Computational approaches to the explanation of universal properties of meaning

Fausto Carcassi & Jakub Szymanik

https://thelogicalgrammar.github.io/ESSLLI22_langevo

Recap

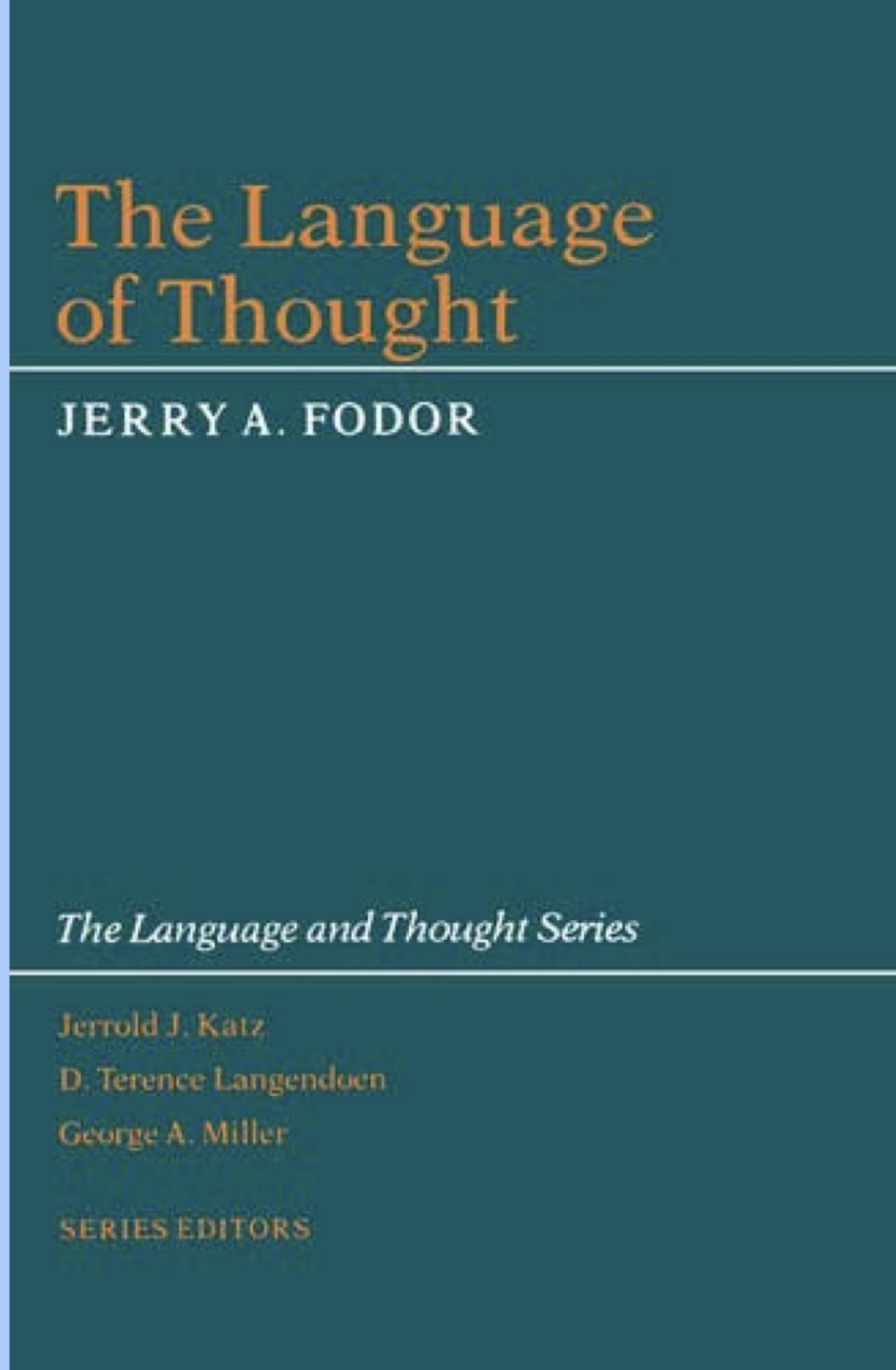
1. Learnability can explain the presence of universals.
2. But is it the only (or the best) such explanation?
3. A natural idea: some notion of complexity explains both the universals and the learnability facts.

What's the right measure of complexity?

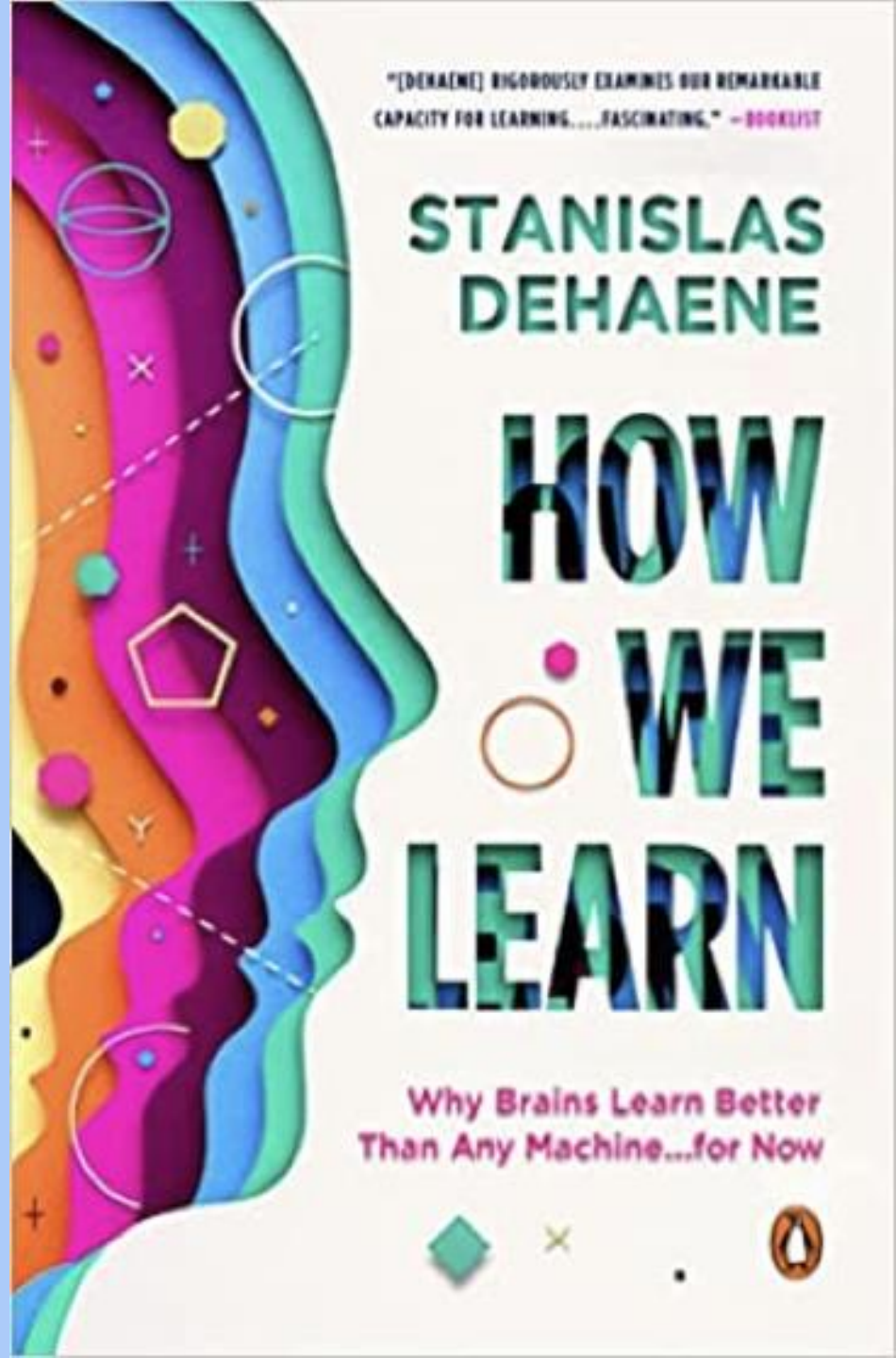
- Previous attempts fail to capture the distinctions:
  - Automata theory (van Benthem 1986)
  - Computational complexity (Szymanik 2016)
  - Formal learning theory (Tiede 1999; Gierasimczuk 2007; Gierasimczuk 2009a)

Let's try a philosophical idea!

- The Language of Thought Hypothesis is the hypothesis that thinking happens in a mental language, the Language of Thought.
- Generally assumed to look somewhat like a logic: predicates get combined with logical operators.
- Modern version popularised by Jerry Fodor.
- The classical picture of LoTs that philosophers developed is meant to be an account of thinking, explaining phenomena such as learning from few examples, decision-making, perception, etc.



