Word-Object Learning via Visual Exploration in Space (WOLVES): A Neural Process Model of Cross-Situational Word Learning

John P. Spencer

with...

Ajaz A. Bhat

Larissa K. Samuelson

University of East Anglia

SCHOOL OF PSYCHOLOGY

DEVELOPMENTAL DYNAMICS LABORATORY

Words are the building blocks of language

How do people learn the meanings of words when there are an infinite number of possible referents?

- One possibility: Track word-object co-occurrences (cross-situational statistical learning)

“Try some of the *banana*”

Words are the building blocks of language

How do people learn the meanings of words when there are an infinite number of possible referents?

- One possibility: Track word-object co-occurrences (cross-situational statistical learning)
- But what is the nature of this type of statistical learning?

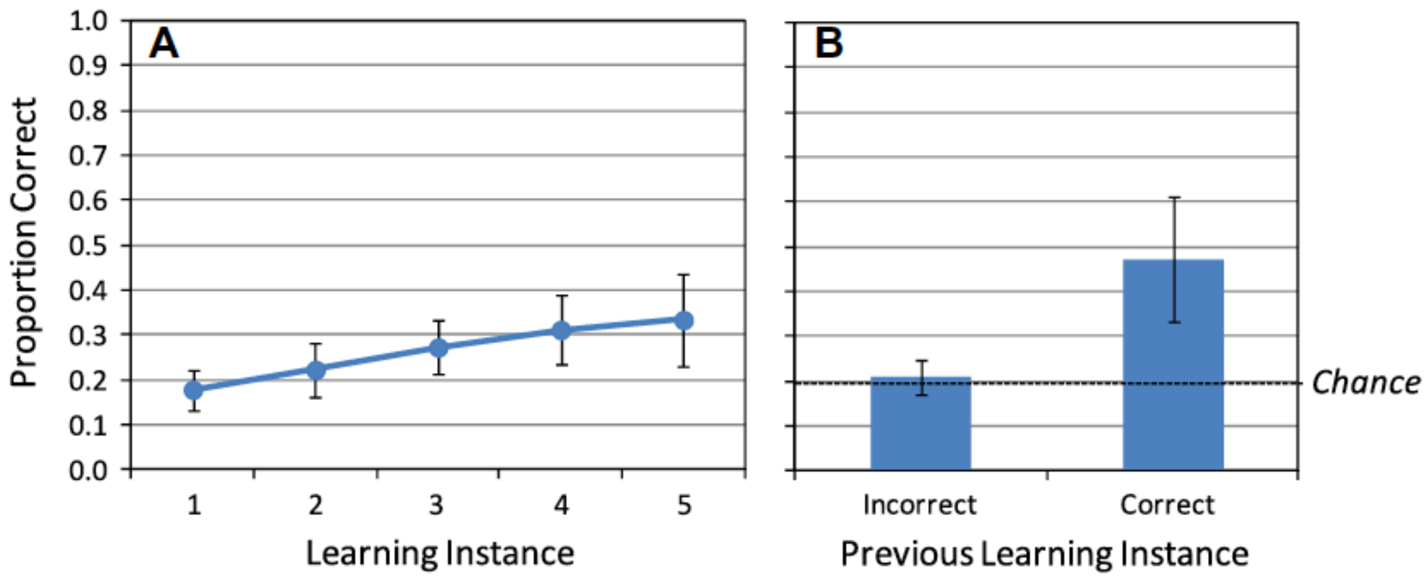
Two classes of theories

- Hypothesis testing accounts
- Associative learning

Hypothesis Testing

- Encounter a novel word
- Make a single hypothesis about the word-object mapping
- If later evidence shows that this hypothesis is wrong, form a new one and proceed to verification…

Hypothesis Testing



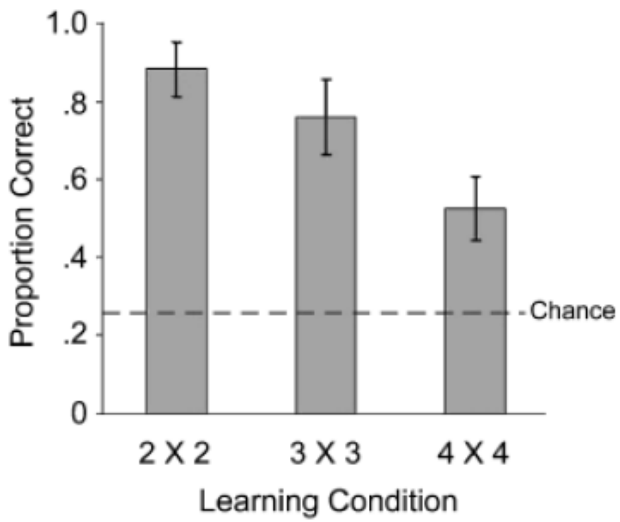
Trueswell et al. (2013). *Cognitive Psychology.*

Associative Learning

- When encounter a novel word, form multiple associations between word and available objects
- Over time, refine these associations based on available co-occurrences
- Strongest association wins (as correct word usage should always drive you to one strong association)

Associative Learning

- 2 words x 2 pictures on each training trial; 3 x 3; 4 x 4
- Taught same 18 words
- 6 repetitions of each target word-ref pairing – so same exposure in each condition but different erroneous mappings (5.09 incorrect mappings in 2x2, 8.78 in 3x3, 12.22 in 4x4)
- 4 AFC test with one word on each test trial (foils from 18)



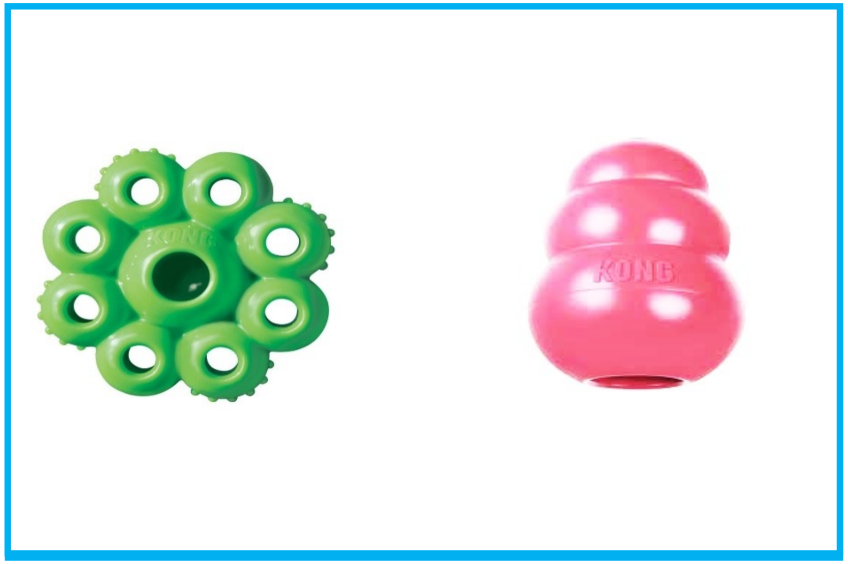
Yu & Smith (2007). *Psychological Science.*

Limitations of existing theories

- Both types of theories have been used to explain the same data; Yu and Smith (2012) used this to call for implementation-level theories
- Current theories are not comprehensive (tend to explain only a subset of data from specific tasks)
- Current theories fail to take time seriously despite evidence that how processes unfold in real time, over learning, and over development matter…

Yu & Smith (2012). *Psychological Review.*

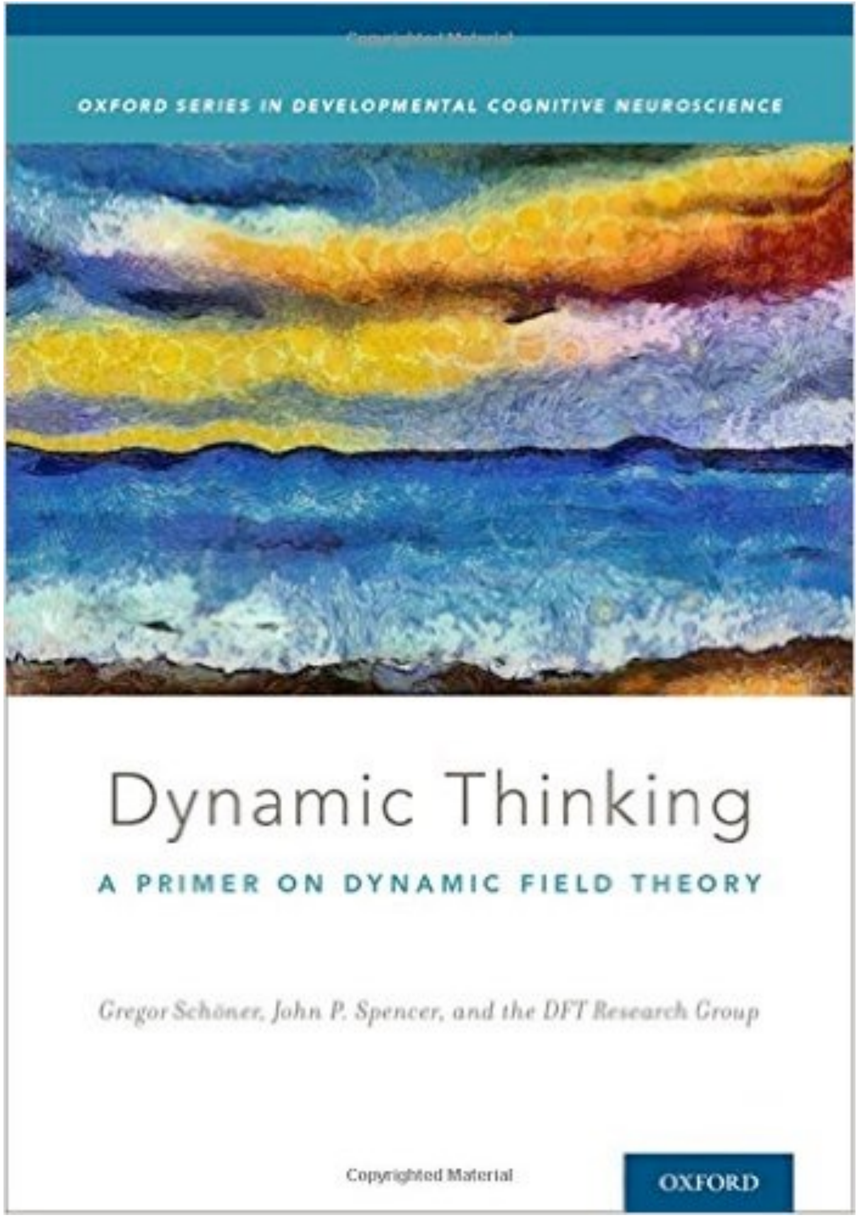
Cross-Situational Word Learning



- 12-14 month old children can learn 4 words (Smith & Yu, 2008, Yu & Smith, 2011). Older kids and adults can learn up to 9-16 words.
  - What is changing over development?
- Individual differences: ‘strong’ vs ‘weak’ learners.
- Moment-by-moment variation in looking matters – strong learners have fewer, longer fixations.

