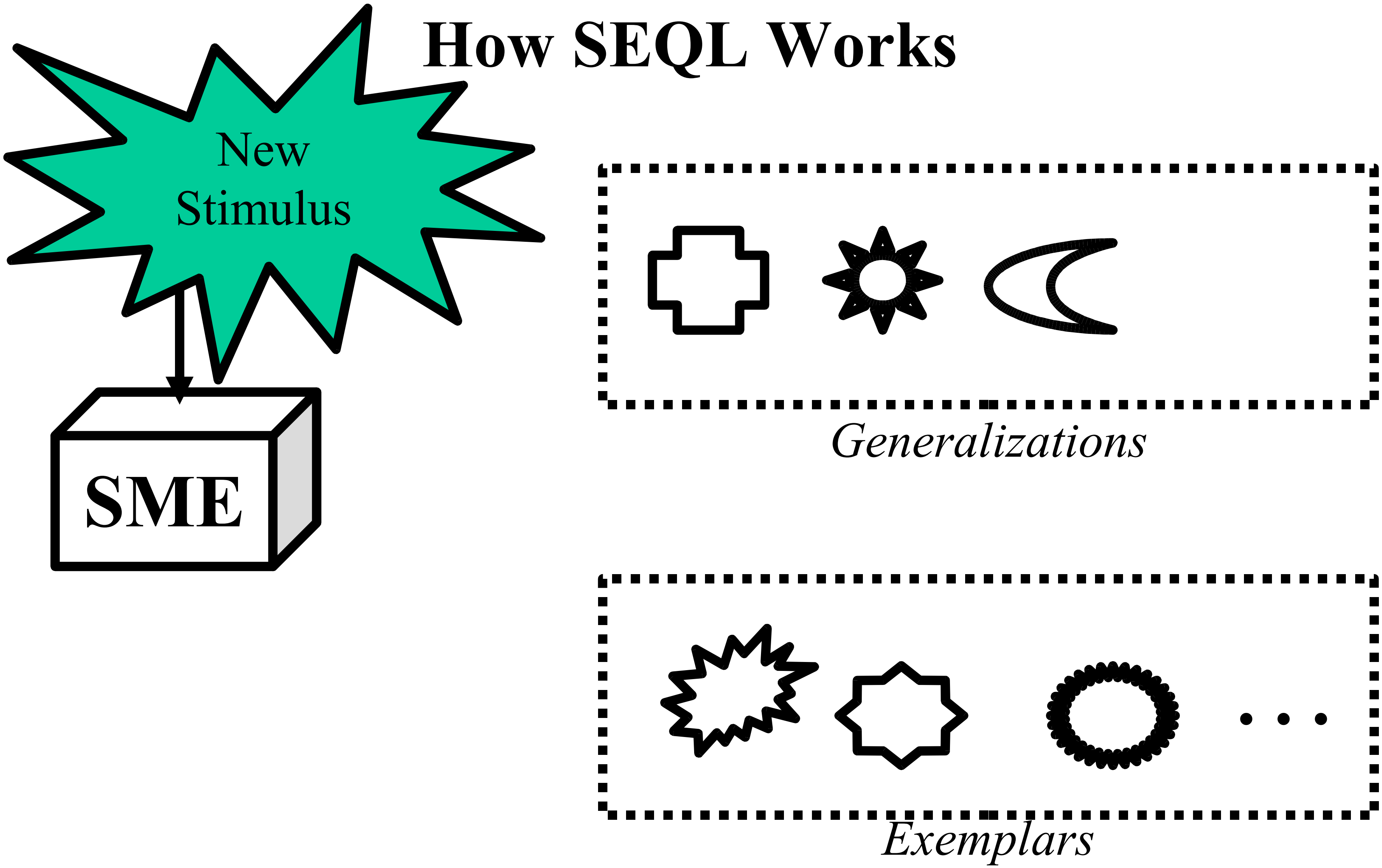
Incremental Abstraction and Rule Generation



