The Neural Basis of Thought and Language

Final

Review Session

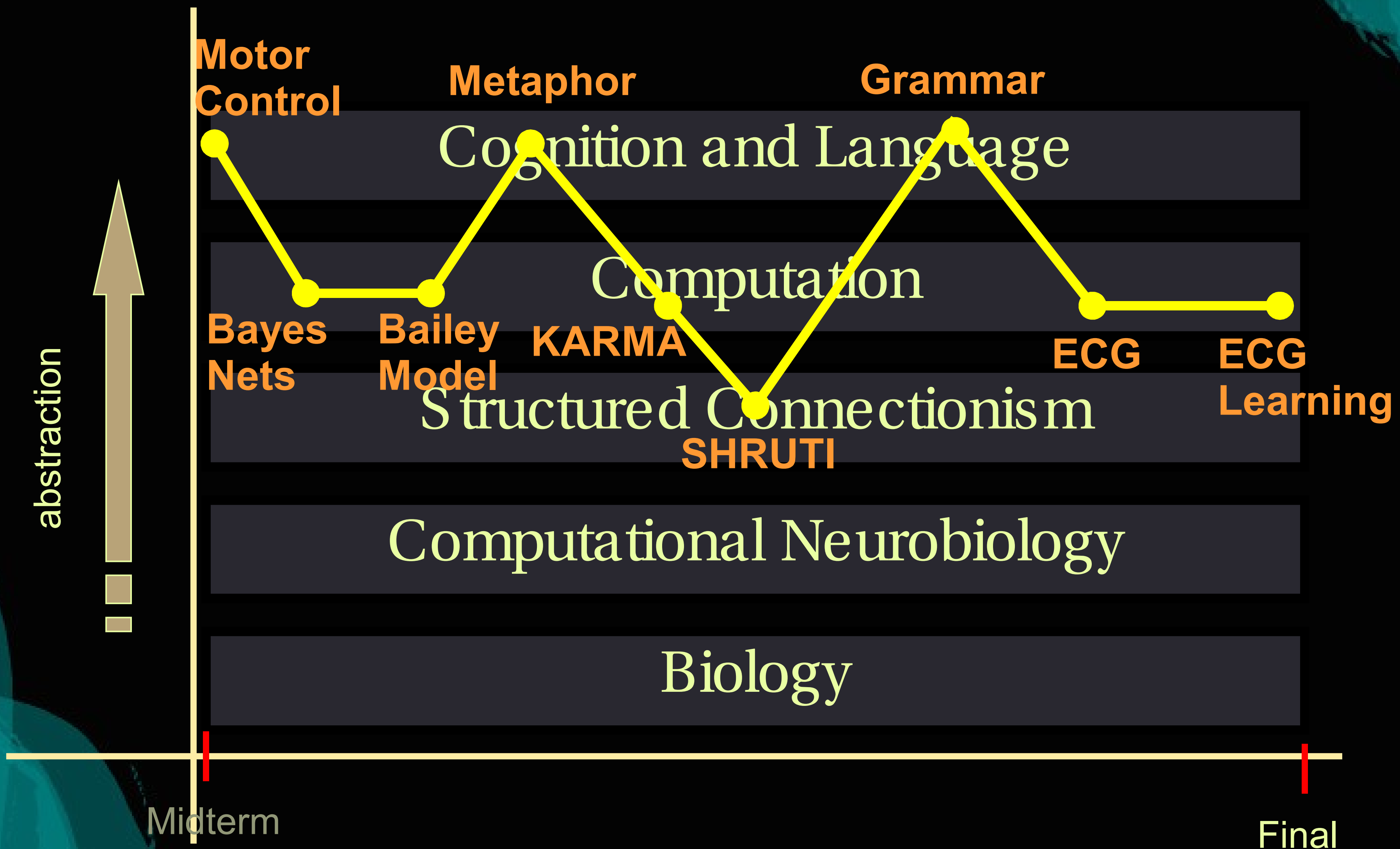
Administrivia

- Final in class next Tuesday, May 8th
- Be there on time!
- Format:
  - closed books, closed notes
  - short answers, no blue books
- Final paper due on bSpace on Friday, May 11

Resources

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

The Second Half



Overview

- Bailey Model
  - feature structures
  - Bayesian model merging
  - recruitment learning
- KARMA
  - X-schema, frames
  - aspect
  - event-structure metaphor
  - inference
- Grammar Learning
  - parsing
  - construction grammar
  - learning algorithm
- SHRUTI
- FrameNet
- Bayesian Model of Human Sentence Processing

Important topics

- Regier's model of spatial relation learning
- Bailey's model of verb learning
- KARMA model of metaphor
- Binding and inference
  - SHRUTI, short signatures
- Grammars and learning
- ECG
  - Learning ECG
- Bayes nets
- Model merging, MDL
- Petri nets
- Language
  - Metaphor
  - Aspect
  - Grammars
  - Schemas
  - Frames
  - SemSpec