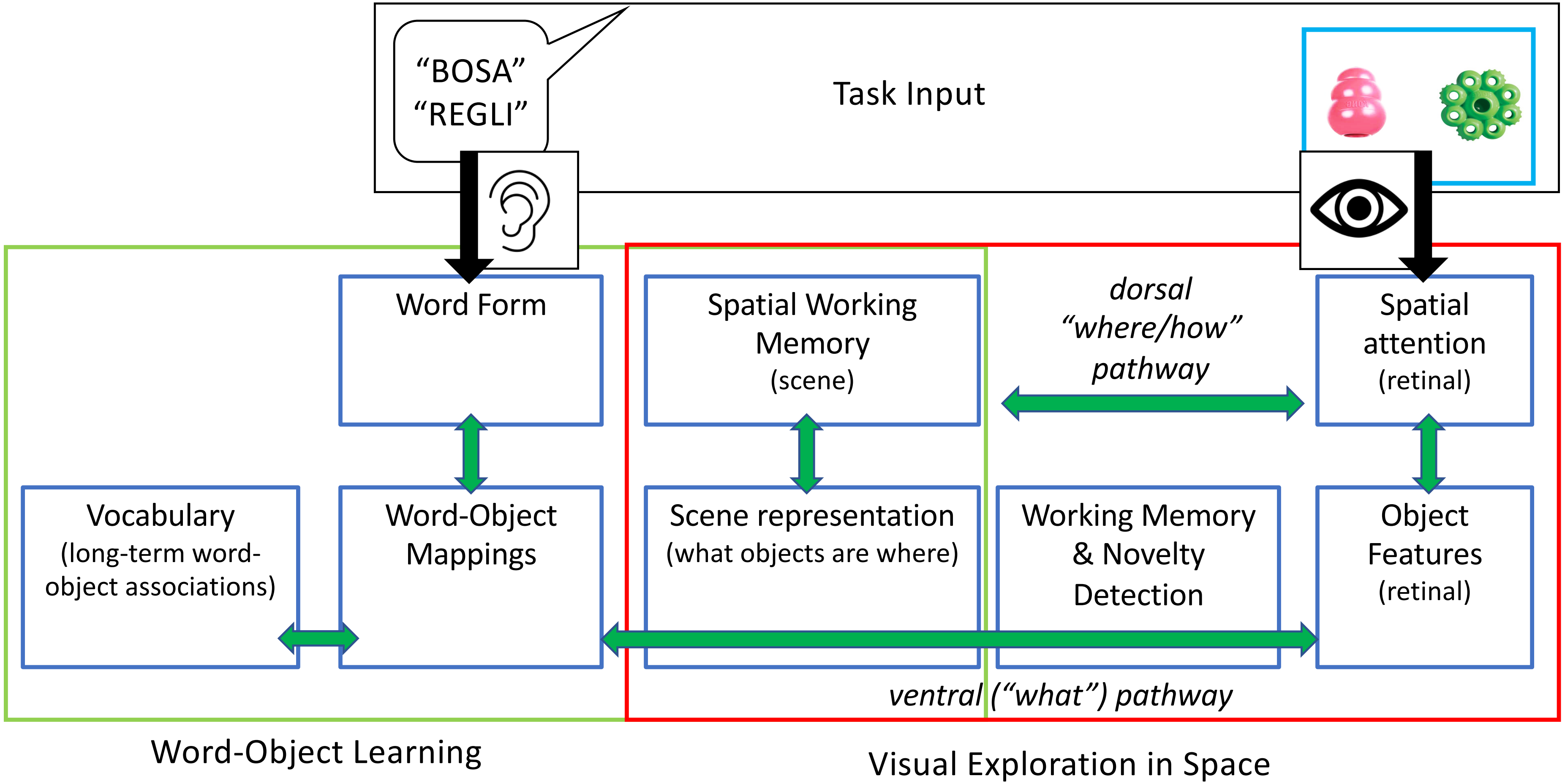
Today's talk focuses on a new theory of CSWL

- WOLVES
  - Overview of model & demonstrate that it is a good model.
- Timescale of the task
  - Simulations that highlight role of attention and learning processes.
- Timescale of development
  - Present the first developmental account of CSWL highlighting the role of memory processes.
- Model evaluation
  - Is the theory comprehensive?
  - How does it fare relative to competitor models?



www.dynamicfieldtheory.org

WOLVES



Samuelson, Smith, Perry & Spencer (2011);
Samuelson, Jenkins & Spencer (2013)

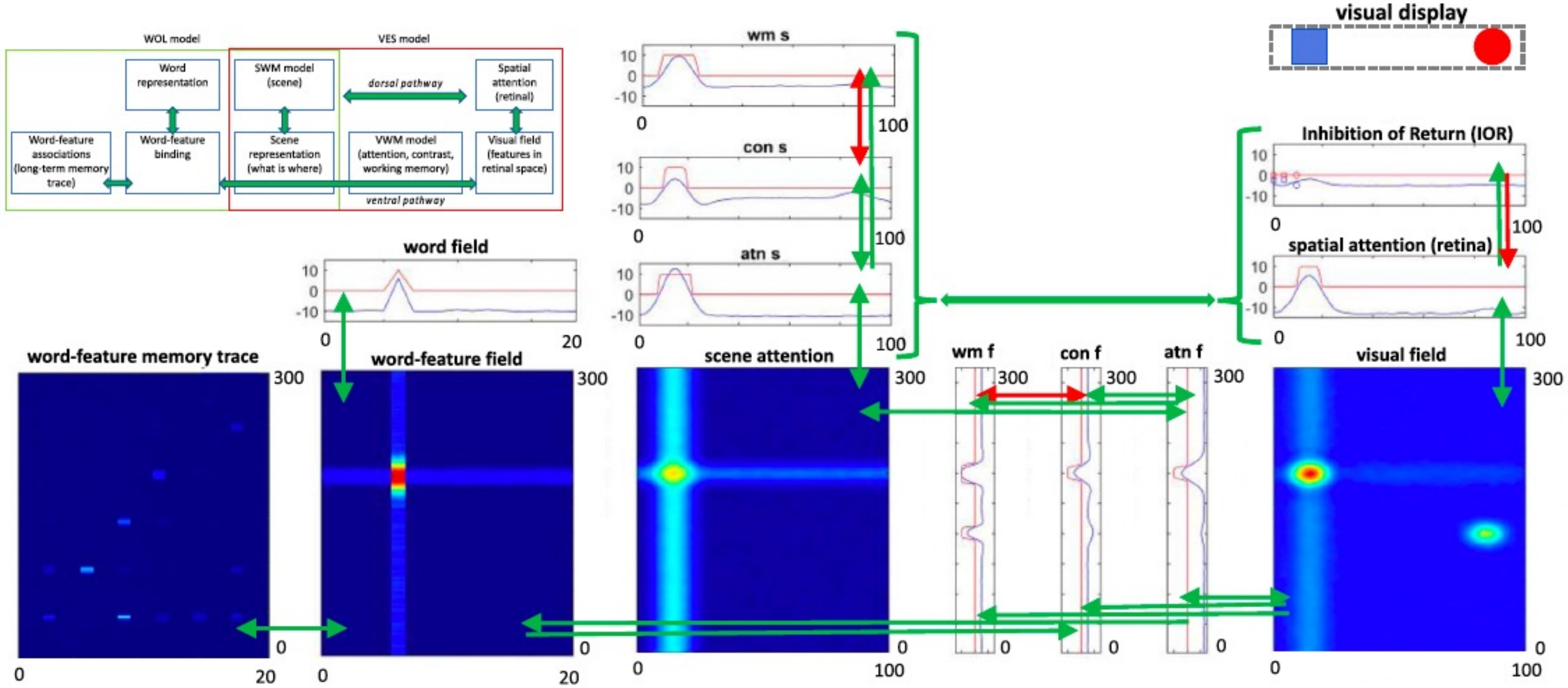
Johnson, Spencer & Schöner (2009);
Perone & Spencer (2013b)

Bhat, Spencer, Samuelson (2021). *Psychological Review.*

WOLVES

Figure 4

The Overall Architecture of WOLVES

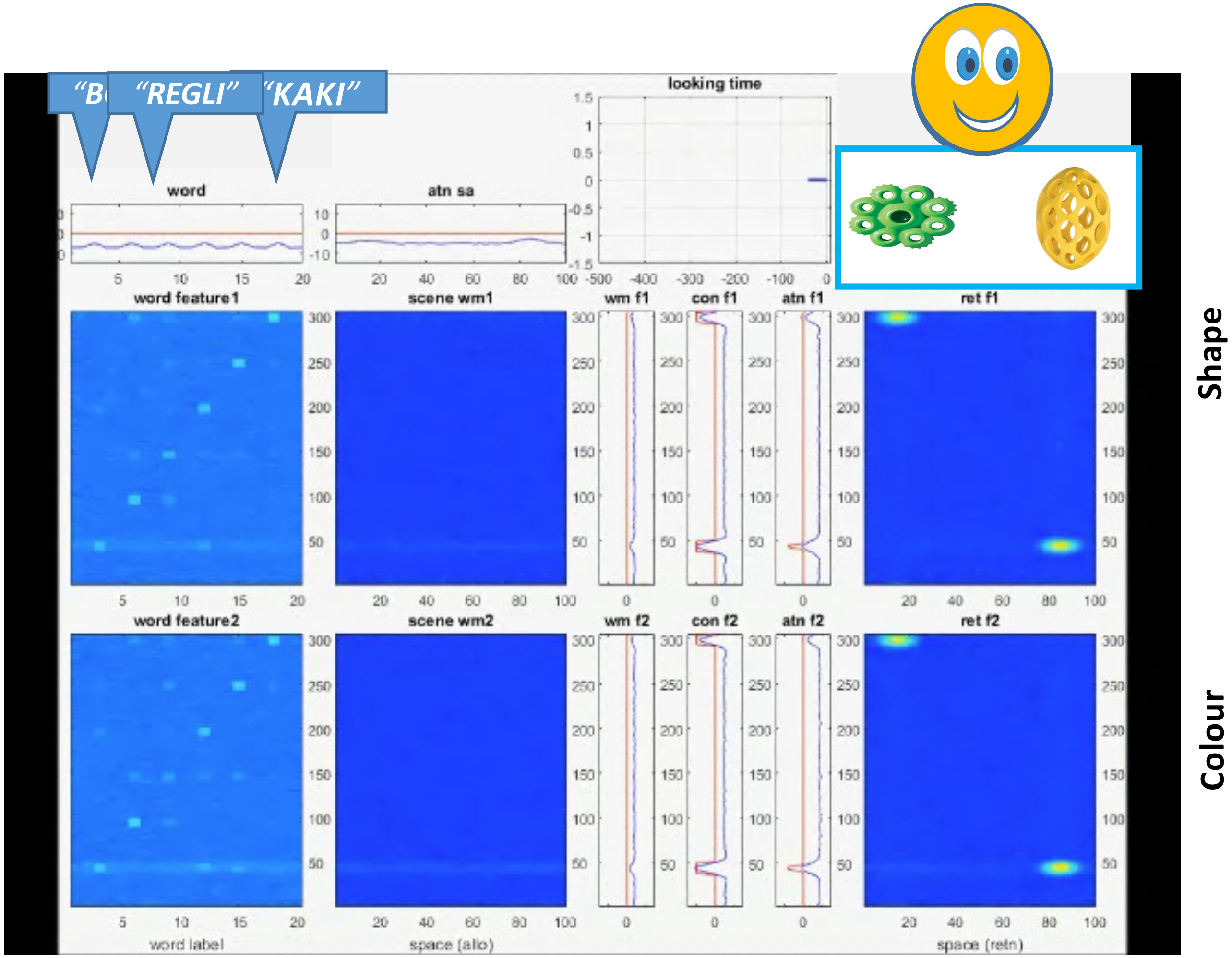


Note. Scene WMs and memory traces are not shown for representational simplicity. Arrows represent uni/bidirectional (green: excitatory, red: inhibitory) connectivity in the model. See text for additional details. WOLVES = word-object learning via visual exploration in space. See the online article for the color version of this figure.