- Leibniz, 1677, Boole, 1854, Fodor, 1975,…, Rescorla, 2019, and others
- Recently, revived interest in cognitive science:
- Feldman, 2003, Tenenbaum and Griffiths, 2001, Tenenbaum and Xu, 2007, Piantadosi, 2016, etc.
- Perhaps a missing peace to make AI more human-like



Boolean Categorization

- Boolean relations are a way to create new concepts:
  - `cousin' is a child of an uncle or aunt
  - 'beer' is an alcoholic beverage usually made from malted cereal grain and flavored with hops, and brewed by slow fermentation
  - in basketball, `travel' is illegally moving the pivot foot or taking three or more steps without dribbling
  - `depression' is a mood disorder characterized by persistent sadness and anxiety, or feeling of hopelessness and pessimism, or …

- How people acquire, represent, and use concepts?
- E.g., concepts depending on and are easier to learn than those depending on or (Bruner et al. `65).
- But the data seems more puzzling (see next slide).
- What's the logical theory of complexity here?

The *wudsy* objects in each set are surrounded by a square:

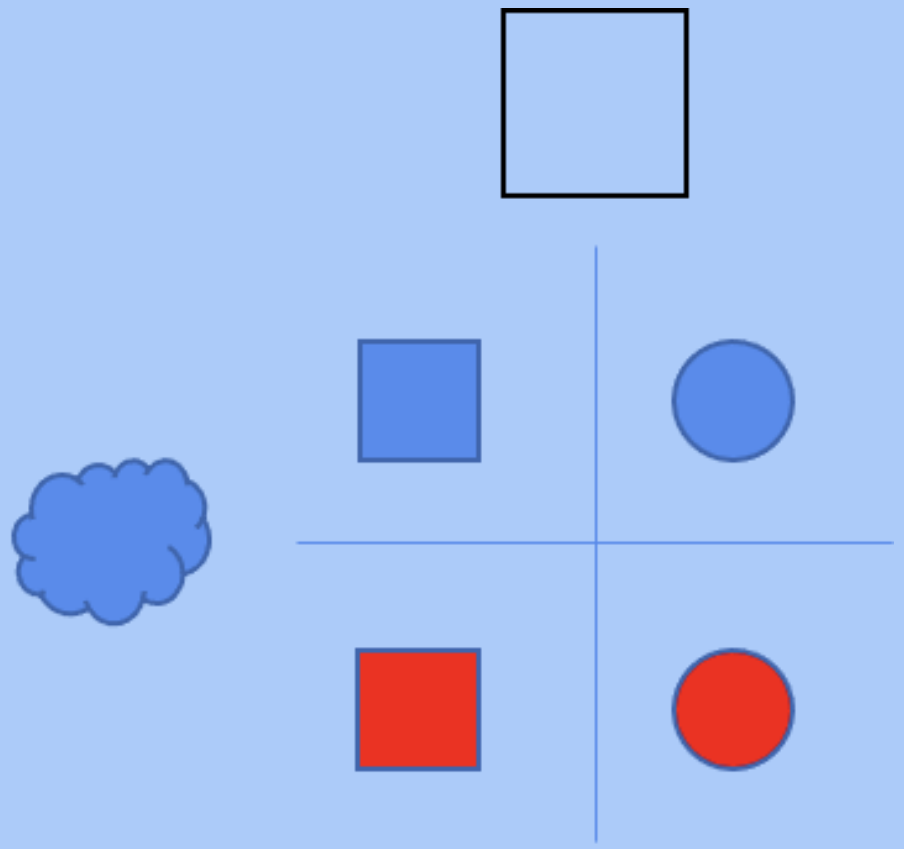
Example 1

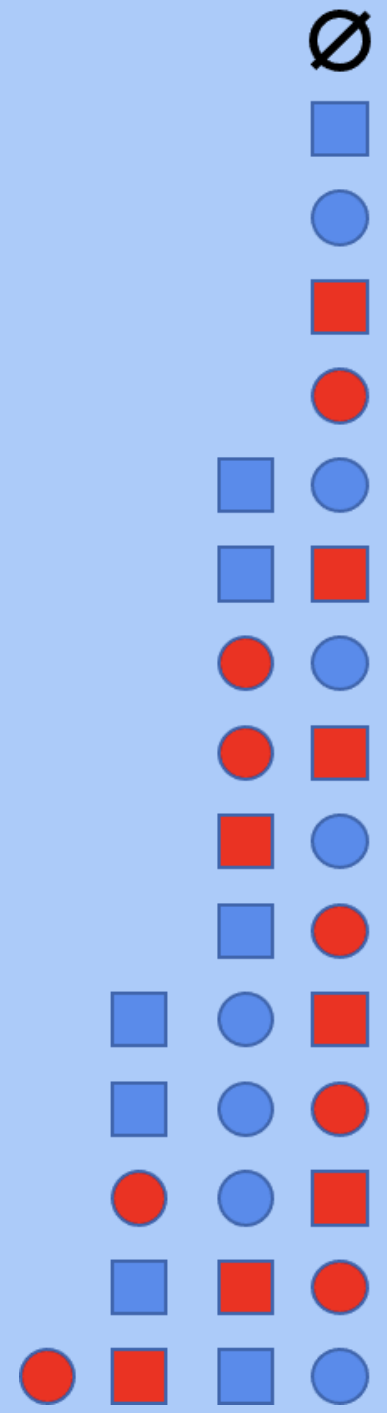
Example 2

Given the above examples, which of the objects in this set are wudsy?

Please respond to each object

Piantadosi et al. 2016

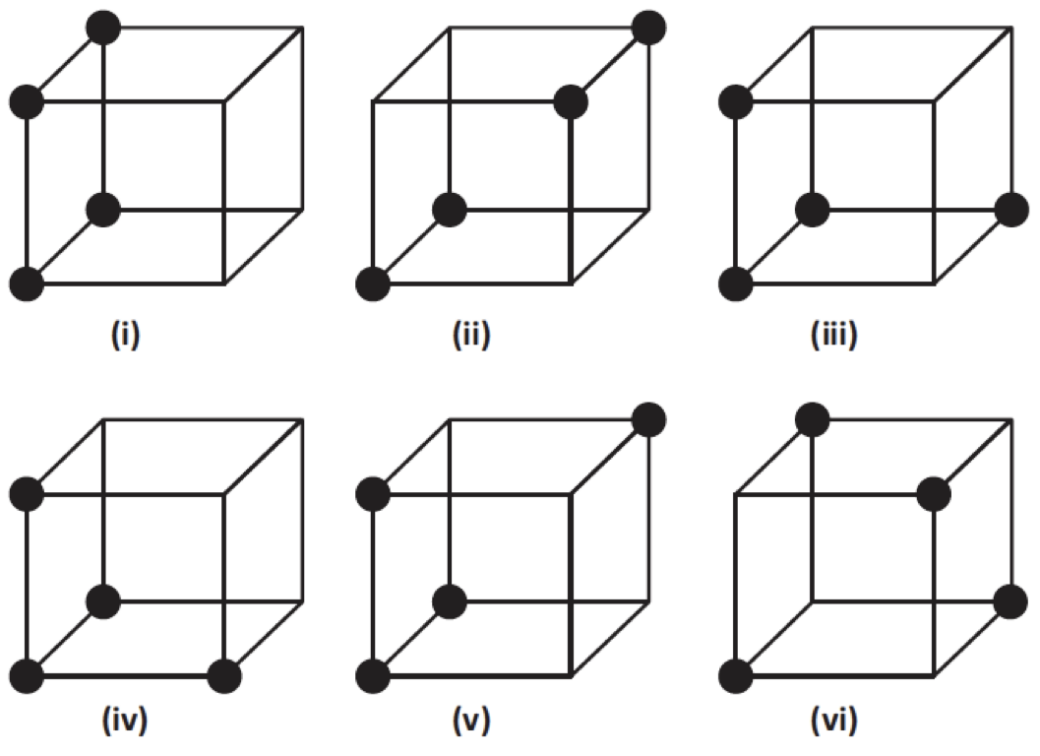




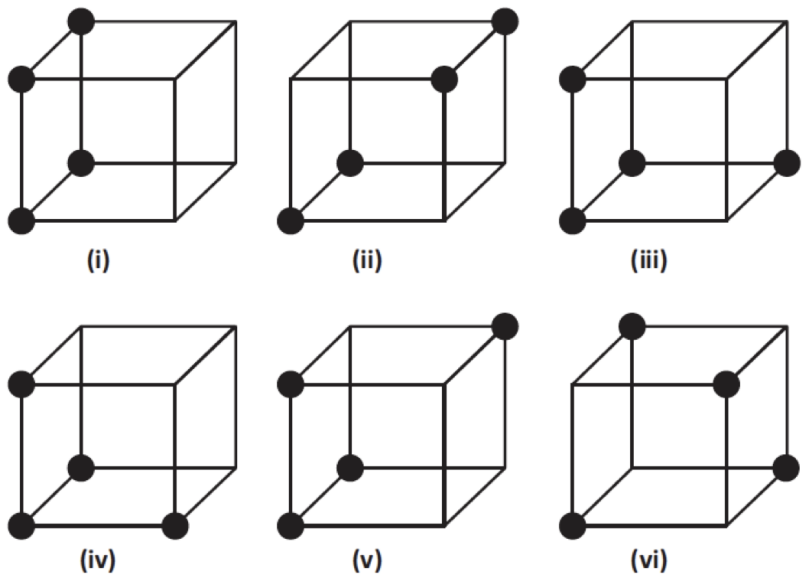
Shepard's trend

- Six different sorts of concept based on three binary variables
- Each concept: 4 instances and 4 non-instances in 8 possibilities
- Different presentations methods: sequentially, simultaneously, etc.
- Dependent variables: errors, latencies, accuracy of descriptions, etc.
- I < II < III, IV, V < VI

