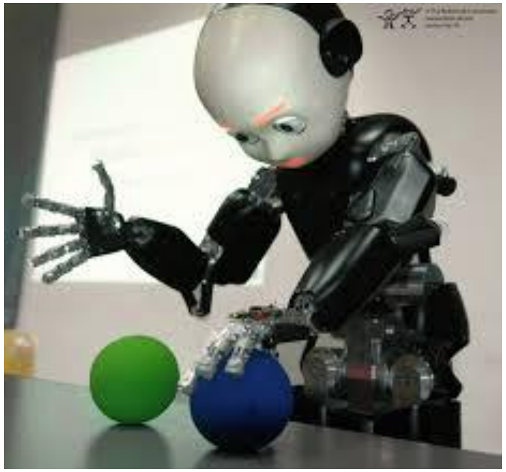
Theories, imagination, and the generation of new ideas

Ullman vs. Schulz

August 27th, 2015

Outline of Debate



Child as intuitive scientist → $P(T \mid D) \propto P(D \mid T)\ P(T)$

↓

Large theory spaces

↓

Stochastic search algorithms?! ←

Stochastic search algorithms!

Outline of Debate

Background (Tomer)

What good are theories?

Representing a good theory

Finding a good theory – stochastic search

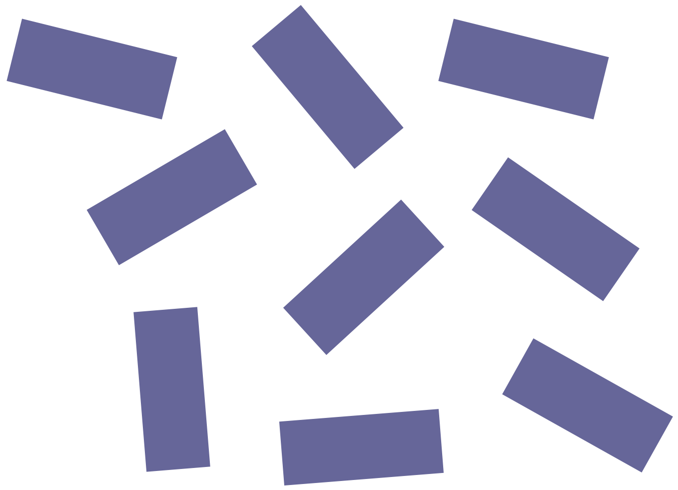
Imagination and issues with stochastic search (Laura)

Response (Tomer)

Response and summary (Laura)

What Good is a Theory?

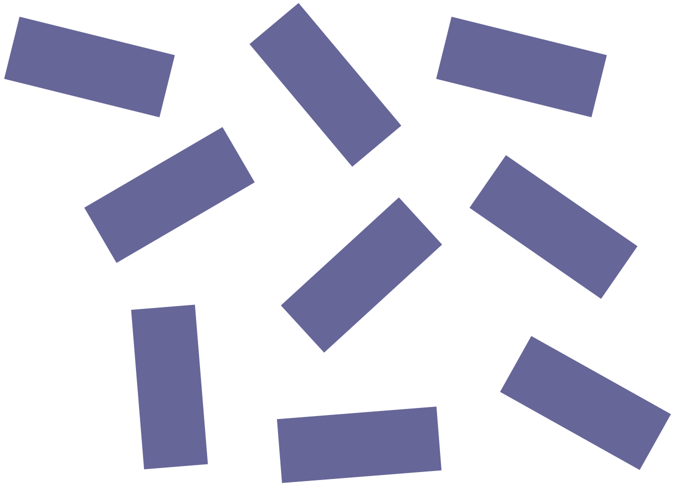
Structured knowledge, “theories”



(Magnets, metals and non-magnetic)

