Constraining the search space in cross-situational learning:
Different models make different predictions

Giovanni Cassani, Robert Grimm, Steven Gillis, Walter Daelemans
name.surname@uantwerpen.be

Universiteit Antwerpen
COMPUTATIONAL LINGUISTICS & PSYCHOLINGUISTICS RESEARCH CENTER CLiPS

1. Cross-situational learning

How do children map sound sequences to objects?

The hypothesis is that children track word-referent co-occurrences over many different situations, which allows to filter noise. However, **missing co-occurrences are equally important:** knowing that a word doesn't occur with an object matters as much as knowing that it does.

We took 4 prominent computational models of cross-situational learning and checked whether their predictions **fit with the behavioral data from children and adults**.

2. Computational Models

Hebbian learner: two layer neural network that learns word-referent associations using Hebb's rule

Naïve Discriminative Learner (NDL): two layer neural network that learns word-referent associations using Rescorla-Wagner equations

Probabilistic Learner: computes a probability distribution over words for each referent

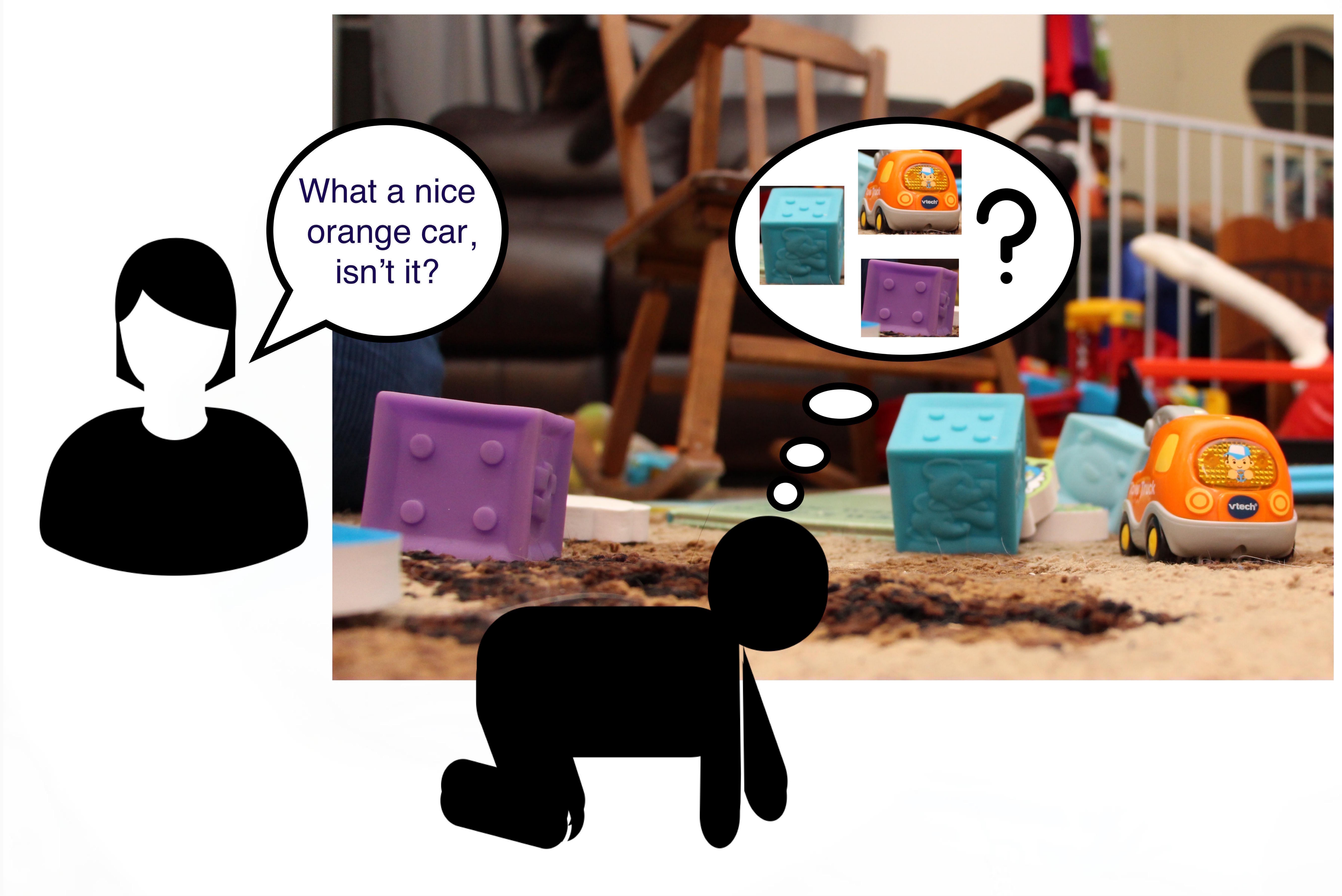
Hypothesis Testing Model (HTM): on the first trial, it picks a single word-referent hypothesis. On subsequent trials, it retrieves a formed hypothesis (with probability p, increasing on every successful retrieval) and checks whether the current trial supports it. If it does not, a new hypothesis is formed at random.