Shepard et al.'61



The instances of the concepts

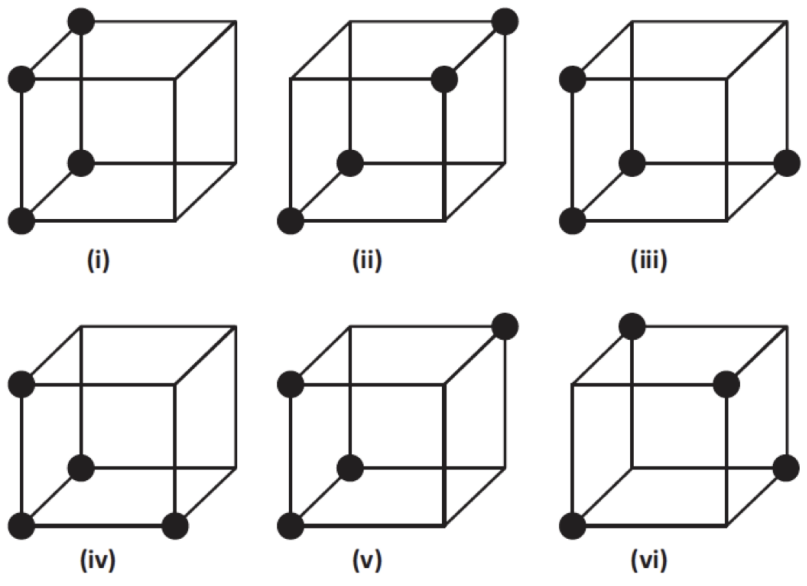


Concept number	Instances
I	not-a b c not-a b not-c not-a not-b c not-a not-b not-c
II	a b c a b not-c not-a not-b c not-a not-b not-c
III	a not-b c not-a b not-c not-a not-b c not-a not-b not-c
IV	a not-b not-c not-a b not-c not-a not-b c not-a not-b not-c
V	a b c not-a b not-c not-a not-b c not-a not-b not-c
VI	a b not-c a not-b c not-a b c not-a not-b not-c

Boolean Complexity

- The length of the shortest Boolean formula logically equivalent to the concept, e.g., expressed in terms of the number of literals (positive or negative variables).
- Btw, finding the shortest formula is intractable.
- *(a and b) or (a and not b) or (not a and b)* reduces to *(a or b)*

BC captures the trend

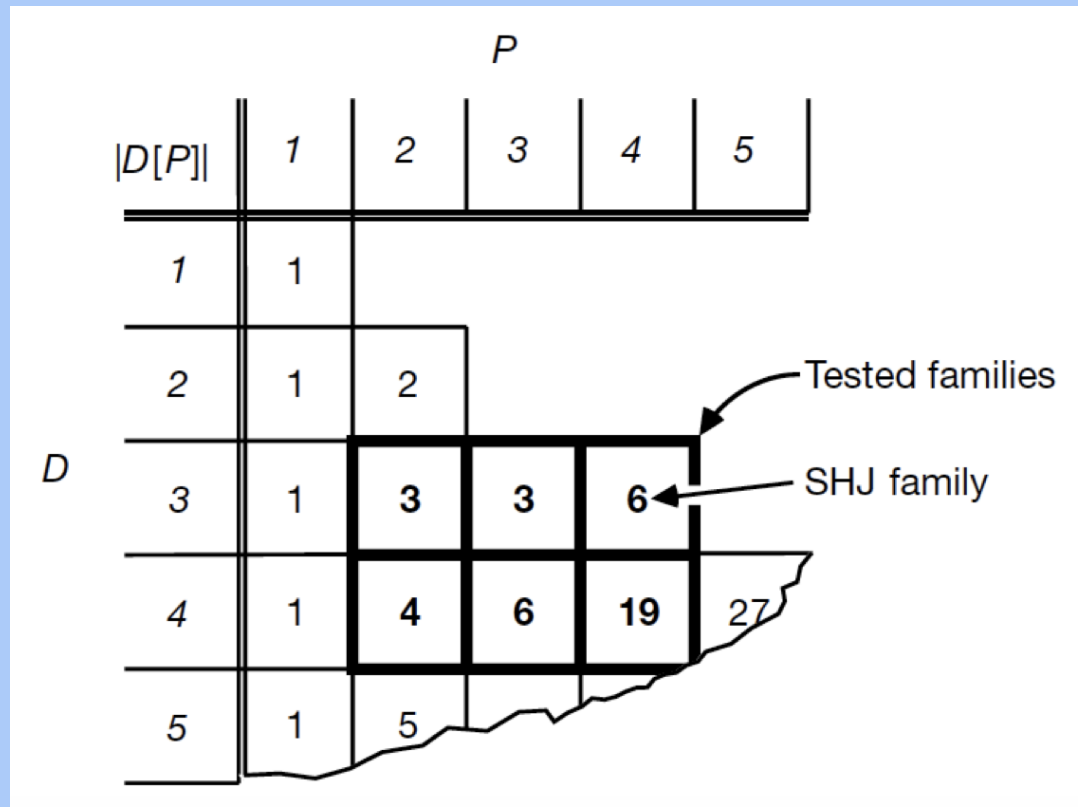


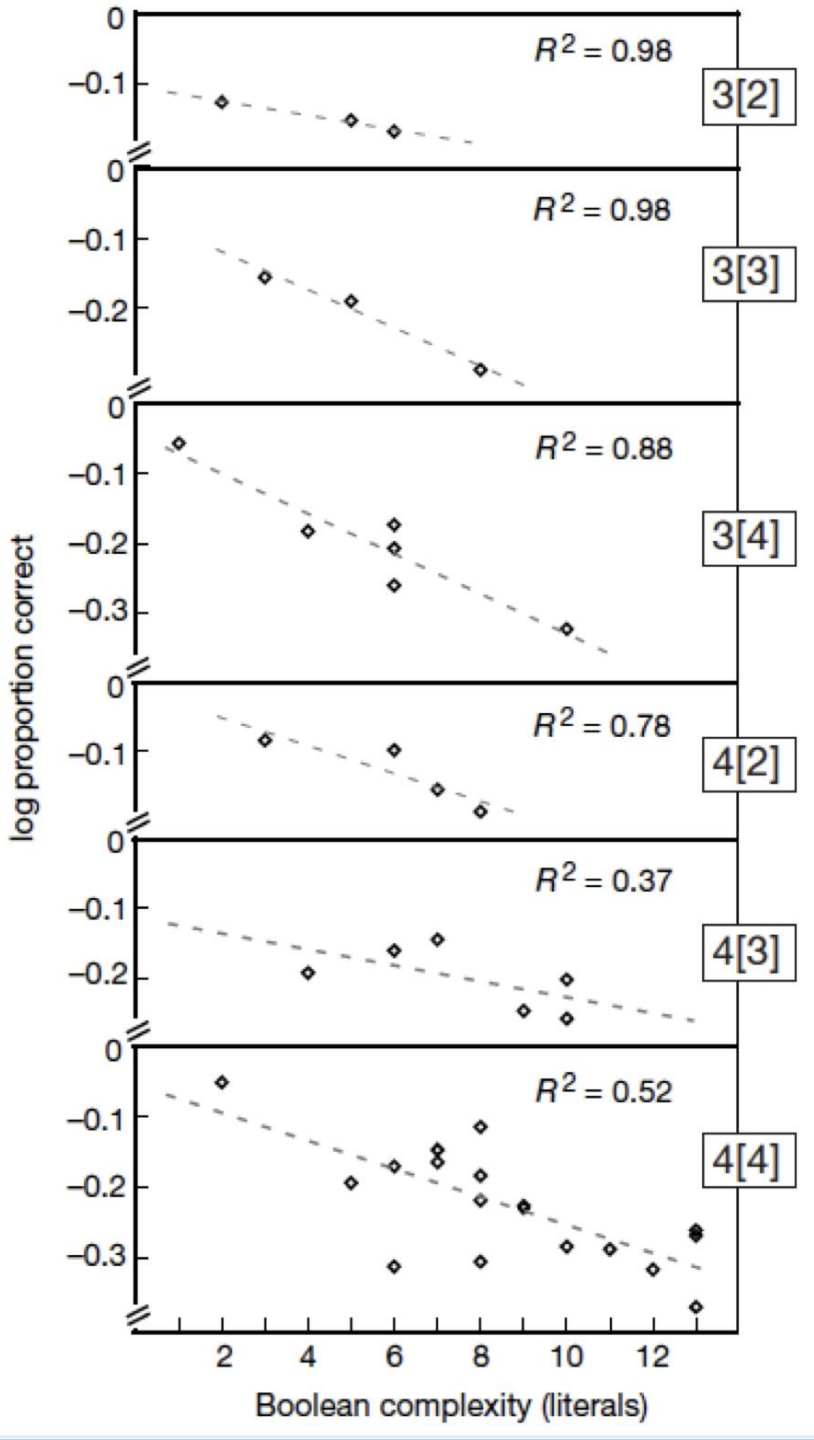
Concept number	Instances	Minimal description
I	not-a b c not-a b not-c not-a not-b c not-a not-b not-c	*not a* (1)
II	a b c a b not-c not-a not-b c not-a not-b not-c	*(a and b) or (not a and not b)* (4)
III	a not-b c not-a b not-c not-a not-b c not-a not-b not-c	*(not a and not c) or (not b and c)* (4)
IV	a not-b not-c not-a b not-c not-a not-b c not-a not-b not-c	*(not c or (not a and not b)) and (not a or not b)* (5)
V	a b c not-a b not-c not-a not-b c not-a not-b not-c	*(not a and not (b and c)) or (a and (b and c))* (6)
VI	a b not-c a not-b c not-a b c not-a not-b not-c	(a and ((not b and c) or (b and not c))) or (not a and ((not b and not c) or (b and c))) (10)