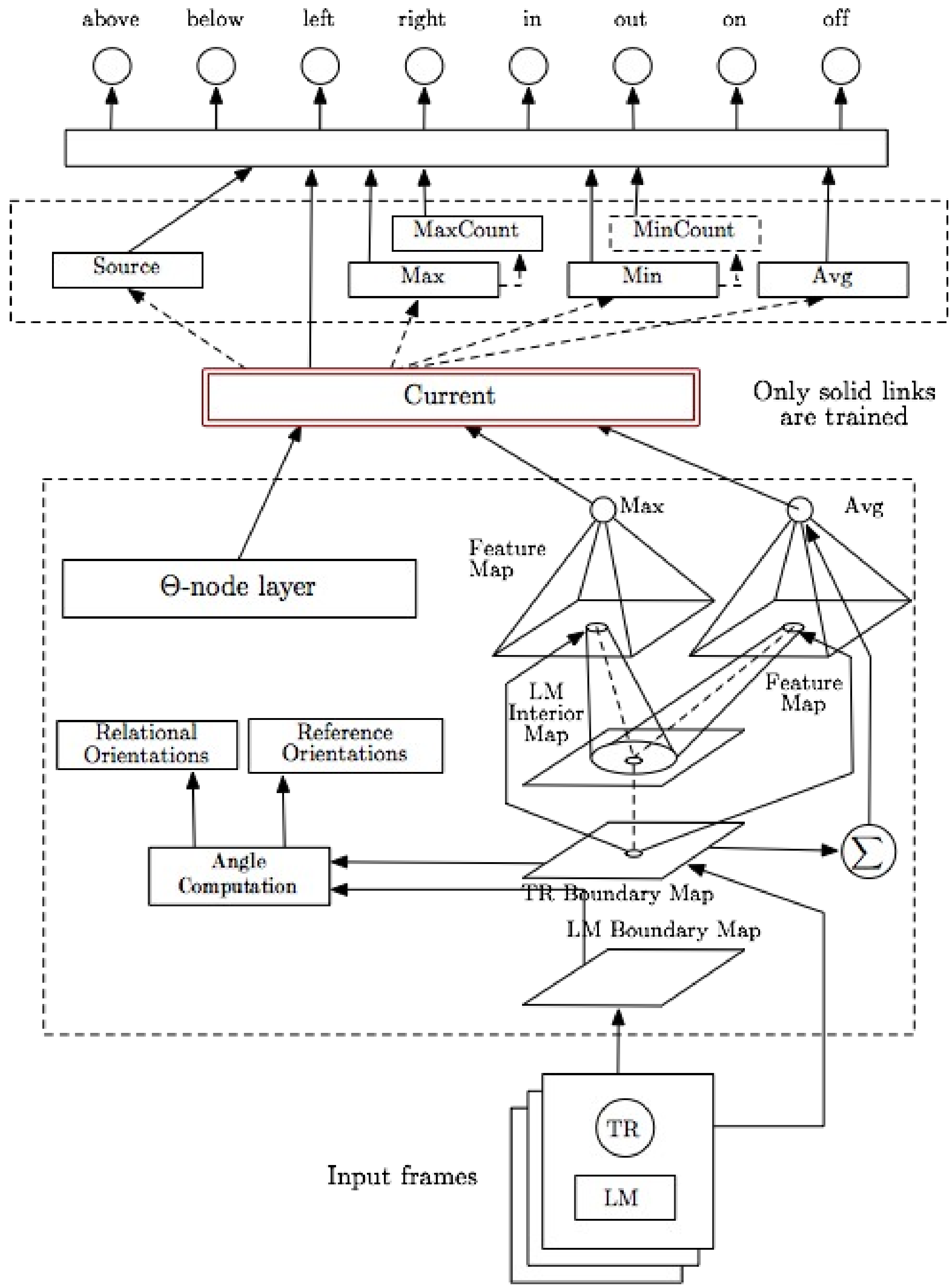
Q & A

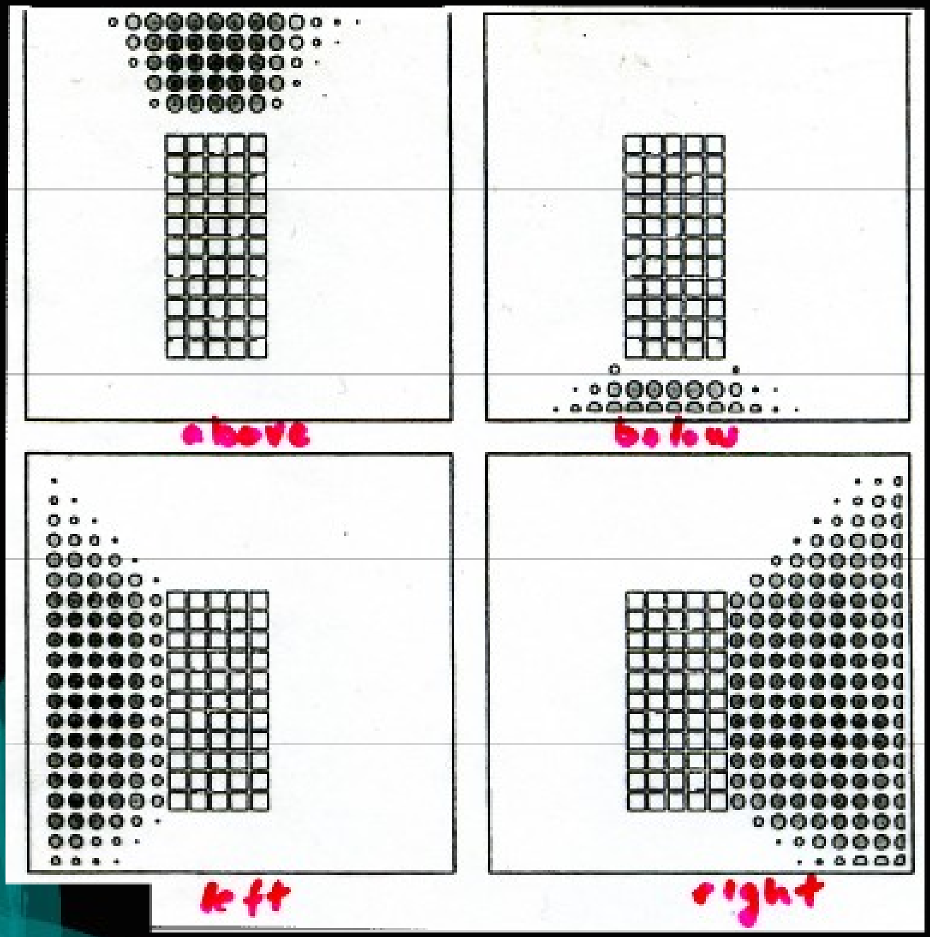
Bayes Nets

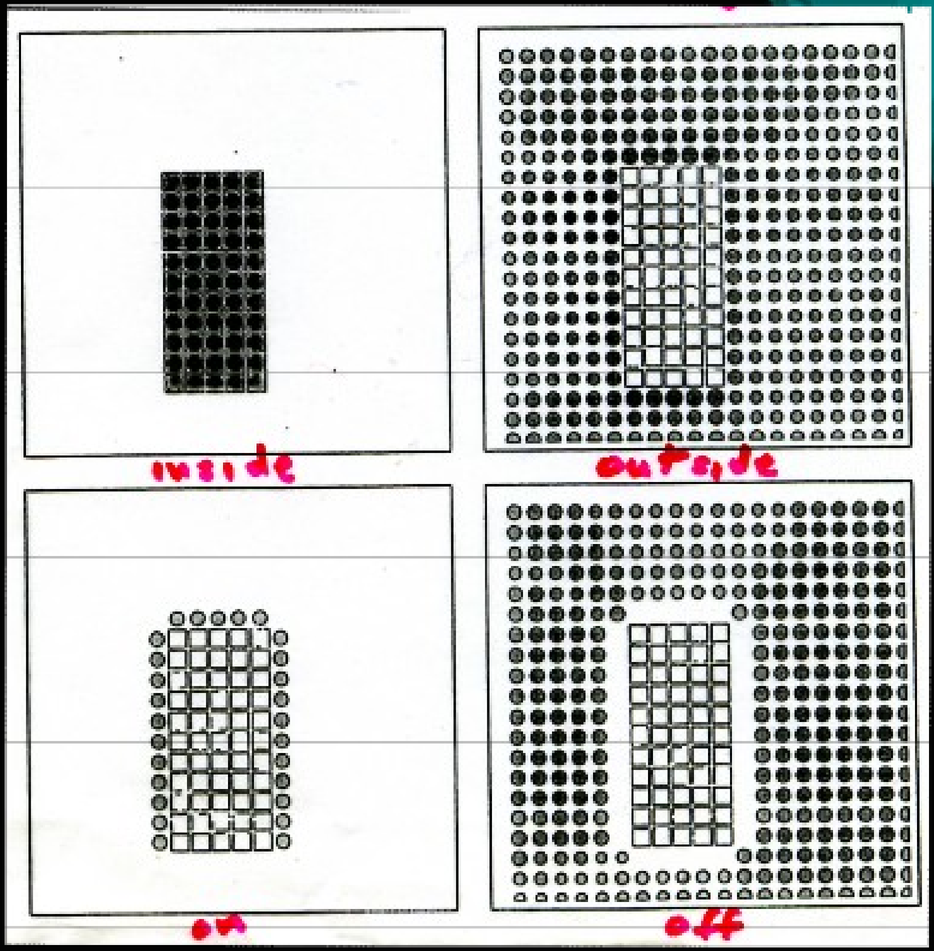
- Bayes' Rule / Product Rule
  - P(A|B) = P(A,B) / P(B)
  - P(A,B) = P(A|B) P(B)
  - P(B|A) = P(A|B) P(B) / P(A)
  - All the same!
- Variables have distributions
- Variables depend on other variables

Regier's model

- Learn spatial relation terms
  - e.g. in, on, above
- Neural network + hand-designed “vision” parts

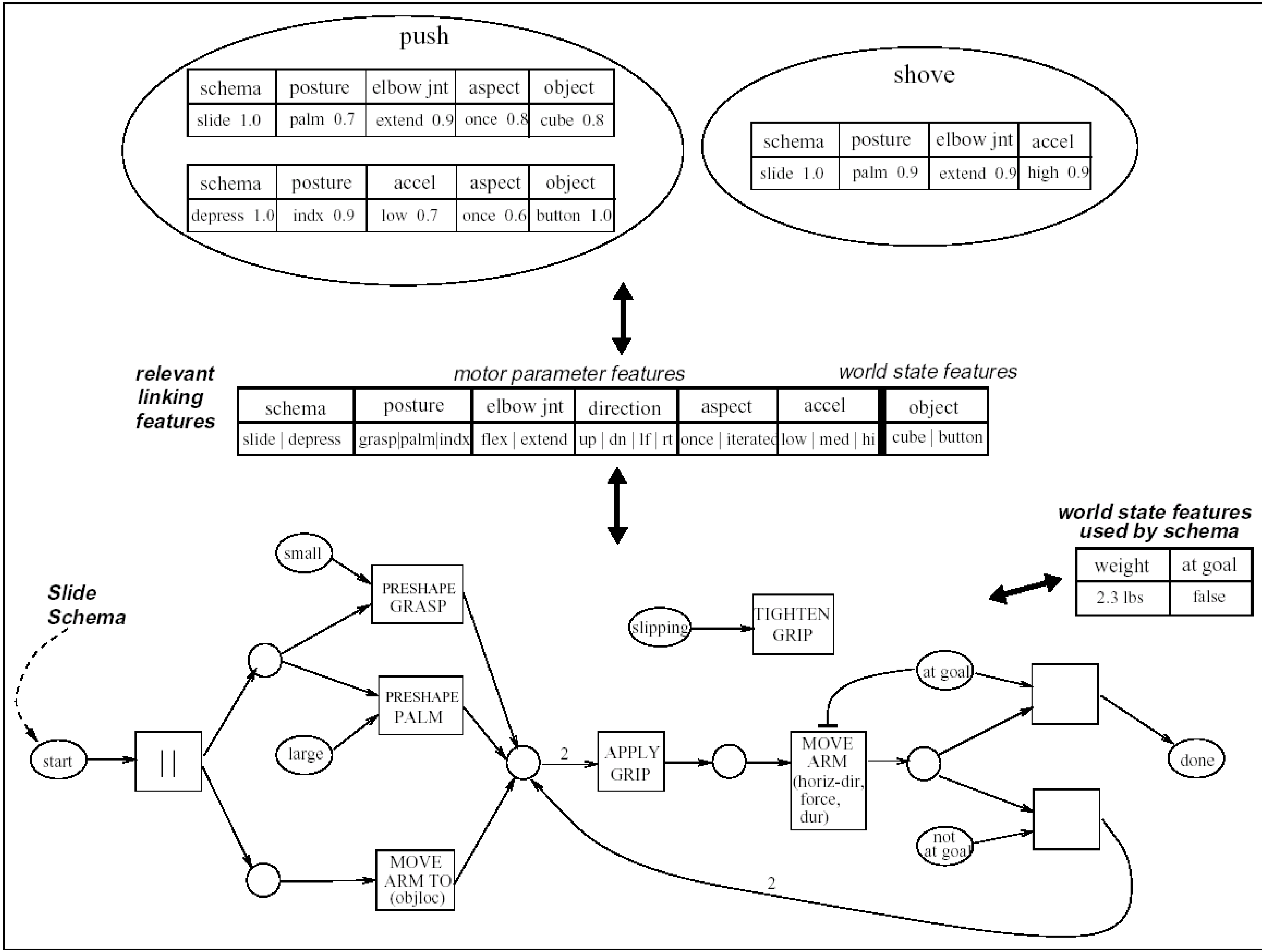

above
below
left
right
in
out
on
off
MaxCount
MinCount
Source
Max
Min
Avg
Current
Only solid links are trained
Max
Avg
Feature Map
Θ-node layer
LM Interior Map
Feature Map
Relational Orientations
Reference Orientations
Angle Computation
Σ
TR Boundary Map
LM Boundary Map
TR
LM
Input frames


above
below
left
right


inside
outside
on
off

Bailey's model

- Verb learning
- Learn parameters matched to words
  - word senses
  - can connect to simulator
- Model merging!


