The Neural Basis of Thought and Language

Final

Review Session

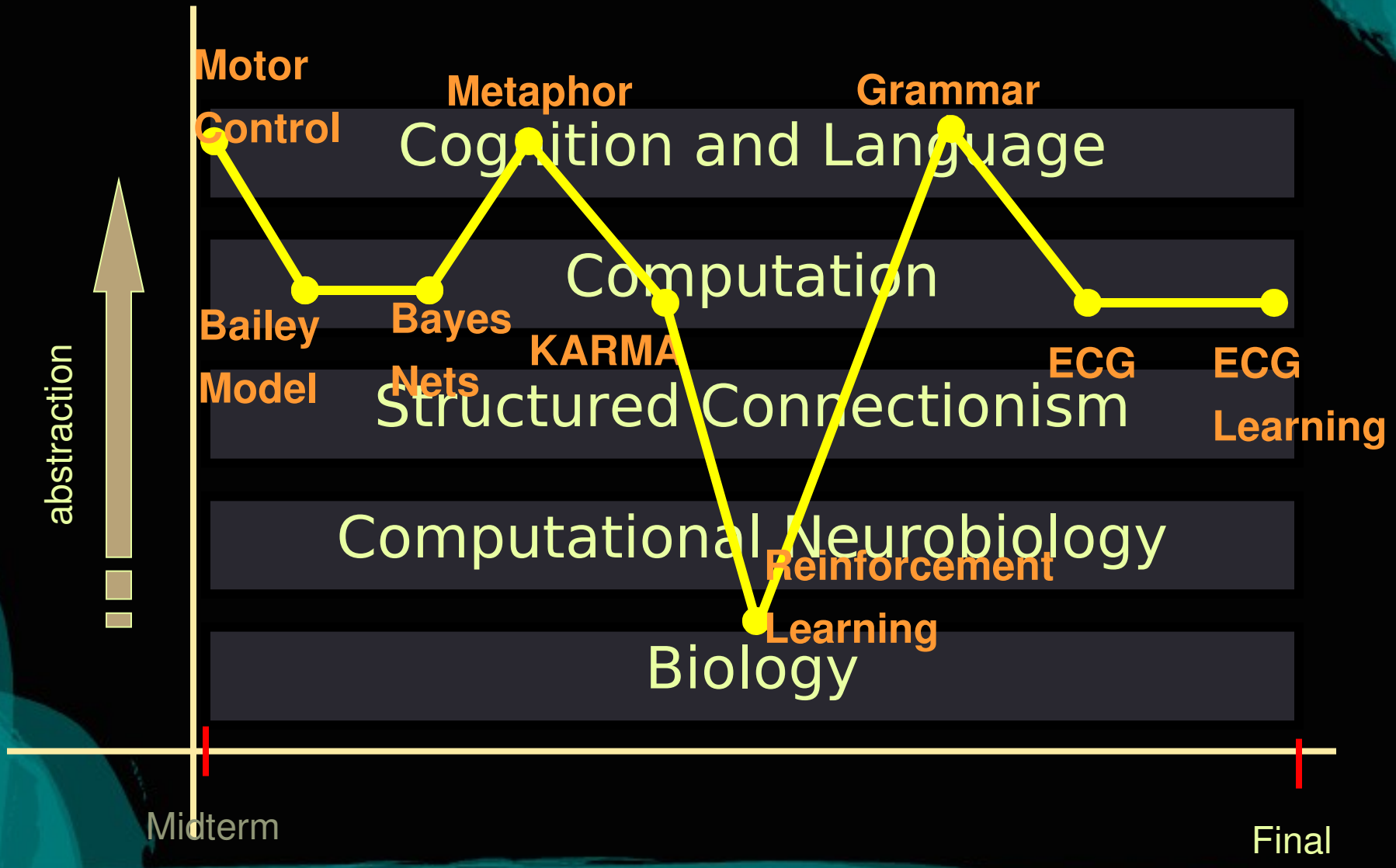
Administrivia

- Final in class Thursday, May 8th
  - Be there on time!
  - Format:
    - closed books, closed notes
    - short answers, no blue books
- A9 due tonight
- Final paper due on Monday, May 12

Resources

- Textbook!
- Class slides
- Section slides
- **Joe Makin's class notes from 2006**
  - on notes page