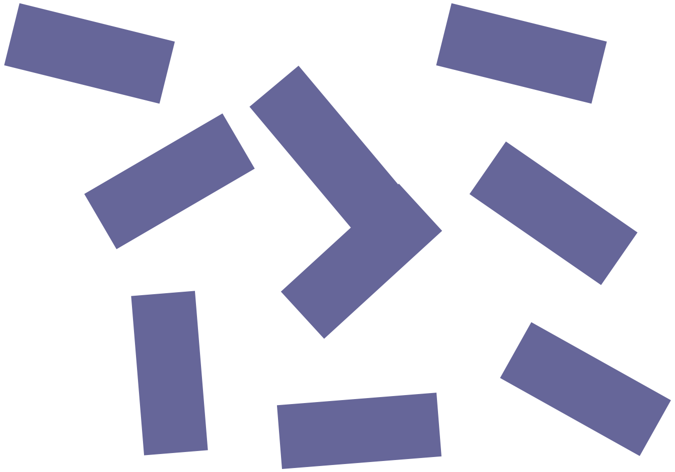
What Good is a Theory?

Begin collecting observations



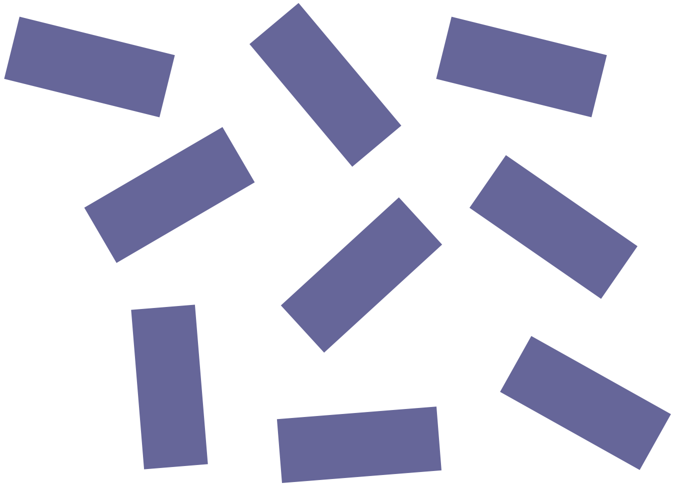
What Good is a Theory?

Sometimes nothing happens



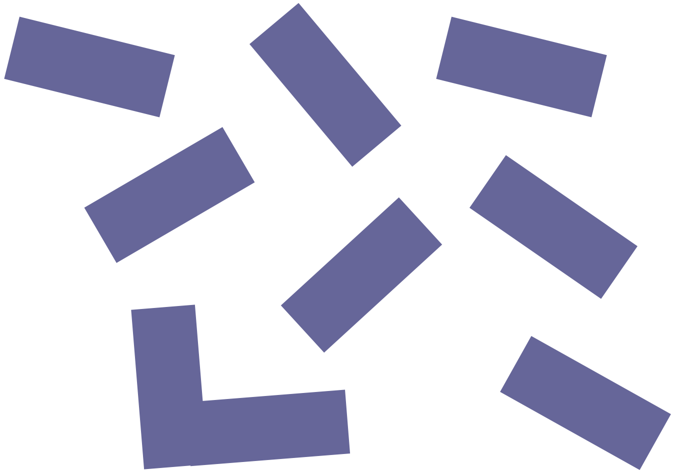
What Good is a Theory?

Sometimes nothing happens



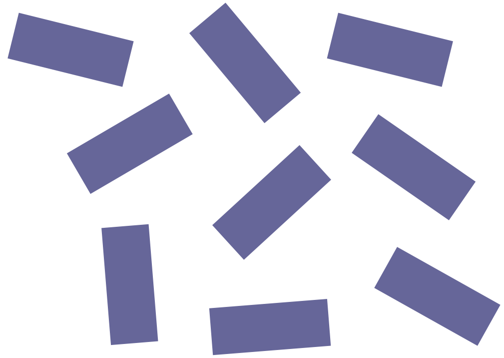
What Good is a Theory?

Sometimes objects **stick**



What Good is a Theory?

Explanation: Bag of data?



