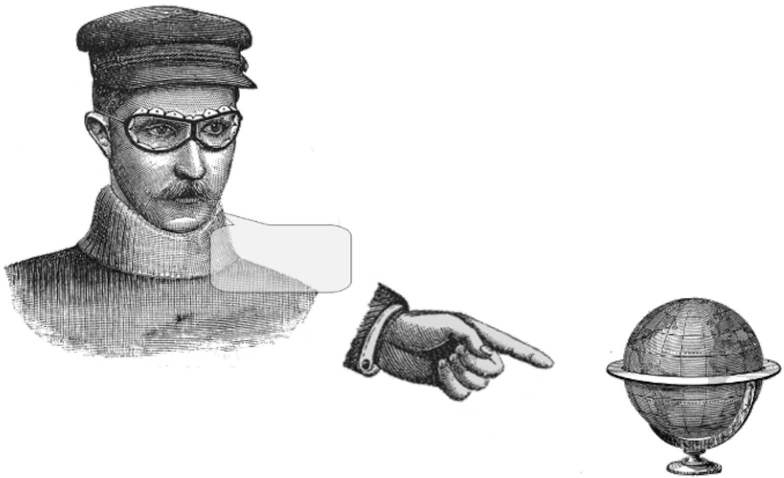
A simulation-based model of semantic working memory

Alistair Knott, Martin Takac

Department of Computer Science, University of Otago

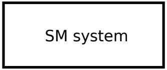
How can we use language to talk about the world?

How can we use language to talk about the world?

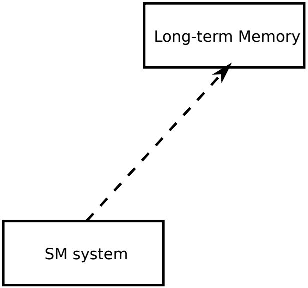


What's working memory?

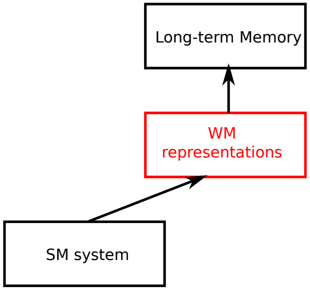
What's working memory?



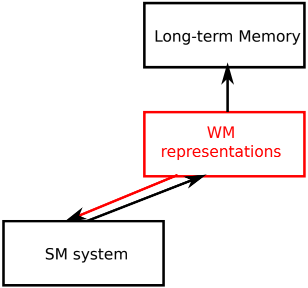
What's working memory?



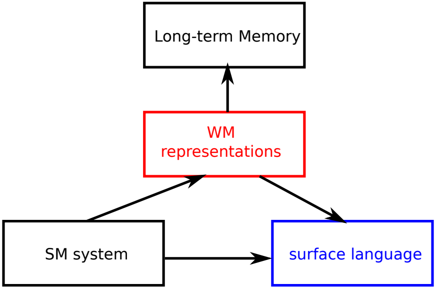
What's working memory?



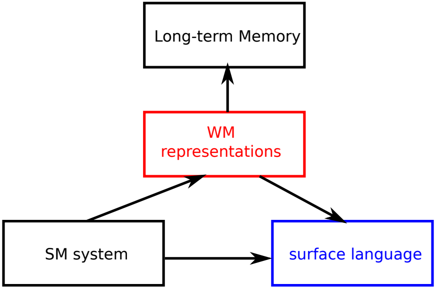
What's working memory?



What's working memory?

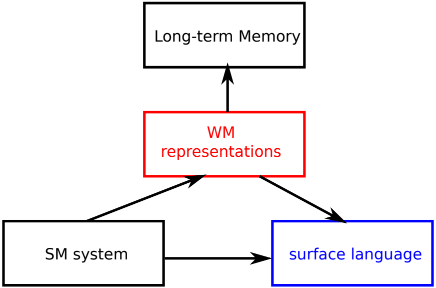


Our main proposal



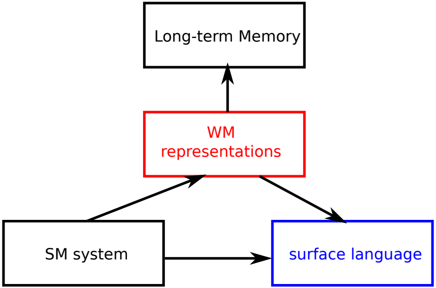
Our main proposal

- We experience the world through SM routines with well-defined *sequential structure*.



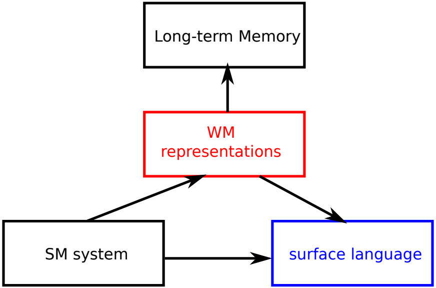
Our main proposal

- We experience the world through SM routines with well-defined *sequential structure*.
- This sequential structure is retained in WM representations of episodes and individuals. . .



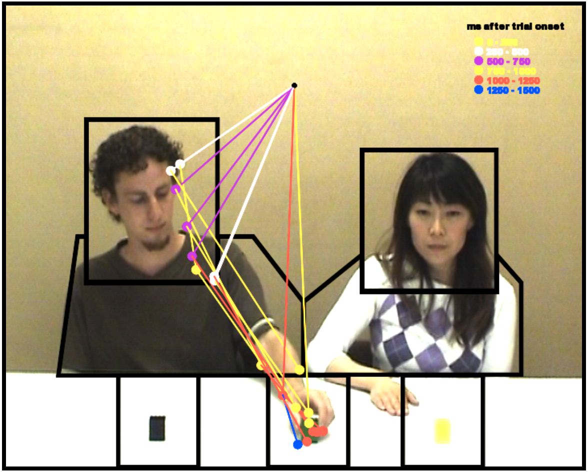
Our main proposal

- We experience the world through SM routines with well-defined *sequential structure*.
- This sequential structure is retained in WM representations of episodes and individuals. . .
- And also in the syntactic structure of sentences.



- We experience the world through SM routines with well-defined *sequential structure*.

Example: the SM routine for perceiving an episode



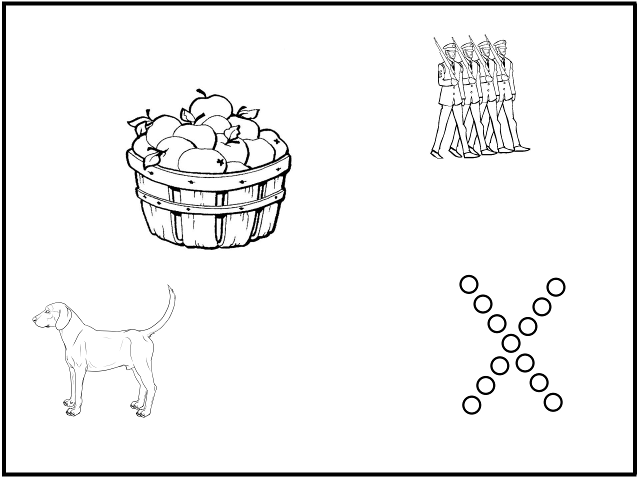
Webb, Knott and MacAskill, ‘Eye movements during transitive action observation have sequential structure’
Acta Psychologica 2010

Example: the SM routine for perceiving an episode

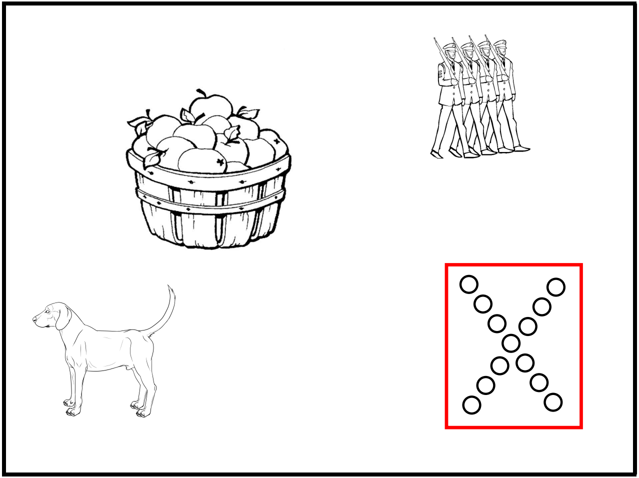
We argue: perceiving (or performing) a transitive action involves a canonical sequence of SM operations.

Step	SM operation
1	*identify_agent*
2	*identify_patient*
3	*identify_action*

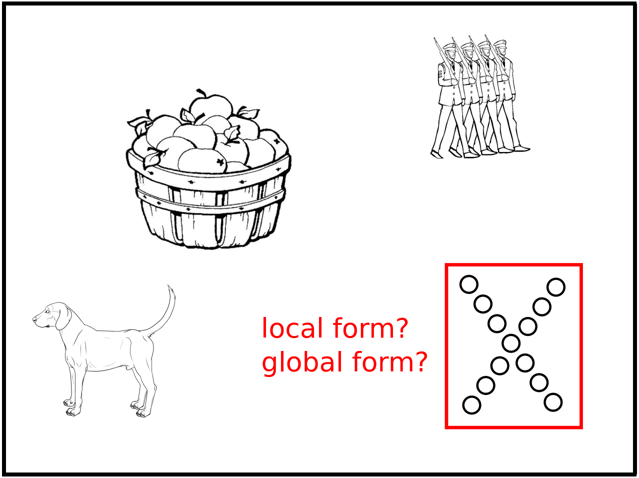
Example: the SM routine for perceiving an individual



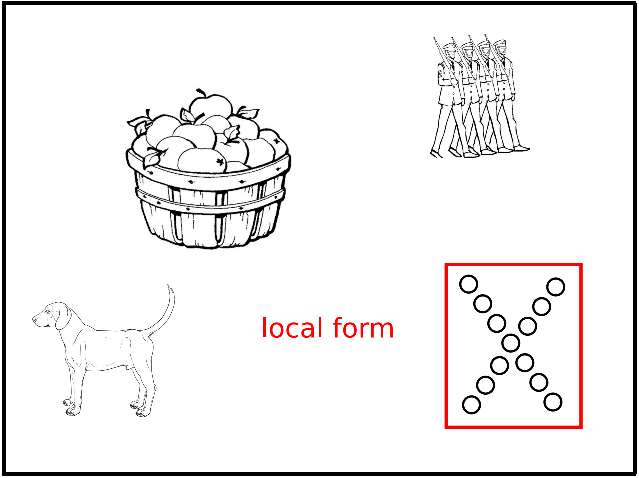
Example: the SM routine for perceiving an individual