Full model includes ventral pathways for colour *and* shape as well as memory traces for all field except visual field, attention fields and IOR

WOLVES in action

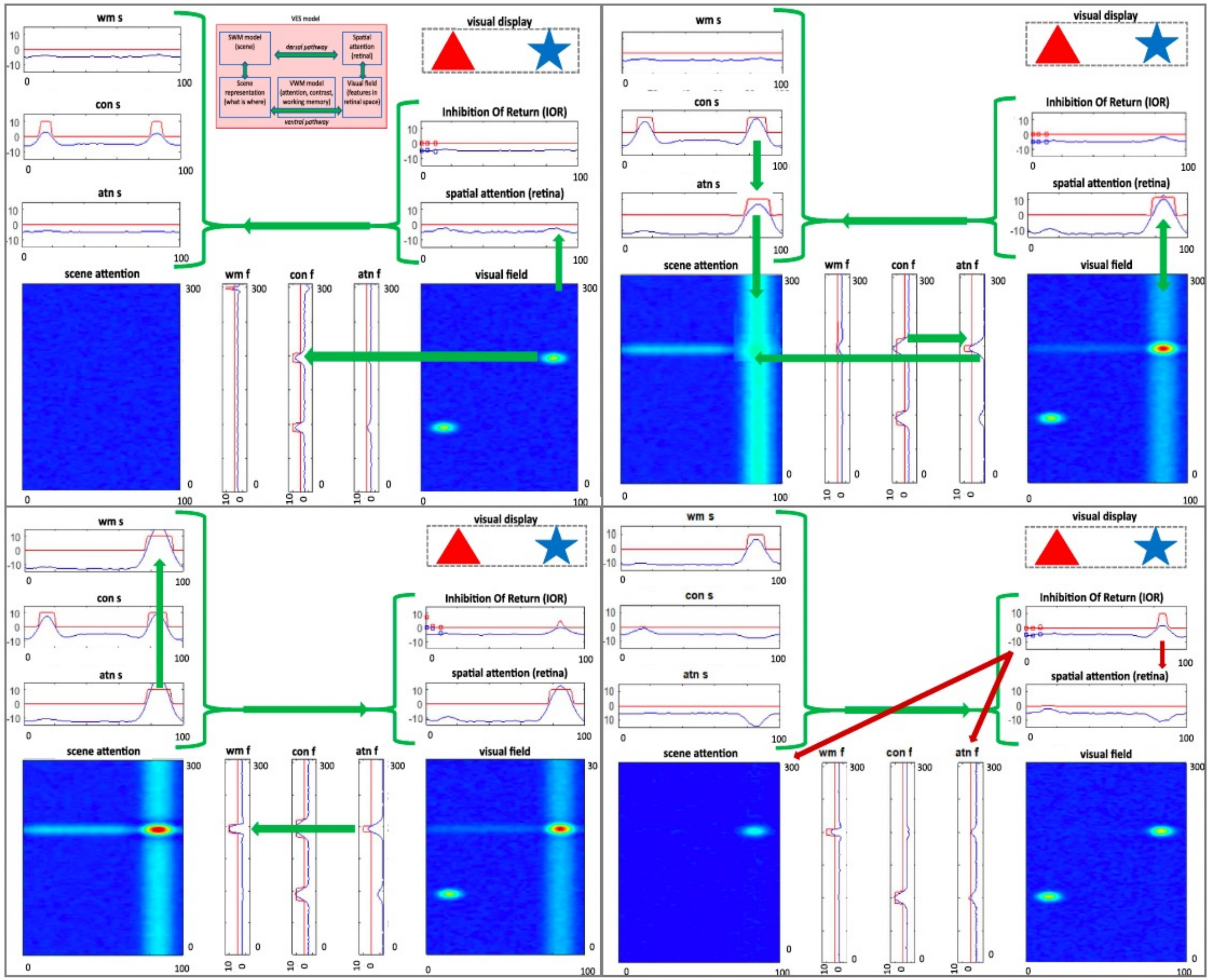
- VES cycles of novelty detection, consolidation in working memory, and release from fixation.
- WOL cycles of associative learning that is non-linear as memory traces evolve
- TDA cycles of top-down memory driven attention



VES Cycle

Figure 3

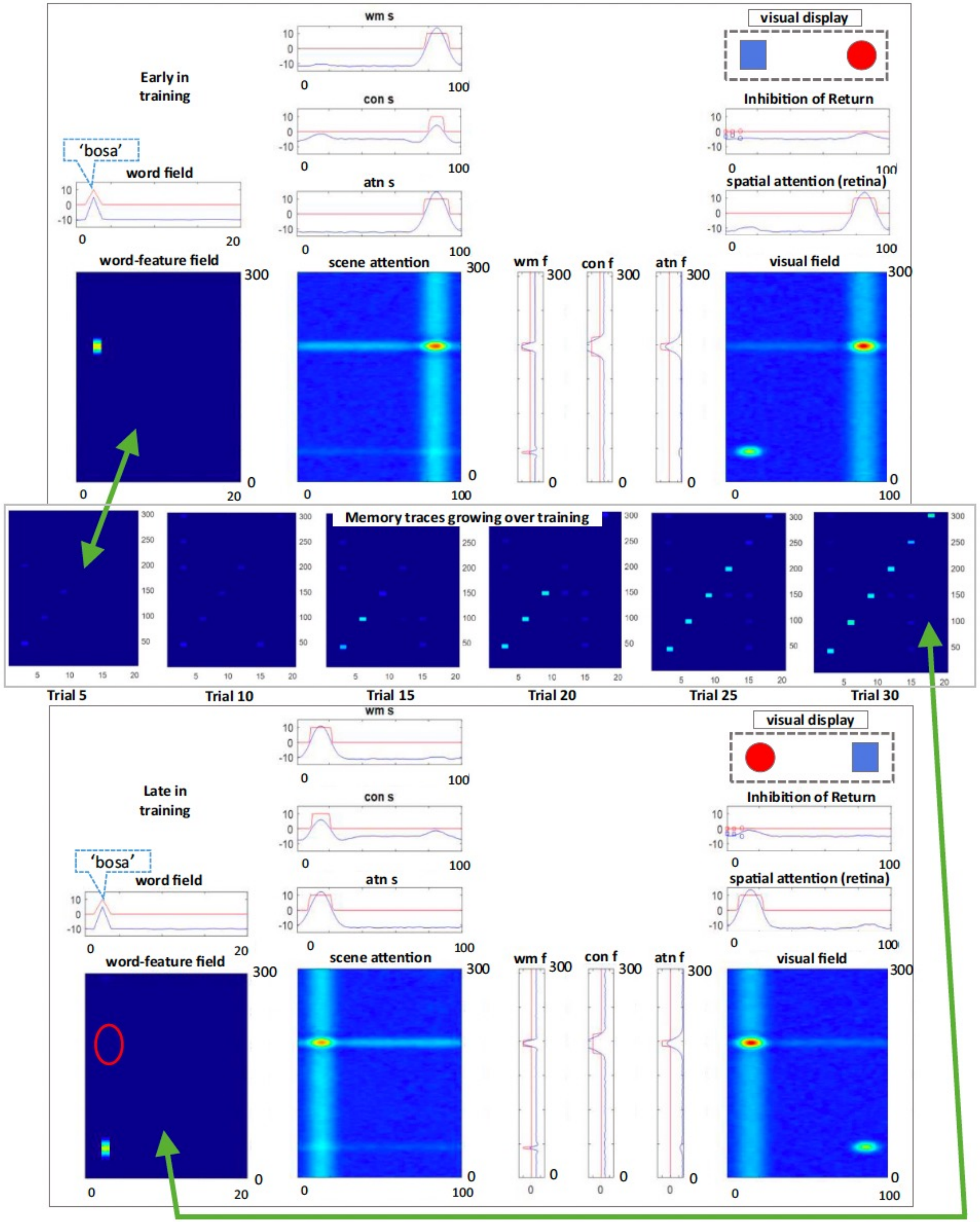
Visual Exploration in Space Model in Four Stages of an Autonomous Looking Cycle



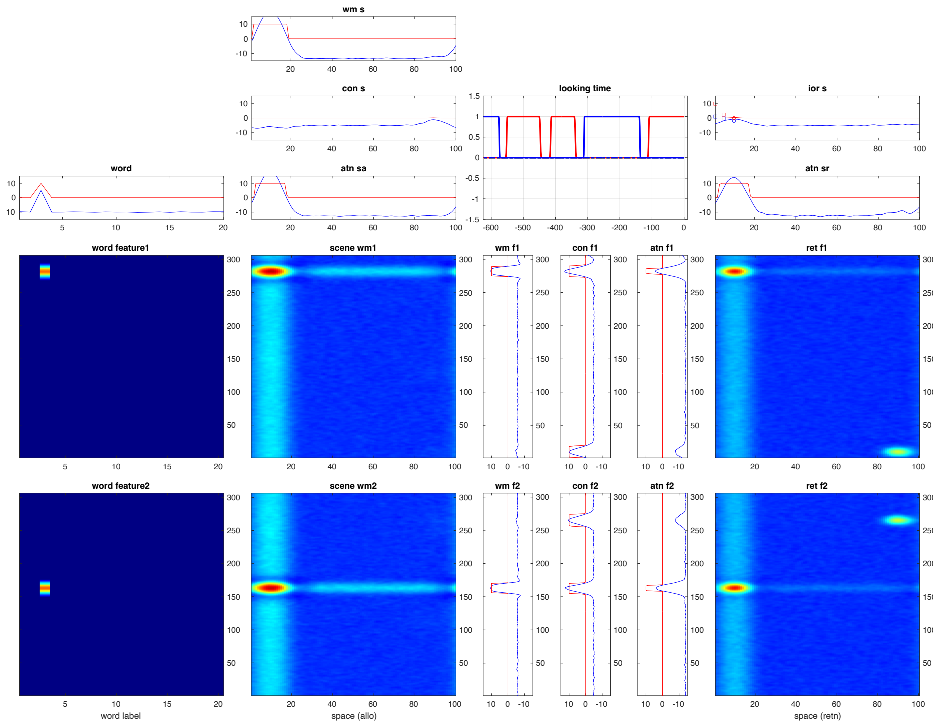
Note. The top-left panel shows the model detecting novel objects in the scene. The top-right panel shows the model attending to one object. The bottom-left panel shows the model having consolidated the object in working memory. The bottom-right panel shows model releasing attention to begin a new looking cycle. VES = visual exploration in space. See the online article for the color version of this figure.