New Data Set

- Consider an arbitrary Boolean concept defined by P positive examples over D binary features.
- For Shepard types D=3 and P=4.
- Feldman studies 76 Boolean concepts.

Feldman `01




log proportion correct
0
−0.1
$R^2 = 0.98$
3[2]
0
−0.1
−0.2
$R^2 = 0.98$
3[3]
0
−0.1
−0.2
−0.3
$R^2 = 0.88$
3[4]
0
−0.1
$R^2 = 0.78$
4[2]
0
−0.1
−0.2
$R^2 = 0.37$
4[3]
0
−0.1
−0.2
−0.3
$R^2 = 0.52$
4[4]
2
4
6
8
10
12
Boolean complexity (literals)

- **Simplicity: a unifying principle in cognitive science?**
- (Chater and Vitanyi, 2013), (Hsu, Chater, Vitanyi, 2013),…
- MDL one of the ways to operationalize simplicity.
- Other options: computational complexity (van Rooij, 2008, Szymanik, 2016), Kolmogorov complexity (see below), machine learning (cf. Carcassi & Szymanik, 2022)

Applying pure complexity idea to quantifier universals

We need a more expressive logical grammar

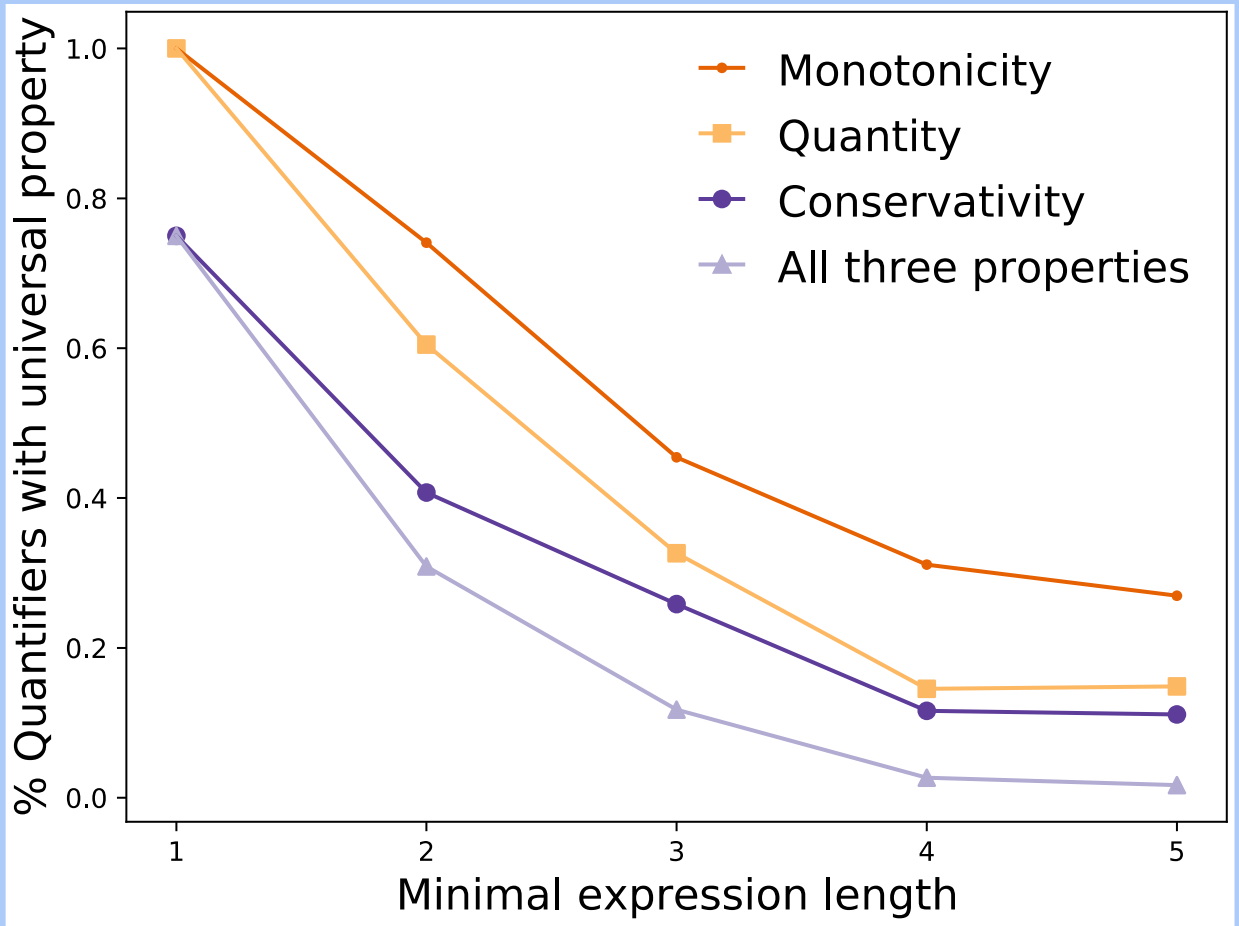
operator	type	gloss
$\cup$	$\text{SET} \times \text{SET} \to \text{SET}$	union
$\cap$	$\text{SET} \times \text{SET} \to \text{SET}$	intersection
$\setminus$	$\text{SET} \times \text{SET} \to \text{SET}$	setminus
$\iota(\cdot,\cdot)$	$\text{INT} \times \text{SET} \to \text{SINGLETON SET}$	'object at index'
$\lvert\cdot\rvert$	$\text{SET} \to \text{INT}$	cardinality
$\subseteq$	$\text{SET} \times \text{SET} \to \text{BOOL}$	subset equal
$=$	$\text{INT} \times \text{INT} \to \text{BOOL}$	integer equality
$>$	$\text{INT} \times \text{INT} \to \text{BOOL}$	integer larger than
$\neg$	$\text{BOOL} \to \text{BOOL}$	negation
$\wedge$	$\text{BOOL} \times \text{BOOL} \to \text{BOOL}$	and
$\vee$	$\text{BOOL} \times \text{BOOL} \to \text{BOOL}$	or

- We generate all expression of length up to 5 (or 7): **solving the minimal pair problem**
- Length = the number of operators
- The minimal expression length of Q is the length of the shortest expression for this quantifier.
- At most 1 = (2>|A⋂B|) or ¬(|A⋂B|)>1
- We generated ~25k unique quantifiers (up to model size 8)

	YES	NO	%
monotonicity	-0.12	0.05	0.28
quantity	-0.15	0.03	0.15
conservativity	-0.16	0.02	0.12
all	-0.79	0.02	0.02