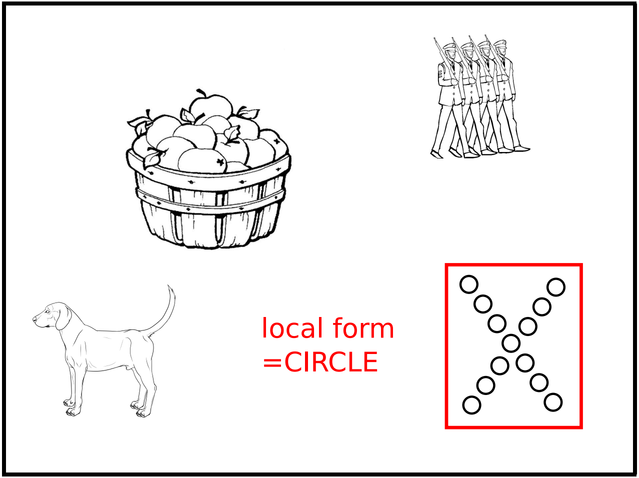
Example: the SM routine for perceiving an individual



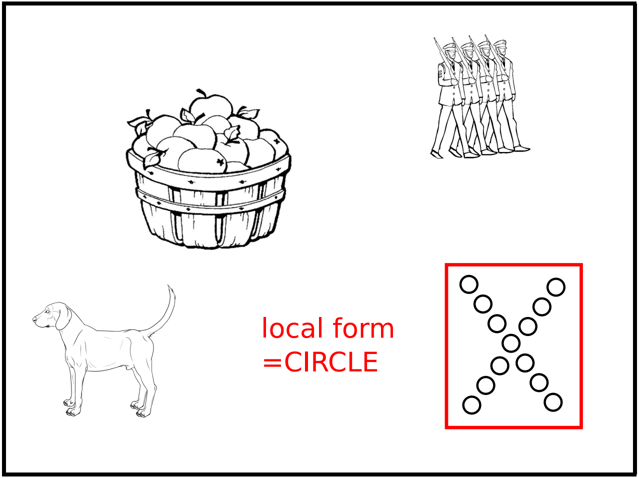
Example: the SM routine for perceiving an individual



Example: the SM routine for perceiving an individual



Example: the SM routine for perceiving an individual



Walles, Robins and Knott, 'A perceptually grounded model of the singular-plural distinction'
Language and Cognition 2014

Example: the SM routine for perceiving an individual

We argue: identifying an individual also involves a canonical sequence of SM operations.

Step	SM operation
1	*attend_location*
2	*establish_scale (=identify singular or plural)*
3	*activate_class*

A model of WM representations

How are episodes and individuals represented in WM?

A model of WM representations

How are episodes and individuals represented in WM?

Our proposal: they are stored as prepared SM routines.

Benefit 1: a road map

We know lots about how SM sequences are stored in WM.

Benefit 1: a road map

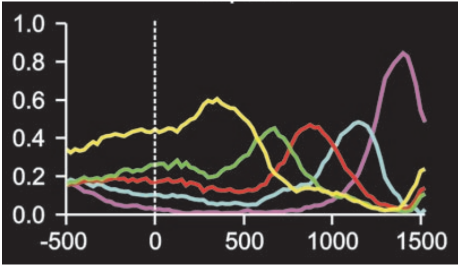
We know lots about how SM sequences are stored in WM.

- Prefrontal cortex is heavily involved.

Benefit 1: a road map

We know lots about how SM sequences are stored in WM.

- Prefrontal cortex is heavily involved.
- In the prefrontal assembly representing a planned sequence of actions, representations of the actions are active *in parallel*.



Averbeck *et al.*, 'Parallel processing of serial movements in prefrontal cortex' PNAS 2002

Benefit 2: support for simulation

If experiences are stored as prepared SM routines, they're *executables*.

Benefit 2: support for simulation

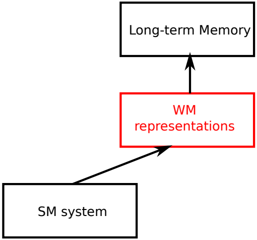
If experiences are stored as prepared SM routines, they're *executables*.

- So there's a natural account of how we replay or simulate experiences.

Benefit 2: support for simulation

If experiences are stored as prepared SM routines, they're *executables*.

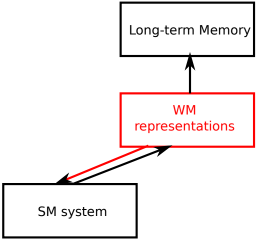
- So there's a natural account of how we replay or simulate experiences.



Benefit 2: support for simulation

If experiences are stored as prepared SM routines, they're *executables*.

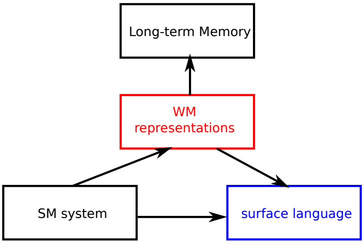
- So there's a natural account of how we replay or simulate experiences.



Benefit 3: support for a model of sentence generation

If experiences are stored as prepared SM routines, they're *executables*.

- Our idea: generating a sentence involves replaying an episode held in WM *in a special mode*, where SM signals trigger output words.



Takac, Benuskova and Knott, 'A connectionist model of language acquisition and sentence generation' Cognition 2012

Benefit 4: a new account of semantic role-binding

Benefit 4: a new account of semantic role-binding

An episode representation has to identify the semantic role played by each of its participants (e.g. AGENT, PATIENT.)

- This is hard to do in a neural network model!

Benefit 4: a new account of semantic role-binding

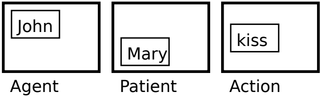
An episode representation has to identify the semantic role played by each of its participants (e.g. AGENT, PATIENT.)

- This is hard to do in a neural network model!
- We can't just activate all concepts in the episode simultaneously.

Benefit 4: a new account of semantic role-binding

An episode representation has to identify the semantic role played by each of its participants (e.g. AGENT, PATIENT.)

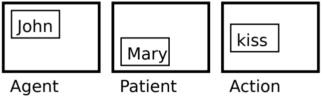
- This is hard to do in a neural network model!
- We can't just activate all concepts in the episode simultaneously.
- One old idea is to code semantic roles *by place*.



Benefit 4: a new account of semantic role-binding

An episode representation has to identify the semantic role played by each of its participants (e.g. AGENT, PATIENT.)

- This is hard to do in a neural network model!
- We can't just activate all concepts in the episode simultaneously.
- One old idea is to code semantic roles *by place*.
- This is a bad idea. . . but it can be resurrected if episodes are structured as sequences.

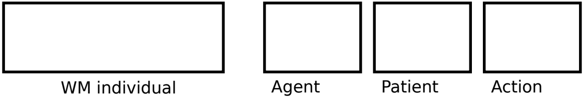


Benefit 4: a new account of semantic role-binding

An episode representation has to identify the semantic role played by each of its participants (e.g. AGENT, PATIENT.)

- This is hard to do in a neural network model!
- We can't just activate all concepts in the episode simultaneously.
- One old idea is to code semantic roles *by place*.

In our model, there's a (single) medium holding WM representations of individuals.

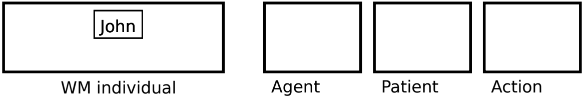


Benefit 4: a new account of semantic role-binding

An episode representation has to identify the semantic role played by each of its participants (e.g. AGENT, PATIENT.)

- This is hard to do in a neural network model!
- We can't just activate all concepts in the episode simultaneously.
- One old idea is to code semantic roles *by place*.

When perceiving an episode, we attend to the agent. . .

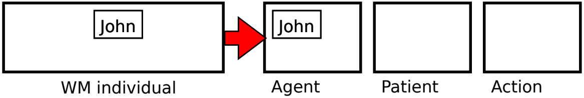


Benefit 4: a new account of semantic role-binding

An episode representation has to identify the semantic role played by each of its participants (e.g. AGENT, PATIENT.)

- This is hard to do in a neural network model!
- We can't just activate all concepts in the episode simultaneously.
- One old idea is to code semantic roles *by place*.

Then we *copy* this representation to the WM episode. . .

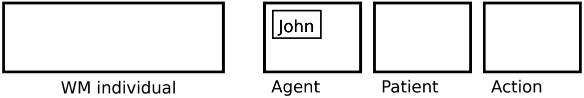


Benefit 4: a new account of semantic role-binding

An episode representation has to identify the semantic role played by each of its participants (e.g. AGENT, PATIENT.)

- This is hard to do in a neural network model!
- We can't just activate all concepts in the episode simultaneously.
- One old idea is to code semantic roles *by place*.

Then we *copy* this representation to the WM episode...

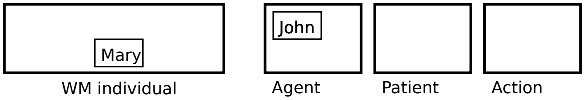


Benefit 4: a new account of semantic role-binding

An episode representation has to identify the semantic role played by each of its participants (e.g. AGENT, PATIENT.)

- This is hard to do in a neural network model!
- We can't just activate all concepts in the episode simultaneously.
- One old idea is to code semantic roles *by place*.

Then we attend to the patient...

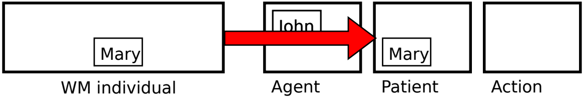


Benefit 4: a new account of semantic role-binding

An episode representation has to identify the semantic role played by each of its participants (e.g. AGENT, PATIENT.)

- This is hard to do in a neural network model!
- We can't just activate all concepts in the episode simultaneously.
- One old idea is to code semantic roles *by place*.

Then we *copy* this representation to the WM episode...

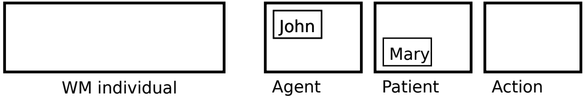


Benefit 4: a new account of semantic role-binding

An episode representation has to identify the semantic role played by each of its participants (e.g. AGENT, PATIENT.)