WOL Cycle

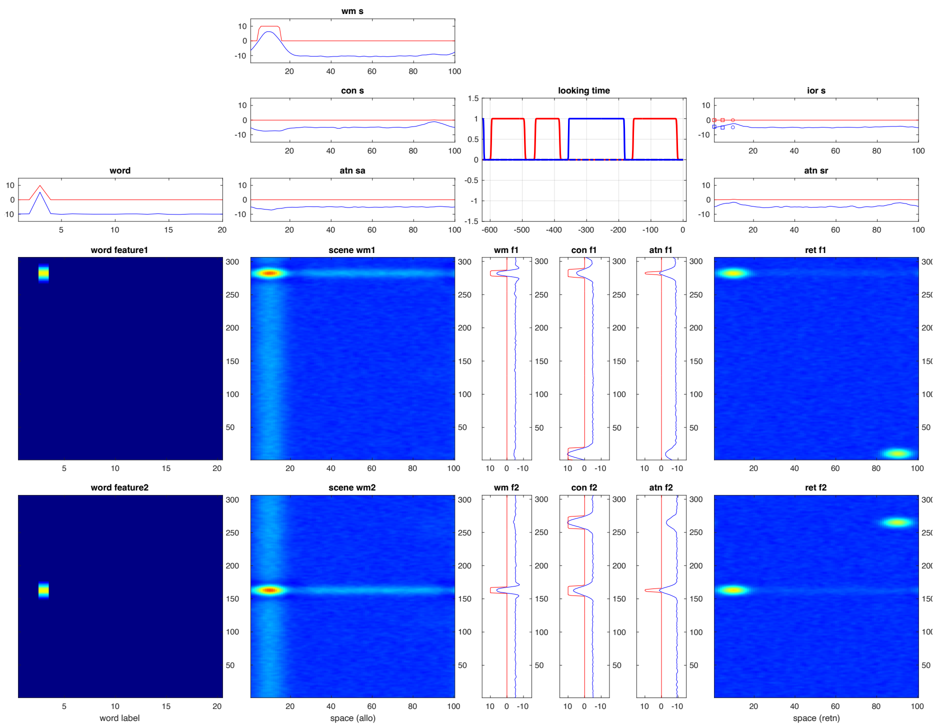
Figure 5
Processing in WOLVES During Smith and Yu's (2008) Cross-Situational Word Learning Task



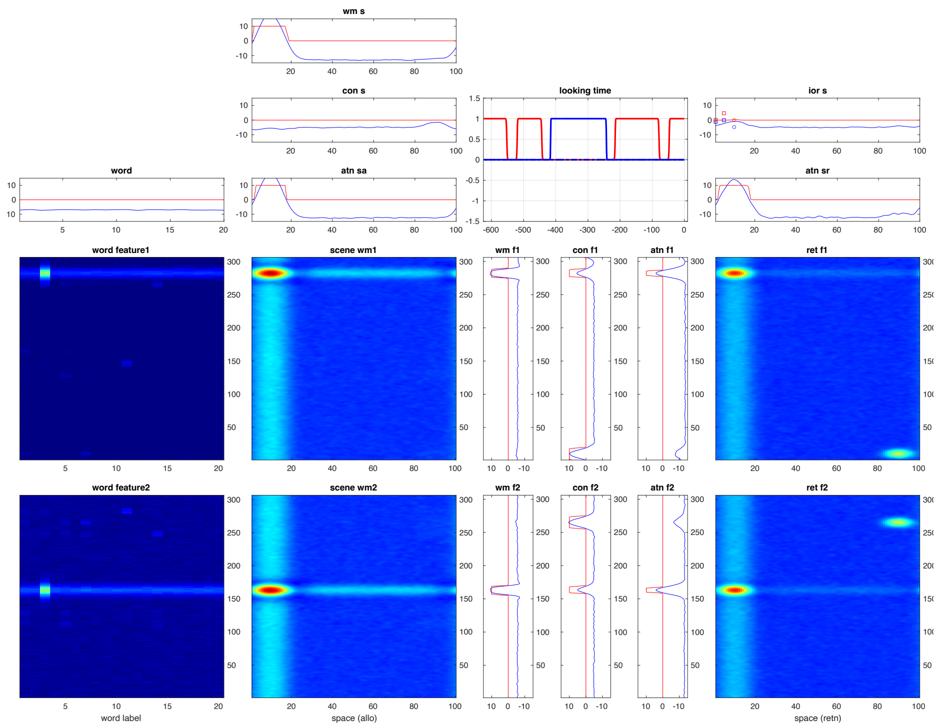
TDA Cycle



wm s
con s
looking time
ior s
word
atn sa
atn sr
word feature1
scene wm1
wm f1
con f1
atn f1
ret f1
word feature2
scene wm2
wm f2
con f2
atn f2
ret f2
word label
space (allo)
space (retn)



wm s
con s
looking time
ior s
word
atn sa
atn sr
word feature1
scene wm1
wm f1
con f1
atn f1
ret f1
word feature2
scene wm2
wm f2
con f2
atn f2
ret f2
word label
space (allo)
space (retn)



wm s
con s
looking time
ior s
word
atn sa
atn sr
word feature1
scene wm1
wm f1
con f1
atn f1
ret f1
word feature2
scene wm2
wm f2
con f2
atn f2
ret f2
word label
space (allo)
space (retn)

Does WOLVES capture – and explain – empirical data?

Will compare WOLVES to Kachergis et al. (2012) as relevant: an AL model that distributes attention between known and novel associations; has memory decay to capture association frequency; one shot computation on each trial.

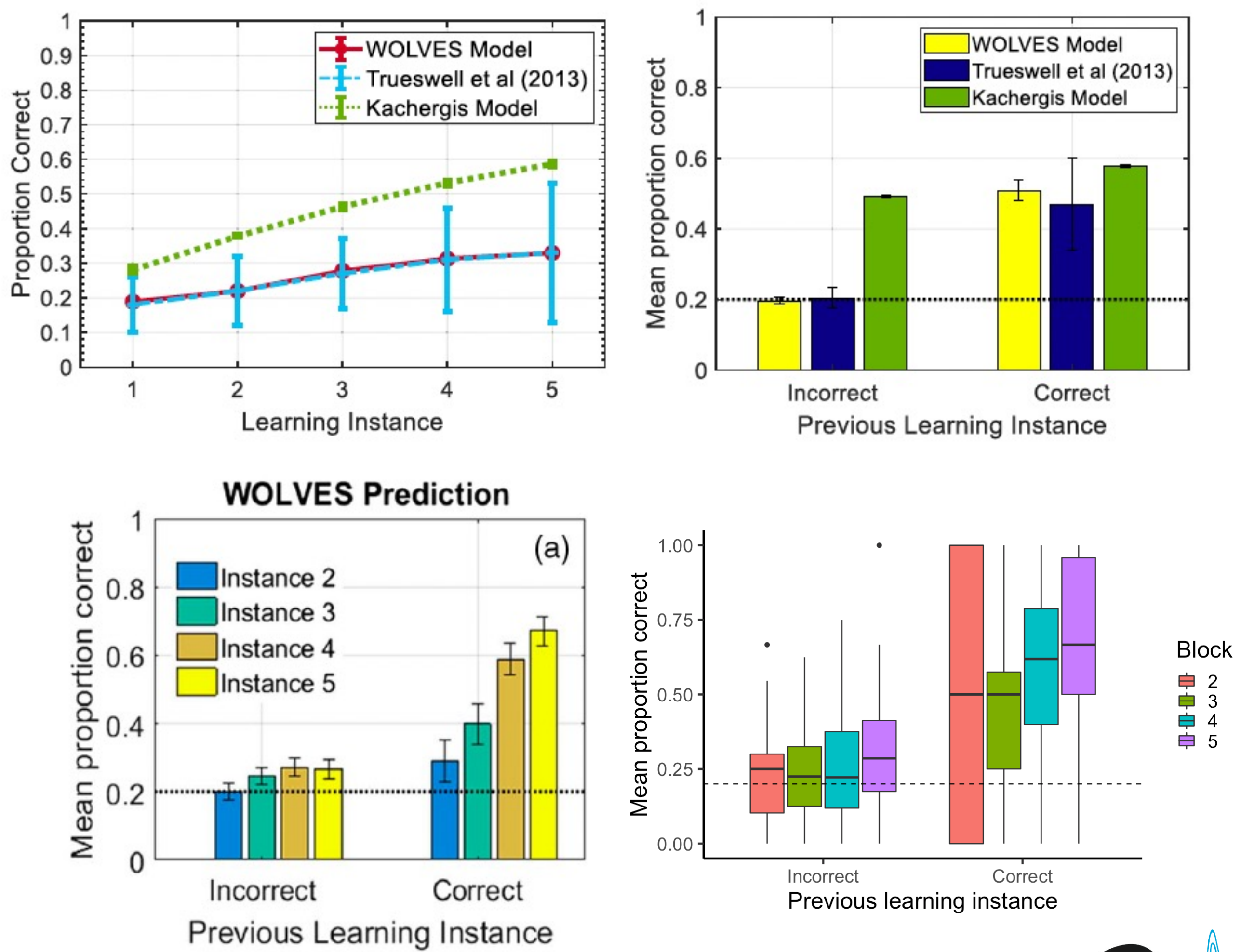
WOLVES explains HT data

WOLVES captures HT data. Why?

- Timing of task means WOLVES typically makes one look on a trial (so only forms one association)
- What if we extend the time?