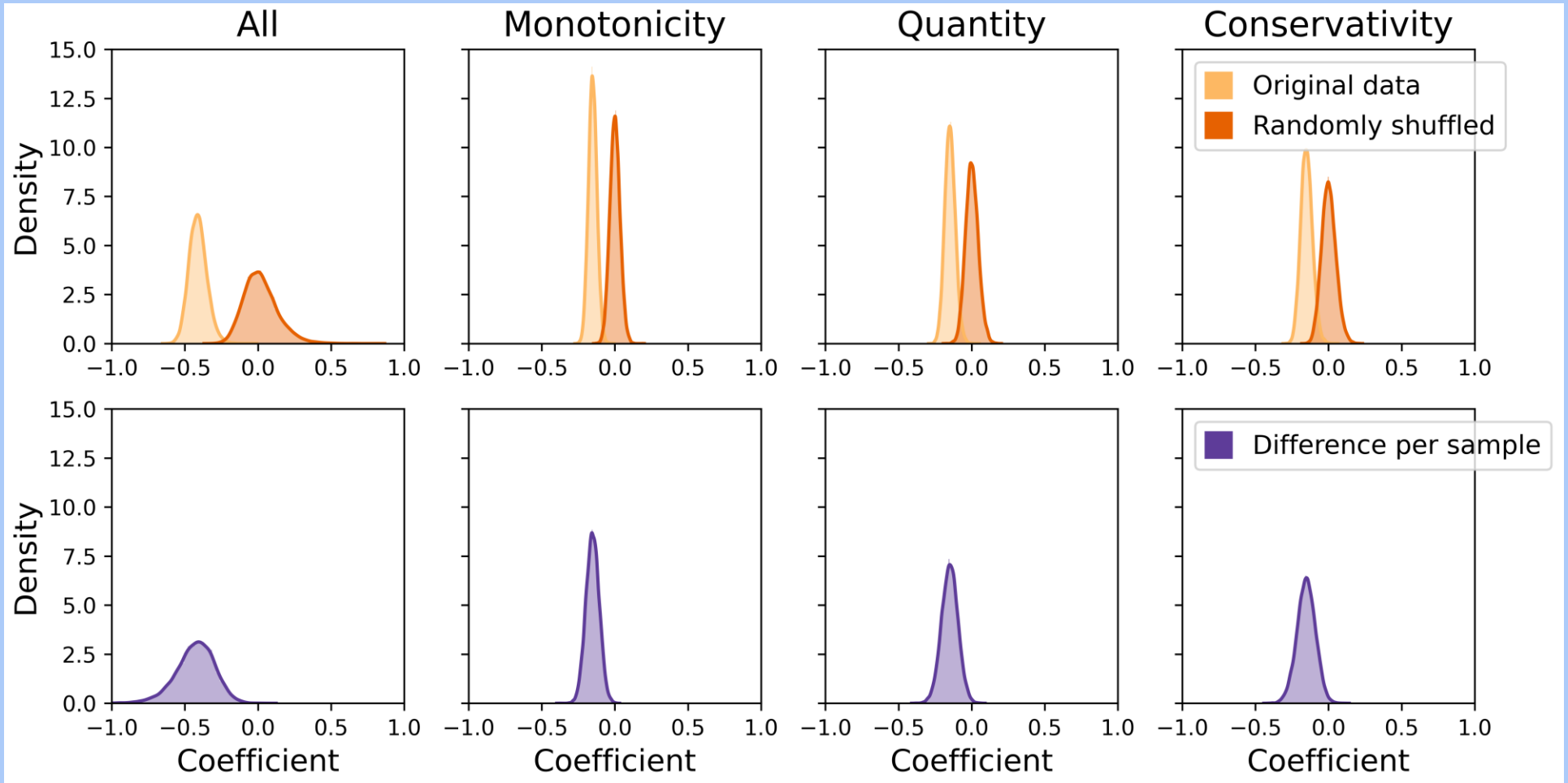
Table 2: Average (standardized) ML scores of quantifiers with (YES) versus without (NO) universal property and the proportion (%) of quantifiers with that universal property. *All* stands for quantifiers that have all three properties. The category *"all, NO"* stands for quantifiers that lack at least one property, i.e., quantifiers that do not have all three properties. For language $\mathcal{L}_{+\iota}$.

van de Pol, Lodder, van Maanen, Steinert-Threlkeld, Szymanik. Quantifiers satisfying semantic universals have shorter minimal description length. Cognition 2022
Data and code: https://github.com/ivdpol/QuantifierComplexity



van de Pol, Lodder, van Maanen, Steinert-Threlkeld, Szymanik. Quantifiers satisfying semantic universals have shorter minimal description length. Cognition 2022

Data and code: https://github.com/ivdpol/QuantifierComplexity



van de Pol, Lodder, van Maanen, Steinert-Threlkeld, Szymanik. Quantifiers satisfying semantic universals have shorter minimal description length. Cognition 2022

Data and code: https://github.com/ivdpol/QuantifierComplexity

- **Meanings satisfying semantic universals are simpler**
- The setup avoids the minimal pair methodology
- Is it robust wrt the chosen LoT?
- So, which one is it: complexity or learnability?

Other measure of complexity

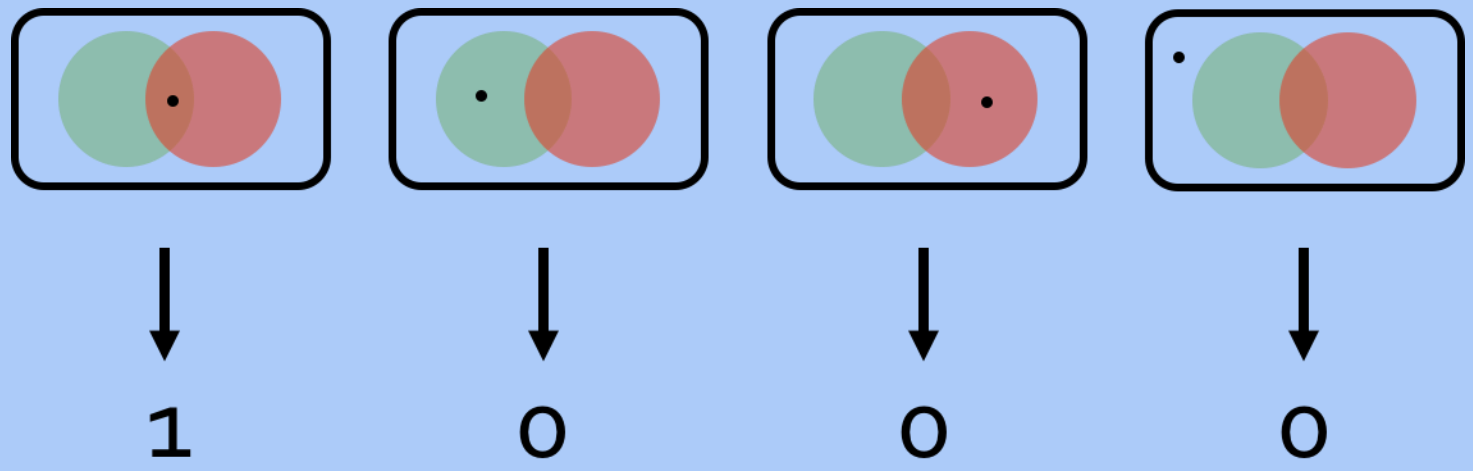
Approximate Kolmogorov Complexity

- K(x)—the length of the shortest program p that outputs x
- Language dependent but the effect is bounded (Invariance Theorem)
- The drawback: K is uncomputable
- LZ(x)—Lempel-Ziv is a tractable approximation of K
- Recent applications: Dingle, Camargo, and Louis 2018; Feldman 2016; Valle-Pérez, Camargo, and Louis 2019