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
at goal
MOVE ARM (horiz-dir, force, dur)
not at goal
done
2

schema	elbow jnt	posture	accel
slide 0.9	extend 0.9	palm 0.9	[6]- 8]
		grasp 0.3	

schema	elbow jnt	posture	accel
depress 0.9	fixed 0.9	index 0.9	[2]

data #1

schema	elbow jnt	posture	accel
slide	extend	palm	6

data #2

schema	elbow jnt	posture	accel
slide	extend	palm	8

data #3

schema	elbow jnt	posture	accel
depress	fixed	index	2

data #4

schema	elbow jnt	posture	accel
slide	extend	grasp	2

Computational Details

- complexity of model + ability to explain data
- maximum a posteriori (MAP) hypothesis

$$\underset{m}{\operatorname{argmax}} \; P(m \mid D)$$

wants the best model given data

$$= \underset{m}{\operatorname{argmax}} \; P(D \mid m) P(m) \quad \text{by Bayes' rule}$$

how likely is the data given this model?

penalize complex models – those with too many word senses

Model merging

- Start with a simple model
- Merge to refine it
  - “Greedy” merges: reduce cost without thought for future
- Cost metric
  - prefer simple representation
  - prefer to explain data well

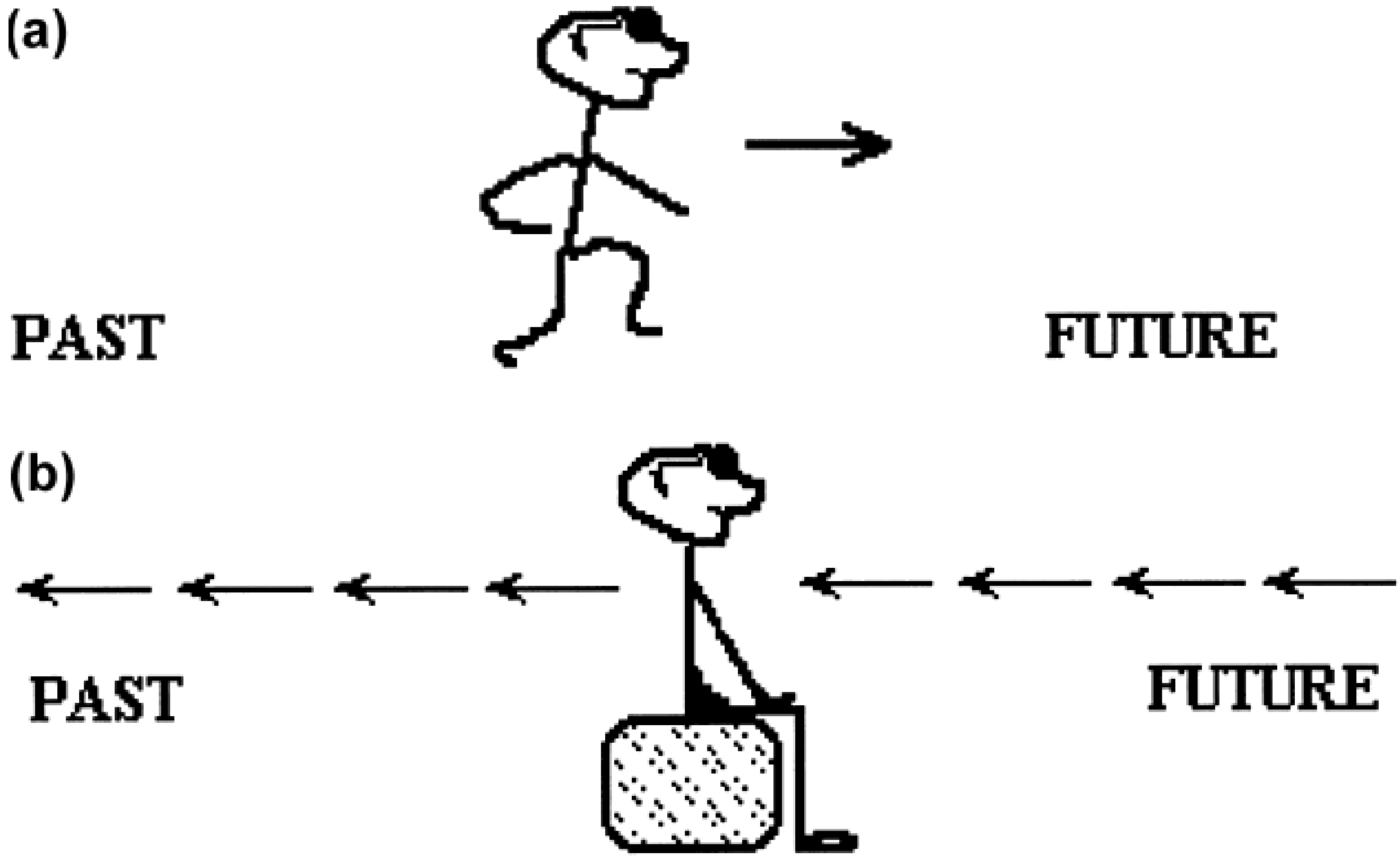
Metaphor

- There are LOTS of metaphors we use
  - Power is size
  - Knowing is seeing
  - Event structure is motion

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



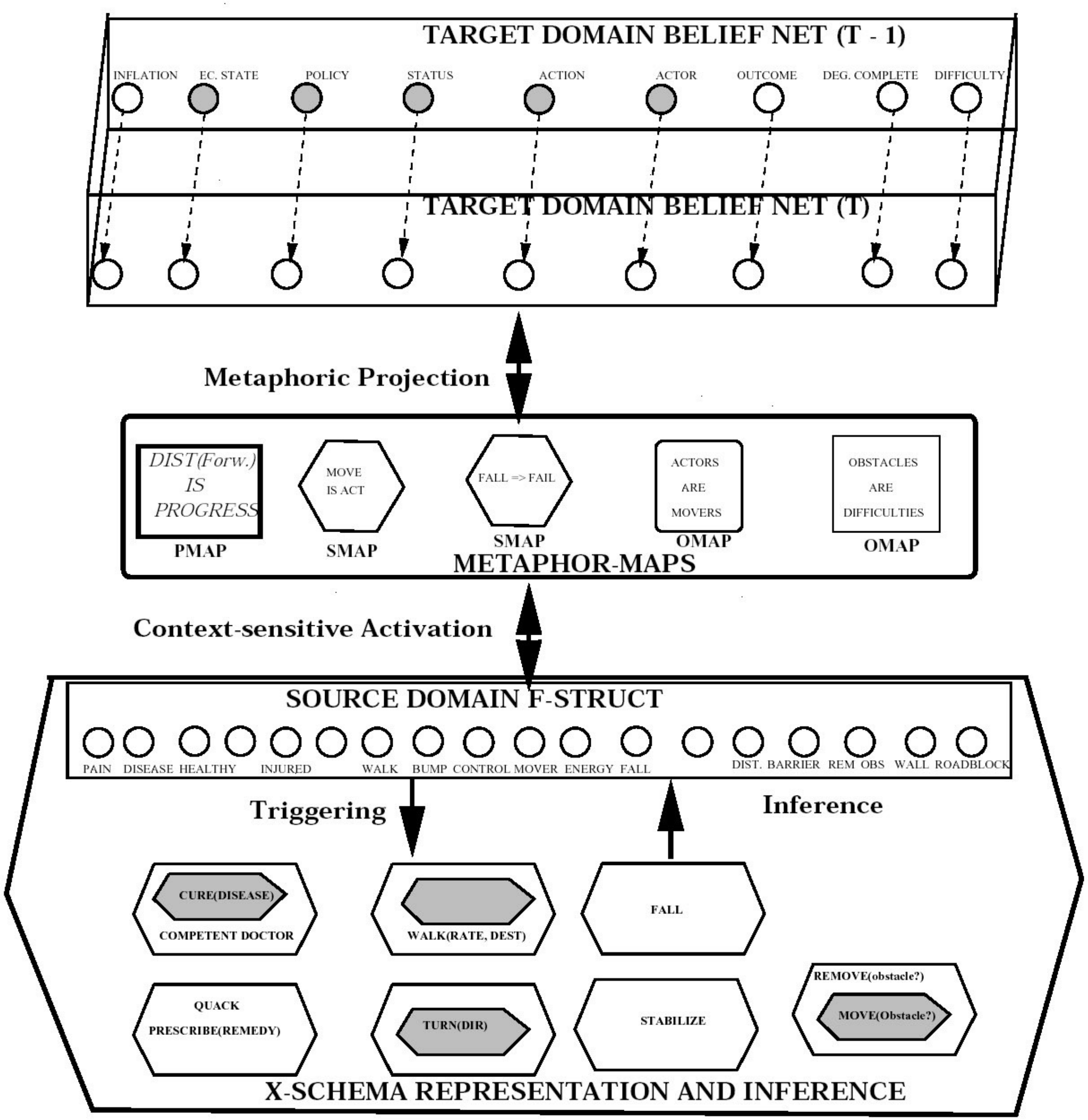
g. 1. (a) Schematic of the ego-moving schema used to organize events in time. (b) Schematic of the
me-moving schema used to organize events in time.

Results

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

KARMA simulator

- Invented by Carson Daly
- Allows metaphor understanding
  - Event structure metaphor
    - Source domain is Petri net
  - Target domain is Bayes net
  - Metaphor maps connect



KARMA

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

Temporal synchrony and SHRUTI

- Binding problem
  - bind properties to objects
  - don't mix them up!
- Reflexive reasoning
  - understand implied information
  - not conscious of this