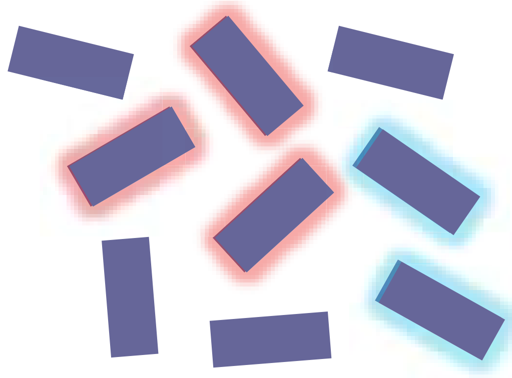
	A	B	C	D	E	F	G	H	I
A		✓							
B	✓		✓		✓				
C		✓			✓				
D									
E		✓	✓					✓	
F									
G									
H					✓				
I									

What Good is a Theory?

Explanation: Theory

Concepts: "schmagnet", "schmetal"

Rules:



Rule 1: interacts(X,Y) ← schmagnet(X) ∧ schmagnet(Y)

Rule 2: interacts(X,Y) ← schmagnet(X) ∧ schmetal(Y)

Rule 3: interacts(X,Y) ← interacts(Y,X)

⬇

Assign "schmagnets" & "schmetals"

⬇

Predict observed data

Finding a Good Theory

Rational inference problem

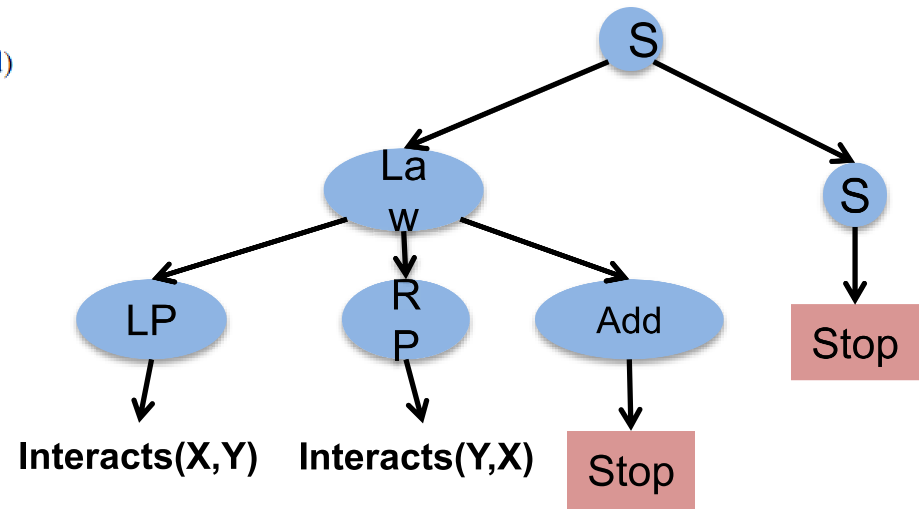
Out of all possible theories, find the one that 'best' explains the observed data