- This is hard to do in a neural network model!
- We can't just activate all concepts in the episode simultaneously.
- One old idea is to code semantic roles *by place*.

Then we *copy* this representation to the WM episode...

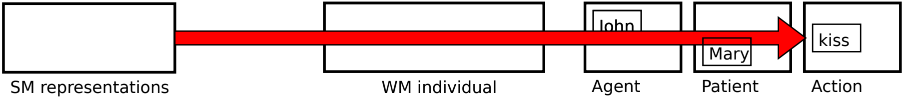


Benefit 4: a new account of semantic role-binding

An episode representation has to identify the semantic role played by each of its participants (e.g. AGENT, PATIENT.)

- This is hard to do in a neural network model!
- We can't just activate all concepts in the episode simultaneously.
- One old idea is to code semantic roles *by place*.

Then we identify the action, and copy that to the WM episode.

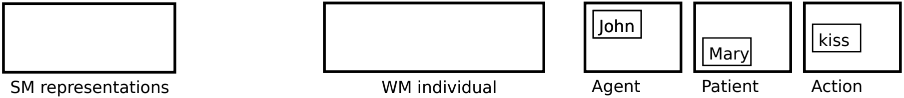


Benefit 4: a new account of semantic role-binding

An episode representation has to identify the semantic role played by each of its participants (e.g. AGENT, PATIENT.)

- This is hard to do in a neural network model!
- We can't just activate all concepts in the episode simultaneously.
- One old idea is to code semantic roles *by place*.

Then we identify the action, and copy that to the WM episode.

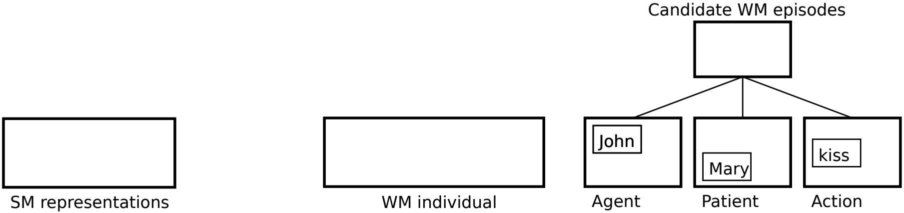


Benefit 4: a new account of semantic role-binding

An episode representation has to identify the semantic role played by each of its participants (e.g. AGENT, PATIENT.)

- This is hard to do in a neural network model!
- We can't just activate all concepts in the episode simultaneously.
- One old idea is to code semantic roles *by place*.

We can now have localist representations of whole episodes.

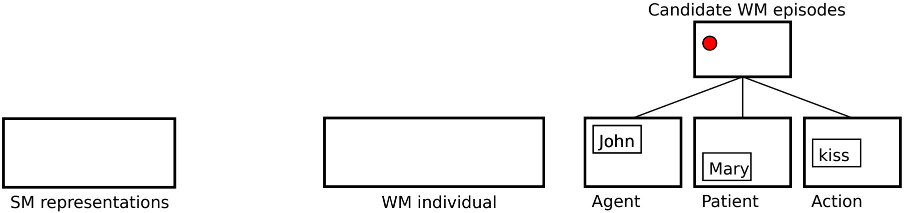


Benefit 4: a new account of semantic role-binding

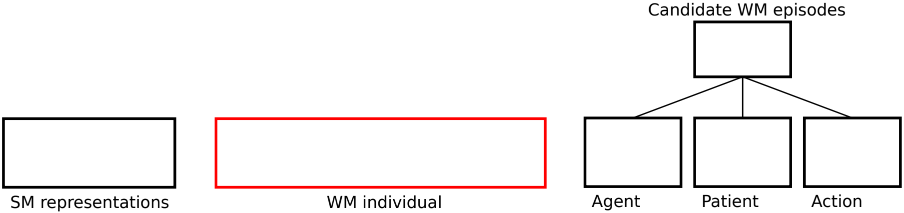
An episode representation has to identify the semantic role played by each of its participants (e.g. AGENT, PATIENT.)

- This is hard to do in a neural network model!
- We can't just activate all concepts in the episode simultaneously.
- One old idea is to code semantic roles *by place*.

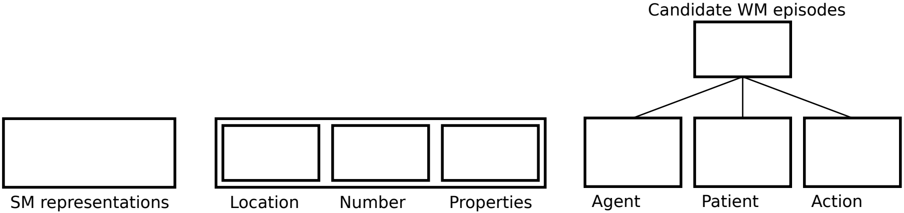
We can now have localist representations of whole episodes.



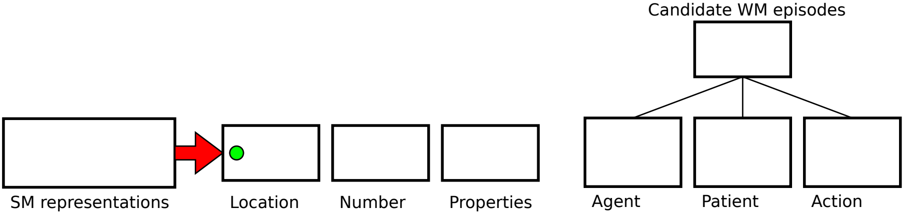
Benefit 5: an account of nested semantic structures



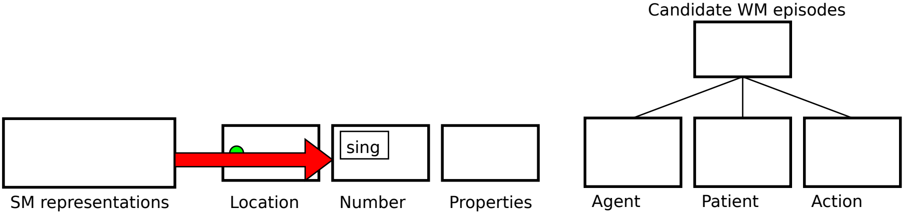
Benefit 5: an account of nested semantic structures



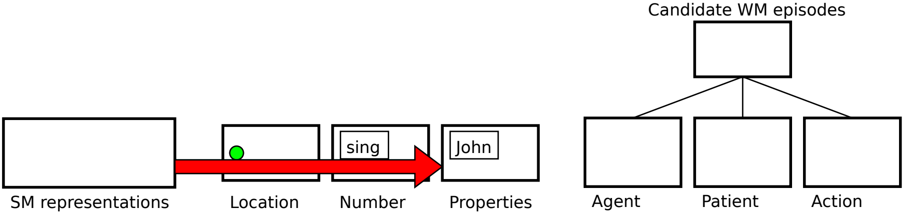
Benefit 5: an account of nested semantic structures



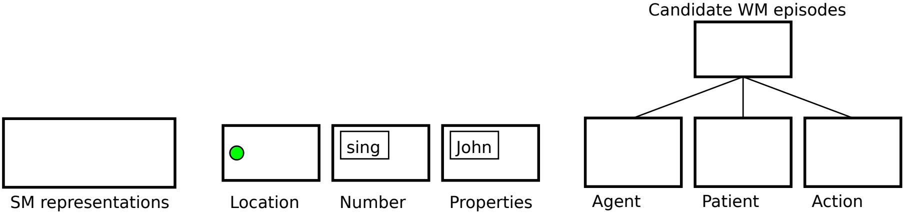
Benefit 5: an account of nested semantic structures



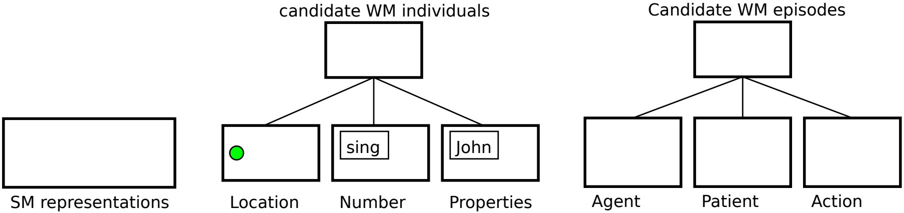
Benefit 5: an account of nested semantic structures



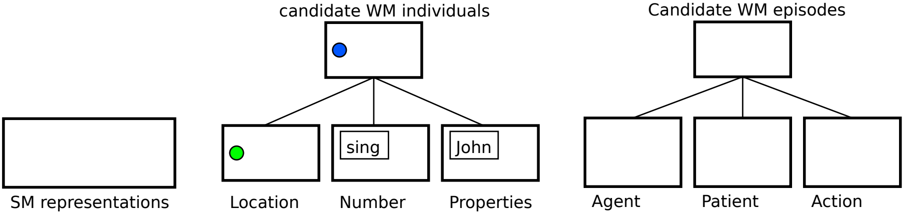
Benefit 5: an account of nested semantic structures



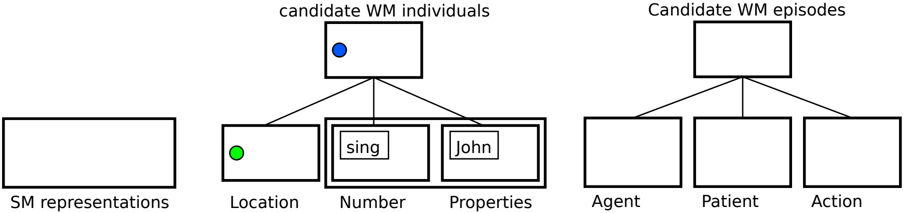
Benefit 5: an account of nested semantic structures



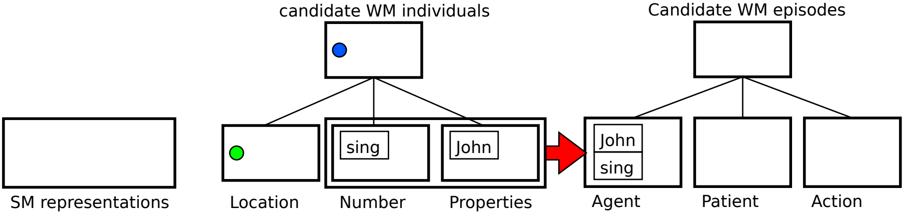
Benefit 5: an account of nested semantic structures