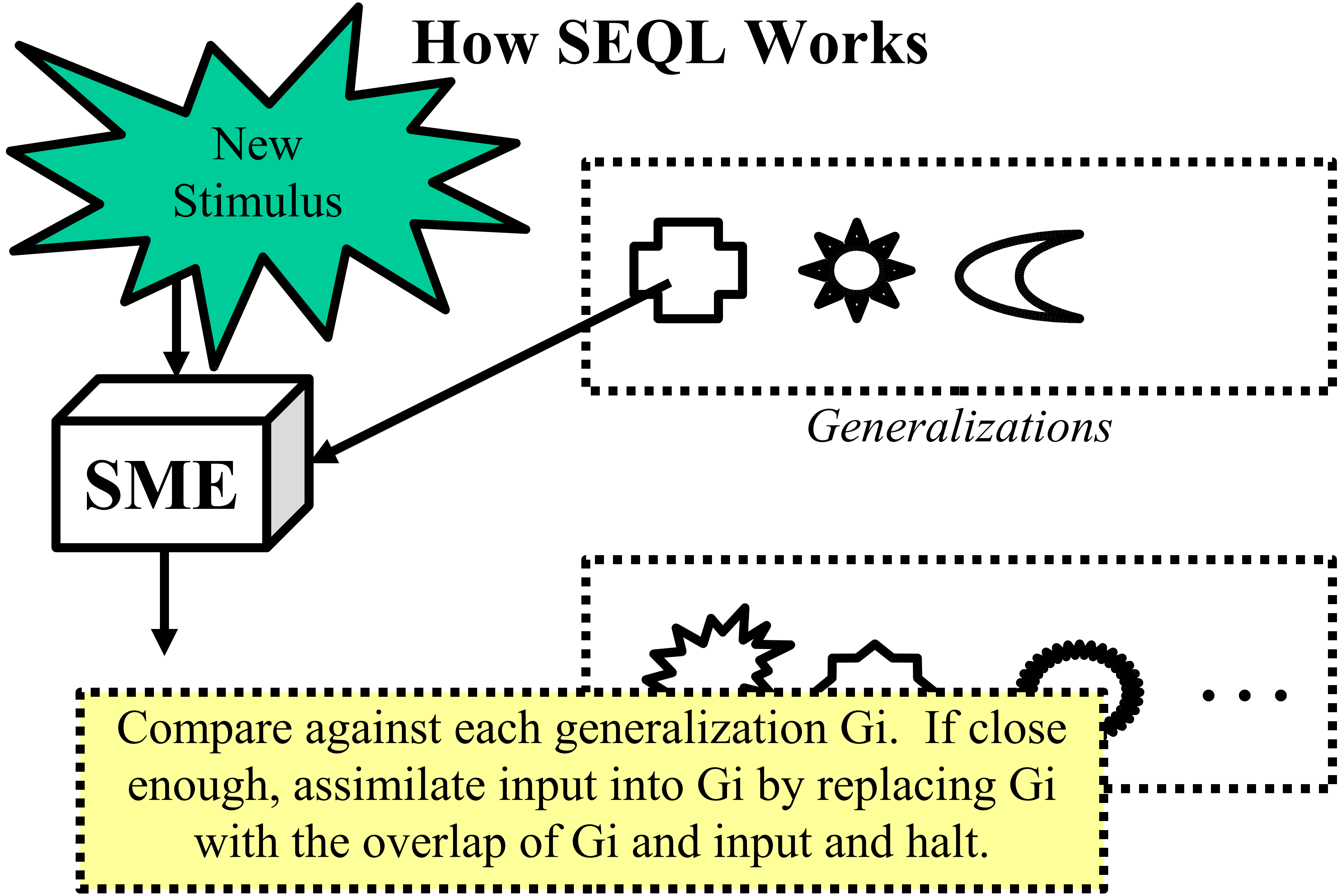
Skorstad, Gentner,
& Medin, 1988

- Uses SME to combine new instances with accumulating categories
- Merges descriptions if overlap sufficient
- Produces abstraction based on commonalities