$$P(T \mid D) \propto P(D \mid T)\ P(T)$$

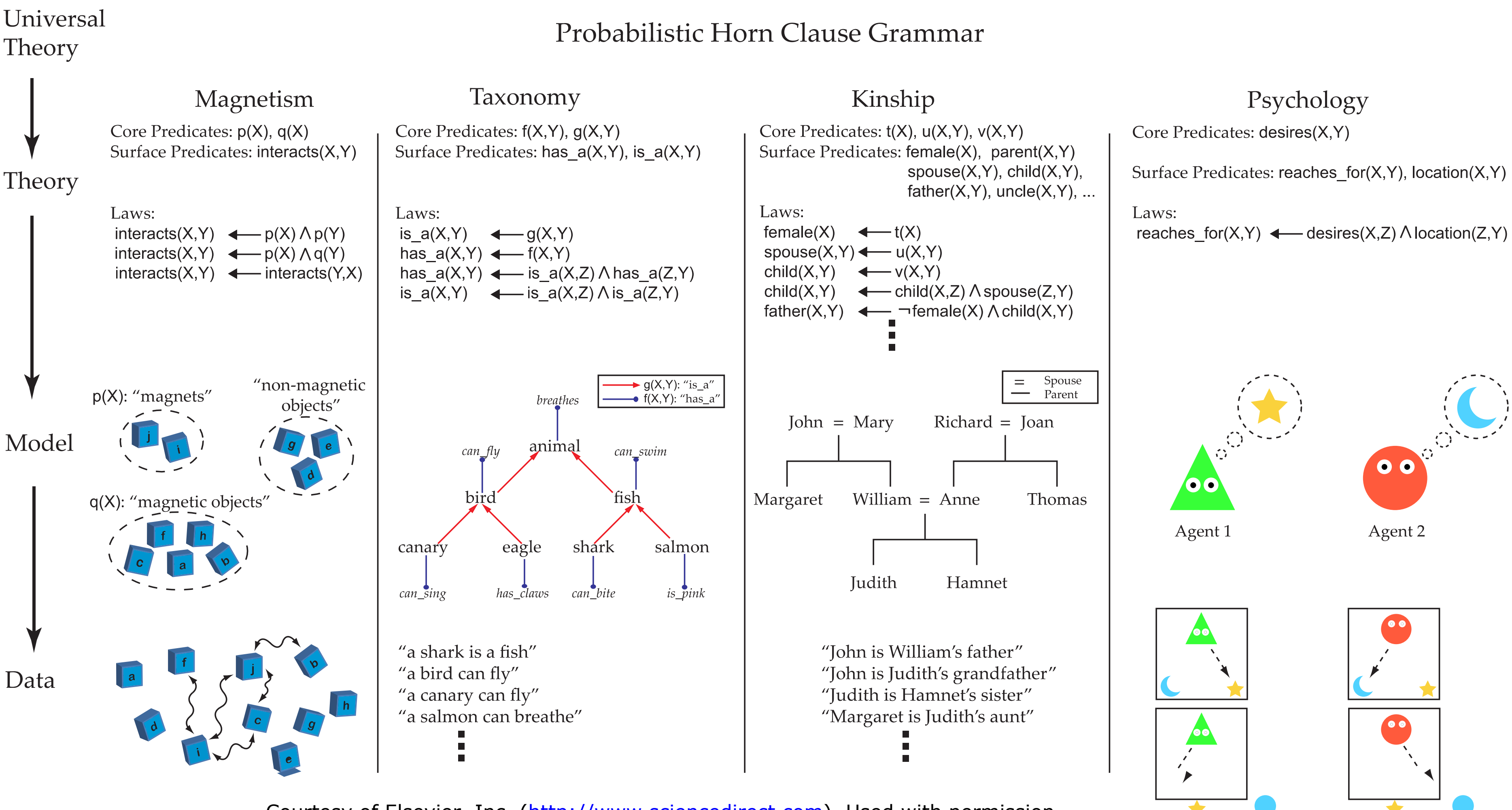
(Tenenbaum, Griffiths, & Kemp, 2006)

Learning a Good Theory - Grammar

Top level theory			
(S1)	S	$\Rightarrow$	(Law) $\wedge$ S
(S2)	S	$\Rightarrow$	(Tem) $\wedge$ S
(S3)	S	$\Rightarrow$	Stop
Random law generation			
(Law)	Law	$\Rightarrow$	($P_{left} \leftarrow P_{right} \wedge$ Add)
(Add1)	A	$\Rightarrow$	P
(Add2)	A	$\Rightarrow$	Stop
Predicate generation			
($P_{left}1$)	P_{left}	$\Rightarrow$	$surface1()$
$\vdots$			
($P_{left}\alpha$)	P_{left}	$\Rightarrow$	$surface\alpha()$
($P_{right}1$)	P_{right}	$\Rightarrow$	$surface1()$
$\vdots$			
($P_{right}\alpha$)	P_{right}	$\Rightarrow$	$surface\alpha()$
($P_{right}(\alpha+1)$)	P_{right}	$\Rightarrow$	$core1()$
$\vdots$			
($P_{right}(\alpha+\beta)$)	P_{right}	$\Rightarrow$	$core\beta()$
Law templates			
(Tem1)	Tem	$\Rightarrow$	template1()
$\vdots$			
(Temγ)	Tem	$\Rightarrow$	templateγ()