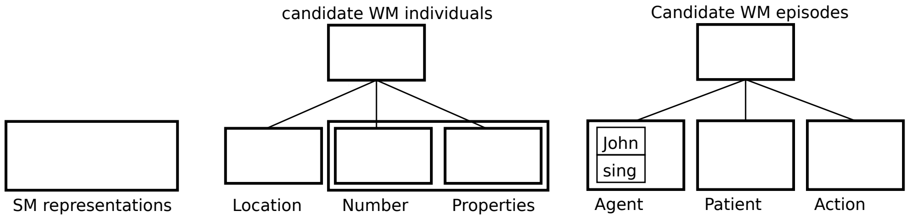
Benefit 5: an account of nested semantic structures



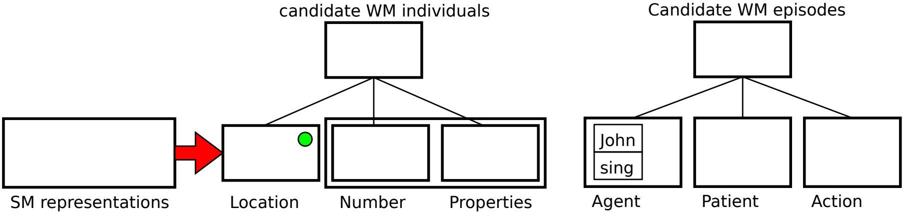
Benefit 5: an account of nested semantic structures



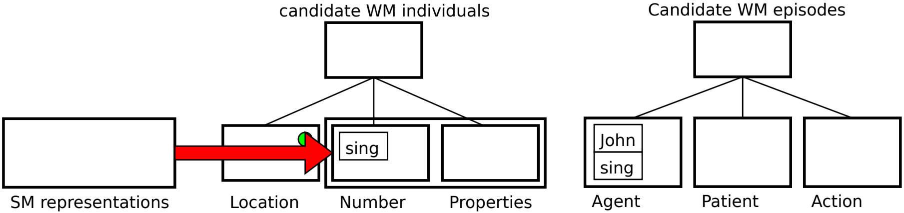
Benefit 5: an account of nested semantic structures



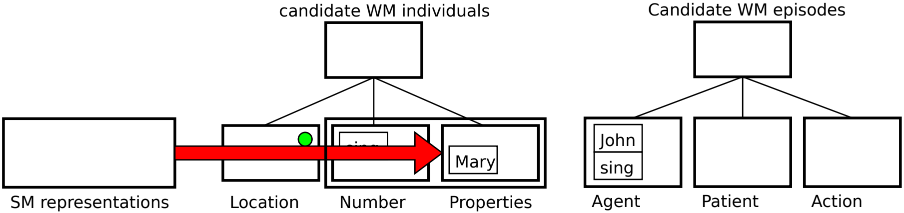
Benefit 5: an account of nested semantic structures



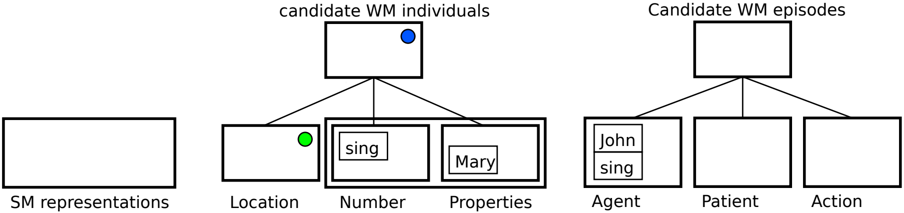
Benefit 5: an account of nested semantic structures



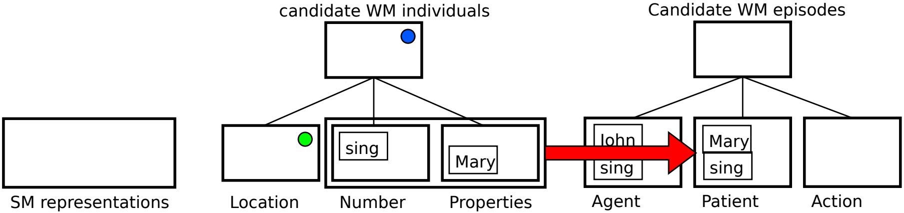
Benefit 5: an account of nested semantic structures



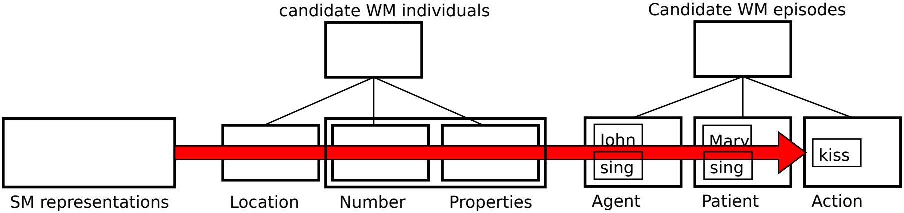
Benefit 5: an account of nested semantic structures



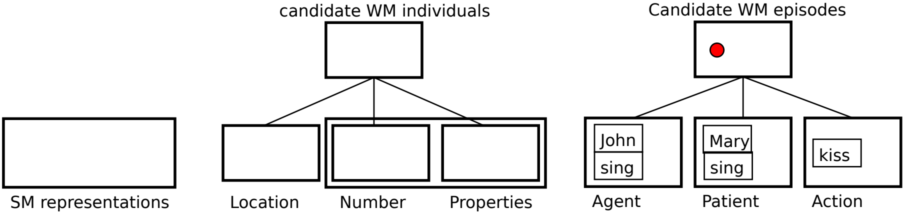
Benefit 5: an account of nested semantic structures



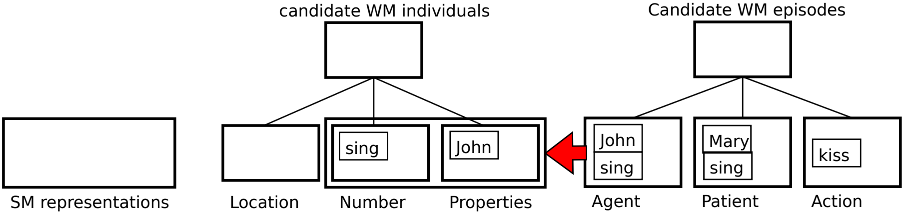
Benefit 5: an account of nested semantic structures



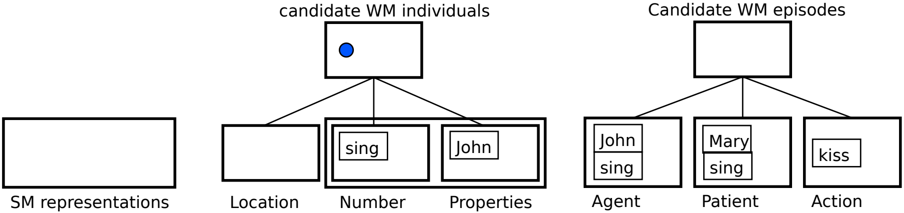
Benefit 5: an account of nested semantic structures



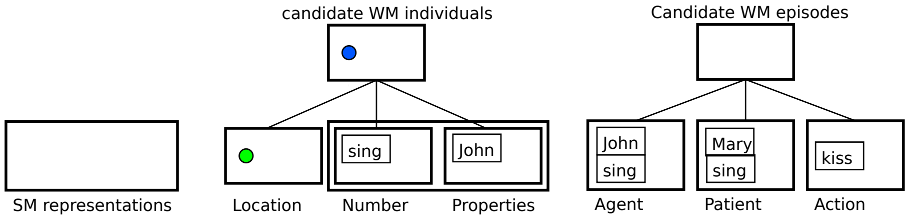
Benefit 5: an account of nested semantic structures



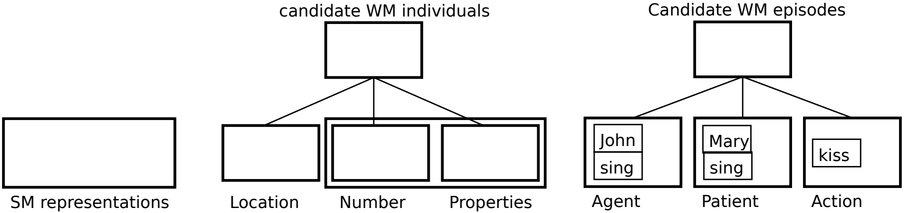
Benefit 5: an account of nested semantic structures



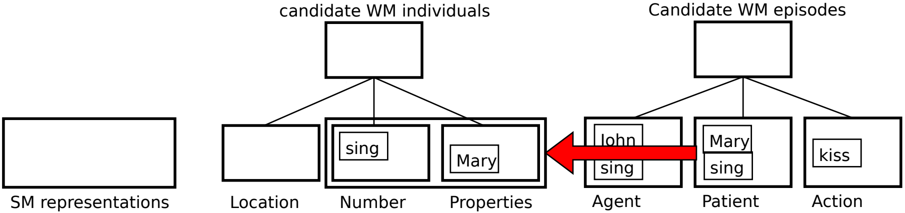
Benefit 5: an account of nested semantic structures



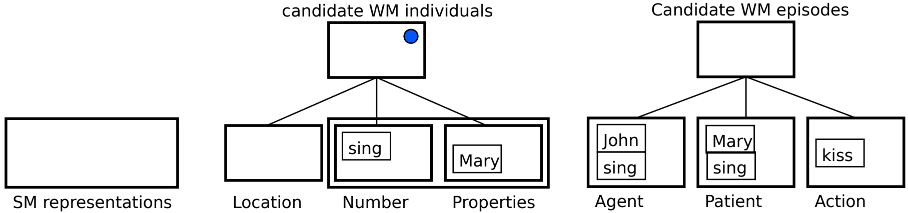
Benefit 5: an account of nested semantic structures



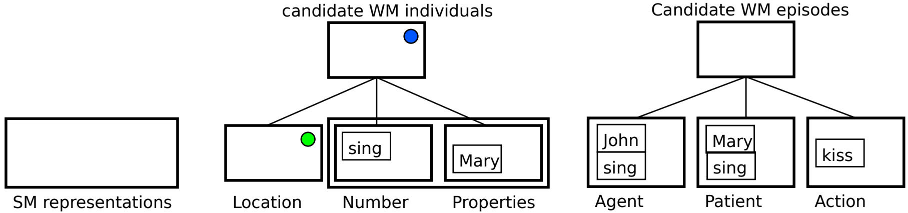
Benefit 5: an account of nested semantic structures