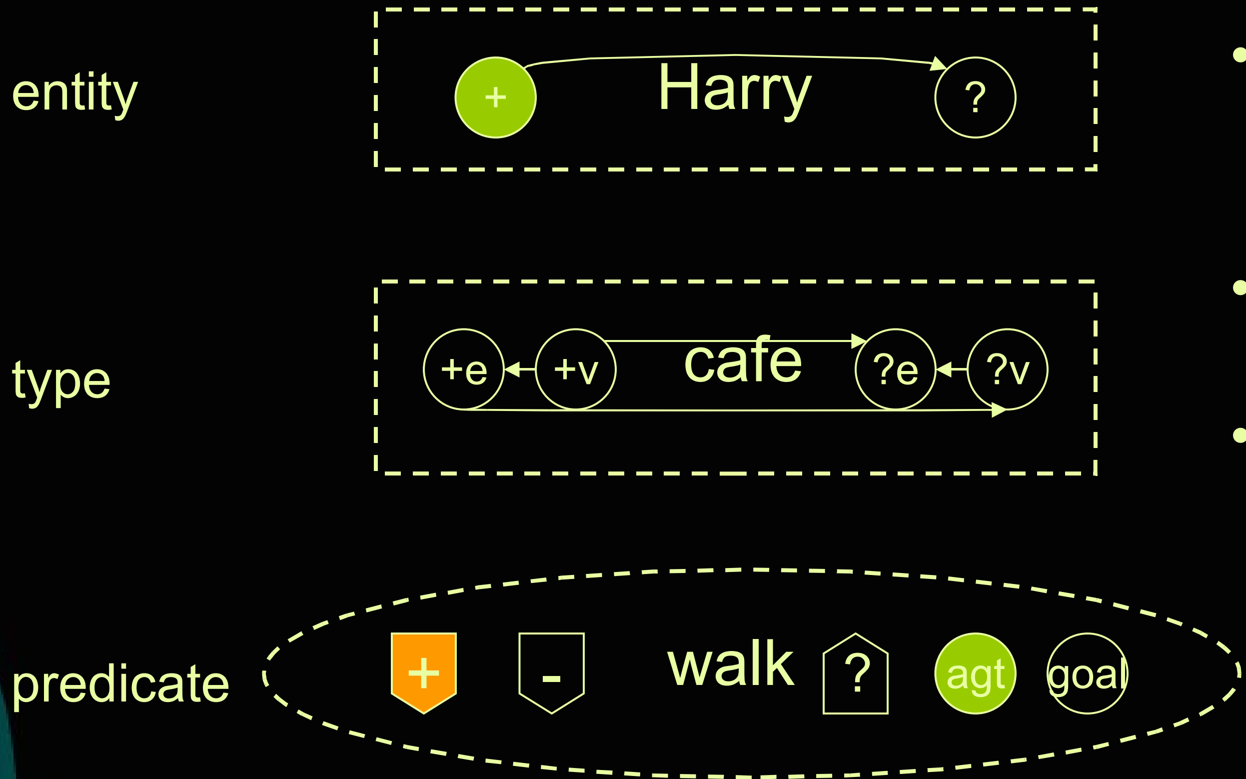
SHRUTI

- entity, type, and predicate focal clusters

- An “entity” is a phase in the rhythmic activity.
- Bindings are synchronous firings of role and entity cells

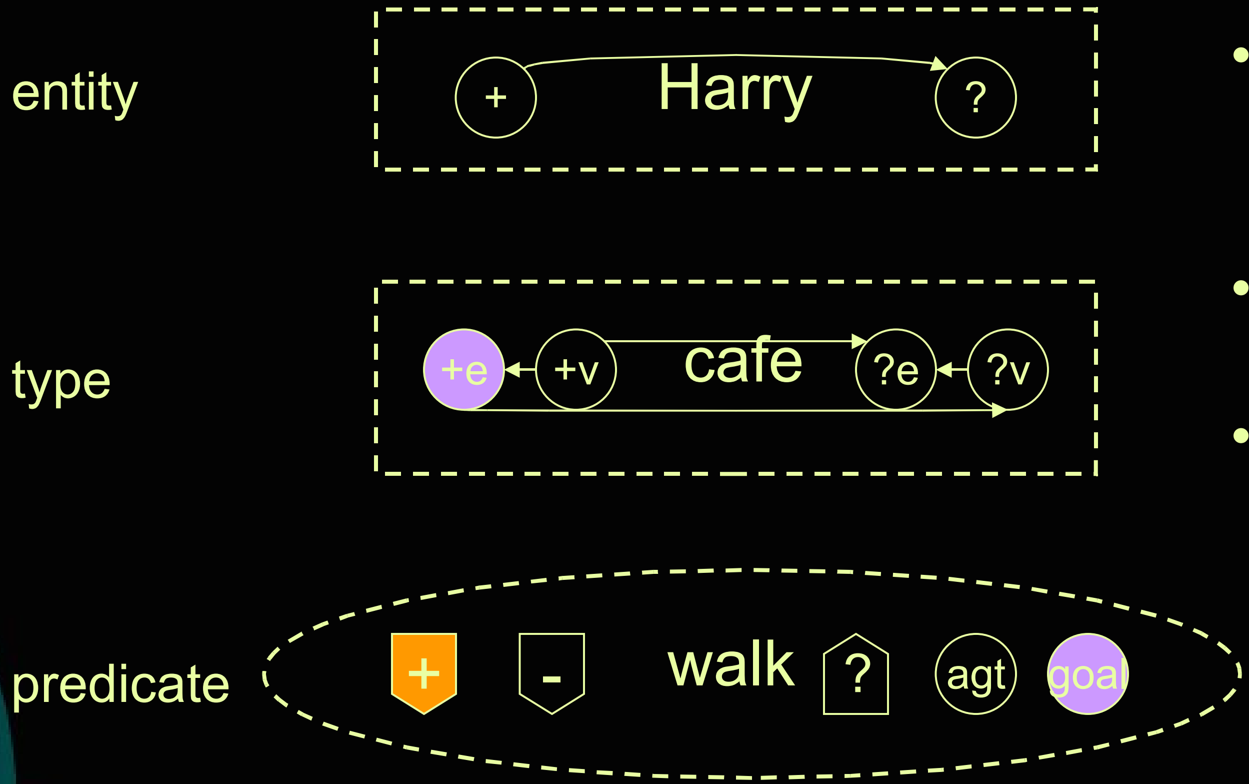
- Rules are interconnection patterns mediated by coincidence detector circuits that allow selective propagation of activity
- An episode of reflexive processing is a transient propagation of rhythmic activity

“Harry walked to the café.”



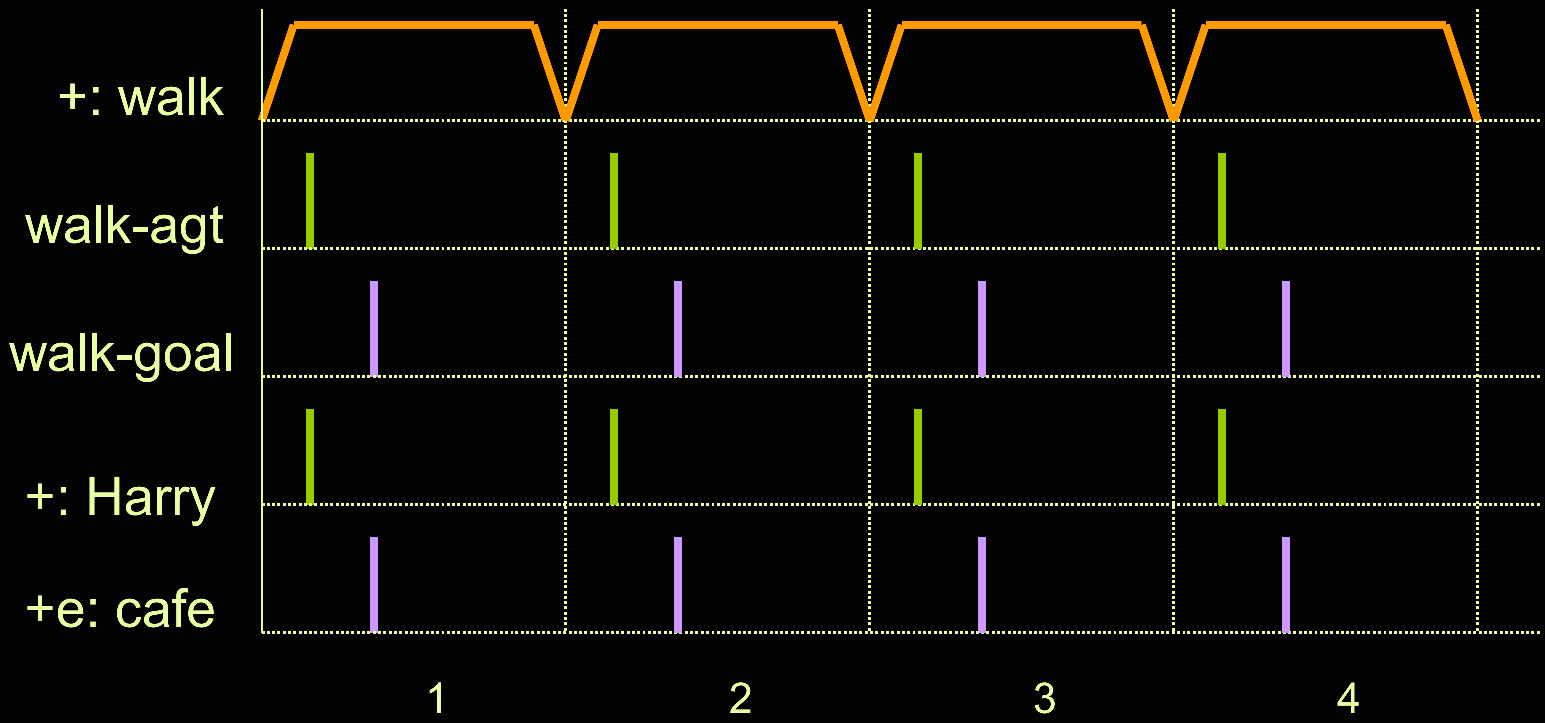
- asserting that walk(Harry, café)
- Harry fires in phase with agent role
- cafe fires in phase with goal role

“Harry walked to the café.”