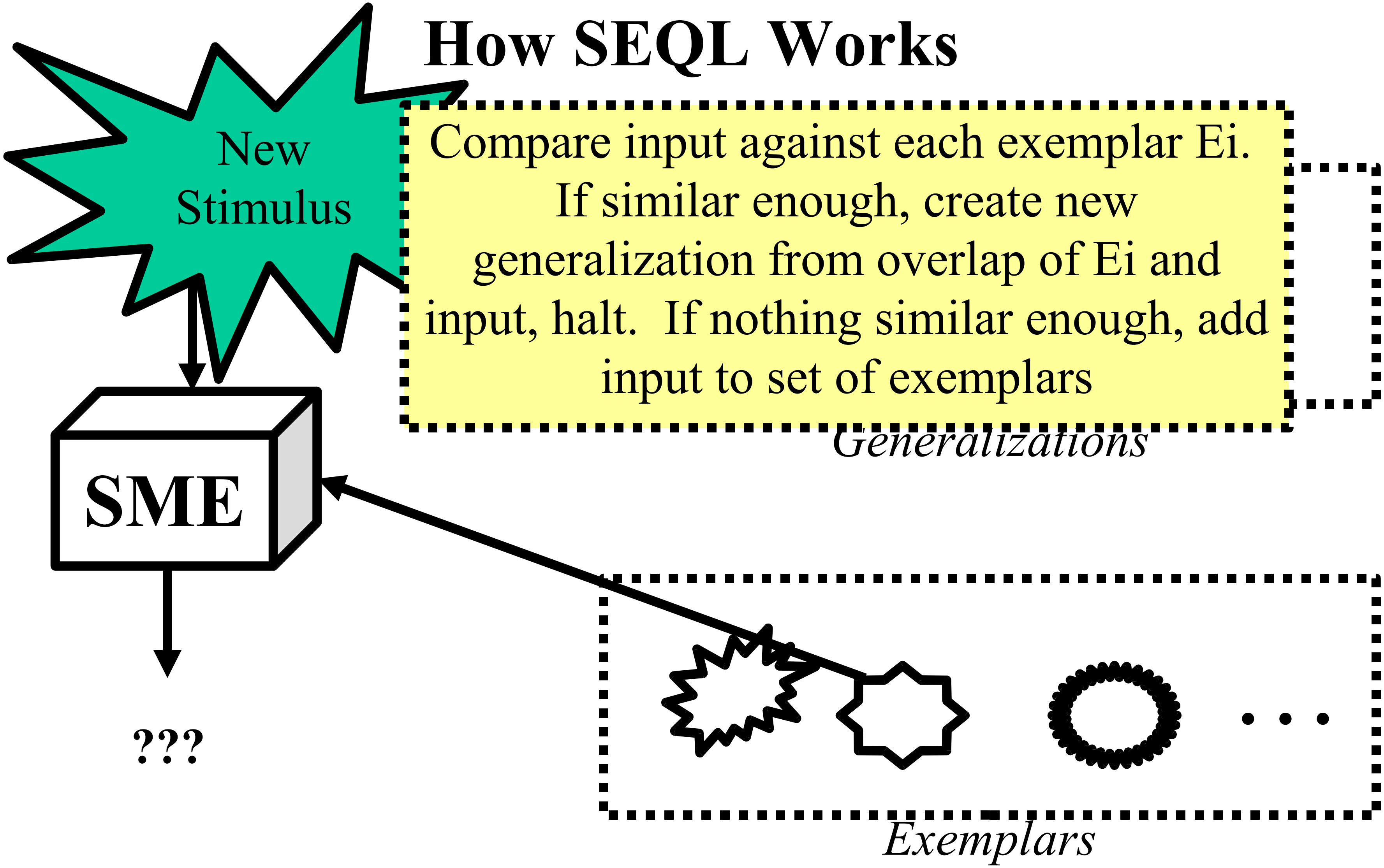
How SEQL Works



How SEQL Works



How SEQL Works



MAGI: A model of symmetry and repetition detection (Ferguson, 1994)

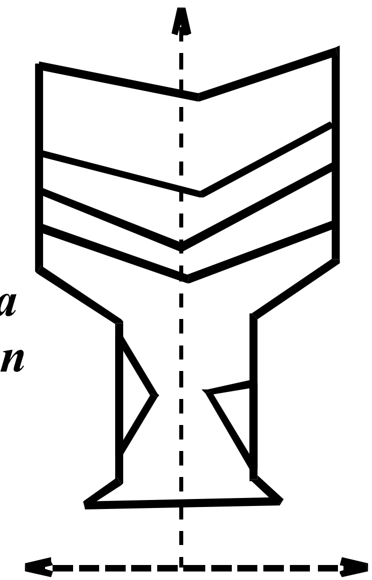
Detects symmetry or repetition by:

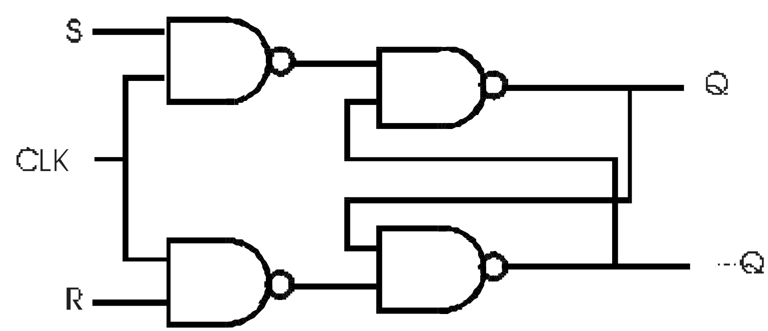
- Generating a structural description
- Finding areas of self-similarity within that description using Structure Mapping

Results:

- Same model operates on visual, functional, conceptual, and mathematical representations
- Makes predictions consistent with human perceptual data