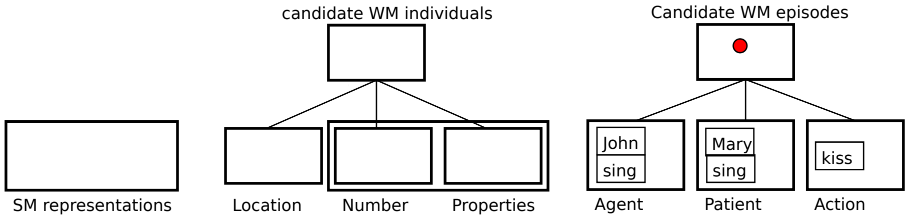
Benefit 5: an account of nested semantic structures



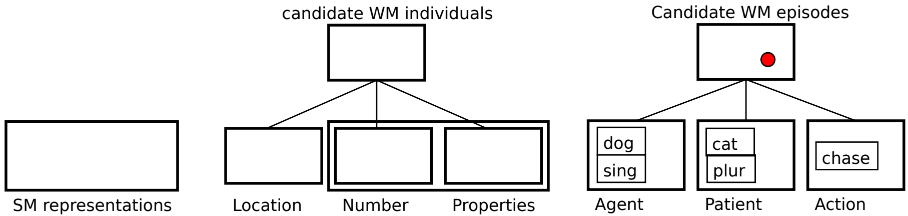
Benefit 5: an account of nested semantic structures



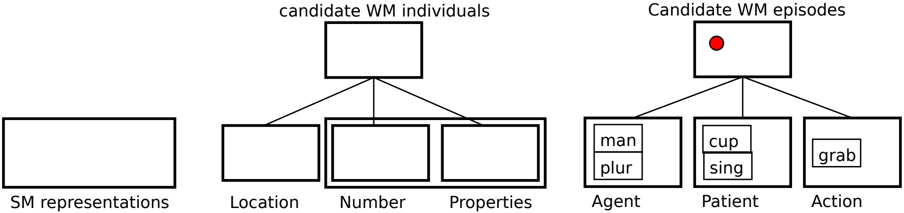
Benefit 6: representations of probability distributions



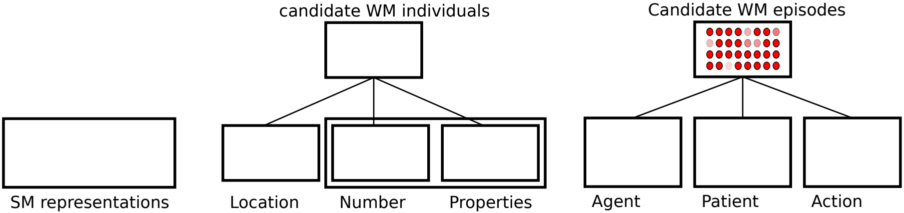
Benefit 6: representations of probability distributions



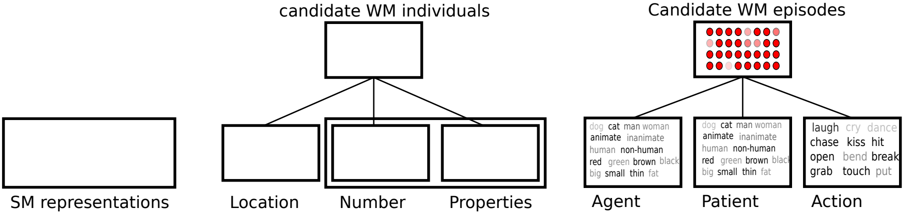
Benefit 6: representations of probability distributions



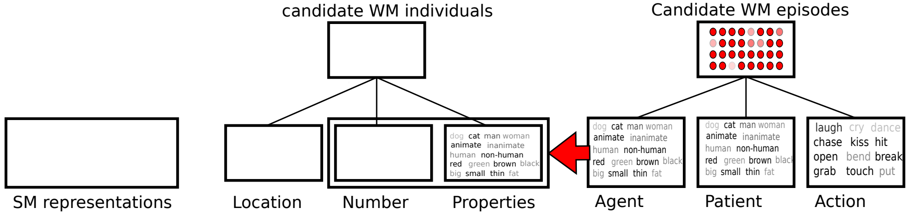
Benefit 6: representations of probability distributions



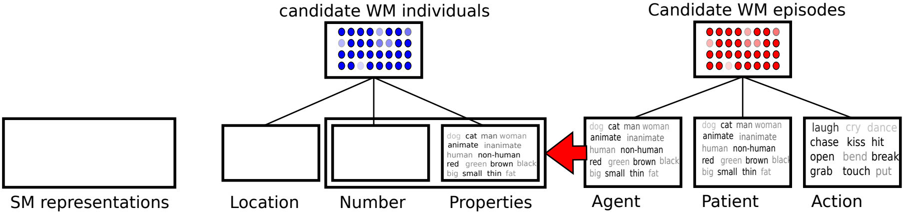
Benefit 6: representations of probability distributions



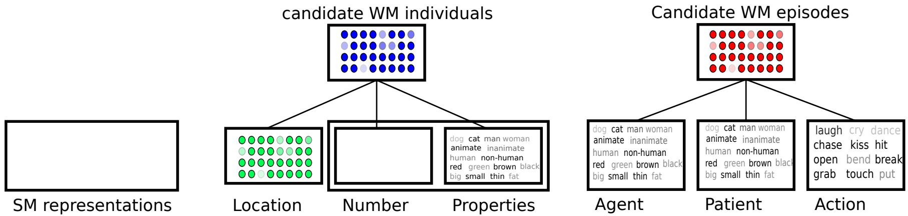
Benefit 7: dynamic expectations during SM experience



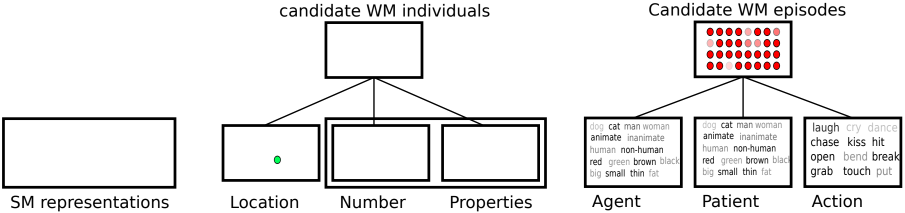
Benefit 7: dynamic expectations during SM experience



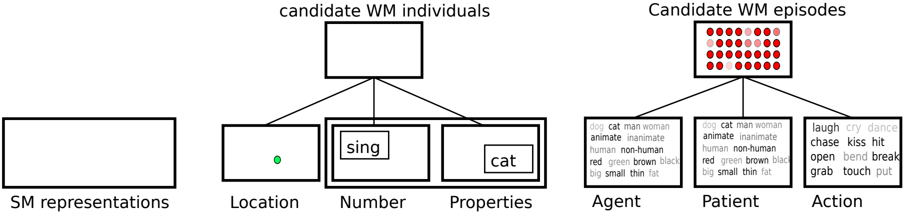
Benefit 7: dynamic expectations during SM experience



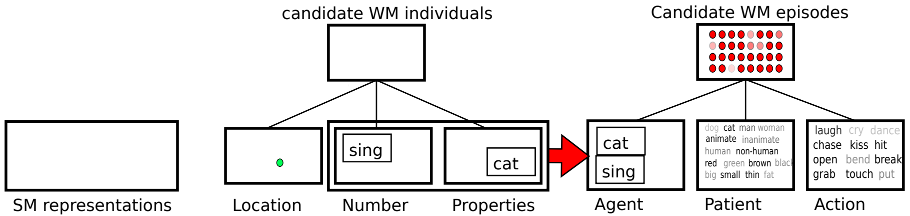
Benefit 7: dynamic expectations during SM experience



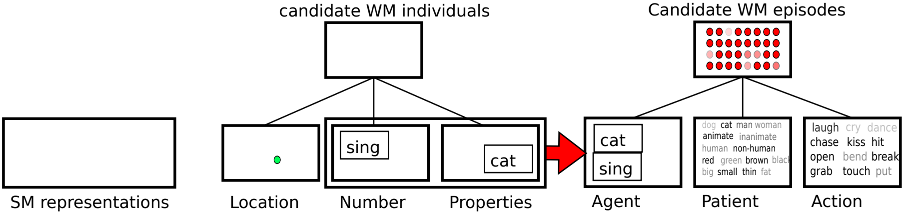
Benefit 7: dynamic expectations during SM experience



Benefit 7: dynamic expectations during SM experience



Benefit 7: dynamic expectations during SM experience



The training world

The world consists of individuals:

- **Animate objects** of 4 types (PERSON, DOG, CAT, BIRD)
- **Things** of 3 types (CUP, BALL, CHAIR).

The training world

The world consists of individuals:

- **Animate objects** of 4 types (PERSON, DOG, CAT, BIRD)
- **Things** of 3 types (CUP, BALL, CHAIR).

Individuals randomly generated—vary in **gender** and **colour** (R, G, B), placed at different **locations** (100x100 grid).

The training world

The world consists of individuals:

- **Animate objects** of 4 types (PERSON, DOG, CAT, BIRD)
- **Things** of 3 types (CUP, BALL, CHAIR).

Individuals randomly generated—vary in **gender** and **colour** (R, G, B), placed at different **locations** (100x100 grid).

100 permanent individuals, non-permanent individuals added with p=0.01 (can be forgotten).

The training world

The world consists of individuals:

- **Animate objects** of 4 types (PERSON, DOG, CAT, BIRD)
- **Things** of 3 types (CUP, BALL, CHAIR).

Individuals randomly generated—vary in **gender** and **colour** (R, G, B), placed at different **locations** (100x100 grid).

100 permanent individuals, non-permanent individuals added with p=0.01 (can be forgotten).

Individuals participate in **episodes**.