Figure 13

Data from Trueswell et al. and the WOLVES and Kachergis Models

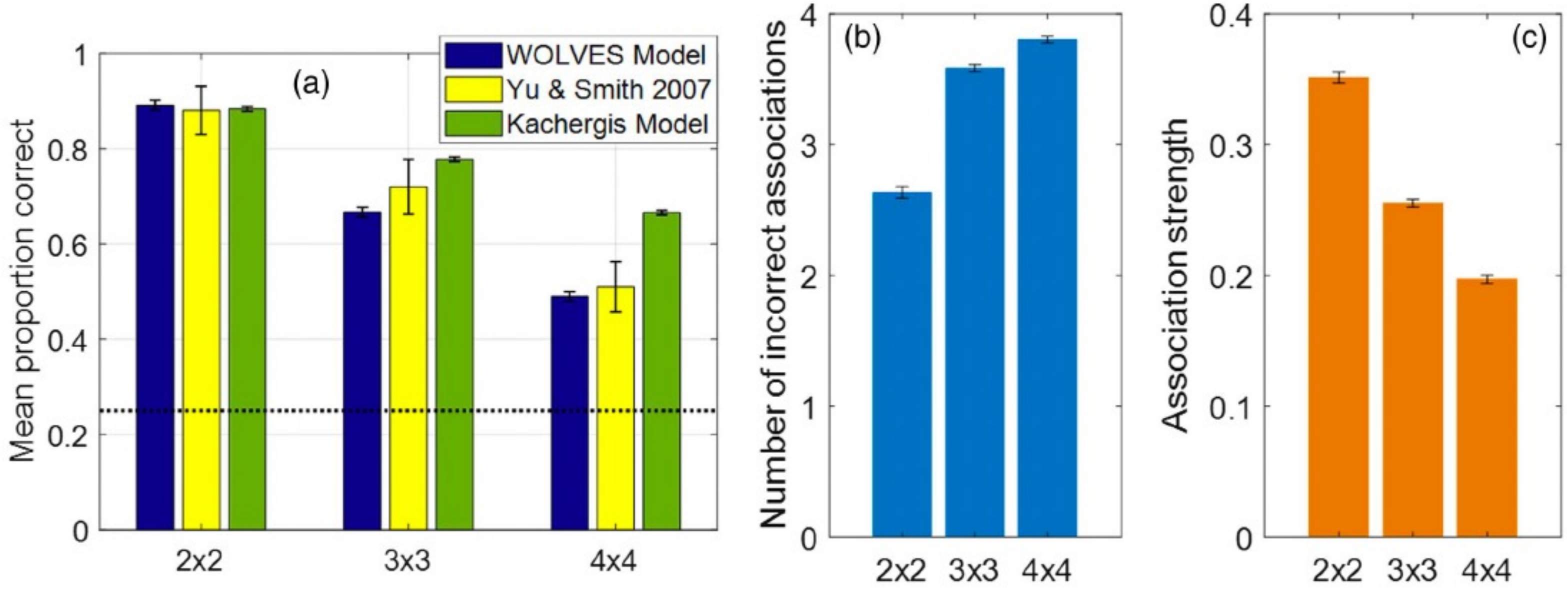


Bhat, Spencer, Samuelson (2021). *Psychological Review.*

WOLVES also explains associative learning data

- With more things to look at, WOLVES forms more incorrect associations with weaker association strengths

Figure 16
Data from Yu and Smith (2007) and the WOLVES and Kachergis Models



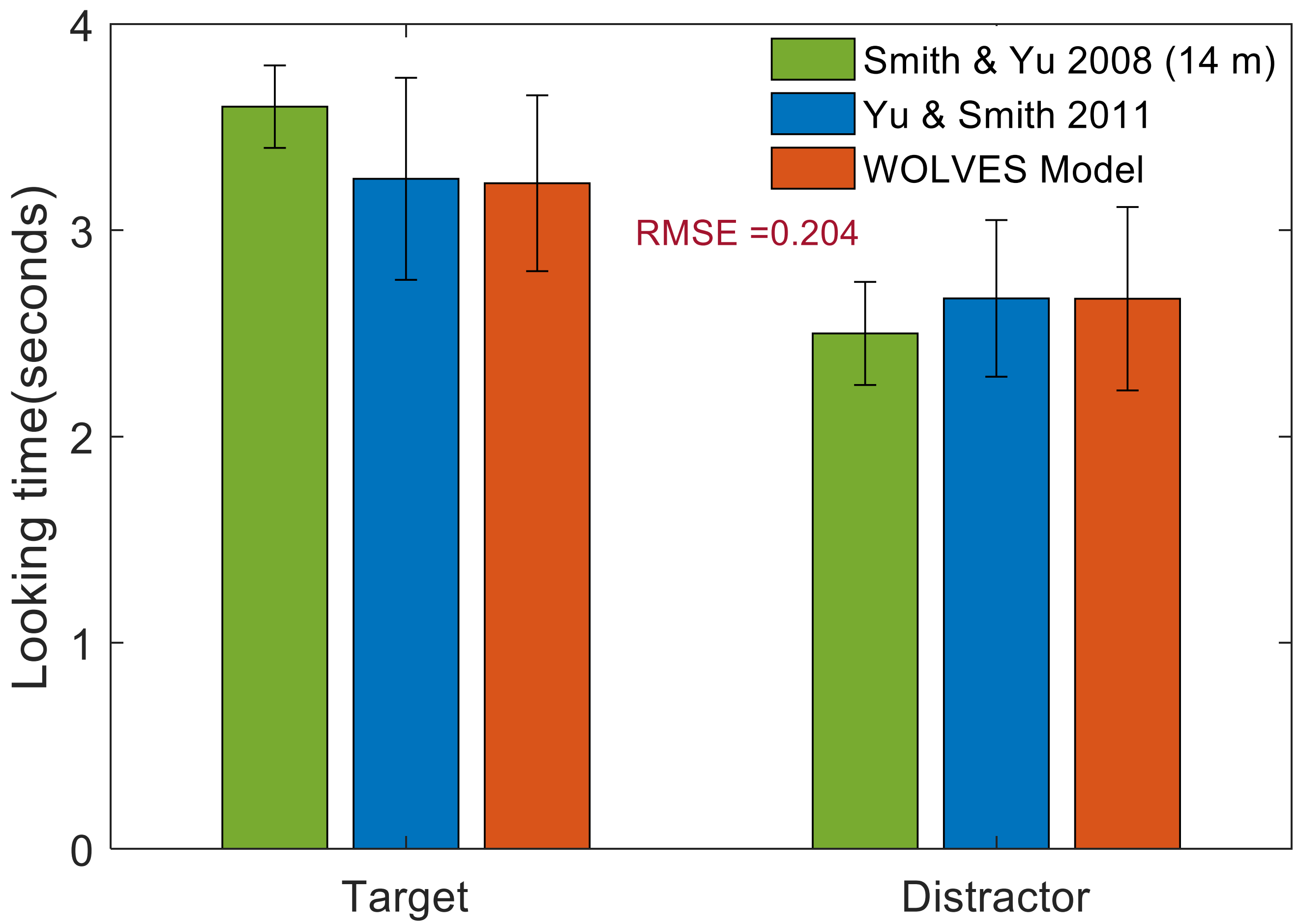
Bhat, Spencer, Samuelson (2021). *Psychological Review.*

Does WOLVES capture – and explain – empirical data?

Yes and successfully generates novel predictions.

What about CSWL in early development?

Smith & Yu (2008) , Yu & Smith (2011)



Bhat, Spencer, Samuelson (2021). *Psychological Review.*

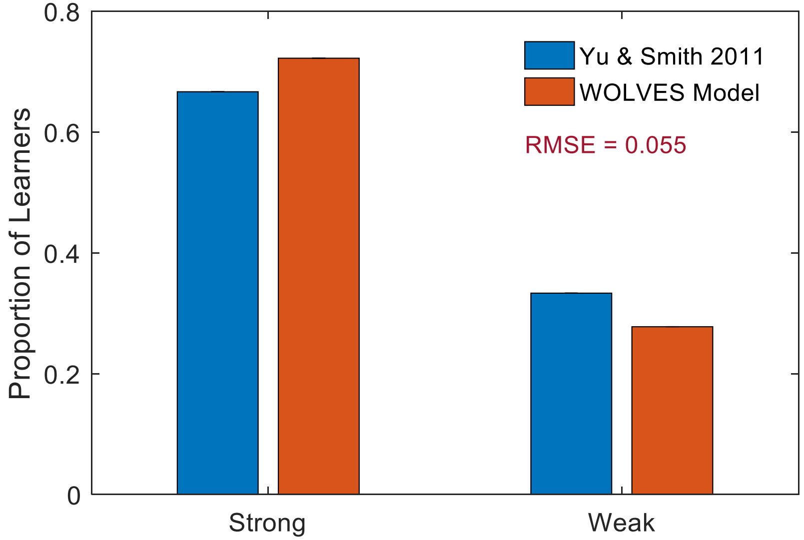
Smith & Yu (2008) , Yu & Smith (2011)

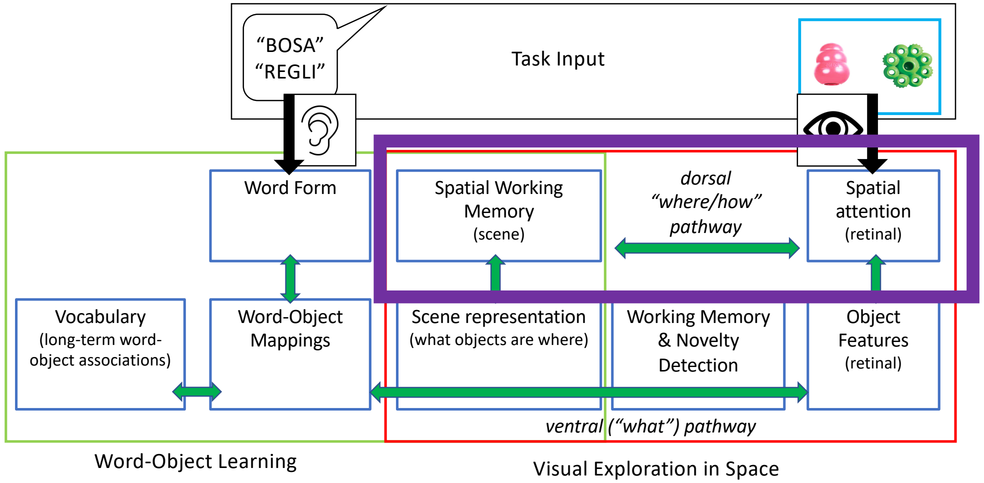
Measure	S & Y (2008)	Y & S (2011)		WOLVES	RMSE	MAPE
Test Trials						
Mean looking per 8s trial	6.10	5.92		6.26	.26	4.22
Pref. looking ratio	.60	.54		.54	.04	6.10
Mean words learned (of 6)	4.0	3.5		4.0	.35	7.14
Prop. Strong/weak learners	NA	.67		.74	.07	10.45
Mean looking to target per trial	3.6	3.25		3.36	.19	5.03
Mean looking to distractor per trial	2.5	2.67		2.89	.32	11.92
Training Trials		S	W			
Mean looking per 4s trial	3.04	2.96	3.07	3.01	.02	.71
Mean fixations per trial	NA	2.75	3.82	2.89	.22	6.98
Mean fixation duration	NA	1.69	1.21	1.31	.22	14.38