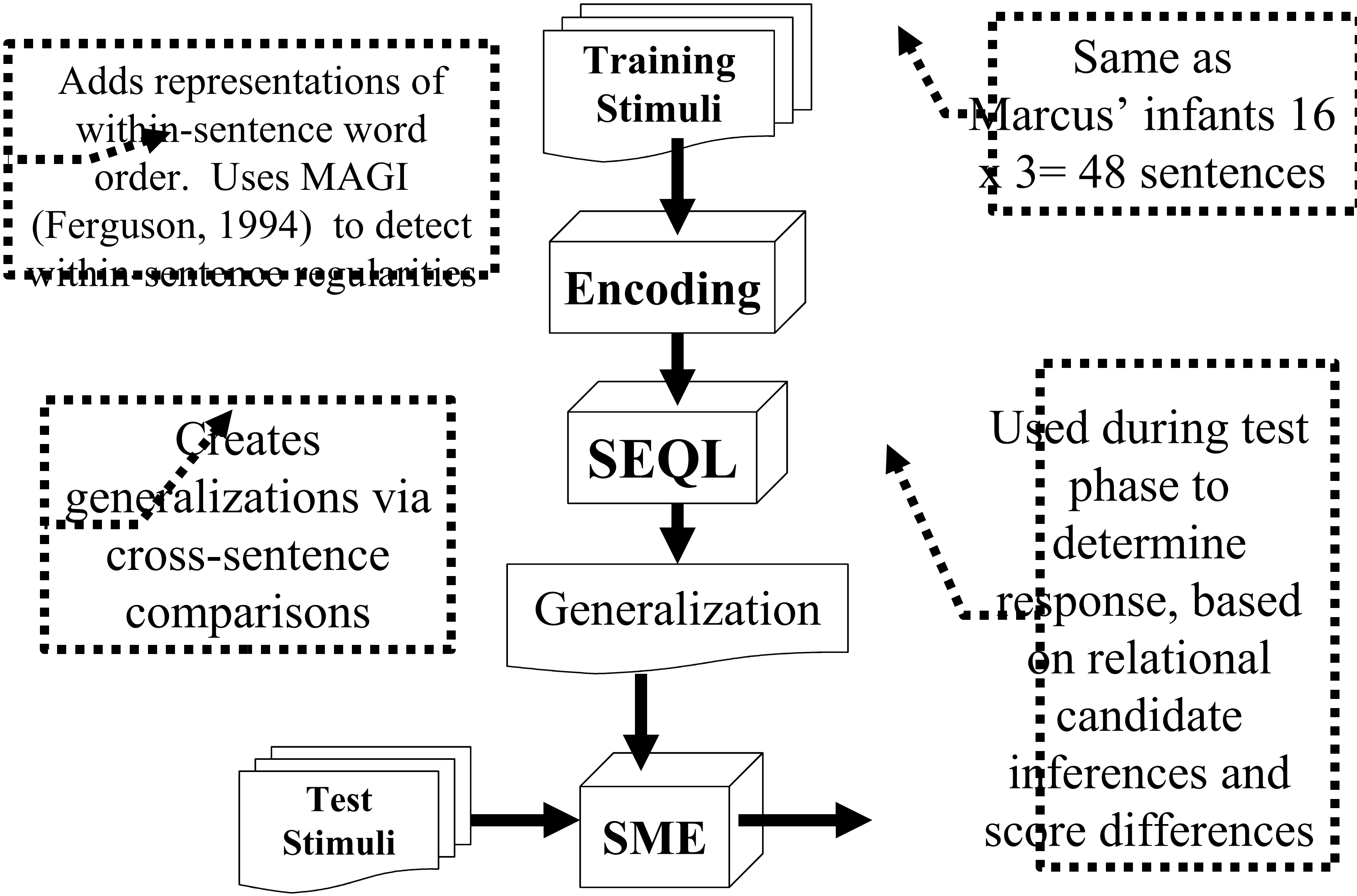
Detecting inexact symmetry in a roughly-drawn figure