Episode

Episode types:

- transitive: agent→patient→trans-action (e.g. MAN CAT STROKE)
- intransitive: agent→intrans-action (e.g. BIRD SING)
- causative: agent→patient→cause-signal→causative-action (e.g. MAN CUP CAUSE-TO-BREAK)

Actions:

- transitive: 10 (GRAB, HIT, PUSH, SEE, HOLD, KICK, HUG, BITE, PAT, STROKE)
- intransitive: 8 (WALK, LIE, SNEEZE, SIT, SLEEP, RUN, SNORE, SING)
- causative: 4 (CAUSE-TO-BREAK, CAUSE-TO-HIDE, CAUSE-TO-STOP,CAUSE-TO-GO)

Episode

Episode types:

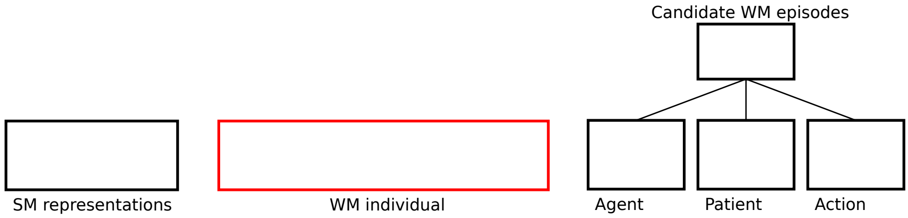
transitive	agent→patient→trans-action (e.g. MAN CAT STROKE)
intransitive	agent→intrans-action (e.g. BIRD SING)
causative	agent→patient→cause-signal→causative-action (e.g. MAN CUP CAUSE-TO-BREAK)

Actions:

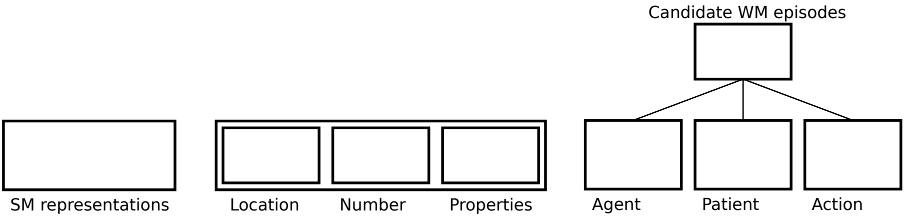
transitive	10 (GRAB, HIT, PUSH, SEE, HOLD, KICK, HUG, BITE, PAT, STROKE)
intransitive	8 (WALK, LIE, SNEEZE, SIT, SLEEP, RUN, SNORE, SING)
causative	4 (CAUSE-TO-BREAK, CAUSE-TO-HIDE, CAUSE-TO-STOP,CAUSE-TO-GO)

The model is exposed to a continuous stream of episodes (not a fixed training set).

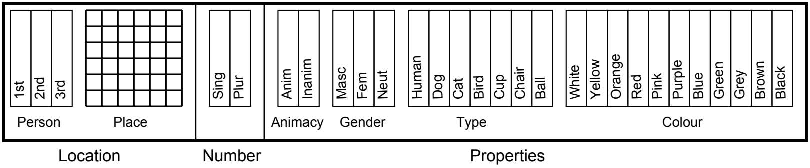
WM individual representation



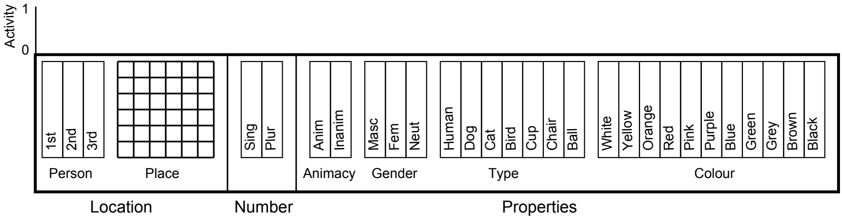
WM individual representation



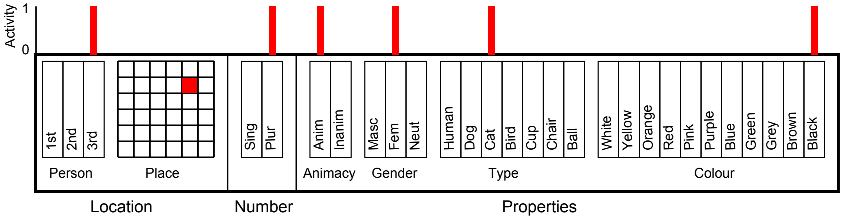
WM individual representation



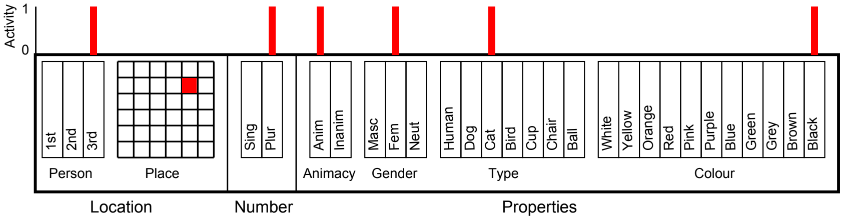
WM individual representation



WM individual representation

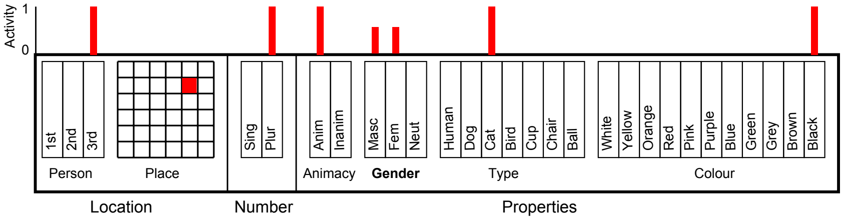


WM individual representation



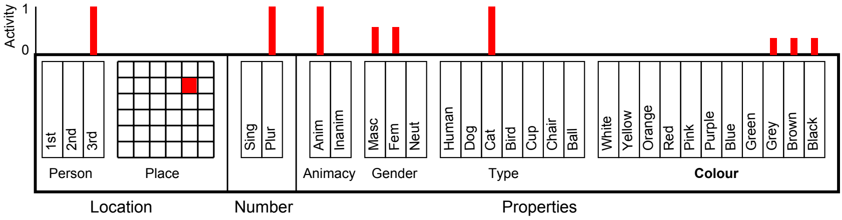
“A group of black female cats.”

WM individual representation



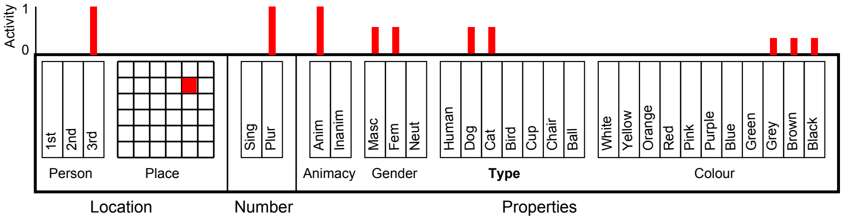
"A group of black cats."

WM individual representation



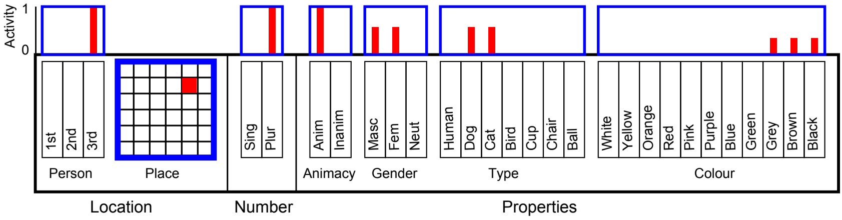
"A group of dark-coloured cats."

WM individual representation



"A group of dark-coloured (dog/cat-like) animals."

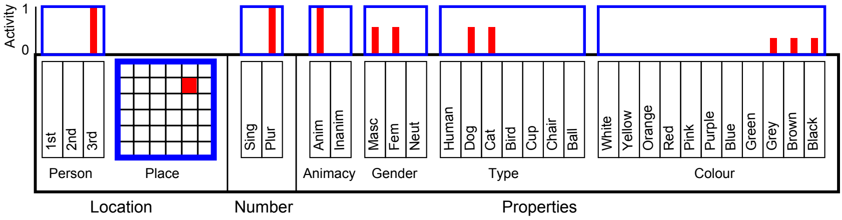
WM individual representation



"A group of dark-coloured (dog/cat-like) animals."

- Probability distributions

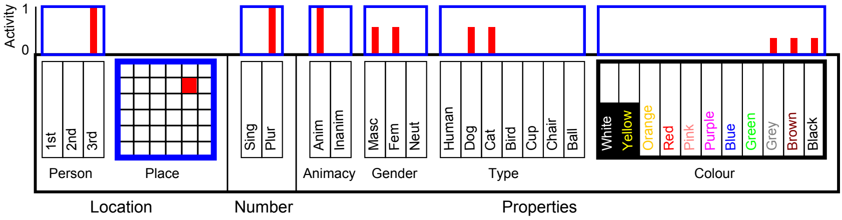
WM individual representation



"A group of dark-coloured (dog/cat-like) animals."

- Probability distributions
- Locally-tuned detectors

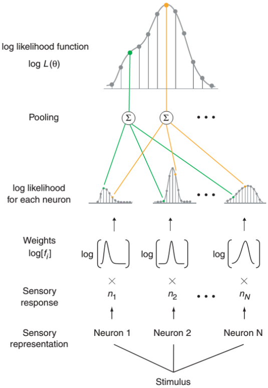
WM individual representation



"A group of dark-coloured (dog/cat-like) animals."

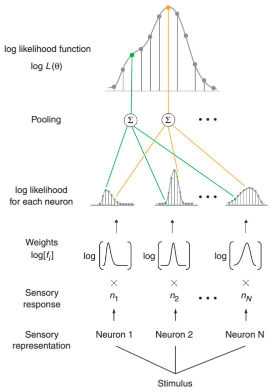
- Probability distributions
- Locally-tuned detectors

Probabilistic interpretation of neuronal activities



Jazayeri and Movshon, 'Optimal representation of sensory information by neural populations'
Nature Neuroscience 2006

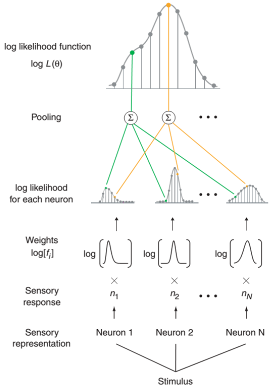
Probabilistic interpretation of neuronal activities



- Sensory information encoded by populations of neurons with local receptive fields.