The Second Half

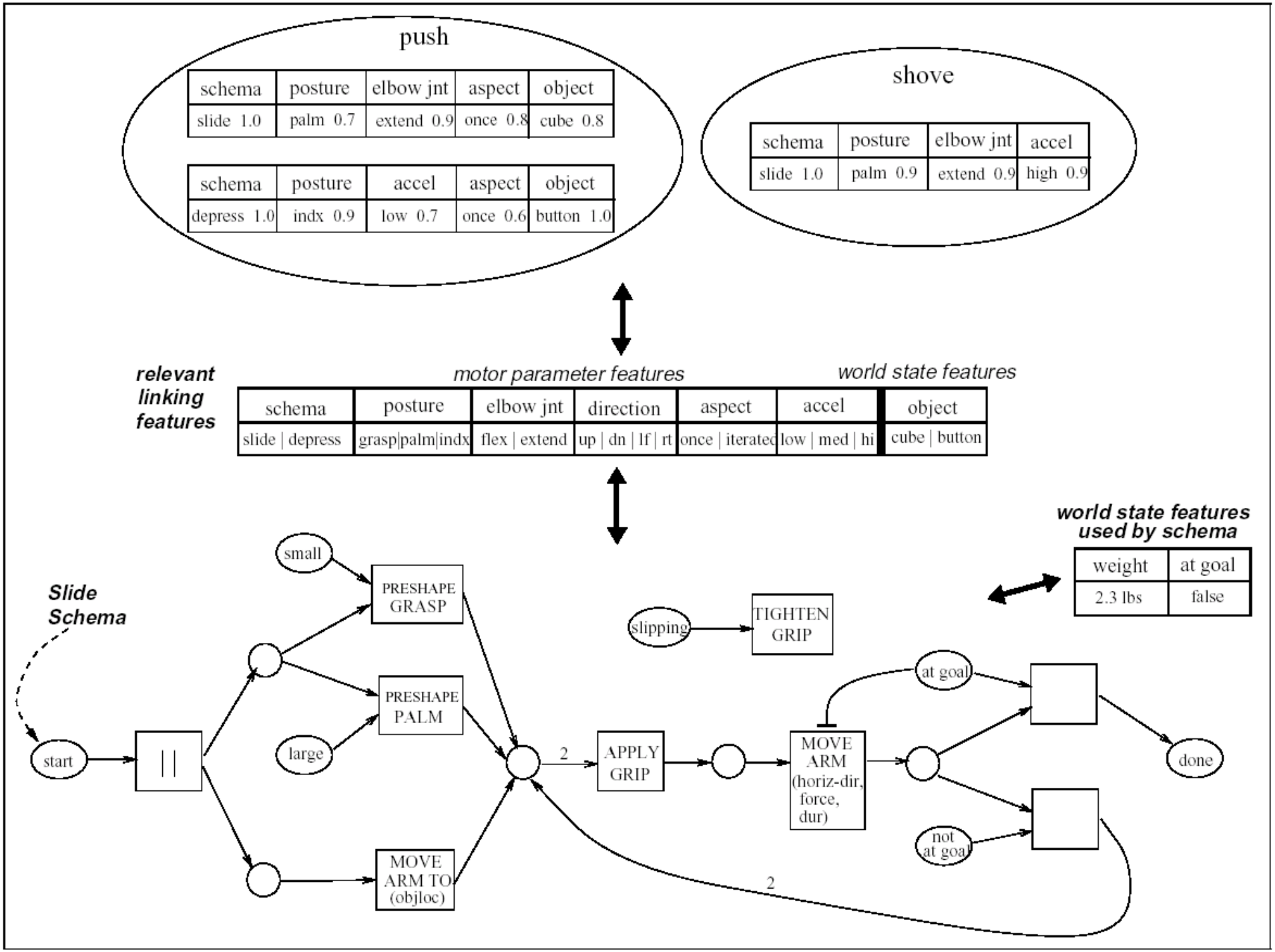


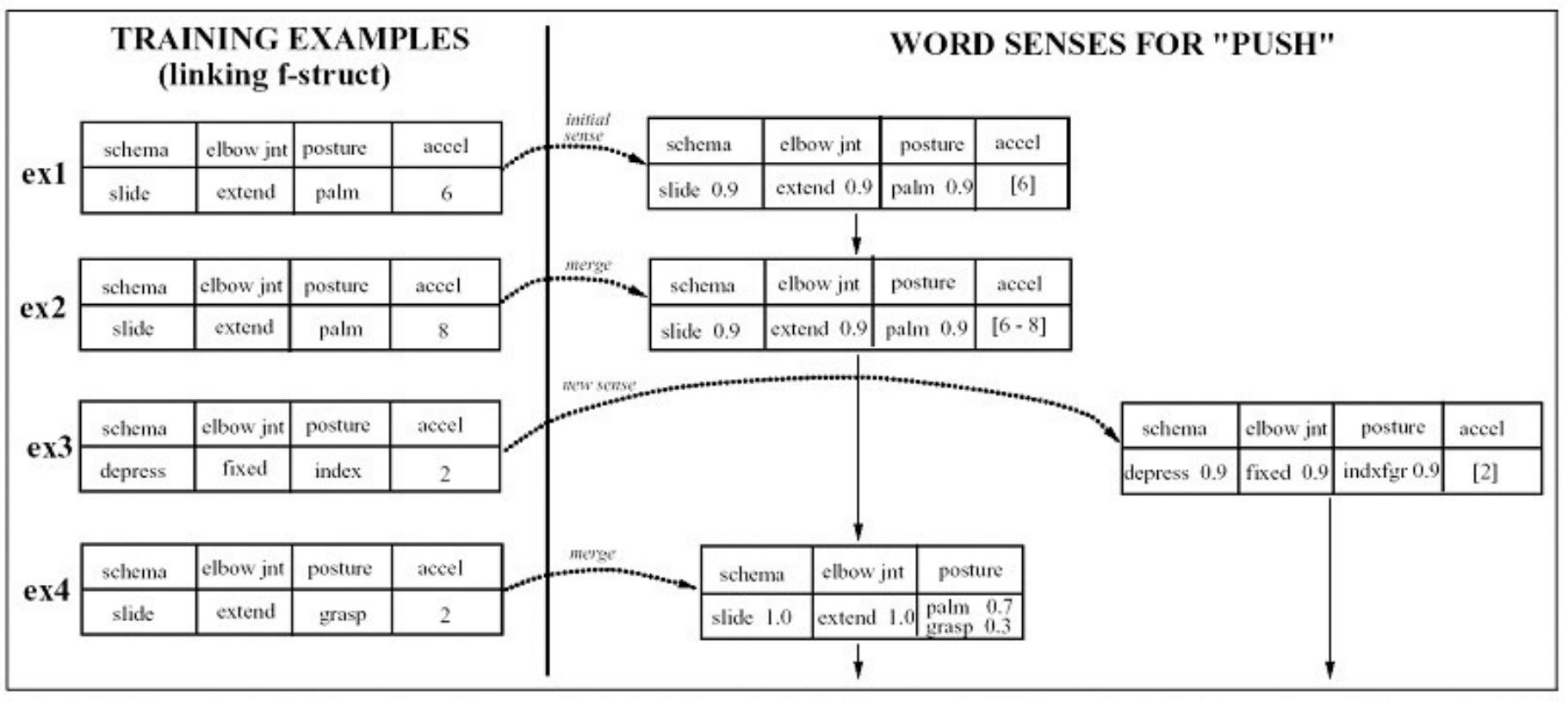
Overview

- Bailey Model
  - feature structures
  - Bayesian model merging
- Bayes Nets
- KARMA
  - X-schema, frames
  - aspect
  - event-structure metaphor
  - inference
- FrameNet
  - frames
  - image schemas
- Reinforcement Learning
- ECG
  - SemSpecs
  - parsing
  - constructions
  - learning algorithm