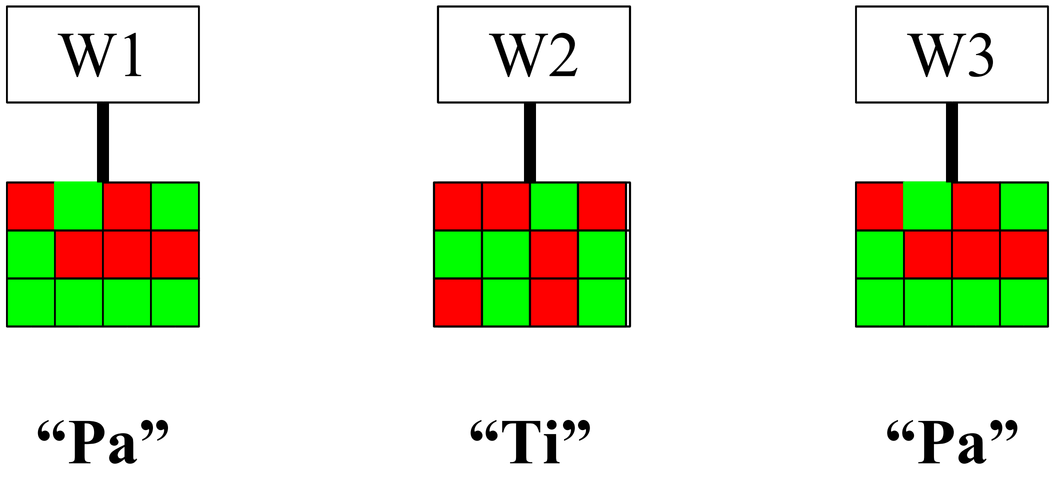


Aligning symmetrically-functional parts of a flip-flop circuit

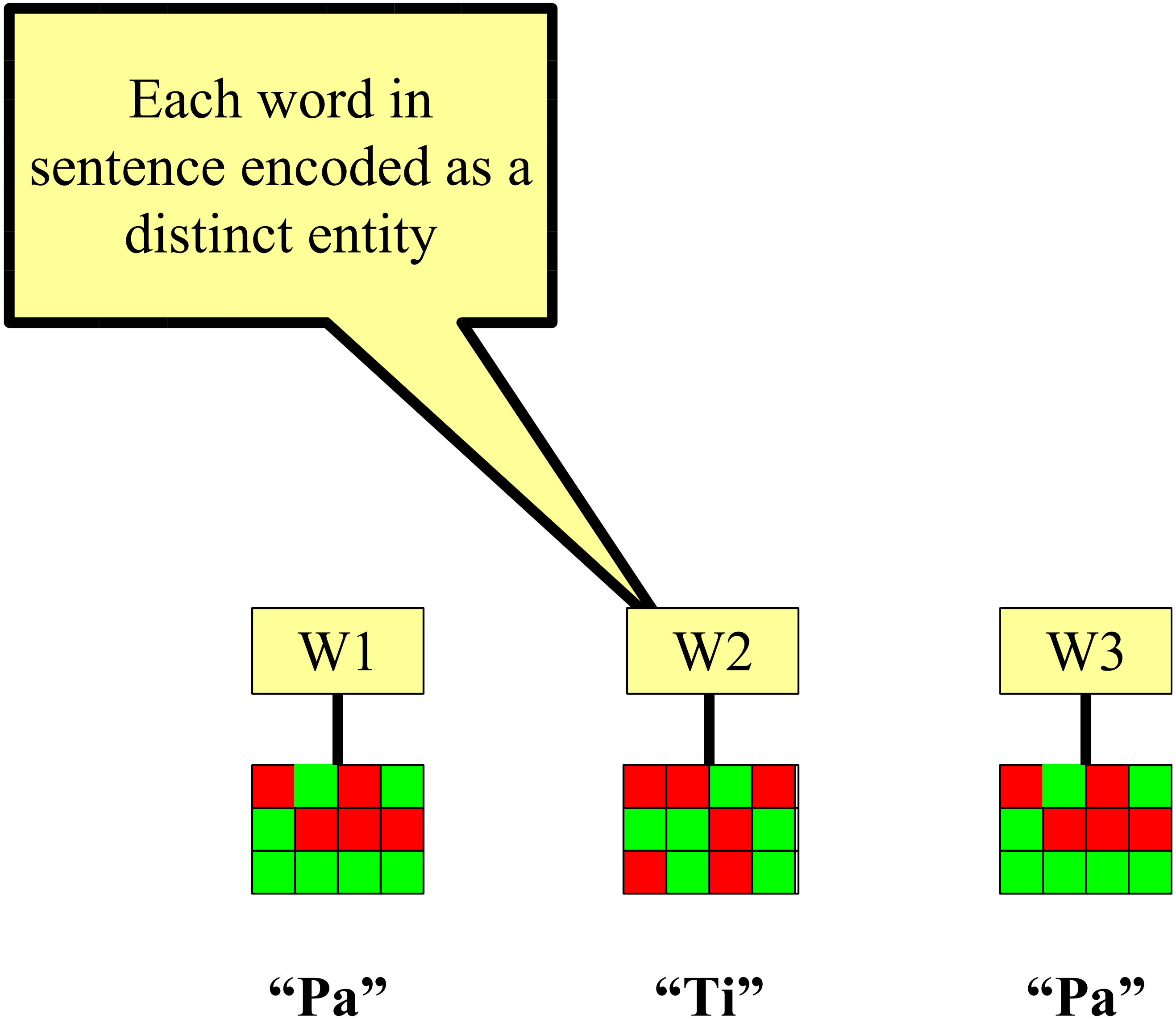
Our simulation



Sentence Encoding

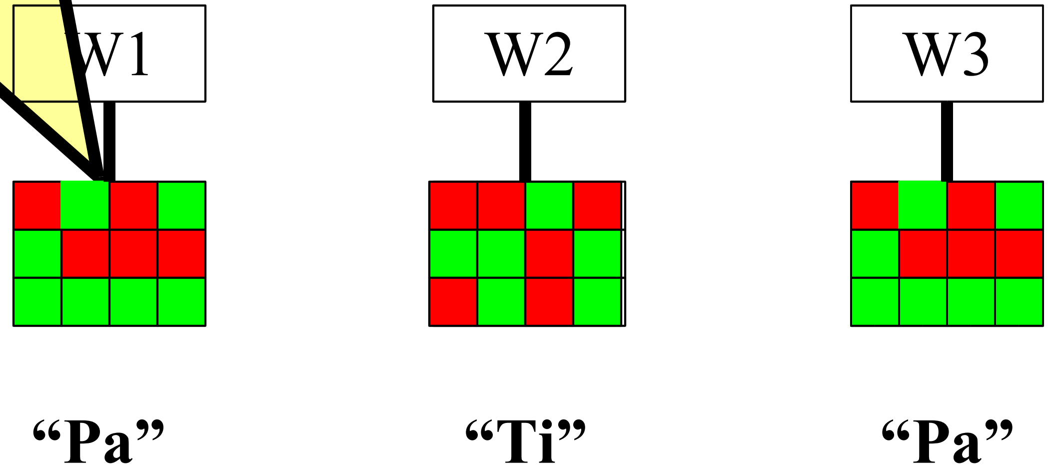


Sentence Encoding