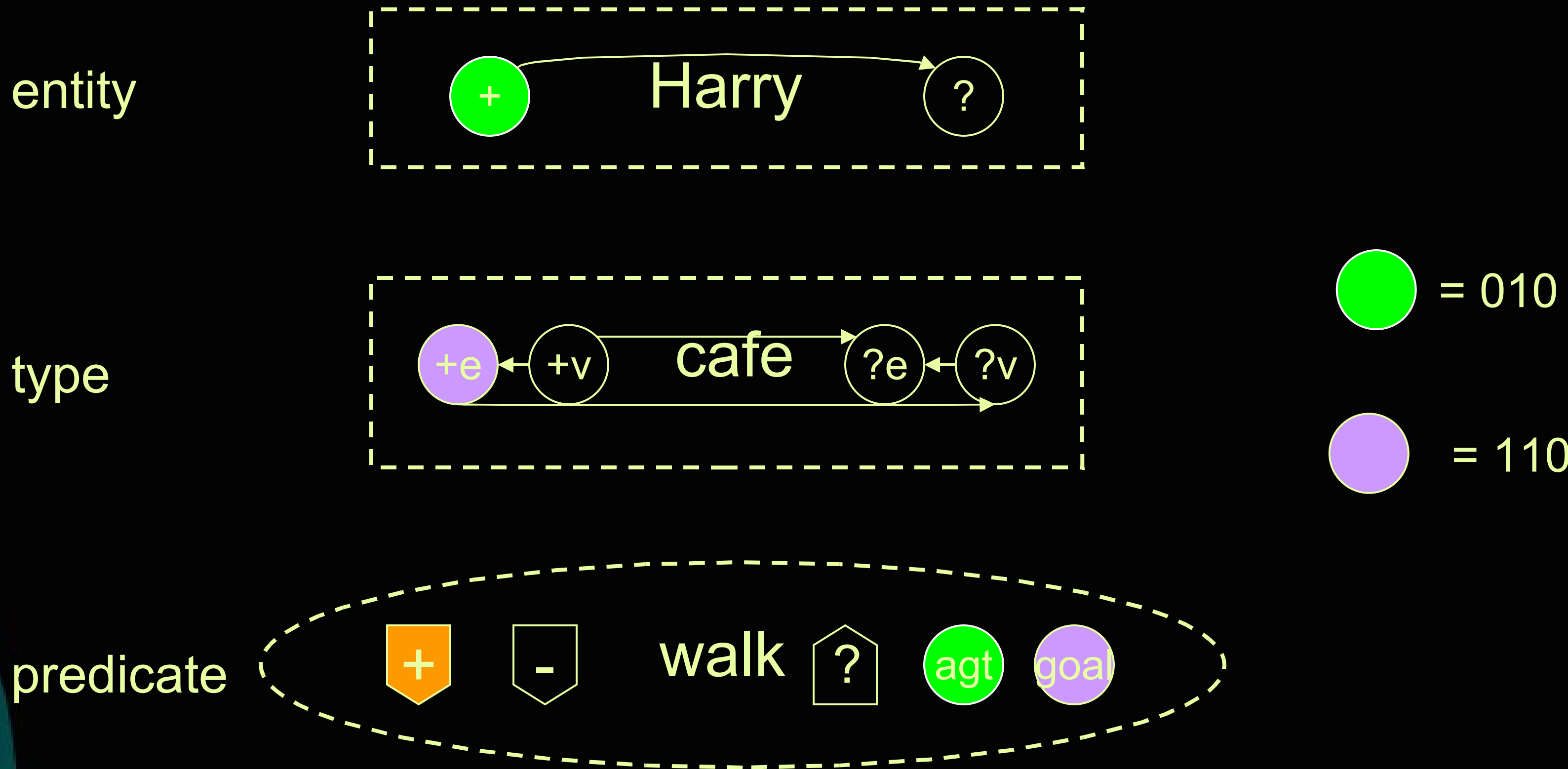


- asserting that walk(Harry, café)
- Harry fires in phase with agent role
- cafe fires in phase with goal role

Activation Trace for walk(Harry, café)



Alternative: short signatures



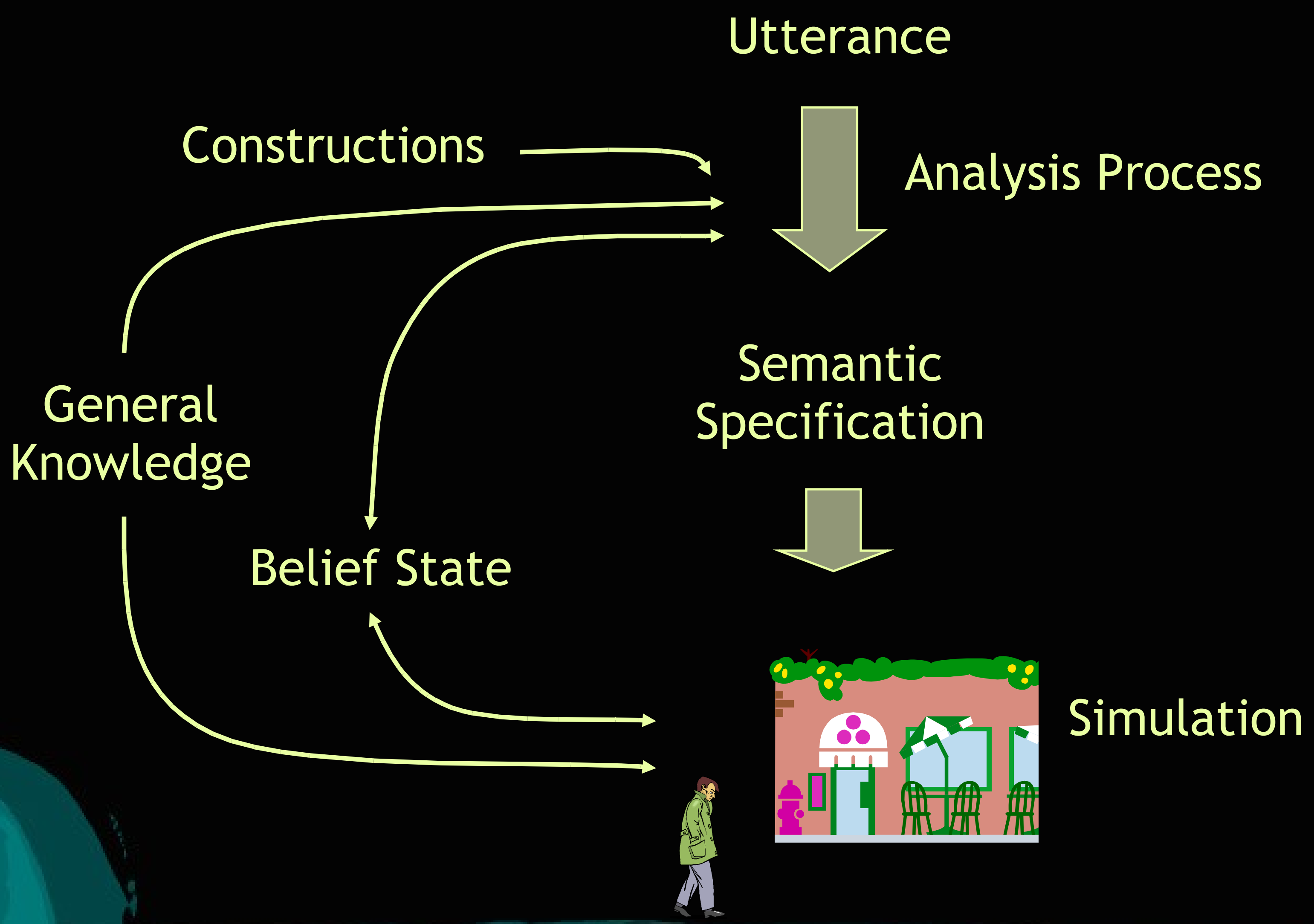
Language

- Grammar
  - Syntax
- Tense
- Aspect
- Semantics
- Metaphor
- Simulation
- Unification

Computer-science style grammar

- Regular grammar
  - X -> a b c Y
- Context-free grammar
  - X -> a Y b Z W

“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

Unification

- Basic idea: capture agreement and semantic features in feature structures

I $\left[\begin{array}{ll} \text{agreement} & \left[\begin{array}{l}\text{number : SG}\\ \text{person : 1st}\end{array}\right]\end{array}\right]$

Pat $\left[\begin{array}{ll} \text{agreement} & \left[\begin{array}{l}\text{number : SG}\\ \text{person : 3rd}\end{array}\right]\end{array}\right]$

Went $\left[\begin{array}{l} \text{agreement} \end{array}\right]$

Shop $\left[\begin{array}{ll} \text{agreement} & \left[\begin{array}{l}\text{number :}\\ \text{person : 1st}\end{array}\right]\end{array}\right]$