Bailey's VerbLearn Model

- 3 Levels of representation
    1. cognitive: words, concepts
    2. computational: f-structs, x-schemas
    3. connectionist: structured models, learning rules
- Input: labeled hand motions (f-structs)
- learning:
    1. the correct number of senses for each verb
    2. the relevant features in each sense, and
    3. the probability distributions on each included feature
- execution: perform a hand motion based on a label


push
schema | posture | elbow jnt | aspect | object
slide 1.0 | palm 0.7 | extend 0.9 | once 0.8 | cube 0.8
schema | posture | accel | aspect | object
depress 1.0 | indx 0.9 | low 0.7 | once 0.6 | button 1.0
shove
schema | posture | elbow jnt | accel
slide 1.0 | palm 0.9 | extend 0.9 | high 0.9
relevant linking features
motor parameter features
world state features
schema | posture | elbow jnt | direction | aspect | accel | object
slide | depress | grasp|palm|indx | flex | extend | up | dn | lf | rt | once | iterated | low | med | hi | cube | button
world state features used by schema
weight | at goal
2.3 lbs | false
Slide Schema
start
small
large
PRESHAPE GRASP
PRESHAPE PALM
MOVE ARM TO (objloc)
2
APPLY GRIP
slipping
TIGHTEN GRIP
MOVE ARM (horiz-dir, force, dur)
at goal
not at goal
done
2


TRAINING EXAMPLES
(linking f-struct)
WORD SENSES FOR "PUSH"
ex1
schema elbow jnt posture accel
slide extend palm 6
ex2
schema elbow jnt posture accel
slide extend palm 8
ex3
schema elbow jnt posture accel
depress fixed index 2
ex4
schema elbow jnt posture accel
slide extend grasp 2
initial sense
schema elbow jnt posture accel
slide 0.9 extend 0.9 palm 0.9 [6]
merge
schema elbow jnt posture accel
slide 0.9 extend 0.9 palm 0.9 [6 - 8]
new sense
schema elbow jnt posture accel
depress 0.9 fixed 0.9 indxfgr 0.9 [2]
merge
schema elbow jnt posture
slide 1.0 extend 1.0 palm 0.7 grasp 0.3

Bayes Nets

- Probability
- Bayes' Rule / Product Rule
  - $P(x,y) = P(x)\ P(y|x)$
    $= P(y)\ P(x|y)$
  - $P(x|y) = P(x)\ P(y|x) / P(y)$
- Write factored distribution $P(x,y,z,...) = P(x)\ P(y|x)$ ...
- Infer distributions over variables given evidence
  - variable elimination (by summation: P(x) = sum_y P(x,y))
- Temporal Bayes' Nets

Event Structure Metaphor

- States are Locations
- Changes are Movements
- Causes are Forces
- Causation is Forced Movement
- Actions are Self-propelled Movements
- Purposes are Destinations
- Means are Paths
- Difficulties are Impediments to Motion
- External Events are Large, Moving Objects
- Long-term, Purposeful Activities are Journeys