Lempel-Ziv algorithm

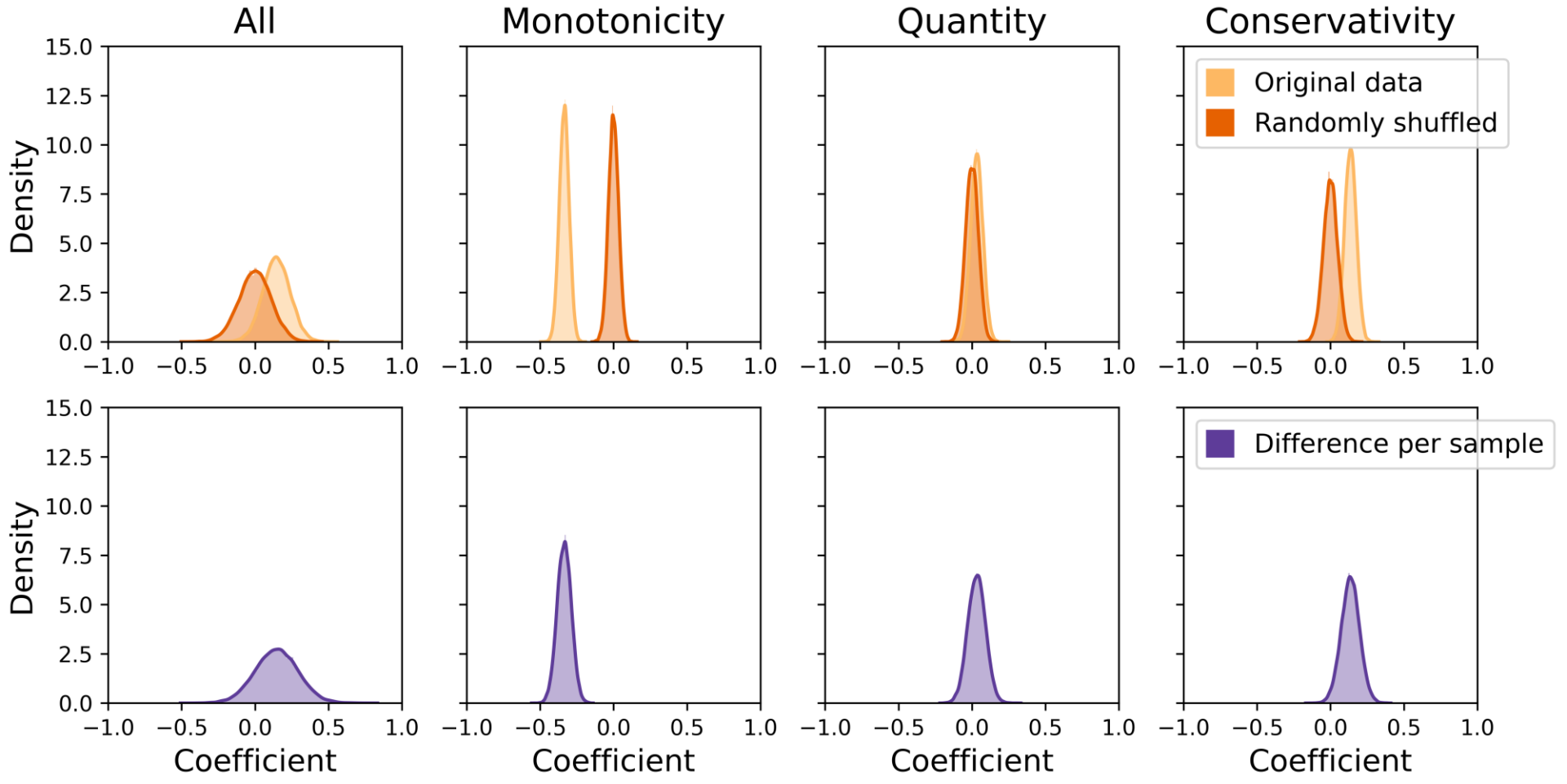
- Loseless compression algorithm
- Approximates K in the limit
- # unique subpaterns of a string
- A measure of structure
- 0|01|1|011|00|0110 = 6
- 1|11|111|1111|111 = 4

Idea: universals induce regularity/structure in the distribution of truth values across models, which aid compressibility.



	YES	NO	%
monotonicity	-0.22	0.08	0.28
quantity	0.02	0.00	0.15
conservativity	0.16	-0.02	0.12
all	0.16	0.00	0.02

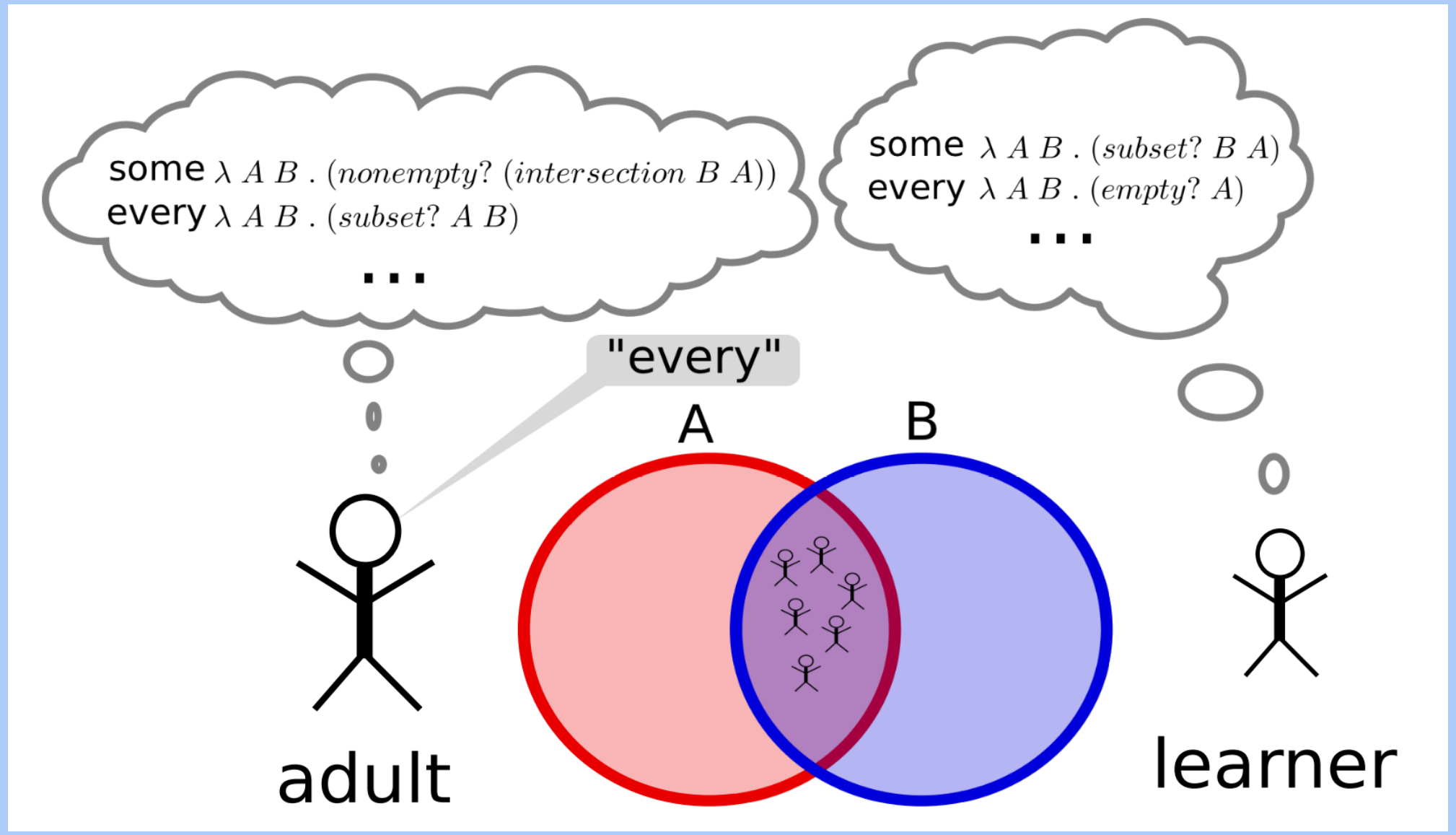
Table 4: Average standardized LZ scores of quantifiers with (YES) versus without (NO) universal property and the proportion (%) of quantifiers with that universal property. *All* stands for quantifiers that have all three properties. The category "all, NO" stands for quantifiers that lack at least one property (i.e. which do not have all three). For language $\mathcal{L}_{+\iota}$.


All
Monotonicity
Quantity
Conservativity
Original data
Randomly shuffled
Difference per sample
Density
Coefficient
15.0
12.5
10.0
7.5
5.0
2.5
0.0
−1.0
−0.5
0.0
0.5
1.0

LoT+learning

Why learnability as fencing-in fails

Piantadosi, Tannenbaum, Goodman, 2013


some λ A B . (nonempty? (intersection B A))
every λ A B . (subset? A B)
...
some λ A B . (subset? B A)
every λ A B . (empty? A)
...
"every"
A
B
adult
learner