RMSE = Root Mean Squared Error, MAPE = Mean Absolute Percentage Error

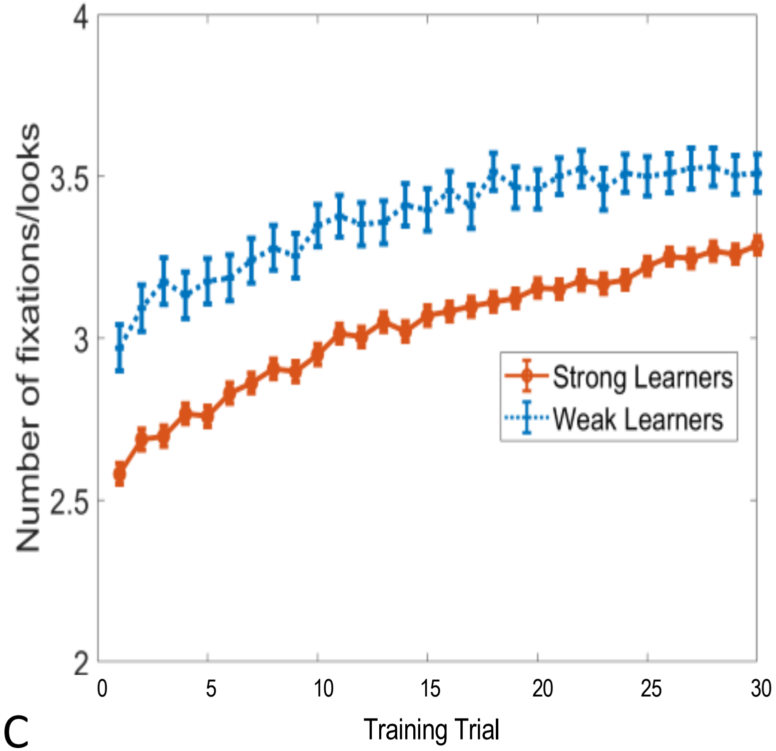
Bhat, Spencer, Samuelson (2021). *Psychological Review.*

The Role of Spatial Attention



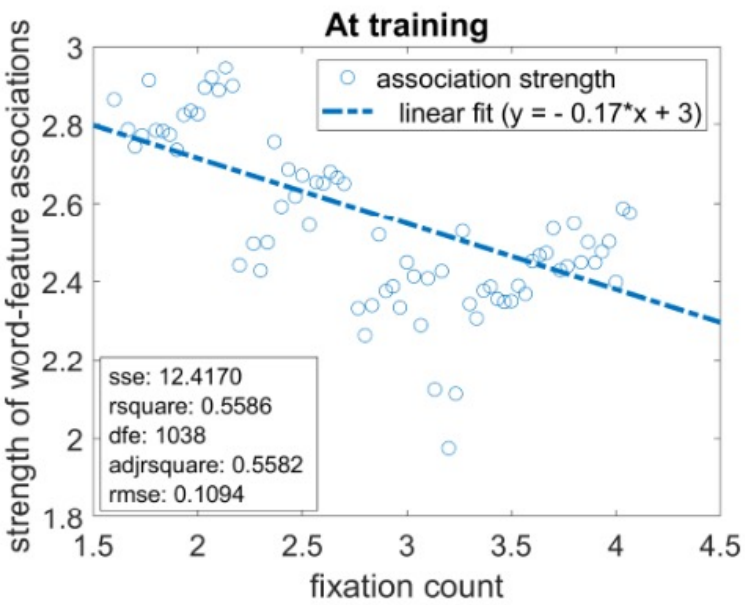


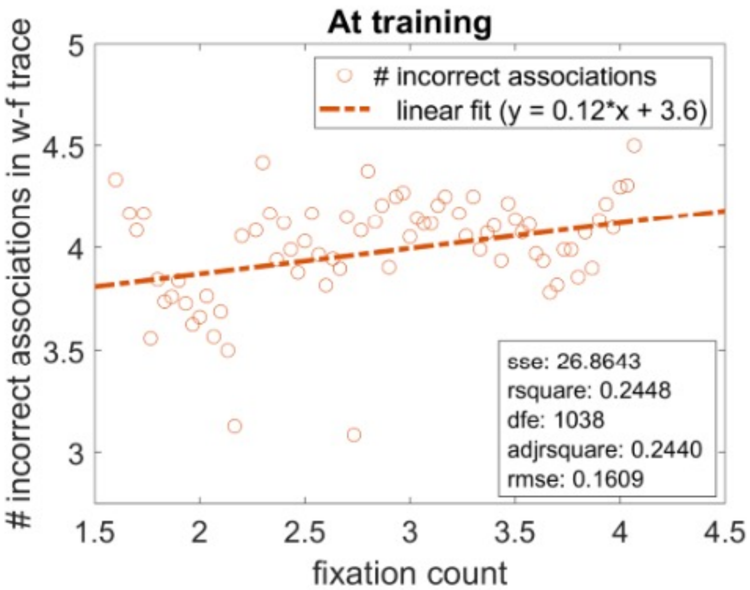
- Spatial attention parameter
  - atn_sa → atn_c
- Sort by strong/weak learners.
- Strong learners have fewer, longer fixations.



C

The Role of Spatial Attention





- As fixations go up, association strength goes down.
- As fixations go up, incorrect associations go up.

Bhat, Spencer, Samuelson (2021). *Psychological Review.*

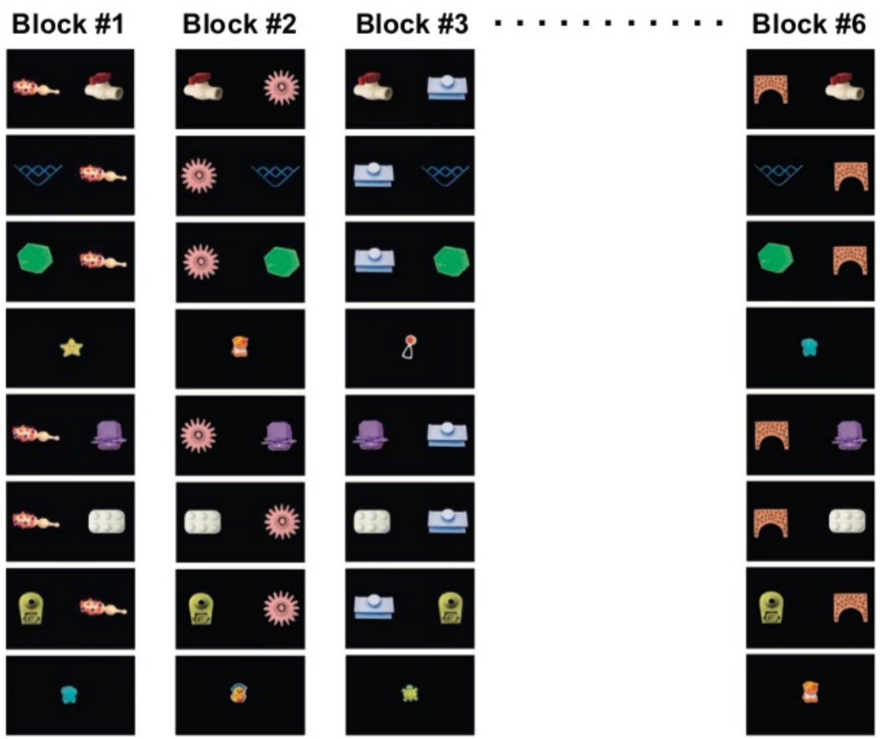
We created the difference between strong and weak learners via manipulation of a particular parameter.

This mechanistically relates variations in spatial attention to learning outcomes and highlights the contribution of real-time looking dynamics to CSWL.

Timescale of Development

- Vlach & Johnson (2013), Vlach & DeBrock (2017, 2019)

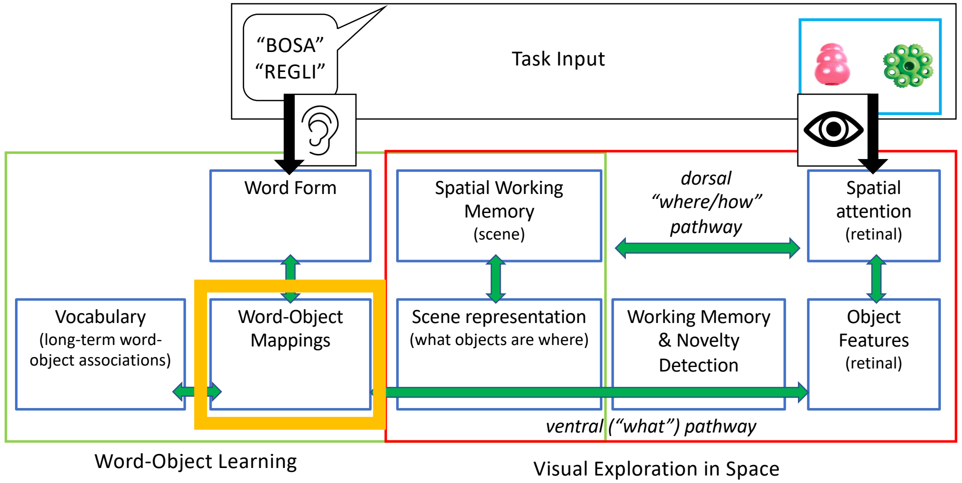
H.A. Vlach, S.P. Johnson / Cognition 127 (2013) 375–382



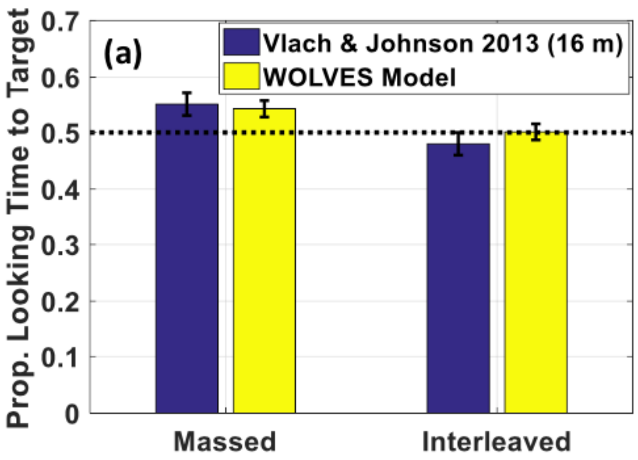
- 16 mo learn words from massed but not interleaved presentation.
- 20 mo learn equally with massed or interleaved.
- Older children learn better with interleaved presentation.