- Enforce constraints on these features using unification rules

$\text{VP} \rightarrow \text{Verb NP}$

$\text{VP.agreement} \leftrightarrow \text{Verb.agreement}$

$\text{S} \rightarrow \text{NP VP}$

$\text{NP.agreement} \leftrightarrow \text{VP.agreement}$

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

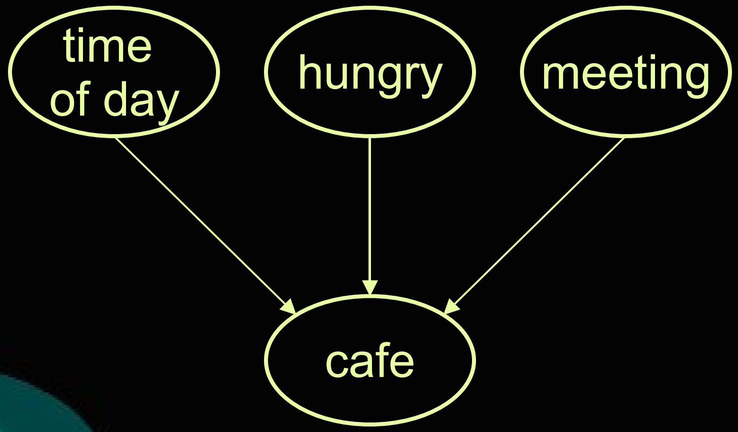
The Directed-Motion construction

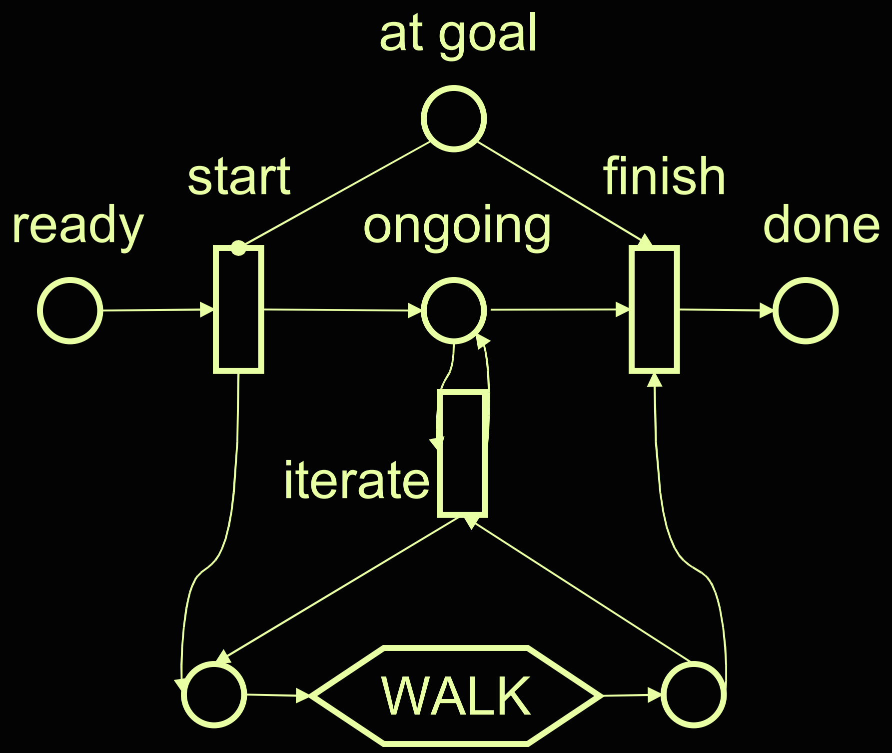
construction DIRECTED-MOTION
  constructional
    constituents
      a : Ref-Exp
      m: Motion-Verb
      p : Spatial-Phrase
  form
    a_f *before* m_f
    m_f *before* p_f
  meaning
    evokes Directed-Motion as dm
    $self_m$.scene $\leftrightarrow$ dm
    dm.agent $\leftrightarrow$ a_m
    dm.motion $\leftrightarrow$ m_m
    dm.path $\leftrightarrow$ p_m

schema Directed-Motion
  roles
    agent : Entity
    motion : Motion
    path : SPG

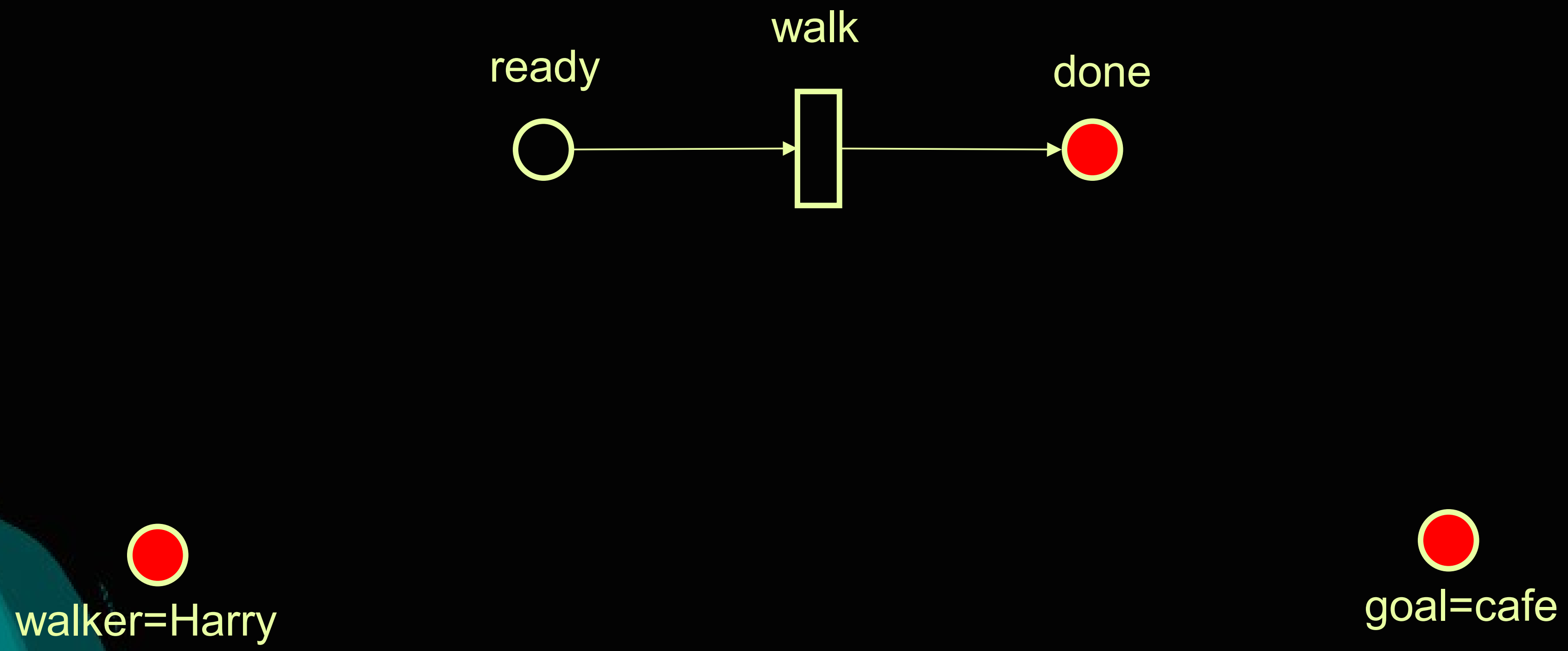
What exactly is simulation?