4. Conclusions & future work

The **Hebbian Learner** and the **HTM** failed to fit the behavioral data. The **NDL model** and the **Probabilistic Learner** learned correct associations.

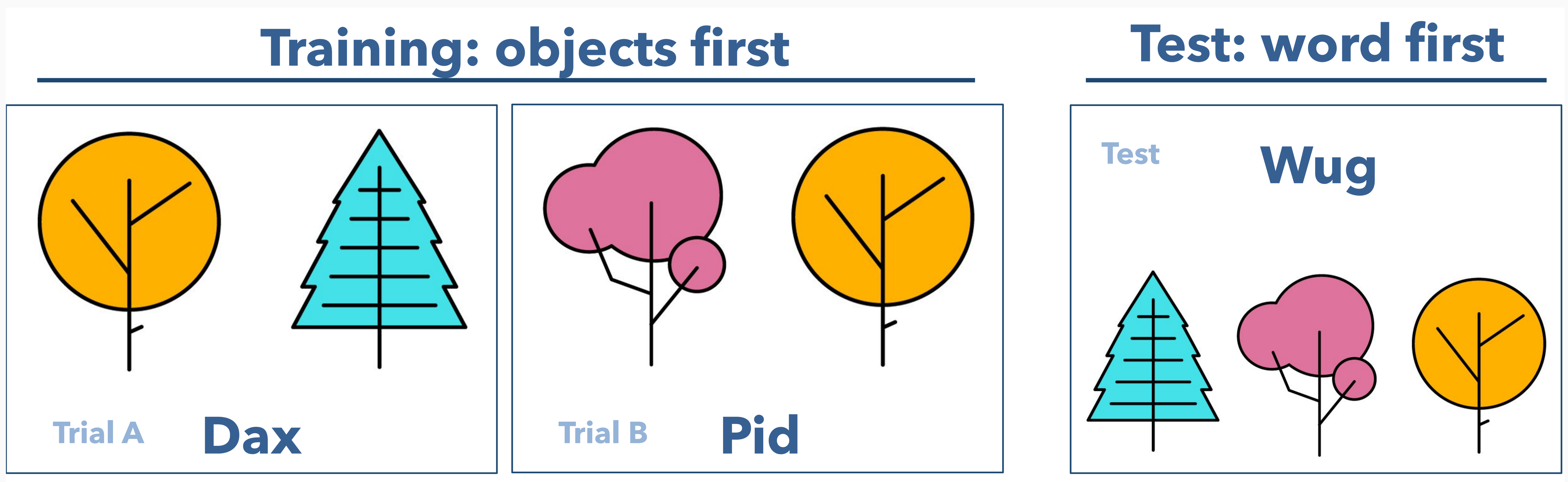
Crucially, the successful models learn from **missing co-occurrences** as well, unlike the unsuccessful ones.

Given that, in the experimental design, spurious word-referent co-occurrences occur as often as correct ones in the data, an unambiguous mapping can only be formed by noting that **certain referents fail to co-occur with certain words**.