Ego Moving versus Time Moving

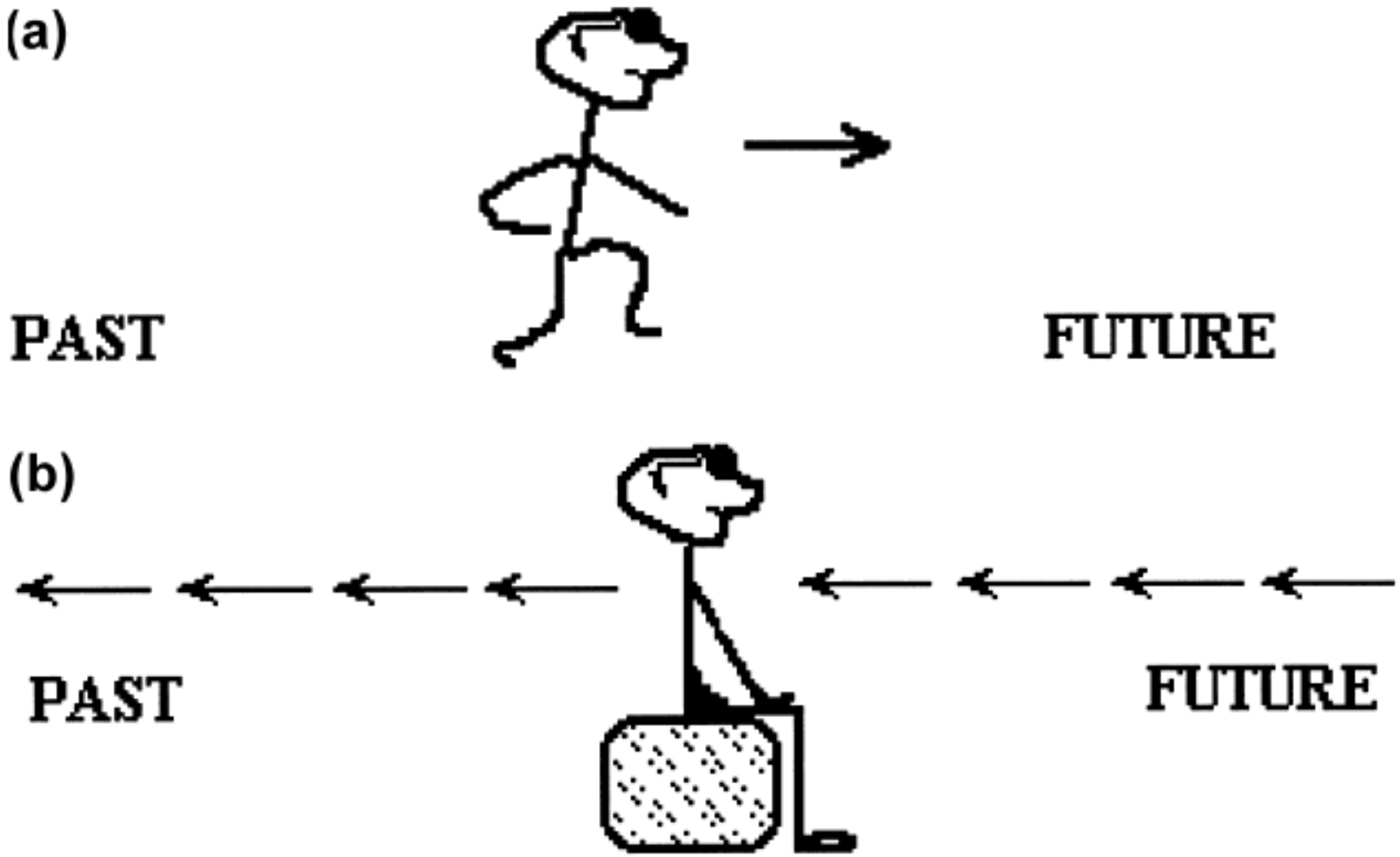
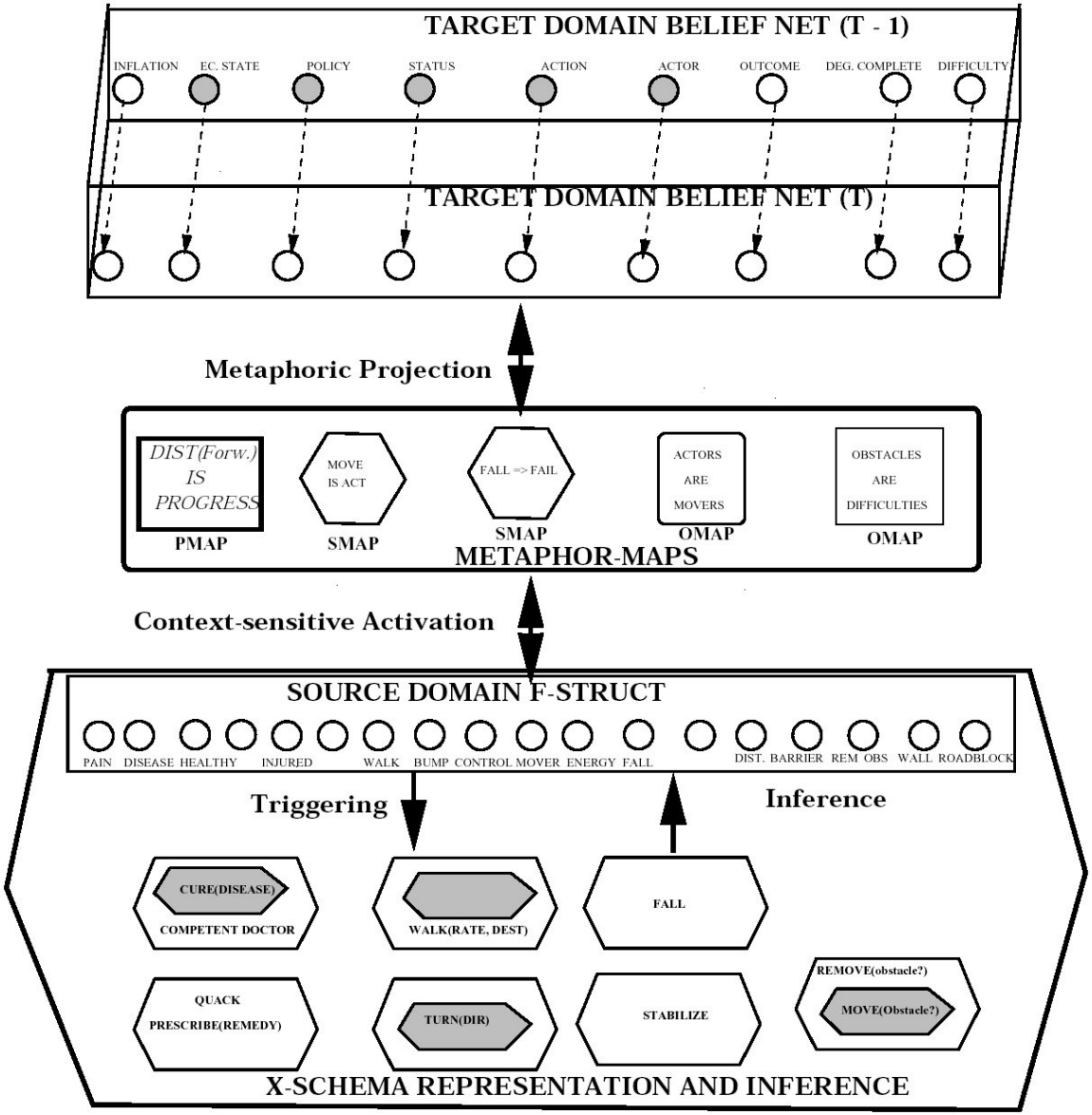


Fig. 1. (a) Schematic of the ego-moving schema used to organize events in time. (b) Schematic of the time-moving schema used to organize events in time.

Results

PRIME	Meeting is Monday	Meeting is Friday
Ego Moving	26.7%	73.3%
Object Moving	69.2%	30.8%



KARMA

- DBN to represent target domain knowledge
- Metaphor maps link target and source domain
- X-schema to represent source domain knowledge

X-Schemas

- Active representation
- Has hierarchical actions
  - defined by network structure
- Actions have structure (e.g. ready, iterating, ongoing, failed, complete)
  - defined by network structure
- Properly-designed nets will be goal-directed
  - take best actions to reach goal, given current context
  - related to “reinforcement learning”