- Belief update and/or X-schema execution

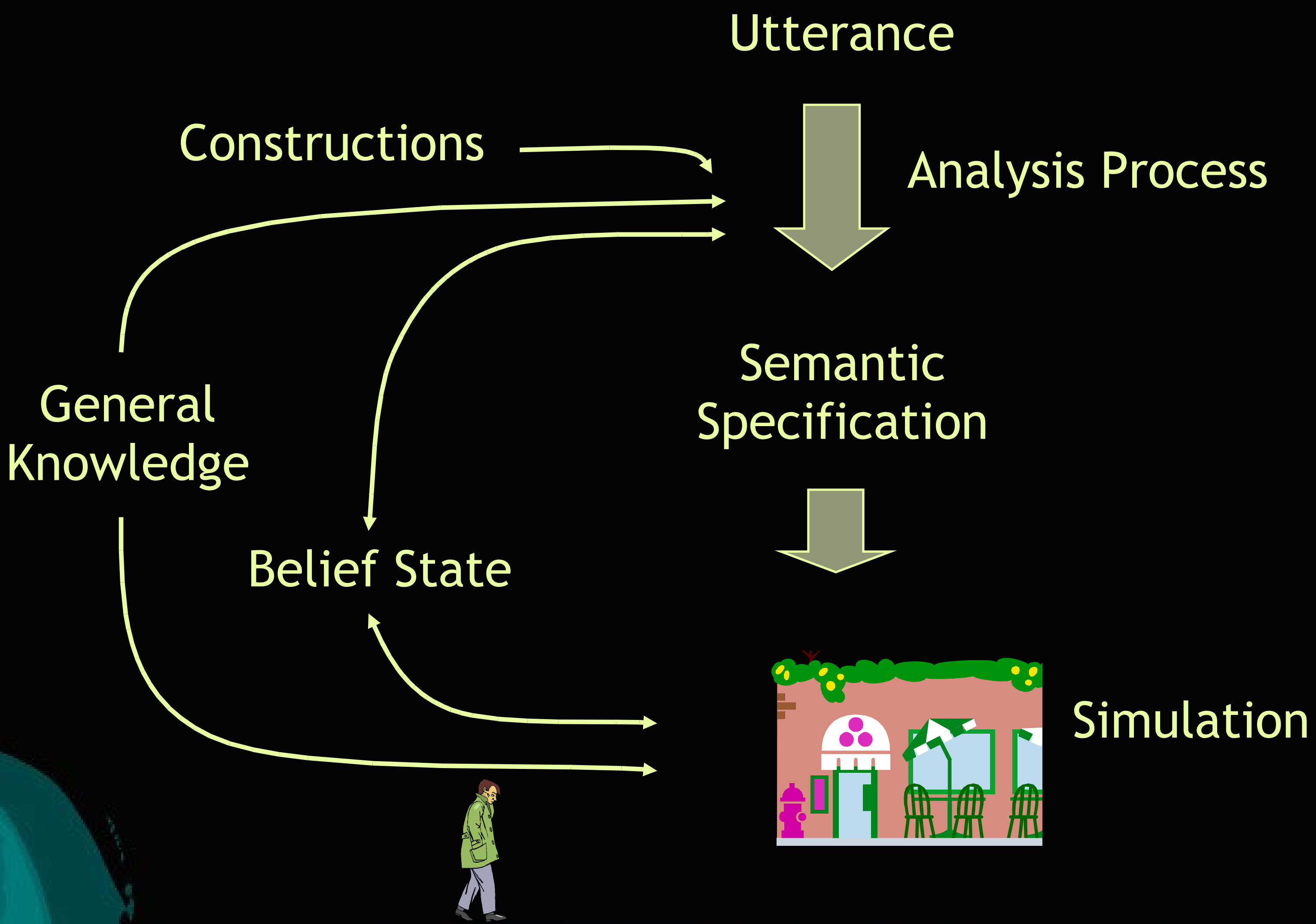




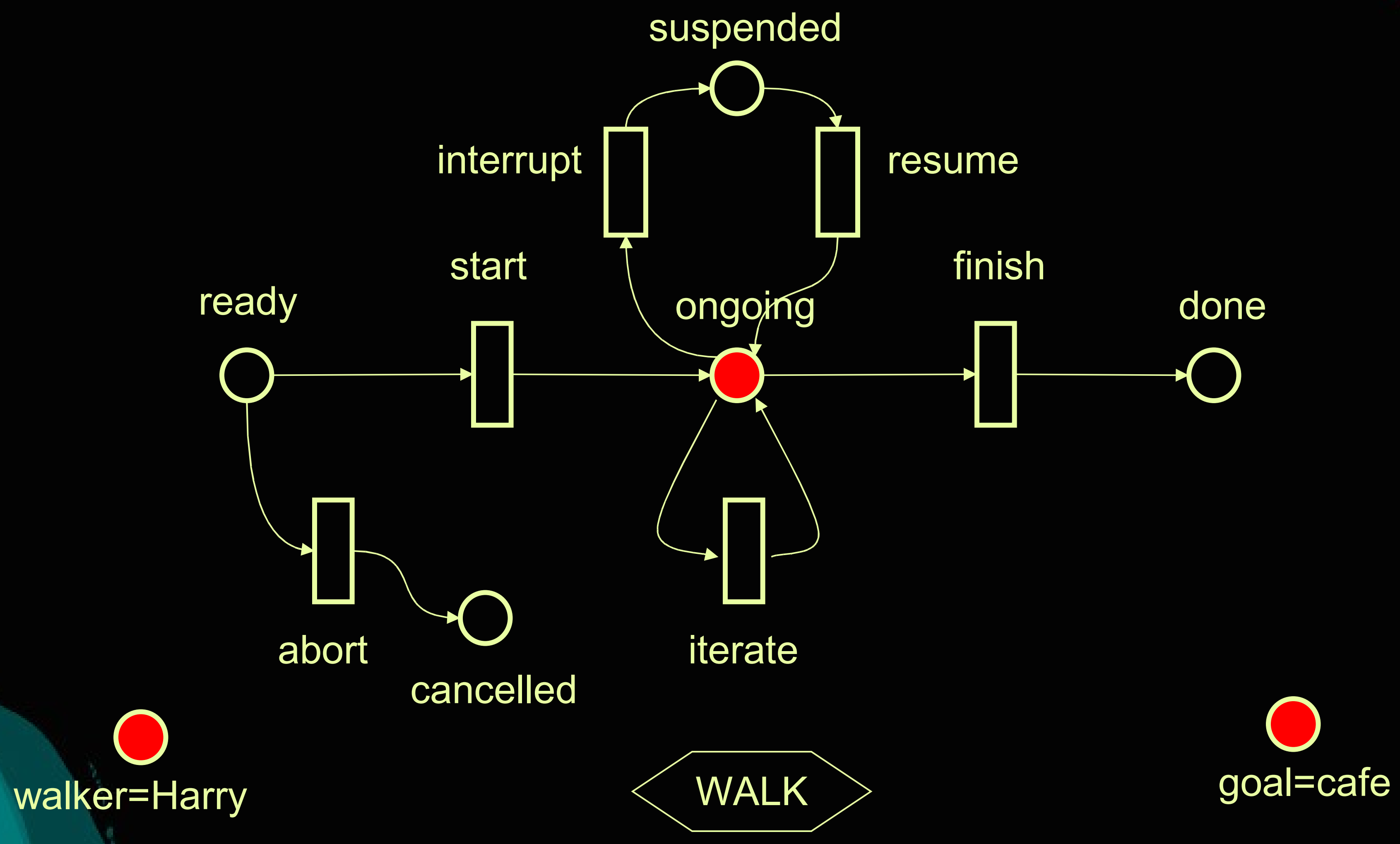
"Harry walked into the café."



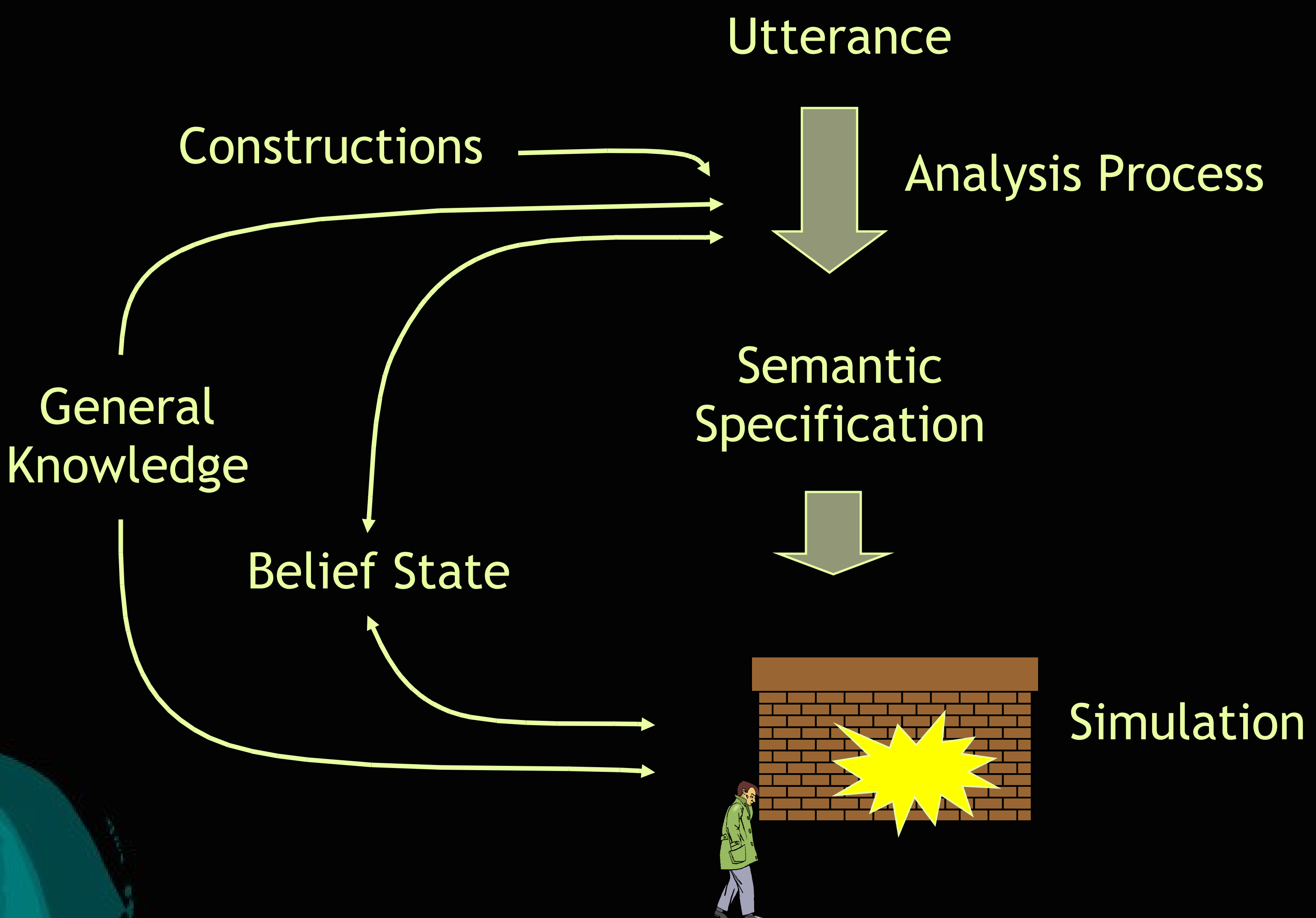
“Harry is walking to the café.”



"Harry is walking to the café."



“Harry has walked into the wall.”

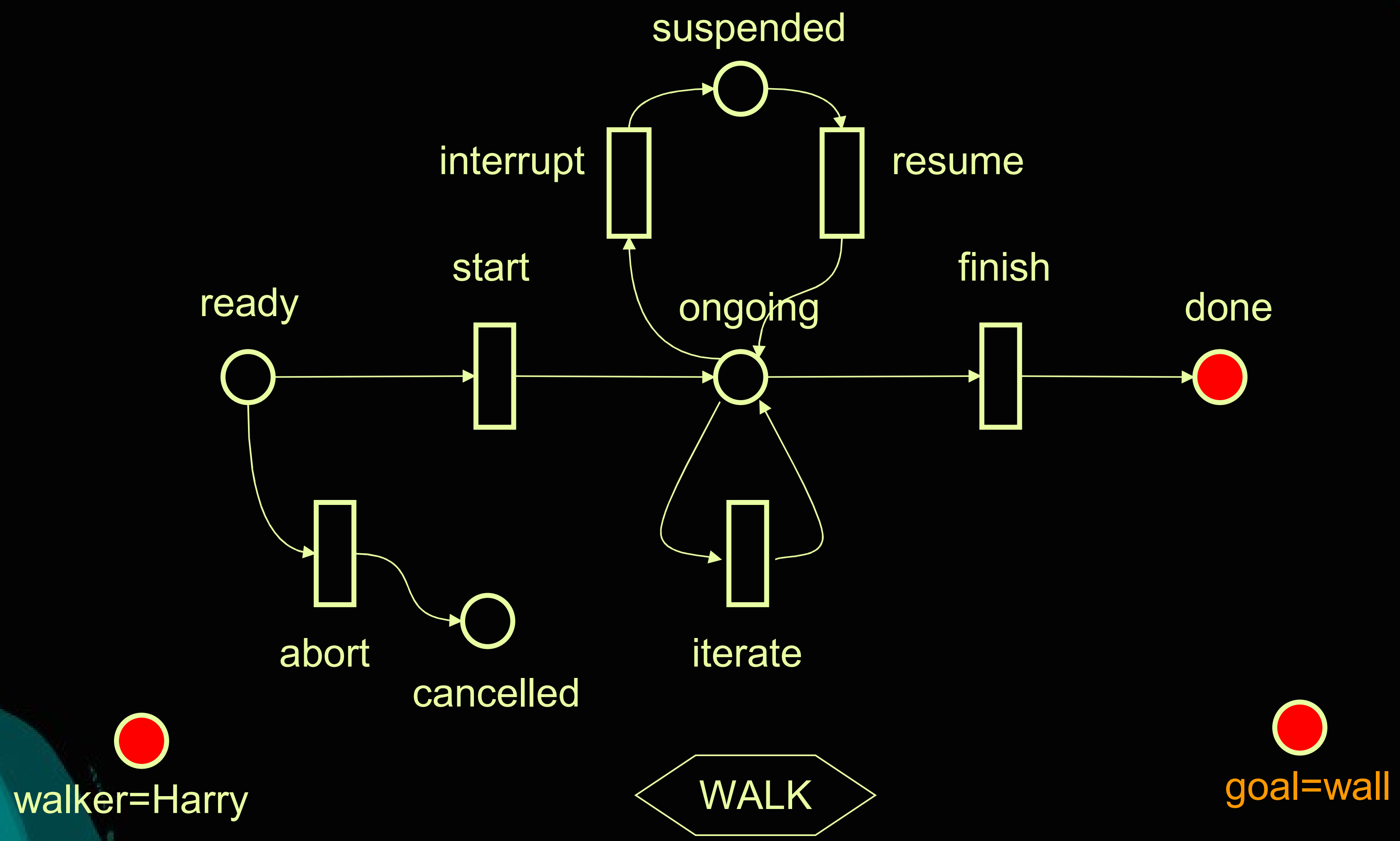


Perhaps a different sense of INTO?