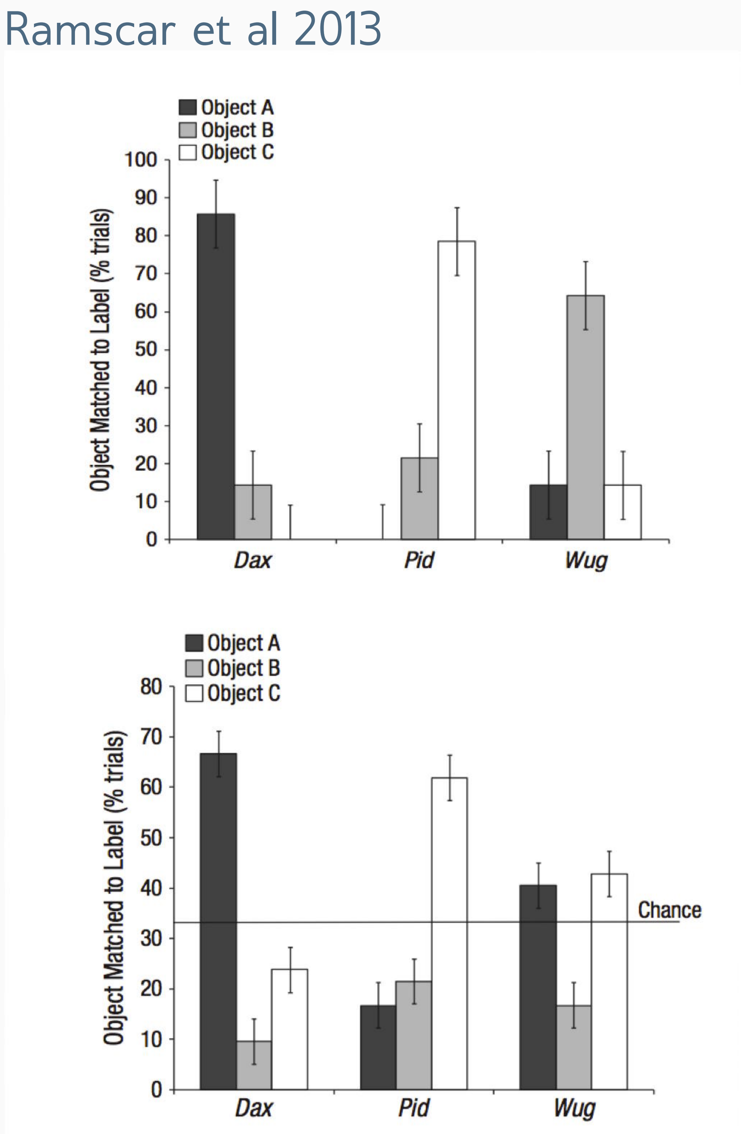


3. Experimental design & results

Experimental design from Ramscar and colleagues (2013): Over 18 trials, object B is always present, 9 times with object A and 9 times with object C. When object A is present, the word *Dax* is always present, while when object C is present, the word *Pid* is also present. A third word, *Wug*, is only presented at test. At test, one of the three words is presented and subjects are asked to retrieve the associated object.
CAVEAT: the images do not depict the objects used during the experiment, we simply use them for descriptive purposes.



Behavioral responses from **undergrads** (top) and **children** (bottom) - plots from Ramscar et al 2013



Model	Cue	DAX	PID
Hebbian Learner	ObjA	**9**	.
	ObjB	**9**	**9**
	ObjC	.	**9**
NDL	ObjA	.134 $\pm$.001	-.021 $\pm$.005
	ObjB	.113 $\pm$.005	.113 $\pm$.005
	ObjC	-.021 $\pm$.005	.134 $\pm$.001
Probabilistic Learner	ObjA	.967 $\pm$.003	.
	ObjB	.483 $\pm$.082	.486 $\pm$.082
	ObjC	.	.967 $\pm$.003
HTM	ObjA	**.455**	.
	ObjB	**.545**	**.485**
	ObjC	.	**.515**

Table 1: Simulated responses The table shows word-referent mappings induced by the 4 simulated learners. In red, we highlighted successful matches, i.e. the highest value column-wise; in **bold** we highlighted learners that would have to guess at random when retrieving the object matching a given label.

Acknowledgements

This research was supported by a **BOF/TOP grant (ID 29072) of the Research Council of the University of Antwerp.**. We thank Michael Ramscar and Konstantin Sering for discussion about the NDL model; Aida Nematzadeh for her help in better understanding the Probabilistic Learner; Chen Yu for sharing his latest research with us. Mommy icon by Gabriel Ciccariello, child icon by Anna Wang, both from The Noun Project. The poster was designed on Overleaf using the baposter template by Brian Amberg.