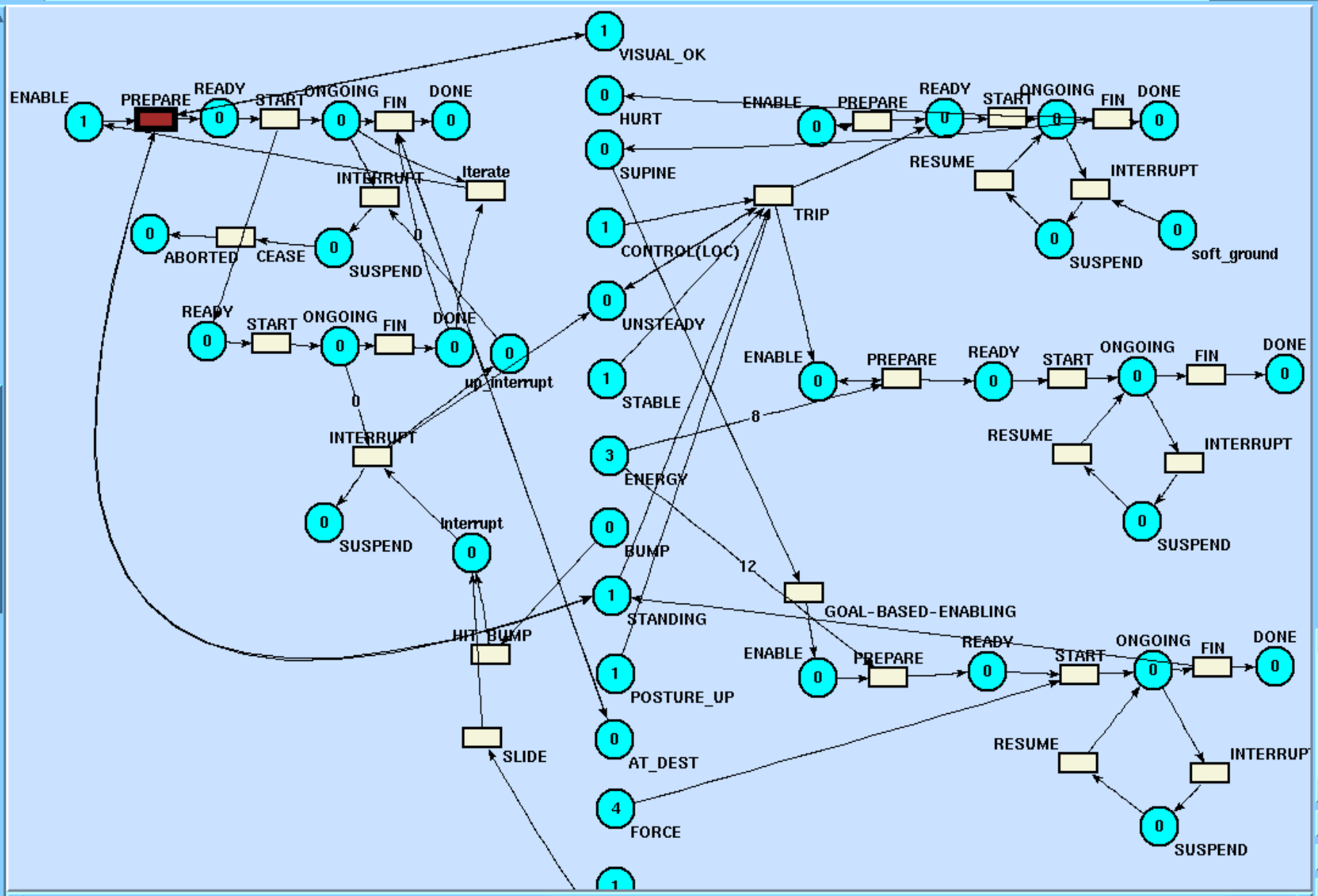
KARMASIM

Net Analysis State Display Options Help

loaded at 842 1123 1918 1893

Name: /u/snarayan/code/simulator/compose

Move Copy Arc



State

Fire

Init

Save

Reinforcement Learning

- unsupervised learning
- learn behaviors
- reward
  - discounts
- use estimated future value

$$V(s) = max_a Q(s,a)$$
$$Q(s,a) = E[r(s,a) + \gamma V(s')]$$

Reinforcement Learning

- Learning methods
  - Value iteration
  - Q-learning
- Biology
  - dopamine = reward difference
    - only for reward, not punishment
  - non-exponential discounting
    - preference switching

Language

- Grammar
  - Syntax
- Semantics
- Metaphor
- Simulation
- Unification

Grammar

- A grammar is a set of rules defining a formal language
- a common example is Context-Free Grammar

$$\alpha \rightarrow \beta$$

$\forall$ α : single non-terminal

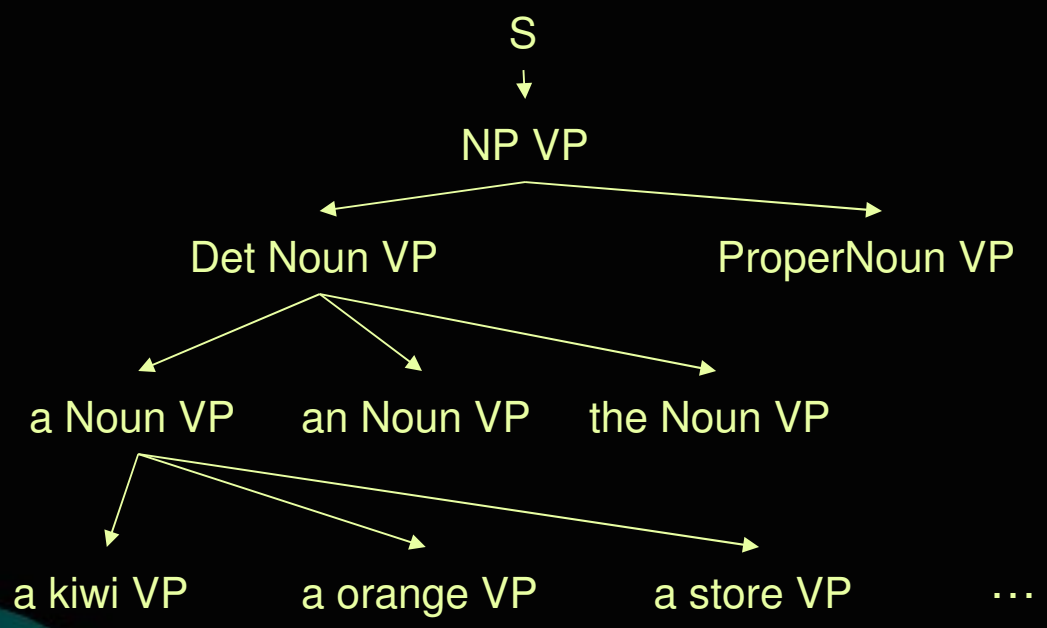
- β : any combination of terminals and non-terminals

S $\rightarrow$ NP VP
NP $\rightarrow$ Det Noun | ProperNoun
VP $\rightarrow$ Verb NP | Verb PP
PP $\rightarrow$ Preposition NP