Timescale of Development

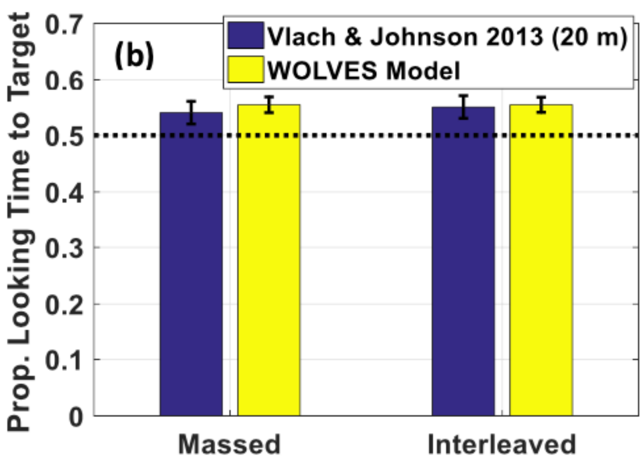
- Memory: Tau_Decay defines how fast a memory trace deteriorates.



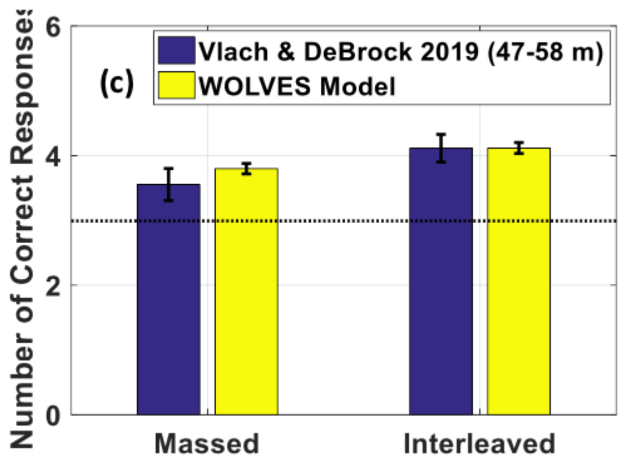
tau_decay = 1000



tau_decay = 1500

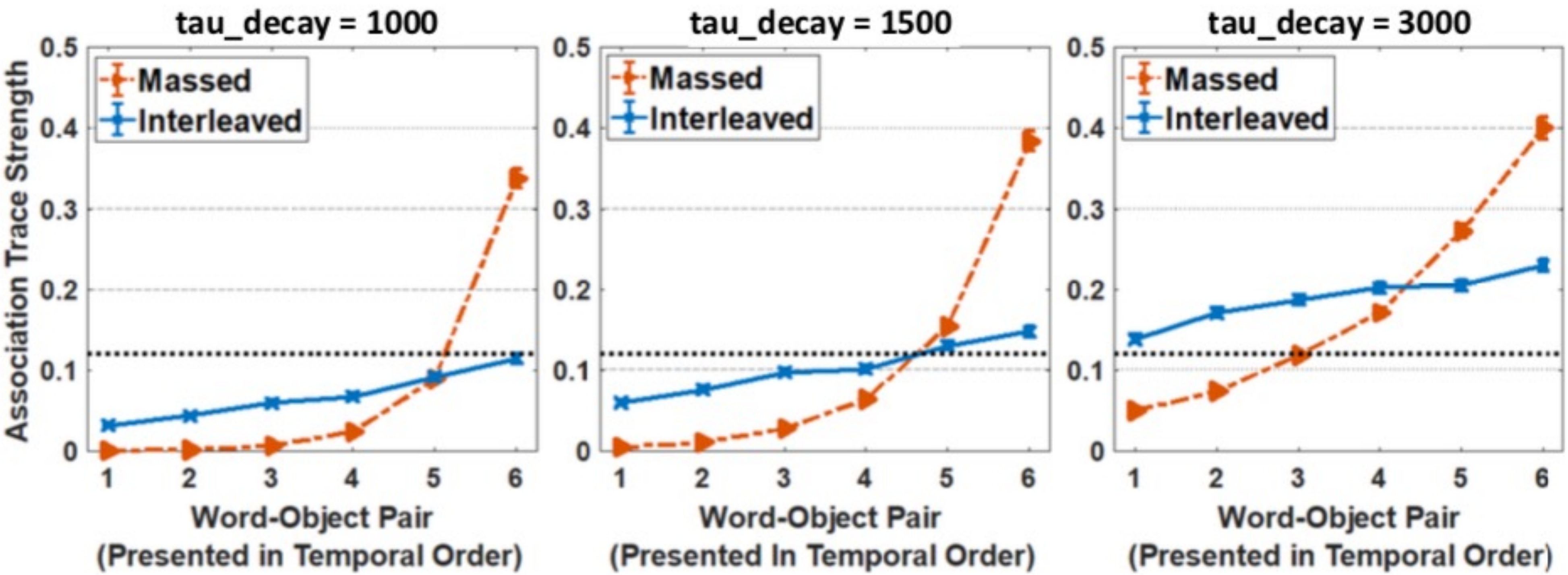


tau_decay = 3000



- Unified developmental account of CSWL

Timescale of Development



Bhat, Spencer, Samuelson (2021). *Psychological Review.*

We captured 60 datapoints from 12 months to 5 years with a change to just one parameter.

WOLVES is a powerful *developmental* model

This is because it has rich real-time and learning dynamics.

Is WOLVES a comprehensive theory?

Compared WOLVES to 2 competitor models:

- Kachergis et al. (2012)
- Stevens et al. (2017) – Pursuit: an HT model that uses an AL mechanism to weigh different hypotheses. Only adds a word to the lexicon if the conditional probability of hypothesis exceeds a threshold.

Model Validation; coverage & comparison

- 5 CSWL studies with adults
  - Trueswell et al. (2013), Yu & Smith (2007), Yu, Zhong & Fricker (2012), Yurovsky et al. (2012), Kachergis et al. (2012)
- 7 CSWL studies with infants, toddlers & children
  - Smith & Yu (2008), Yu & Smith (2011), Smith & Yu (2013), Vlach & Johnson (2013), Vlach & DeBrock (2019), Vlach & DeBrock (2017), Suanda etal. (2014)

Measure	**Data Points**	**WOLVES**		**Kachergis et al.[+]**		**Pursuit***	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Grand Mean Specific tasks	69	.05	13.51	.08	19.95	.20	42.13
Standard Deviations	69	.04	15.79	.07	21.99	.13	25.52
Grand Mean 3 Gen Exp	15	.03	4.05	.21	47.42	.13	23.91
Grand Mean	132	.10	15.80	unable to capture			
Overall AIC	69	-239.67		-295.78		-193.32	

[+]Kachergigs et al. (2012, 2013, 2017); *Stevens et al. (2017)

Bhat, Spencer, Samuelson (2021). *Psychological Review.*

Is WOLVES a comprehensive theory?

Yes.

Also raises interesting questions about metrics for model comparison. AIC lowest for Kachergis model, but WOLVES clearly outperforms this competitor model.

Suggests that the penalty for ‘free’ parameters too steep and/or that other metrics – like model generalisation – are more useful.

Conclusions

- WOLVES
  - Formal neural-process account of CSLW based on autonomous real-time visual exploration and non-linear associative learning.
  - Captures a large range of data and beats other models in direct comparison.
- Timescale of the task
  - Mechanistically related the strength of spatial attention to learning outcomes.
- Timescale of development
  - Presented the first developmental account of CSWL based on changes in memory strength.
- Future Directions
  - Currently exploring how we can use the model to make predictions, understand relations between tasks, and understand individual differences.

Thank you

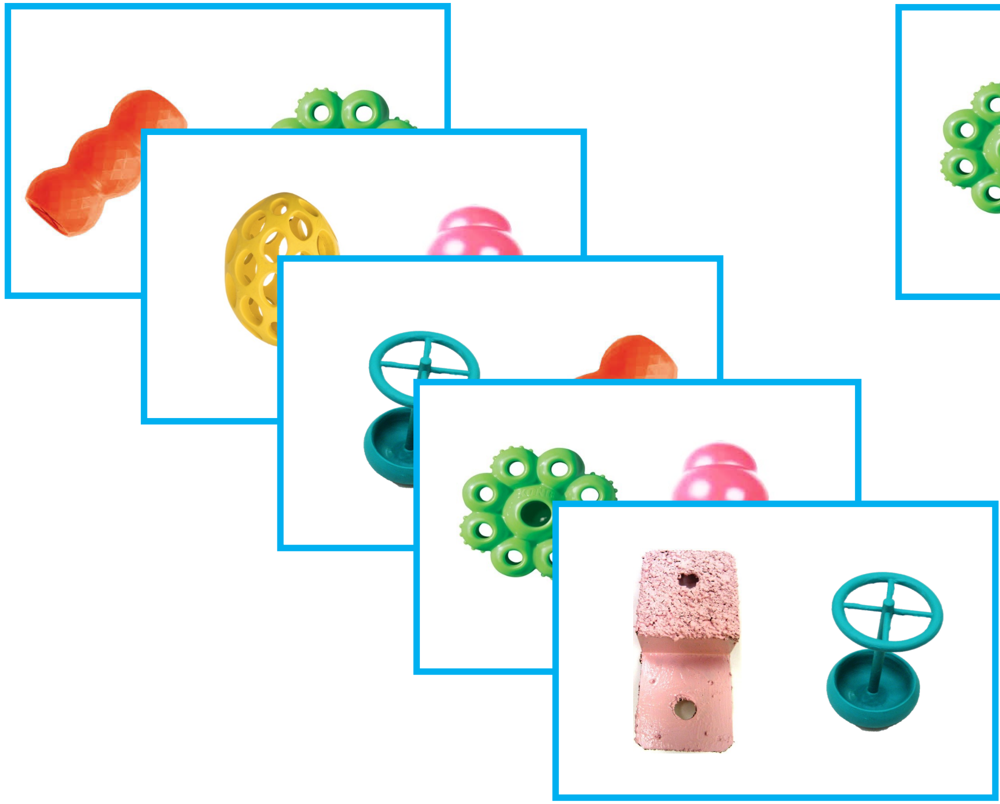
Members of Developmental Dynamics Lab, University of East Anglia