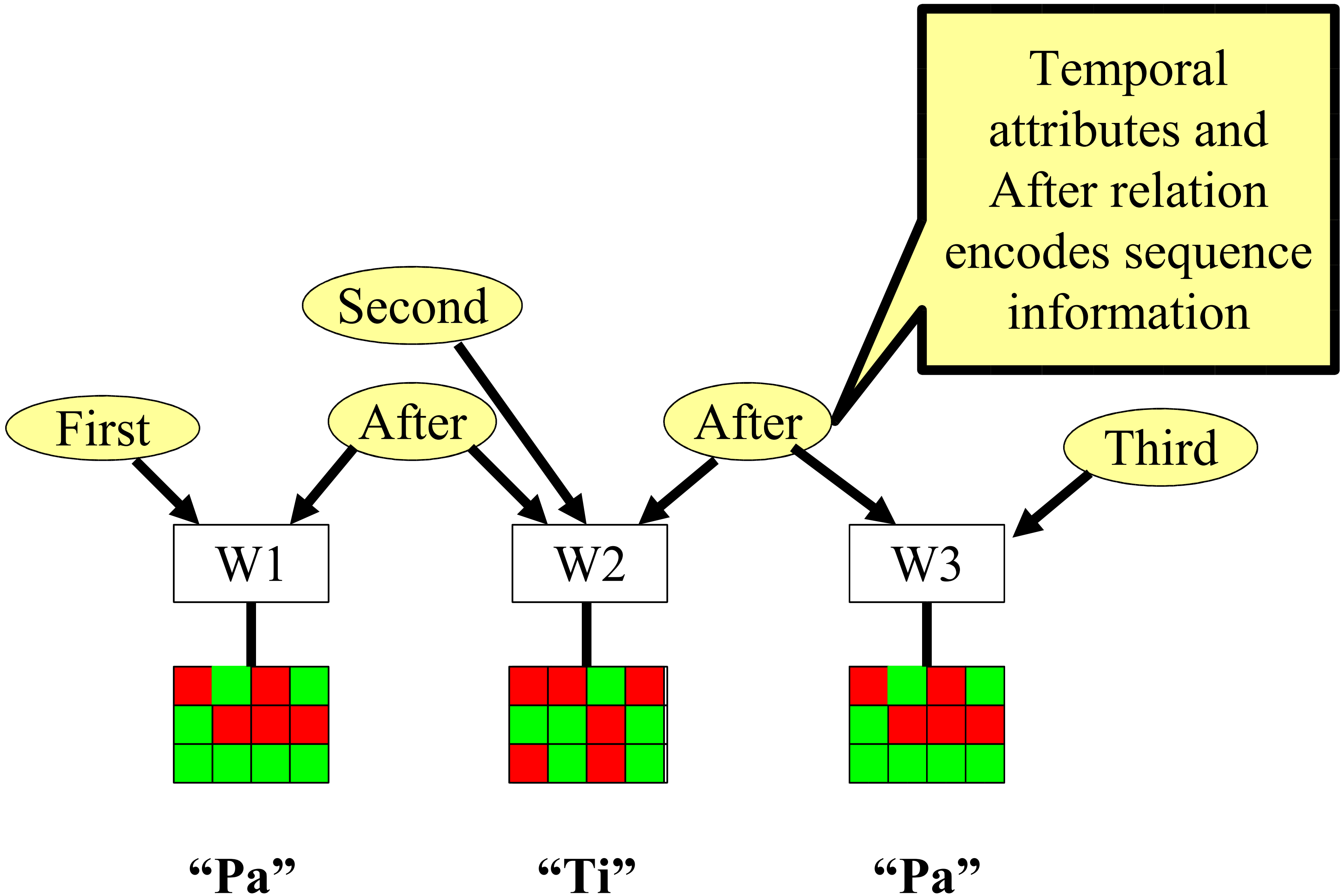


Sentence Encoding

Phonetic features encoded as set of 12 attribute choices, depending on whether present or absent, using Plunket & Marchman's distinctive feature notation (as used by Elman and other sims)