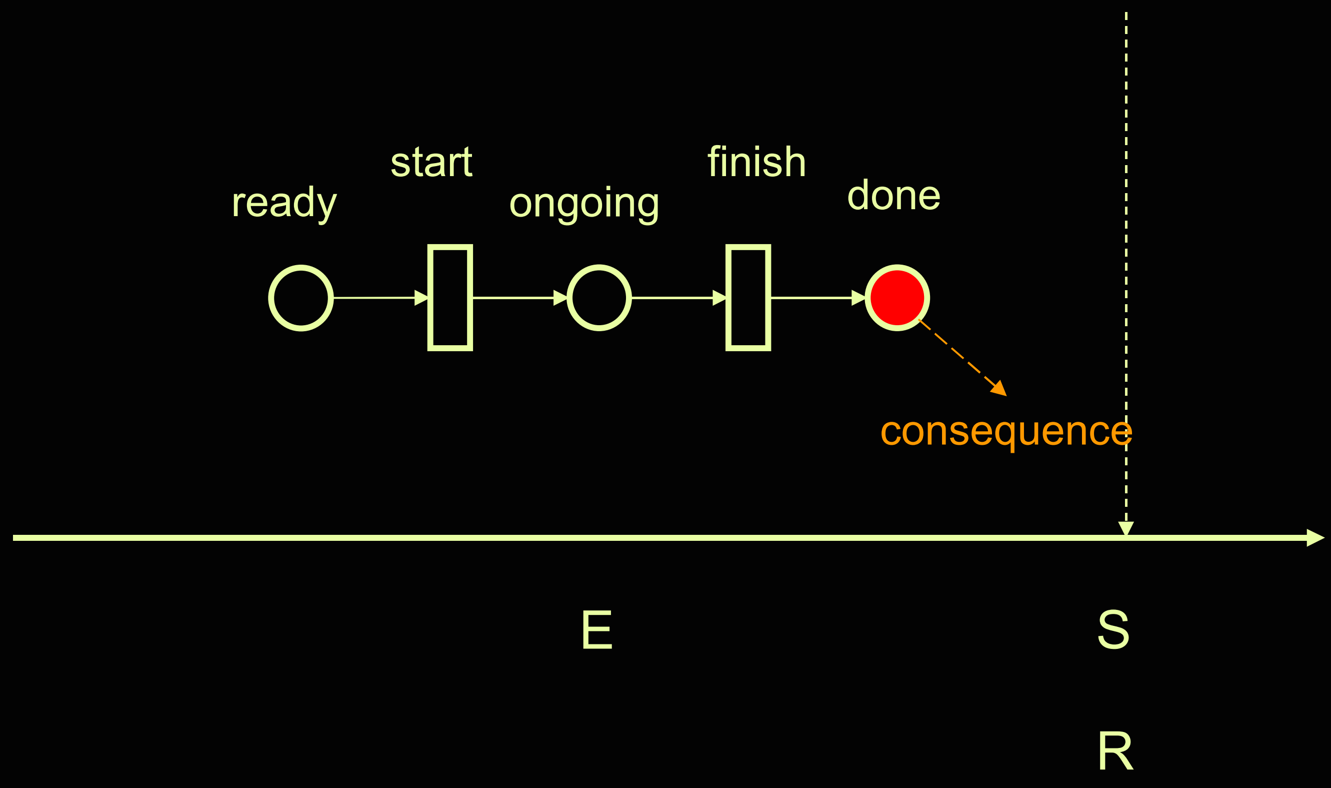
```
construction INTO
  subcase of spatial-prep
    form
      self_f .orth ← "into"
    meaning
      evokes Trajector-Landmark as tl
      evokes Container as cont
      evokes Source-Path-Goal as spg
      tl.trajector ↔ spg.trajector
      tl.landmark ↔ cont
      cont.interior ↔ spg.goal
      cont.exterior ↔ spg.source
```

```
construction INTO
  subcase of spatial-prep
    form
      self_f .orth ← "into"
    meaning
      evokes Trajector-Landmark as tl
      evokes Impact as im
      evokes Source-Path-Goal as spg
      tl.trajector ↔ spg.trajector
      tl.landmark ↔ spg.goal
      im.obj1 ↔ tl.trajector
      im.obj2 ↔ tl.landmark
```

“Harry has walked into the wall.”



Map down to timeline



further questions?

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,

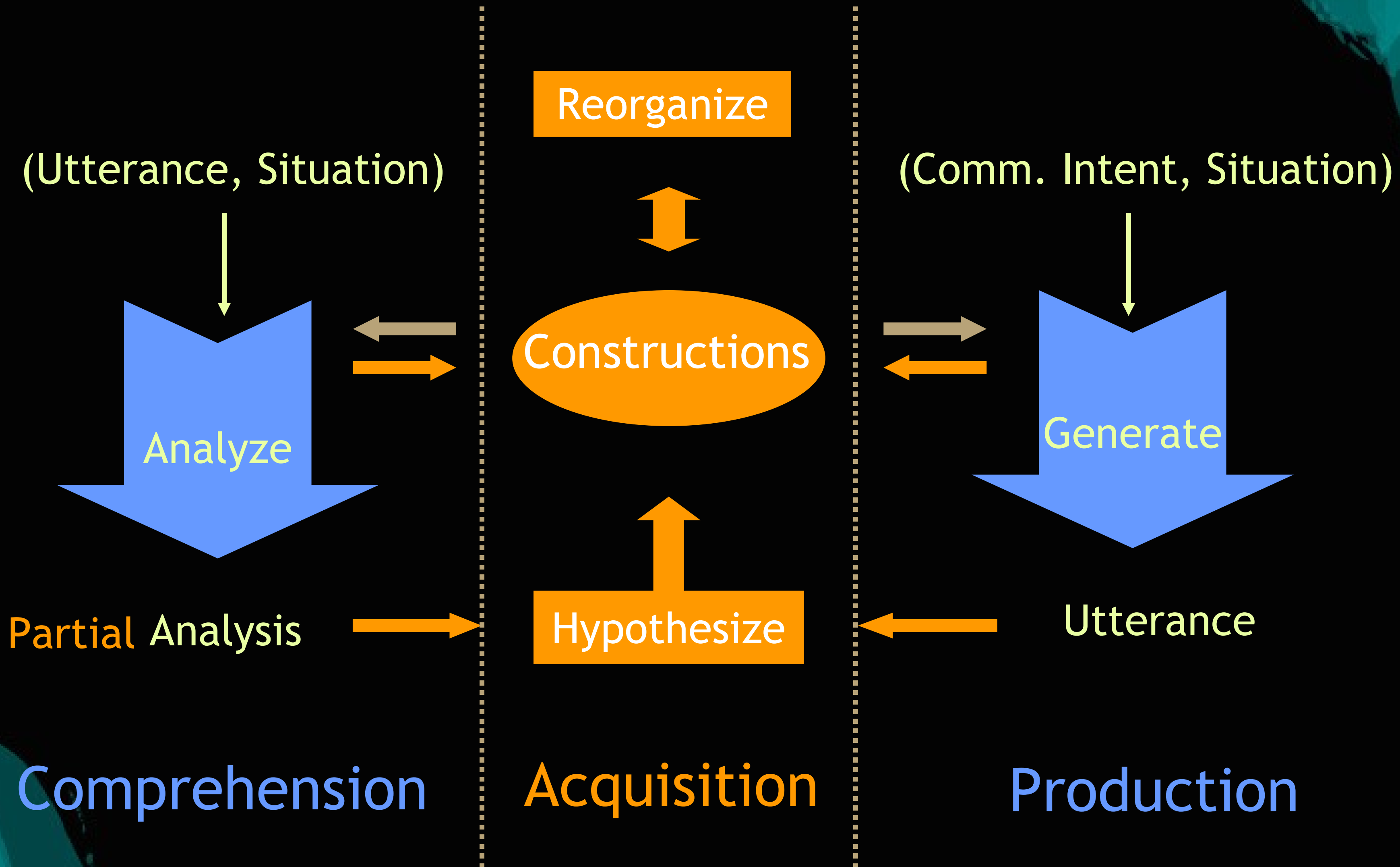
the meanings of verbs,

the metaphors, and

the constructions?

construction learning

Usage-based Language Learning



Main Learning Loop

```
while <utterance, situation> available and cost > stoppingCriterion
    analysis = analyzeAndResolve(utterance, situation, currentGrammar);
    newCxns = hypothesize(analysis);
    if cost(currentGrammar + newCxns) < cost(currentGrammar)
        addNewCxns(newCxns);
    if (re-oganize == true)  // frequency depends on learning parameter
        reorganizeCxns();
```

Three ways to get new constructions

- Relational mapping
  - **throw the ball**

THROW < BALL

- Merging
  - **throw the block**
  - **throwing the ball**

THROW < OBJECT

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

THROW < BALL < OFF

Minimum Description Length

- Choose grammar G to minimize cost(G|D):
  - cost(G|D) = α • size(G) + β • complexity(D|G)
  - Approximates Bayesian learning;
    cost(G|D) ≈ posterior probability P(G|D)

- Size of grammar = size(G) ≈ 1/prior P(G)
  - favor fewer/smaller constructions/roles; isomorphic mappings

- Complexity of data given grammar ≈ 1/likelihood P(D|G)
  - favor simpler analyses
    (fewer, more likely constructions)
  - based on derivation length + score of derivation

further questions?