Funding: NICHD RO1HD045713 to L.K. Samuelson

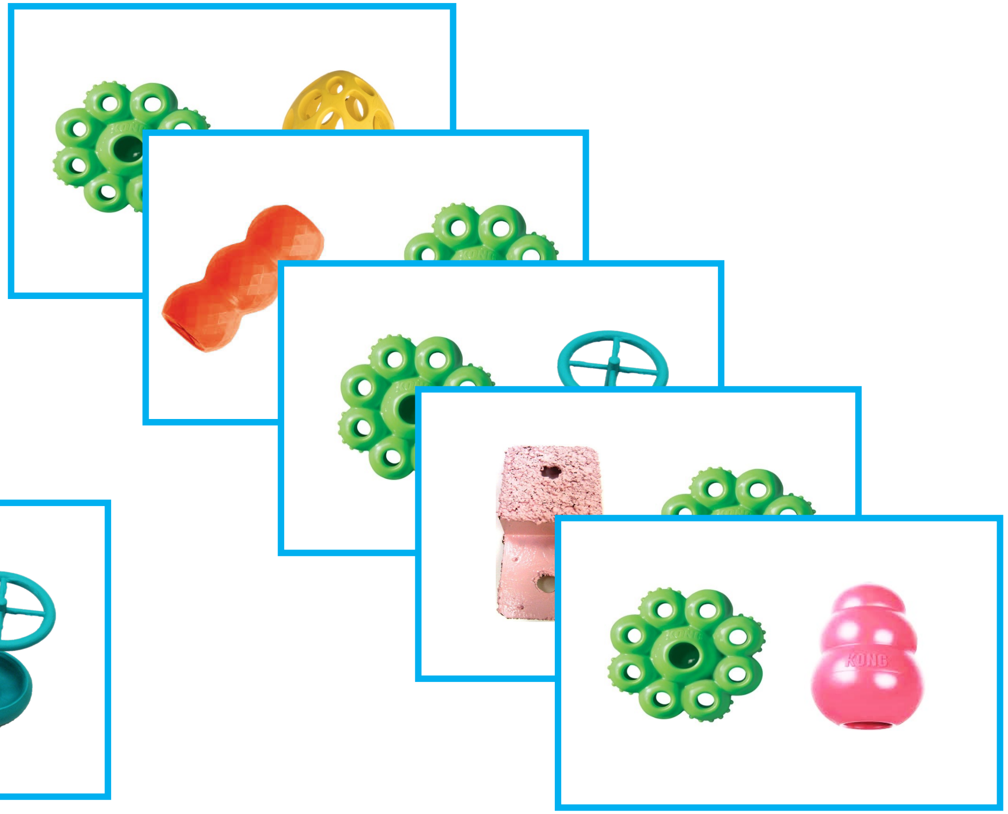
We’ve highlighted the real timescale.
What about the timescale of learning in the task and the timescale of development?

Timescale of the task

Smith & Yu (2008)



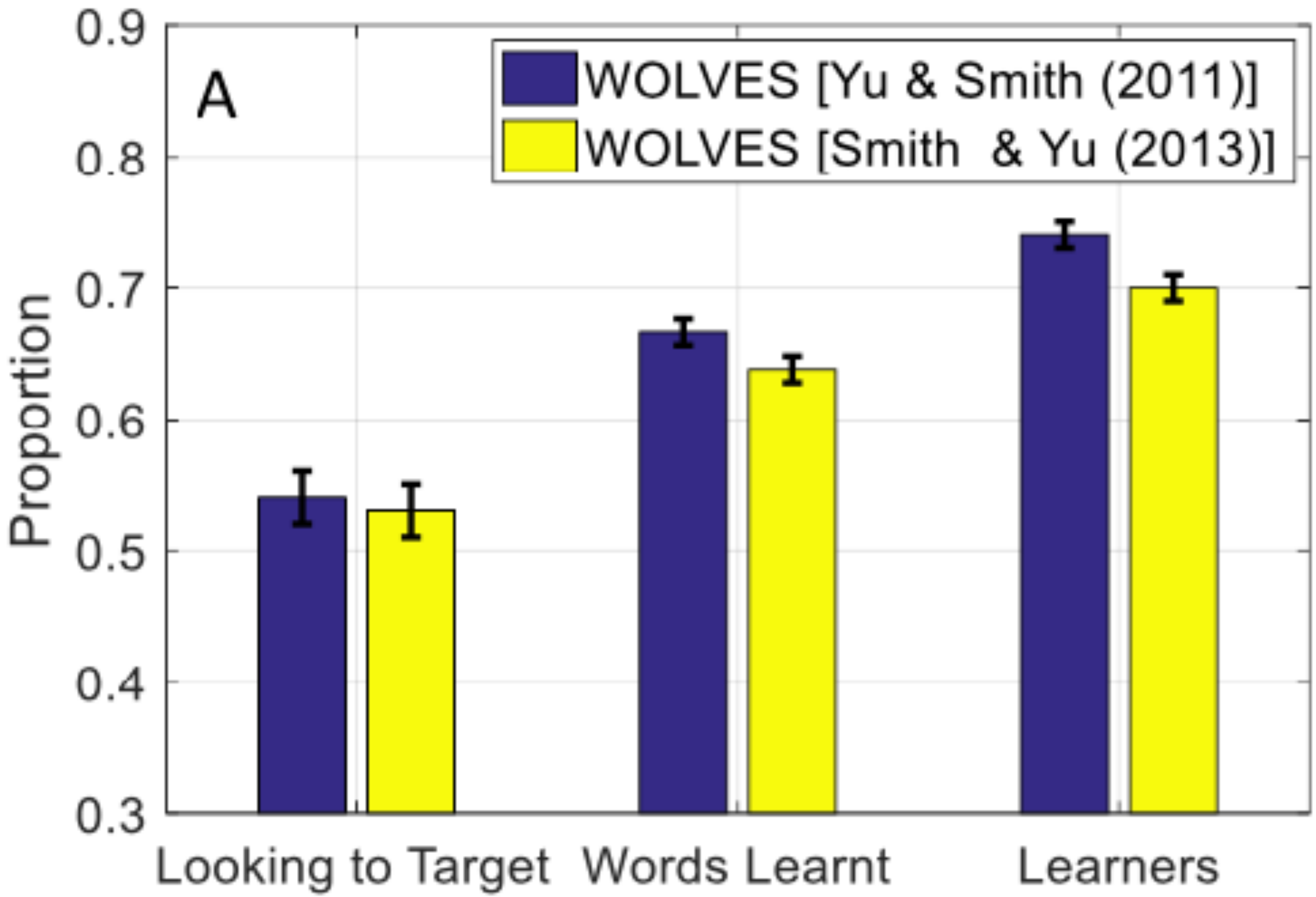
Smith & Yu (2013): Novelty Trap



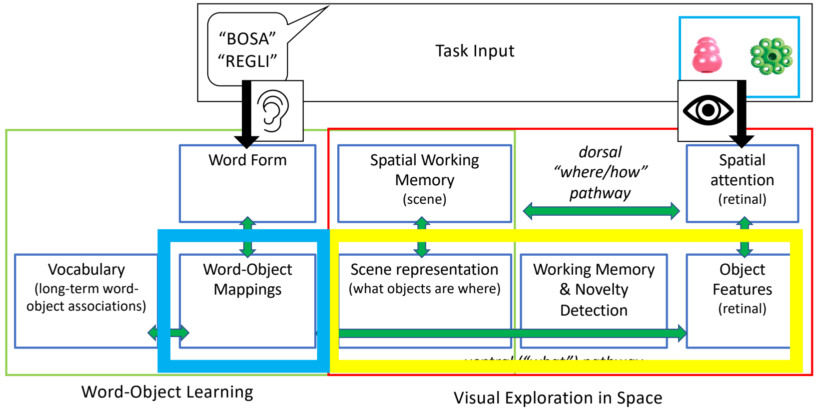
No overall difference in looks to target v. distractor at test

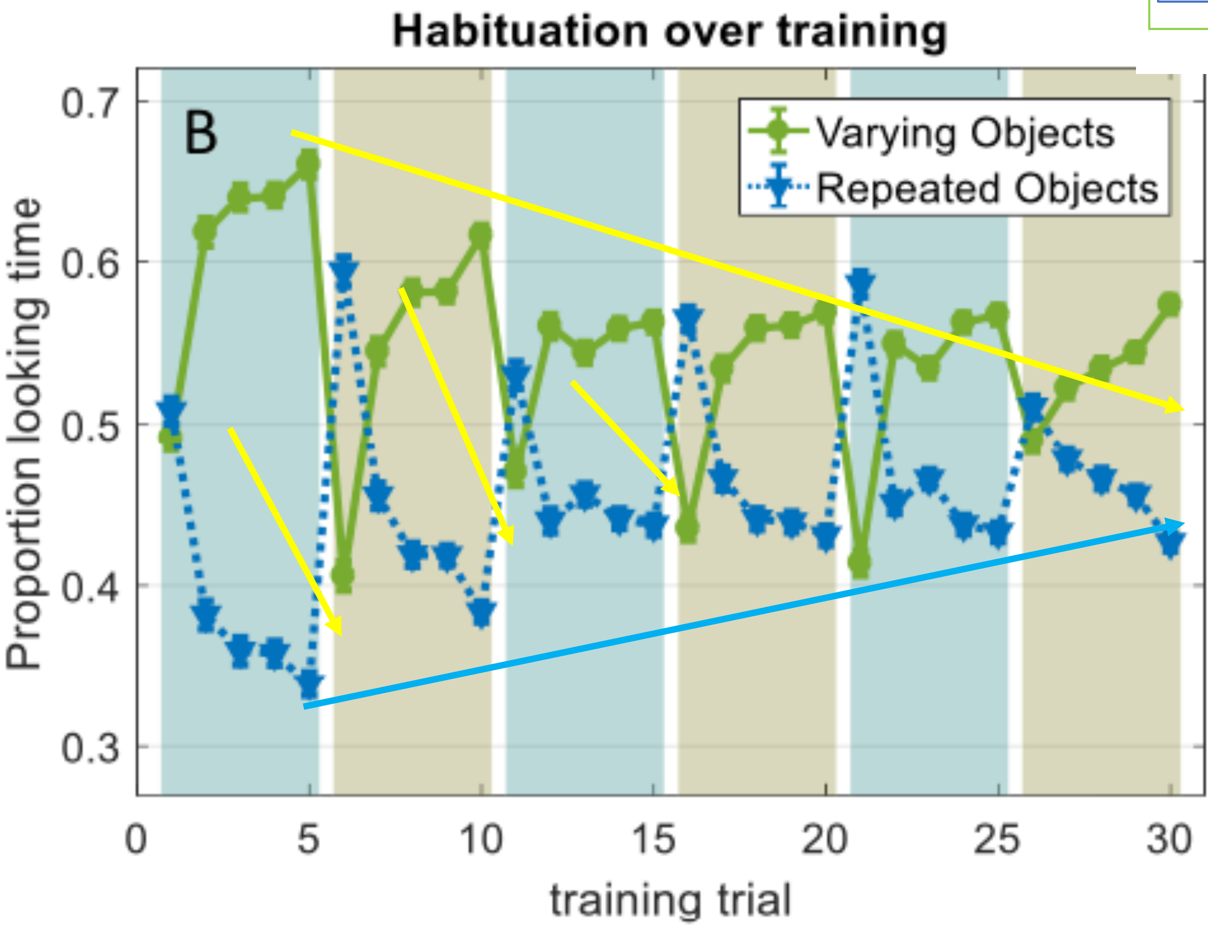
Fewer “learners”

Timescale of the task



Timescale of the task





- Novelty detection & consolidation in working memory
- Top-down attention driven by growing associations

Two types of learning on timescale of the task:

- learning / habituating to visual features
- learning word + object mappings