Noun $\rightarrow$ kiwi | orange | store
ProperNoun $\rightarrow$ Pat | I
Det $\rightarrow$ a | an | the
Verb $\rightarrow$ ate | went | shop
Preposition $\rightarrow$ to | at

Sentence generation: *Pat ate the kiwi*

- start from S and apply any applicable rules
- forward expansion



S → NP VP

NP → Det Noun | ProperNoun

VP → Verb NP | Verb PP

PP → Preposition NP

Noun → kiwi | orange | store

ProperNoun → Pat | I

Det → a | an | the

Verb → ate | went | shop

Preposition → to | at

Unification Grammar

- Basic idea: capture these agreement features for each non-terminal in feature structures

I $\begin{pmatrix} \text{agreement} & \begin{pmatrix} \text{number : SG} \\ \text{person : 1st} \end{pmatrix} \end{pmatrix}$

Pat $\begin{pmatrix} \text{agreement} & \begin{pmatrix} \text{number : SG} \\ \text{person : 3rd} \end{pmatrix} \end{pmatrix}$

Went $\begin{pmatrix} \text{agreement} \end{pmatrix}$

Shop $\begin{pmatrix} \text{agreement} & \begin{pmatrix} \text{number :} \\ \text{person : 1st} \end{pmatrix} \end{pmatrix}$

Enforce constraints on these features using unification rules

VP $\rightarrow$ Verb NP

VP.agreement $\leftrightarrow$ Verb.agreement

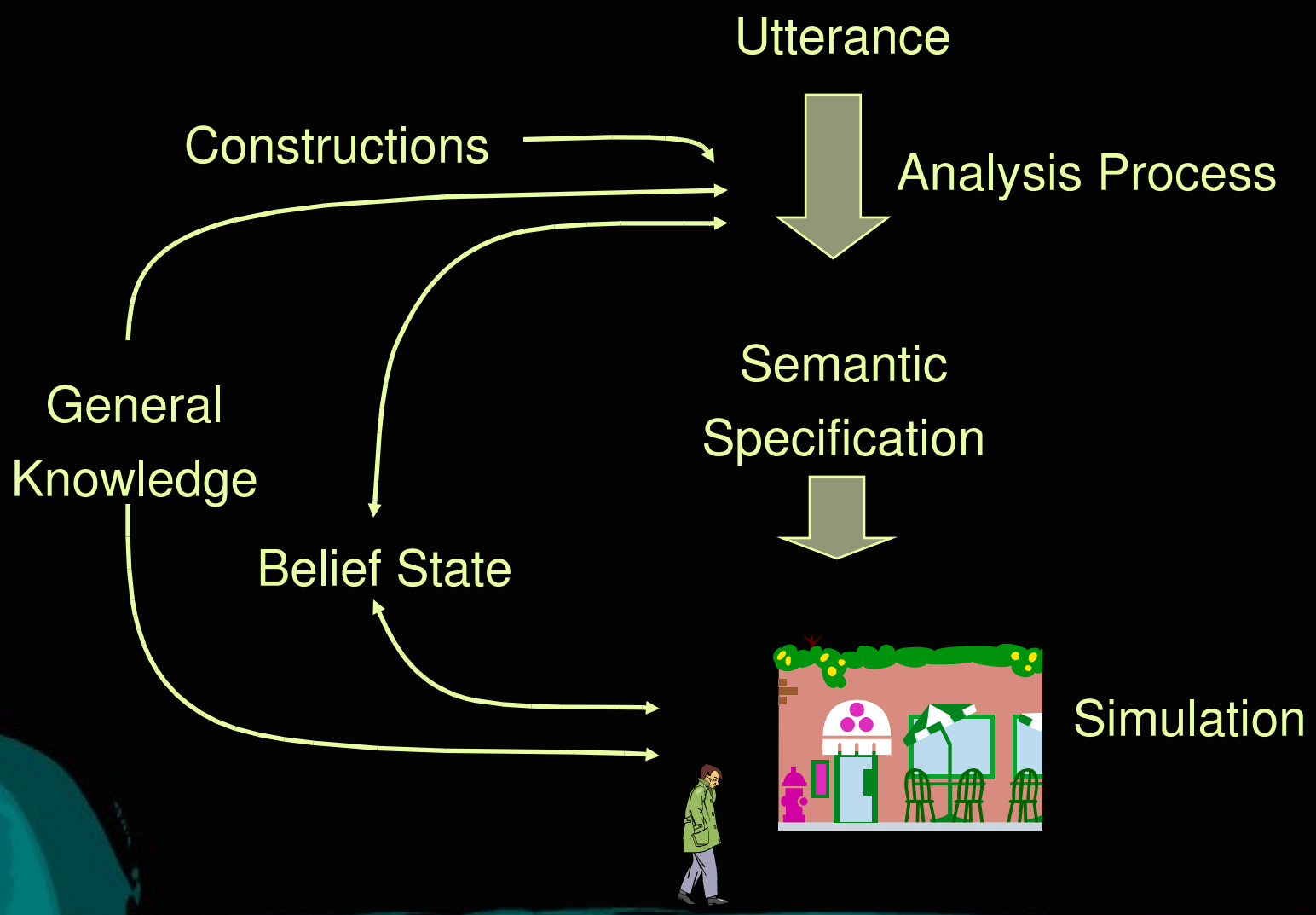
S $\rightarrow$ NP VP

NP.agreement $\leftrightarrow$ VP.agreement

Poverty and Opulence

- Poverty of the stimulus
  - Coined to suggest how little information children have to learn from
- Opulence of the substrate
  - Opulence = “richness”
  - Coined in response to suggest how much background information children have

“Harry walked into the café.”