Jazayeri and Movshon, 'Optimal representation of sensory information by neural populations'
Nature Neuroscience 2006

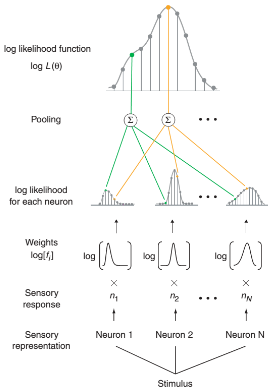
Probabilistic interpretation of neuronal activities



- Sensory information encoded by populations of neurons with local receptive fields.
- From tuning curves of individual neurons, compute for each neuron the (log)likelihood that its response was generated by a particular stimulus.

Jazayeri and Movshon, 'Optimal representation of sensory information by neural populations'
Nature Neuroscience 2006

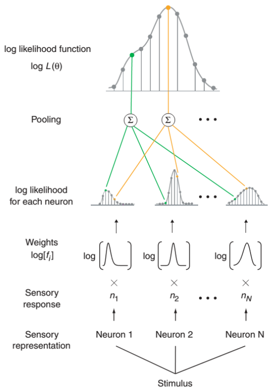
Probabilistic interpretation of neuronal activities



- Sensory information encoded by populations of neurons with local receptive fields.
- From tuning curves of individual neurons, compute for each neuron the (log)likelihood that its response was generated by a particular stimulus.
- Log-likelihoods of neurons additively combined to yield the log-likelihood of the population response to the stimulus.

Jazayeri and Movshon, 'Optimal representation of sensory information by neural populations'
Nature Neuroscience 2006

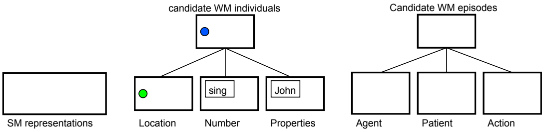
Probabilistic interpretation of neuronal activities



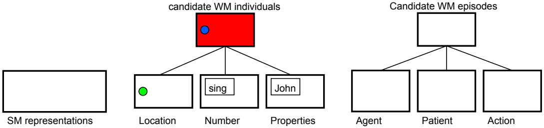
- Sensory information encoded by populations of neurons with local receptive fields.
- From tuning curves of individual neurons, compute for each neuron the (log)likelihood that its response was generated by a particular stimulus.
- Log-likelihoods of neurons additively combined to yield the log-likelihood of the population response to the stimulus.
- By repeating for all stimuli, get the likelihood of every stimulus for the particular observed population response.

Jazayeri and Movshon, 'Optimal representation of sensory information by neural populations' Nature Neuroscience 2006

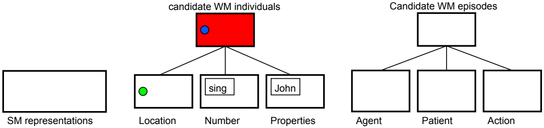
Candidate WM individuals



Candidate WM individuals



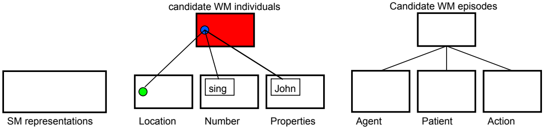
Candidate WM individuals



Functions:

- **Storage:** Store *exact* combinations of Location+Number+Properties for a *short period of time*.

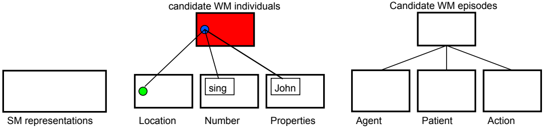
Candidate WM individuals



Functions:

- **Storage:** Store *exact* combinations of Location+Number+Properties for a *short period of time*.

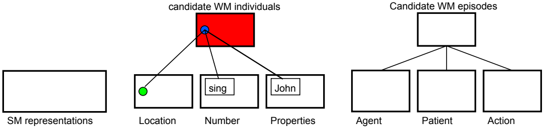
Candidate WM individuals



Functions:

- **Storage:** Store *exact* combinations of Location+Number+Properties for a *short period of time*.
- **Novelty detection:** For a sensory input (in WM individual), decide whether it corresponds to a novel (not recently seen) individual.

Candidate WM individuals



Functions:

- **Storage:** Store *exact* combinations of Location+Number+Properties for a *short period of time*.
- **Novelty detection:** For a sensory input (in WM individual), decide whether it corresponds to a novel (not recently seen) individual.
- **Recognition:** If not novel, tell which of the stored individuals does the sensory input correspond to.

Recognition/novelty in candidate WM individuals

- **Stimulus:** a perceived individual in the world

Recognition/novelty in candidate WM individuals

- **Stimulus:** a perceived individual in the world
- **Neural response:** SM activity stored in WM individual (*I*)

Recognition/novelty in candidate WM individuals

- **Stimulus:** a perceived individual in the world
- **Neural response:** SM activity stored in WM individual (I)
- **Candidate stimuli:** remembered in weights of candidate WM indiv. units (C_i)

Recognition/novelty in candidate WM individuals

- **Stimulus:** a perceived individual in the world
- **Neural response:** SM activity stored in WM individual (I)
- **Candidate stimuli:** remembered in weights of candidate WM indiv. units (C_i)
- **Likelihoods:** $-\log L(C_i) \approx \mathrm{KL}(I, C_i)$

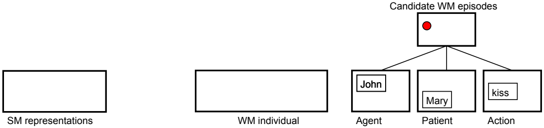
Recognition/novelty in candidate WM individuals

- **Stimulus:** a perceived individual in the world
- **Neural response:** SM activity stored in WM individual (I)
- **Candidate stimuli:** remembered in weights of candidate WM indiv. units (C_i)
- **Likelihoods:** $-\log L(C_i) \approx \text{KL}(I, C_i)$
- **Recognized:** most likely $C_j,\ j = \text{argmin}_i \text{KL}(I, C_i)$

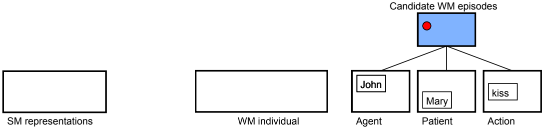
Recognition/novelty in candidate WM individuals

- **Stimulus:** a perceived individual in the world
- **Neural response:** SM activity stored in WM individual (I)
- **Candidate stimuli:** remembered in weights of candidate WM indiv. units (C_i)
- **Likelihoods:** $-\log L(C_i) \approx \mathrm{KL}(I, C_i)$
- **Recognized:** most likely $C_j,\ j = \mathrm{argmin}_i \mathrm{KL}(I, C_i)$
- **Novelty:** if KL of the most likely candidate still too big

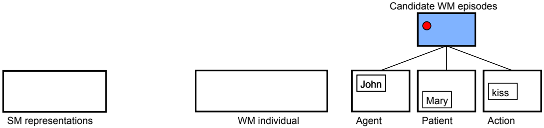
Candidate WM episodes



Candidate WM episodes

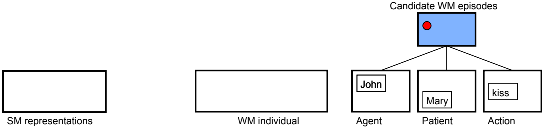


Candidate WM episodes



Functions:

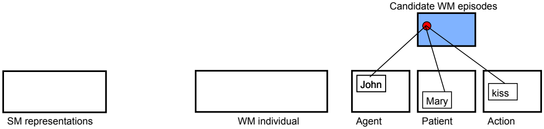
Candidate WM episodes



Functions:

- **Storage:** Store typical combinations of Agent+Patient+Action for a longer period of time.

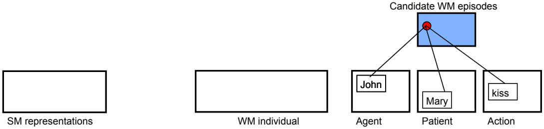
Candidate WM episodes



Functions:

- **Storage:** Store typical combinations of Agent+Patient+Action for a longer period of time.

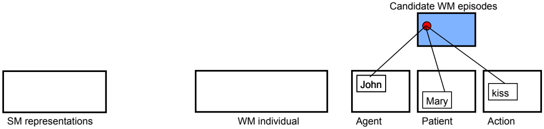
Candidate WM episodes



Functions:

- **Storage:** Store *typical* combinations of Agent+Patient+Action for a *longer period of time*.

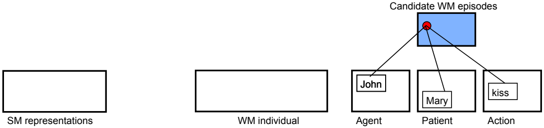
Candidate WM episodes



Functions:

- **Storage:** Store *typical* combinations of Agent+Patient+Action for a *longer period of time*.
- **Associative retrieval:** Retrieve typical episodes most closely resembling the input episode.

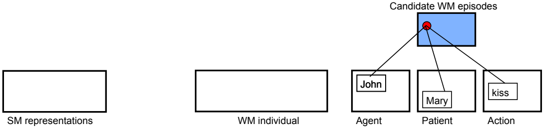
Candidate WM episodes



Functions:

- **Storage:** Store *typical* combinations of Agent+Patient+Action for a *longer period of time*.
- **Associative retrieval:** Retrieve typical episodes most closely resembling the input episode.
- **Topological organization:** Represent similar types of episodes close to each other.

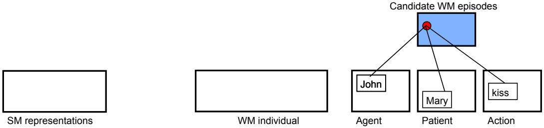
Candidate WM episodes



Functions:

- **Storage:** Store *typical* combinations of Agent+Patient+Action for a *longer period of time*.
- **Associative retrieval:** Retrieve typical episodes most closely resembling the input episode.
- **Topological organization:** Represent similar types of episodes close to each other.
- **Distribution:** Represent multiple episode types in parallel (probability distribution).