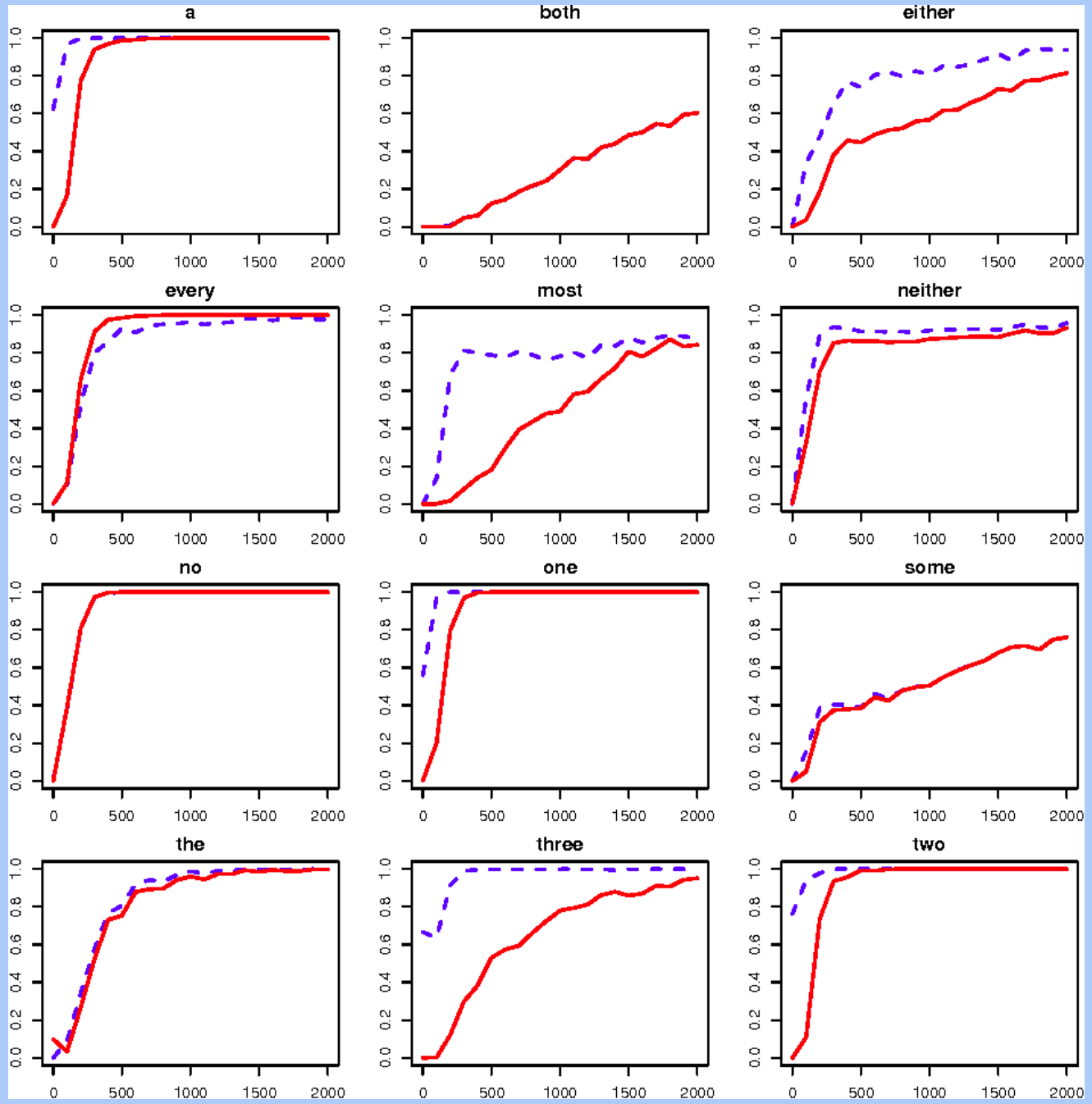
Nonterminal		Expansion	Gloss
START	$\rightarrow$	$\lambda\ A\ B\ .\ BOOL$	Function of A and B
BOOL	$\rightarrow$	$true$	Always true
	$\rightarrow$	$false$	Always false
	$\rightarrow$	$(card{>}\ SET\ SET)$	Compare cardinalities ($>$)
	$\rightarrow$	$(card{=}\ SET\ SET)$	Check if cardinalities are equal
	$\rightarrow$	$(subset?\ SET\ SET)$	Is a subset?
	$\rightarrow$	$(empty?\ SET)$	Is a set empty?
	$\rightarrow$	$(nonempty?\ SET)$	Is a set not empty?
	$\rightarrow$	$(exhaustive?\ SET)$	Is the set the entire set in the context?
	$\rightarrow$	$(singleton?\ SET)$	Contains 1 element?
	$\rightarrow$	$(doubleton?\ SET)$	Contains 2 elements?
	$\rightarrow$	$(tripleton?\ SET)$	Contains 3 elements?
SET	$\rightarrow$	$(union\ SET\ SET)$	Union of sets
	$\rightarrow$	$(intersection\ SET\ SET)$	Intersection of sets
	$\rightarrow$	$(set\text{-}difference\ SET\ SET)$	Difference of sets
	$\rightarrow$	A	Argument A
	$\rightarrow$	B	Argument B

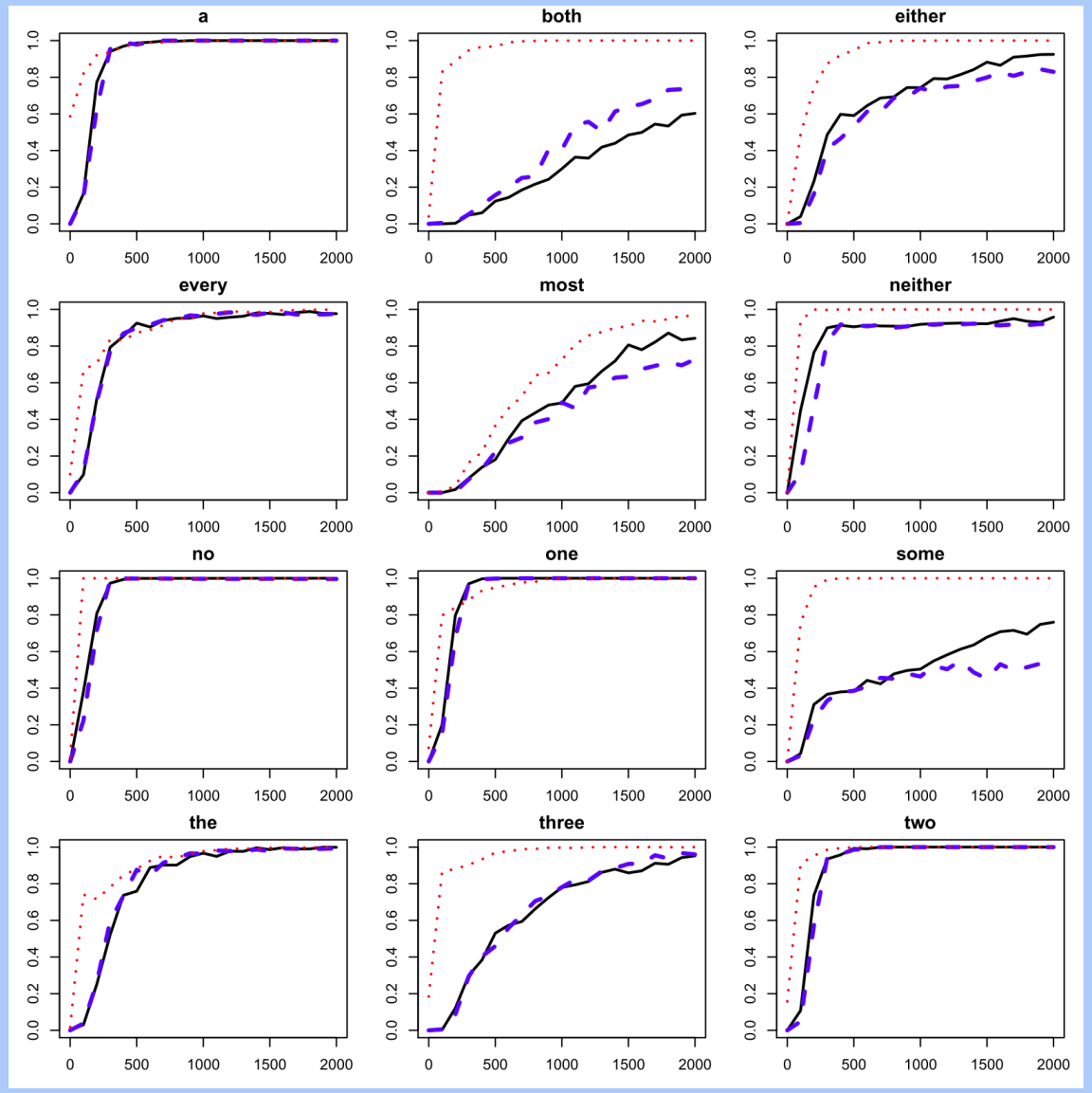
Word	Presupposition	Literal meaning
the	$\lambda\ A\ B\ .\ (singleton?\ A)$	$\lambda\ A\ B\ .\ (nonempty?\ (intersection\ A\ B))$
a/some	$\lambda\ A\ B\ .\ TRUE$	$\lambda\ A\ B\ .\ (nonempty?\ (intersection\ A\ B))$
one	$\lambda\ A\ B\ .\ (nonempty?\ A)$	$\lambda\ A\ B\ .\ (singleton?\ (intersection\ A\ B))$
two	$\lambda\ A\ B\ .\ (nonempty?\ A)$	$\lambda\ A\ B\ .\ (doubleton?\ (intersection\ A\ B))$
three	$\lambda\ A\ B\ .\ (nonempty?\ A)$	$\lambda\ A\ B\ .\ (tripleton?\ (intersection\ A\ B))$
both	$\lambda\ A\ B\ .\ (doubleton?\ A)$	$\lambda\ A\ B\ .\ (doubleton?\ (intersection\ A\ B))$
either	$\lambda\ A\ B\ .\ (doubleton?\ A)$	$\lambda\ A\ B\ .\ (singleton?\ (intersection\ A\ B))$
neither	$\lambda\ A\ B\ .\ (doubleton?\ A)$	$\lambda\ A\ B\ .\ (empty?\ (intersection\ A\ B))$
every	$\lambda\ A\ B\ .\ (nonempty?\ A)$	$\lambda\ A\ B\ .\ (subset?\ A\ B)$
most	$\lambda\ A\ B\ .\ (nonempty?\ A)$	$\lambda\ A\ B\ .\ (card{>}\ (intersection\ A\ B)\ (set\text{-}difference\ A\ B))$
none/no	$\lambda\ A\ B\ .\ (nonempty?\ A)$	$\lambda\ A\ B\ .\ (empty?\ (intersection\ A\ B))$

$$P(m|u_1, c_1, u_2, c_2, \ldots u_n, c_n) \propto P(u_1, u_2, \ldots, u_n|m, c_1, \ldots, c_n) \cdot P(m)$$
$$\propto \prod_{i=1}^{n} P(u_i|m, c_i) \cdot P(m)$$

Prior $P(m)$: determined by the LoT grammar; shorter-to-express meanings are preferred.