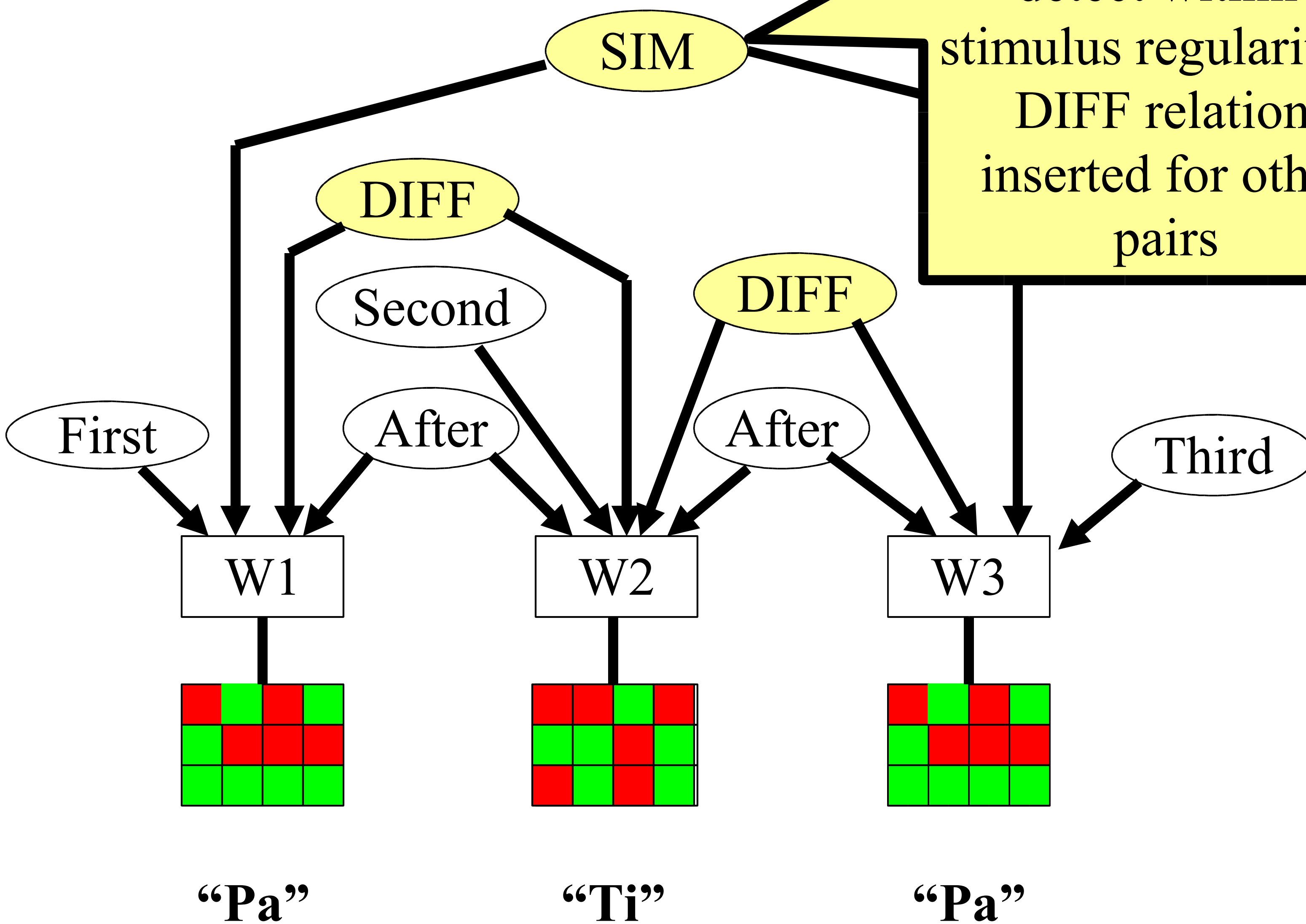


Sentence Encoding

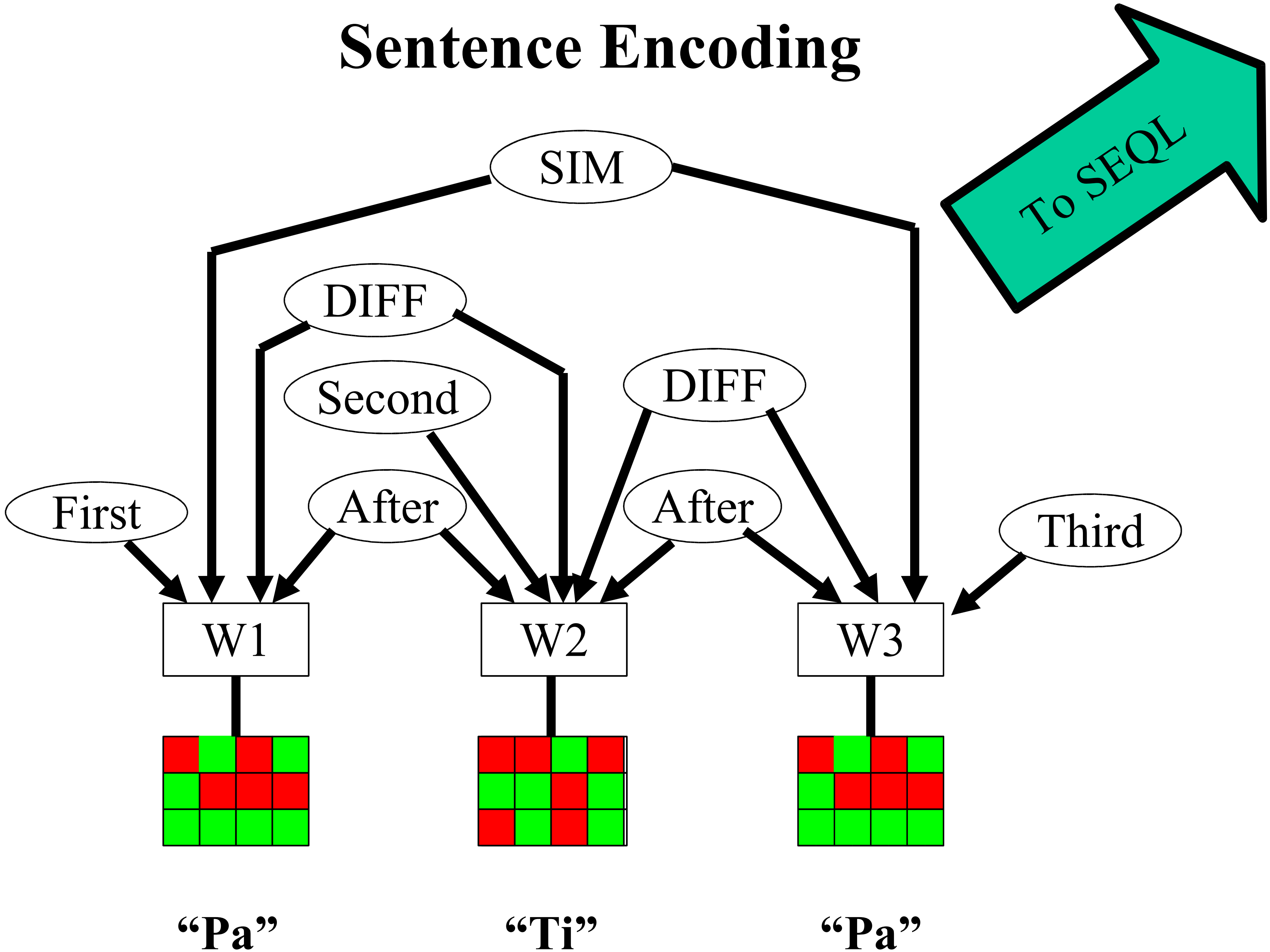


Sentence Encodin

MAGI used to detect within-stimulus regularities. DIFF relations inserted for other pairs



Sentence Encoding



Simulation Experiment 1

- 20 simulation runs
  - Training: Randomly ordered sequence of ABA or ABB training stimuli, from Marcus experiments
  - Test: Stimuli from Marcus experiments
- Results
  - SEQL always produced single generalization
  - When matching in-grammar test stimuli
    - Structural evaluation score > 0.64
    - No relational candidate inferences
  - When matching out-of-grammar stimuli
    - Structural evaluation score < 0.5
    - Relational candidate inferences produced
    - Claim: Their presence and evaluation is what leads infants to process that stimuli more

Simulation Experiment 1

With ABA training

Test Stimulus	Match Score	Inferences
Ba-po-ba	0.66	0
Ko-ga-ko	0.69	0
Ba-po-po	0.48	3
Ko-ga-ga	0.45	3

With ABB training

Test Stimulus	Match Score	Inferences
Ba-po-ba	0.33	2
Ko-ga-ko	0.35	2
Ba-po-po	0.78	0
Ko-ga-ga	0.75	0

Simulation Experiment 2

- Same setup as experiment one, but with noise
  - For each sentence, pick one word at random, then drop or flip one of its attributes at random.
  - In this representation, single flip can change word identity
- Results: Same pattern as previous experiment
  - More variation in match scores, inferences generated