Candidate WM episodes

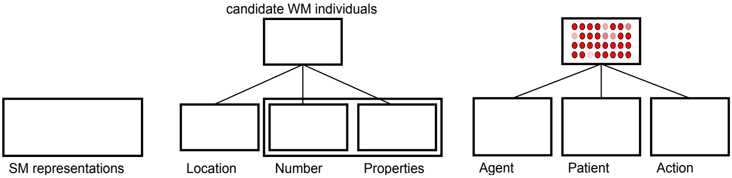


Functions:

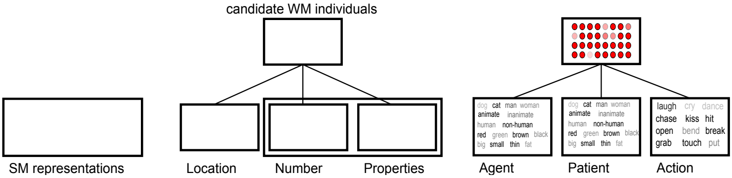
- **Storage:** Store *typical* combinations of Agent+Patient+Action for a *longer period of time*.
- **Associative retrieval:** Retrieve typical episodes most closely resembling the input episode.
- **Topological organization:** Represent similar types of episodes close to each other.
- **Distribution:** Represent multiple episode types in parallel (probability distribution).

Self-organizing map (SOM) (Kohonen, 1982)

Representation of distribution of episodes

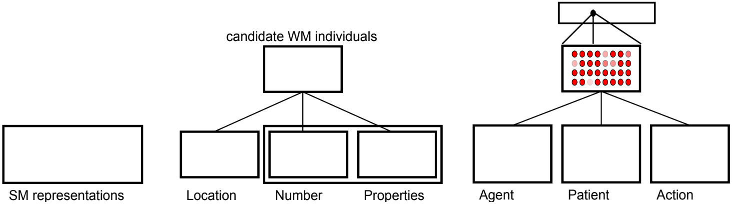


Representation of distribution of episodes

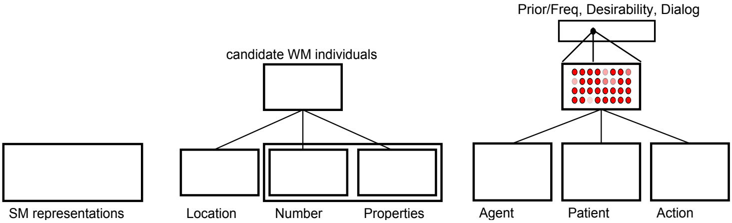


Top-down bias

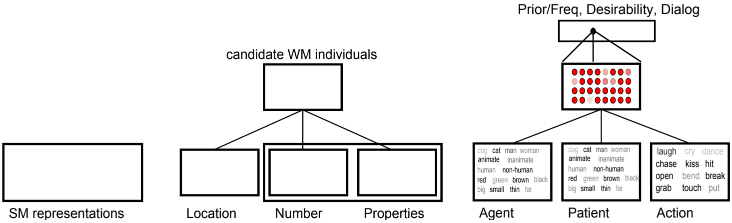
Representation of distribution of episodes



Representation of distribution of episodes



Representation of distribution of episodes



Expectations on episode continuation

Expectations on episode continuation

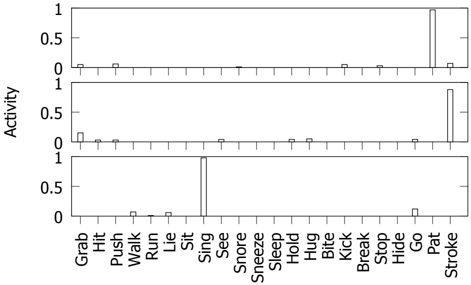
- PERSON+DOG $\rightarrow$ PAT
- PERSON+CAT $\rightarrow$ STROKE

Expectations on episode continuation

- PERSON+DOG $\rightarrow$ PAT
- PERSON+CAT $\rightarrow$ STROKE
- BIRD $\rightarrow$ SING

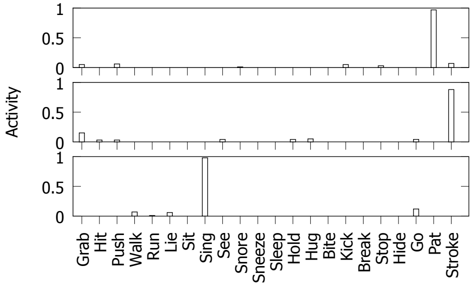
Expectations on episode continuation

- PERSON+DOG $\rightarrow$ PAT
- PERSON+CAT $\rightarrow$ STROKE
- BIRD $\rightarrow$ SING

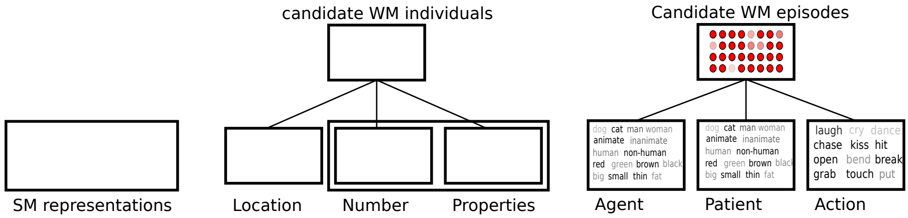


Expectations on episode continuation

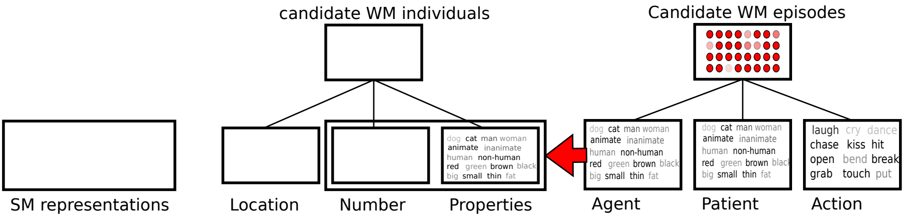
- PERSON+DOG → PAT
- PERSON+CAT → STROKE
- BIRD → SING
- prior expectations



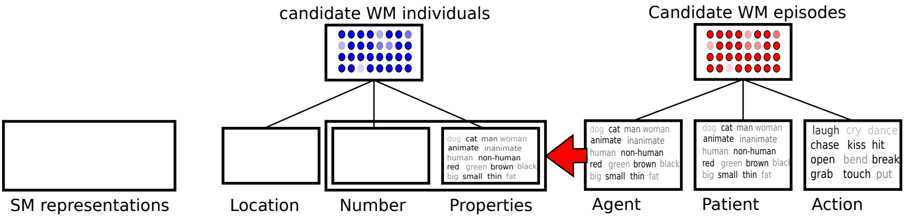
Expectations on properties/locations of individuals



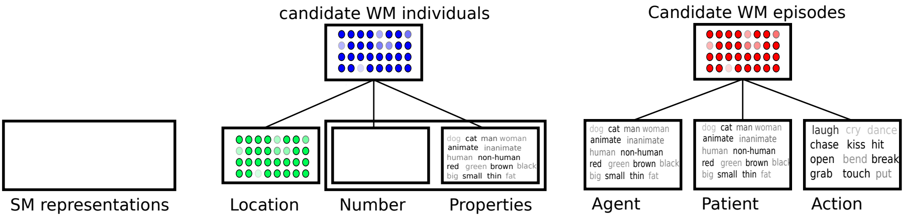
Expectations on properties/locations of individuals



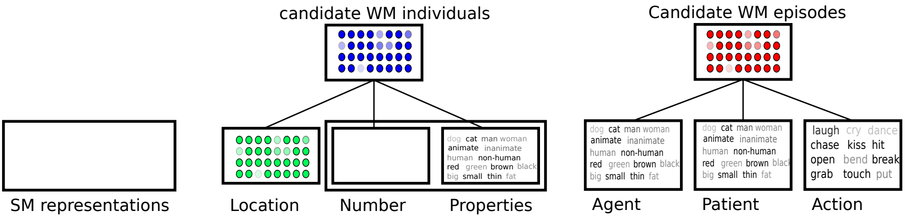
Expectations on properties/locations of individuals



Expectations on properties/locations of individuals

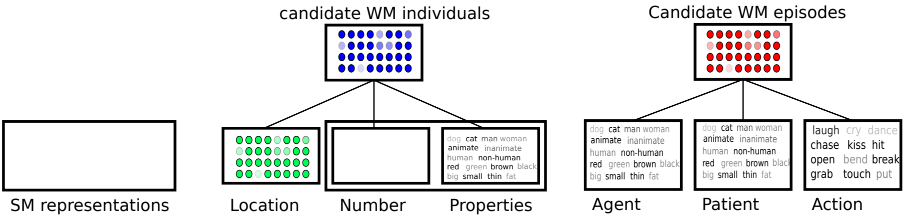


Expectations on properties/locations of individuals



- MAN $\rightarrow$ black DOG

Expectations on properties/locations of individuals



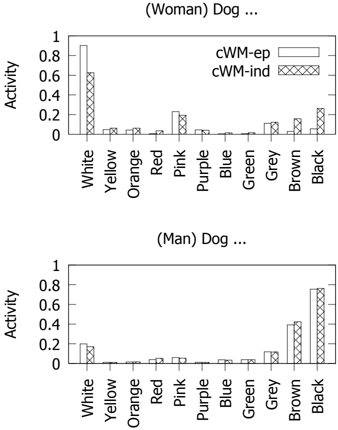
- MAN → black DOG
- WOMAN → white DOG

Expectations on properties/locations of individuals

- MAN → black DOG
- WOMAN → white DOG

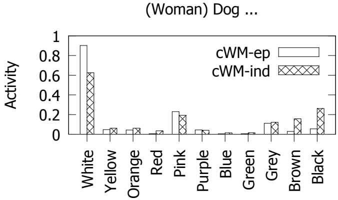
Expectations on properties/locations of individuals

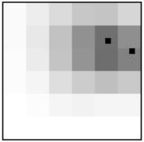
- MAN $\rightarrow$ black DOG
- WOMAN $\rightarrow$ white DOG

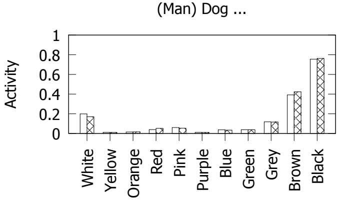


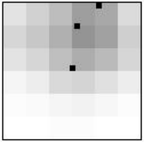
Expectations on properties/locations of individuals

- MAN $\rightarrow$ black DOG
- WOMAN $\rightarrow$ white DOG









Summary

A novel account of semantic working memory that supports:

- simulations of stored episodes,
- binding between roles and fillers,
- nested semantic structures,
- representation of probability distributions of episodes,
- dynamic expectations/predictions.