Likelihood $P(u_i|m, c_i)$: preference for more informative, but with noise.

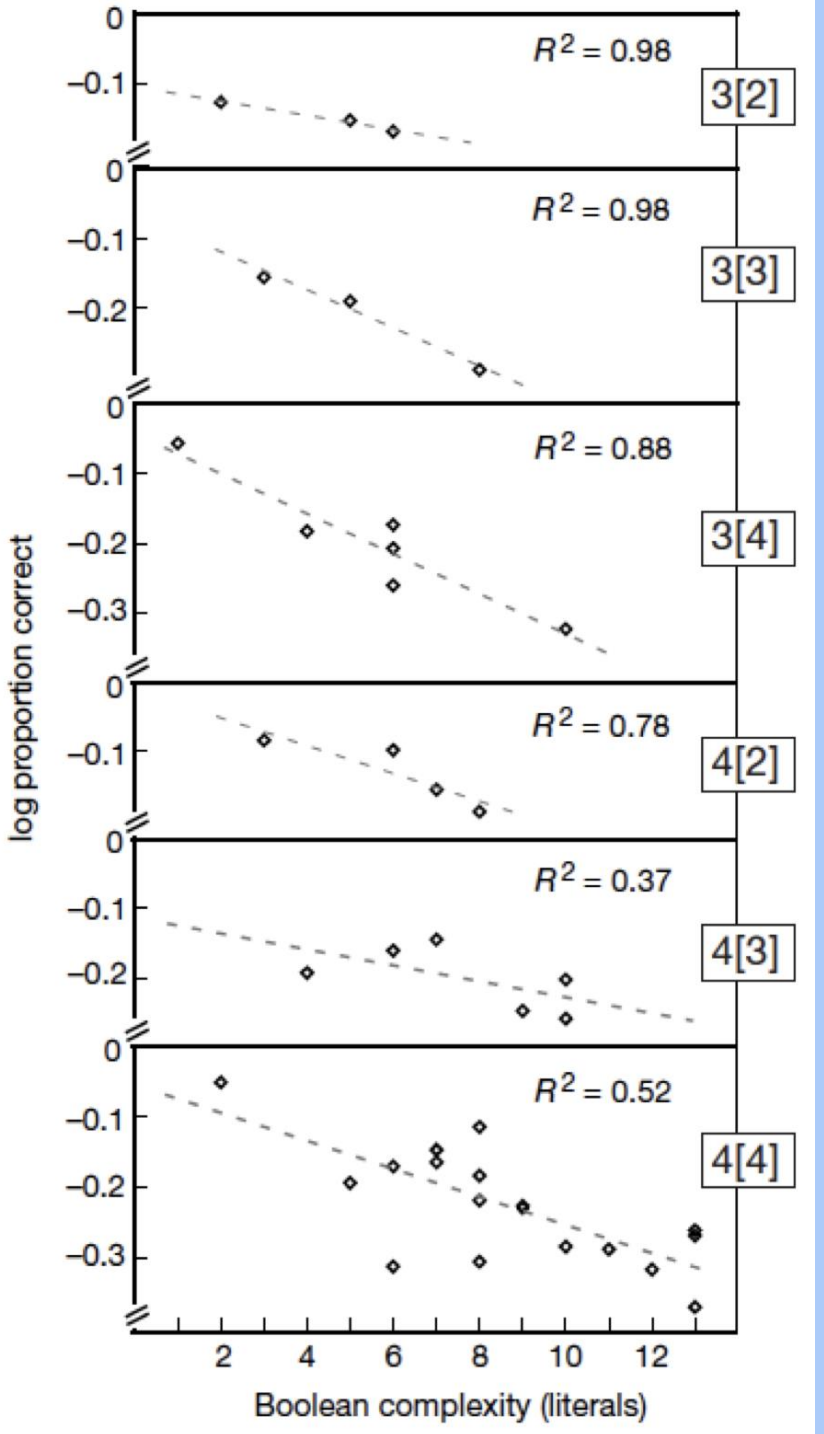

a
both
either
every
most
neither
no
one
some
the
three
two
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000


a
both
either
every
most
neither
no
one
some
the
three
two
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000

“Likely, the unrestricted space has many hypotheses which are so implausible, they can be ignored quickly and do not affect learning. The hard part of learning, may be choosing between the plausible competitor meanings, not in weeding out a large space of potential meanings.”

- Does this model predict human learning curves well?
- How sensitive are the model and its results sensitive to various choices (e.g. primitives, weights, shape of likelihood function)?
- Does it say anything general about e.g. monotone and topic-neutral quantifiers?

Robustness of the results depending on the LoT



Which Boolean connectives?

SIMPLEBOOLEAN		
START	→	*lambda x . BOOL*
BOOL	→	*(and BOOL BOOL)*
		(or BOOL BOOL)
		(not BOOL)
		true
		false
BOOL	→	*(F OBJECT)*
OBJECT	→	*x*
F	→	COLOR
		SHAPE
		SIZE
COLOR	→	*blue?*
		green?
		yellow?
SHAPE	→	*circle?*
		rectangle?
		triangle?
SIZE	→	*size1?*
		size2?
		size3?

NAND		
START	→	*lambda x . BOOL*
BOOL	→	*(nand BOOL BOOL)*
		true
		false
BOOL	→	*(F OBJECT)*
OBJECT	→	*x*
F	→	COLOR
		SHAPE
		SIZE
COLOR	→	*blue?*
		green?
		yellow?
SHAPE	→	*circle?*
		rectangle?
		triangle?
SIZE	→	*size1?*
		size2?
		size3?

DNF		
START	→	*lambda x . DISJ*
DISJ	→	*CONJ*
		(or CONJ DISJ)
CONJ	→	*BOOL*
		(and BOOL CONJ)
BOOL	→	*(F OBJECT)*
		(not (F OBJECT))
OBJECT	→	*x*
F	→	COLOR
		SHAPE
		SIZE
COLOR	→	*blue?*
		green?
		yellow?
SHAPE	→	*circle?*
		rectangle?
		triangle?
SIZE	→	*size1?*
		size2?
		size3?

HORN CLAUSE		
START	→	*lambda x . HORN-CONJ*
HORN-CONJ	→	*HORN-CLAUSE*
		(and HORN-CLAUSE HORN-CONJ)
HORN-CLAUSE	→	*(implies HORN-CONJ PRIM)*
HORN-CLAUSE	→	*(implies HORN-CONJ false)*
PRIM	→	*(F OBJECT)*
OBJECT	→	*x*
F	→	COLOR
		SHAPE
		SIZE
COLOR	→	*blue?*
		green?
		yellow?
SHAPE	→	*circle?*
		rectangle?
		triangle?
SIZE	→	*size1?*
		size2?
		size3?

Which representational system is the most likely, given human responses?

Grammar	H.O.LL	FP	$R^2_{response}$	R^2_{mean}
FULLBOOLEAN	−16296.84	27	.88	.60
BICONDITIONAL	−16305.13	26	.88	.64
CNF	−16332.39	26	.89	.69
DNF	−16343.87	26	.89	.66
SIMPLEBOOLEAN	−16426.91	25	.87	.70
IMPLIES	−16441.29	26	.87	.70
HORNCLAUSE	−16481.90	27	.87	.65
NAND	−16815.60	24	.84	.61
NOR	−16859.75	24	.85	.58
UNIFORM	−19121.65	4	.77	.06
EXEMPLAR	−23634.46	5	.55	.15
ONLYFEATURES	−31670.71	19	.54	.14
RESPONSE-BIASED	−37912.52	4	.03	.04

Piantadosi et al. 2016

- You can also infer LoT from the informativeness-complexity trade-off but again it’s hard to distinguish top candidates (Denić & Szymanik, 2022).
- All reasoning data (Zhai, Titov, Szymanik, 2015).
- In general, you shouldn’t rather expect to find a unique candidate LoT (Carcassi & Szymanik, 2022).
- So, how shall we measure complexity? (relevant for tomorrow)

So, which one is the fundamental notion: complexity or learnability?