  - Same criterion suffices to distinguish in-grammar from out-of-grammar stimuli

Simulation Experiment 2

With ABA training

Test Stimulus	**Ave Match Score**	**Inferences** (min, mean, max)
Ba-po-ba	0.65	0, 0, 0
Ko-ga-ko	0.68	0, 0, 0
Ba-po-po	0.43	2, 2.45, 3
Ko-ga-ga	0.39	2,2.55, 3

With ABB training

Test Stimulus	**Ave Match Score**	**Inferences** (min, mean, max)
Ba-po-ba	0.34	2, 2, 2
Ko-ga-ko	0.35	2, 2.05, 3
Ba-po-po	0.80	0, 0, 0
Ko-ga-ga	0.78	0, 0, 0

Summary of simulation results

- Our simulation accounts for the Marcus data better than any other simulation we know of
  - Learns unsupervised
  - Learns within span of examples given to infants
  - Can handle small amounts of noise
  - Suggests relational candidate inferences involving basic relations (SIM, DIFF) as reason why infants attend to novel stimuli
  - Uses off-the-shelf simulation modules with independent evidence for their psychological plausibility

Issues for learning raised by this simulation

- Intermediate speed of learning
  - Faster than connectionist models
  - Slower than most symbolic models (e.g., EBL)
  - Seems appropriate tradeoff for modeling human learning
  - SEQL may need to incorporate more statistical notions to be robust in more complex and noisier tasks.
- What is appropriate psychological notion of rule?
  - SEQL does not produce “algebraic” rules as per Marcus
  - But correspondence provides variable-like functionality
  - Rules = very abstract generalizations?
    - c.f. Gentner & Medina, 1998