The INTO construction

construction INTO
  subcase of Spatial-Relation
    form
      $self_f$ .orth ← “into”
    meaning: Trajector-Landmark
      evokes Container as cont
      evokes Source-Path-Goal as spg
      trajector ↔ spg.trajector
      landmark ↔ cont
      cont.interior ↔ spg.goal
      cont.exterior ↔ spg.source

The Spatial-Phrase construction

construction SPATIAL-PHRASE

constructional

constituents

sr : Spatial-Relation

lm : Ref-Expr

form

sr_f *before* lm_f

meaning

sr_m.landmark ↔ lm_m

The Directed-Motion construction

construction DIRECTED-MOTION

constructional

constituents

a : Ref-Exp

m: Motion-Verb

p : Spatial-Phrase

form

a_f *before* m_f

m_f *before* p_f

meaning

evokes Directed-Motion as dm

$self_m$.scene ↔ dm

dm.agent ↔ a_m

dm.motion ↔ m_m

dm.path ↔ p_m

schema Directed-Motion

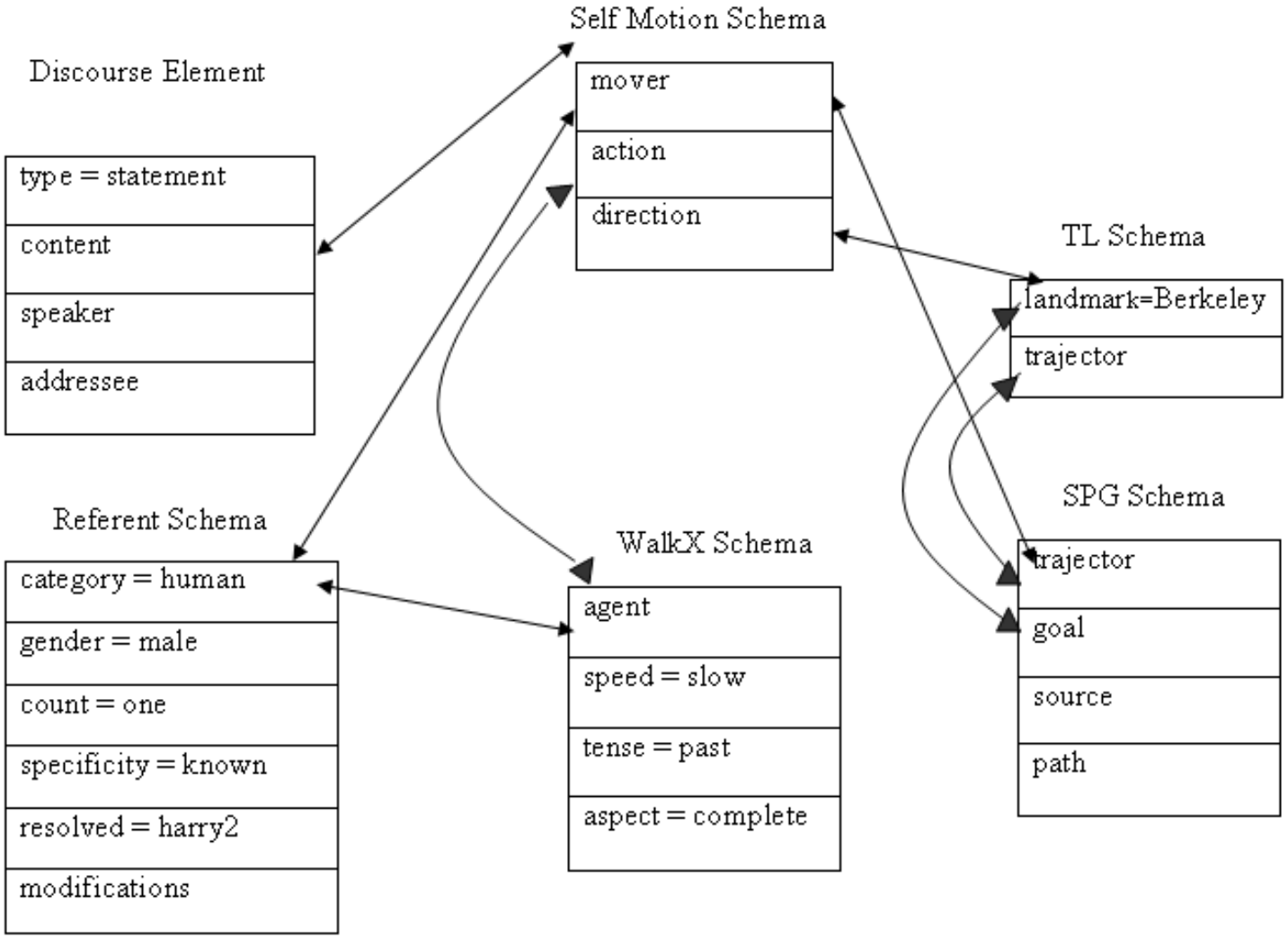
roles

agent : Entity

motion : Motion

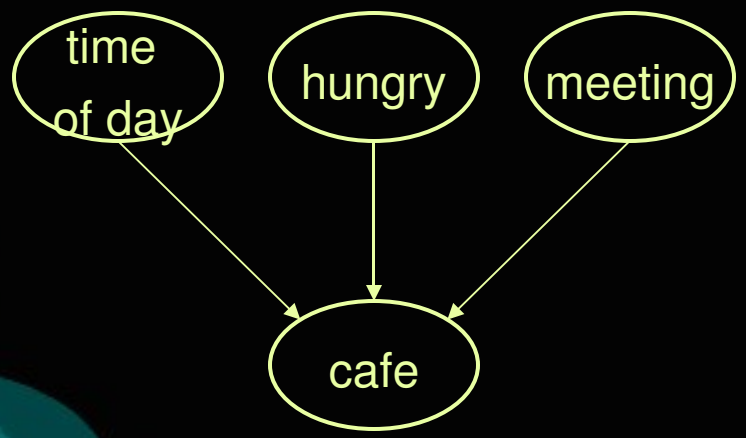
path : SPG

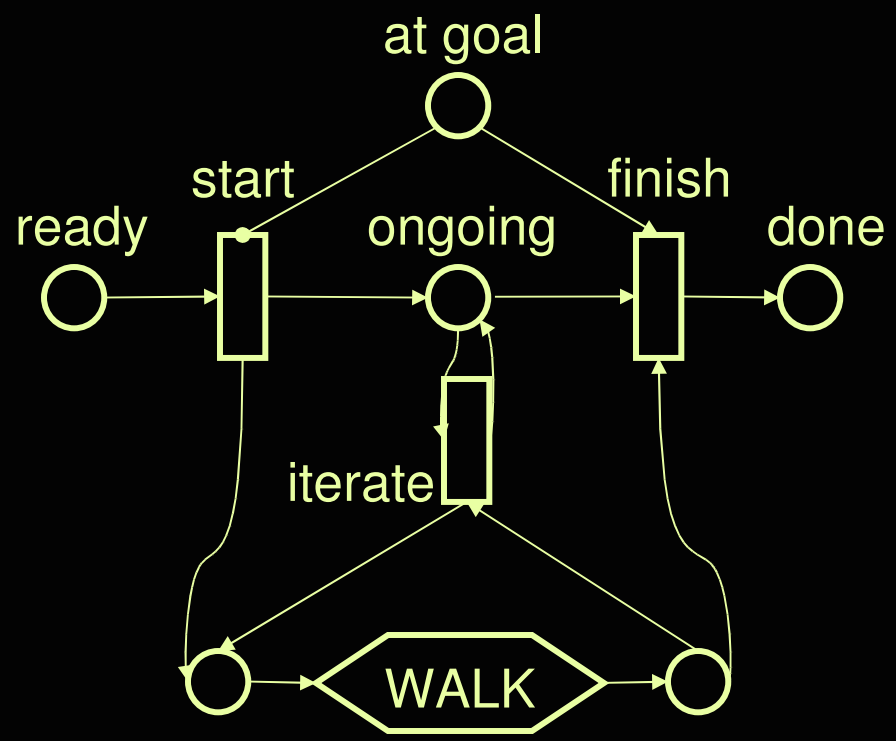
Do not forget the SemSpec!



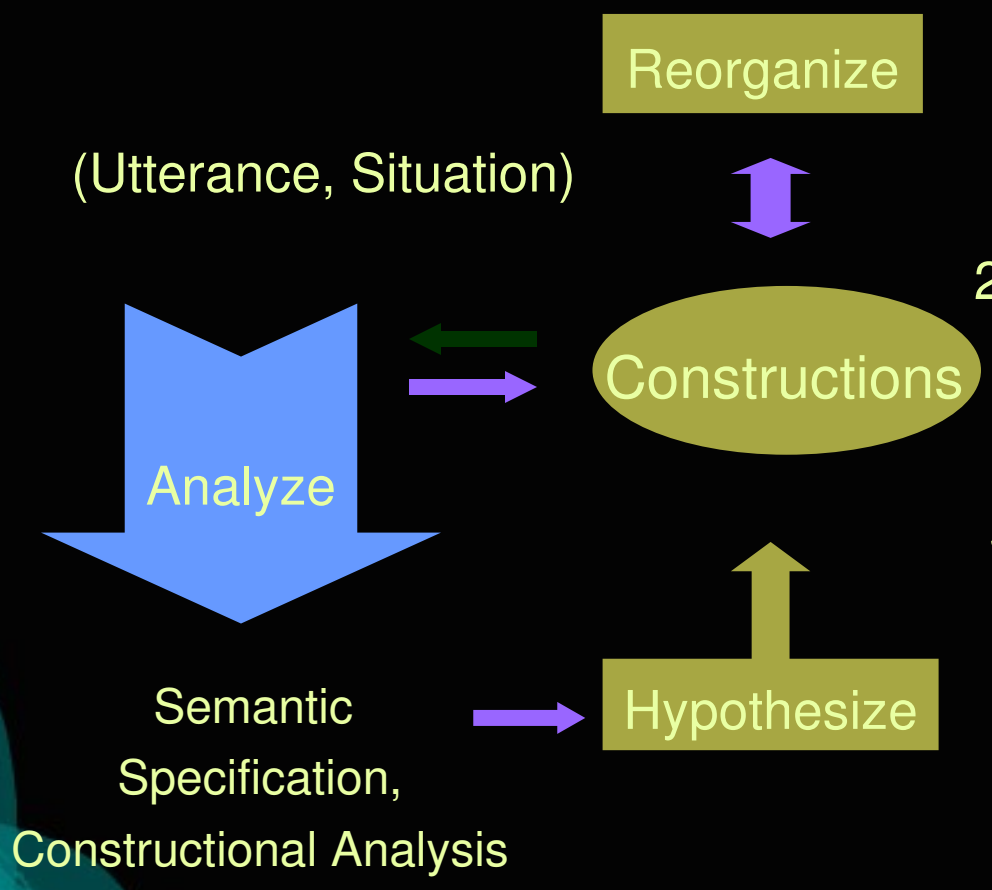
What exactly is simulation?

- Belief update and/or X-schema execution





Learning-Analysis Cycle (Chang, 2004)



1. Learner passes input (Utterance + Situation) and current grammar to Analyzer.
2. Analyzer produces SemSpec and Constructional Analysis.
3. Learner updates grammar:
   a. Hypothesize new map.
   b. Reorganize grammar (merge or compose).
   c. Reinforce (based on usage).

Three ways to get new constructions

- Relational mapping
  - **throw** the **ball**

} THROW < BALL

- Merging
  - **throw** the **block**
  - **throw**ing the **ball**

} THROW < OBJECT

- Composing
  - **throw** the **ball**
  - **ball off**
  - you **throw** the **ball off**

} THROW < BALL < OFF

Grammar merging

- How can we measure description length?
  - complicated rules are bad
  - lots of rules are bad
    - measure “derivation length”
  - alpha * size(rules) + derivationCost(rules, sentences)

How do you learn…

the meanings of spatial relations,

the meanings of verbs,

the metaphors, and

the constructions?

How do you learn…

the meanings of spatial relations,

the meanings of verbs,

the metaphors, and

the constructions?

That’s the Regier model.

How do you learn…

the meanings of spatial relations,

the meanings of verbs,

the metaphors, and

the constructions?

That's Bailey's model

How do you learn…

the meanings of spatial relations,

the meanings of verbs,

the metaphors, and

the constructions?

conflation hypothesis
(primary metaphors)

How do you learn…

the meanings of spatial relations,

the meanings of verbs,

the metaphors, and

the constructions?

construction learning

further questions?