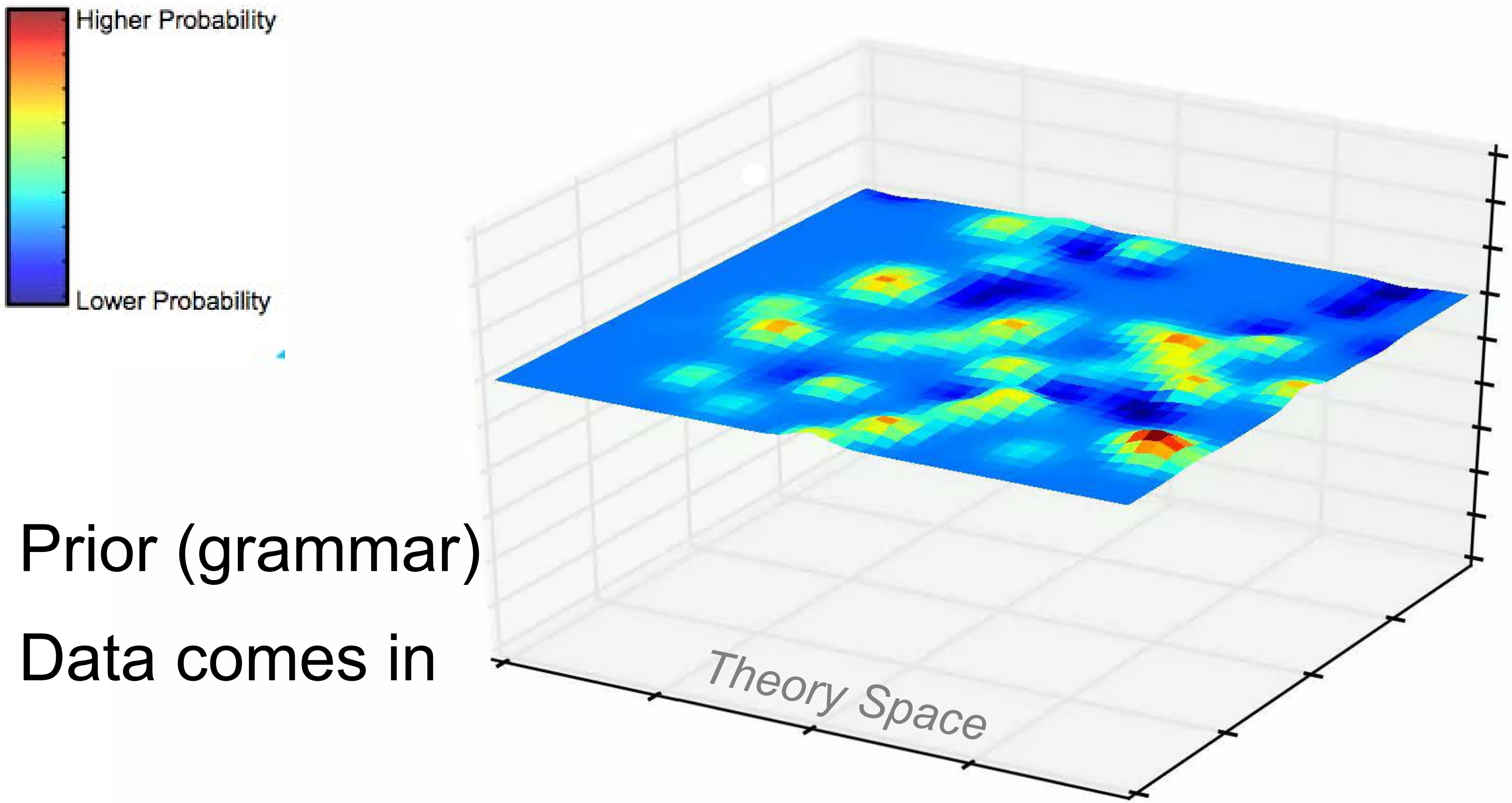


Examples of Theories



Courtesy of Elsevier, Inc. (http://www.sciencedirect.com). Used with permission.
Source: T. Ullman, N. Goodman & J. Tenenbaum. "Theory learning as stochastic search in the language of thought." *Cognitive Development* 27 no. 4 (2012): 455-480.

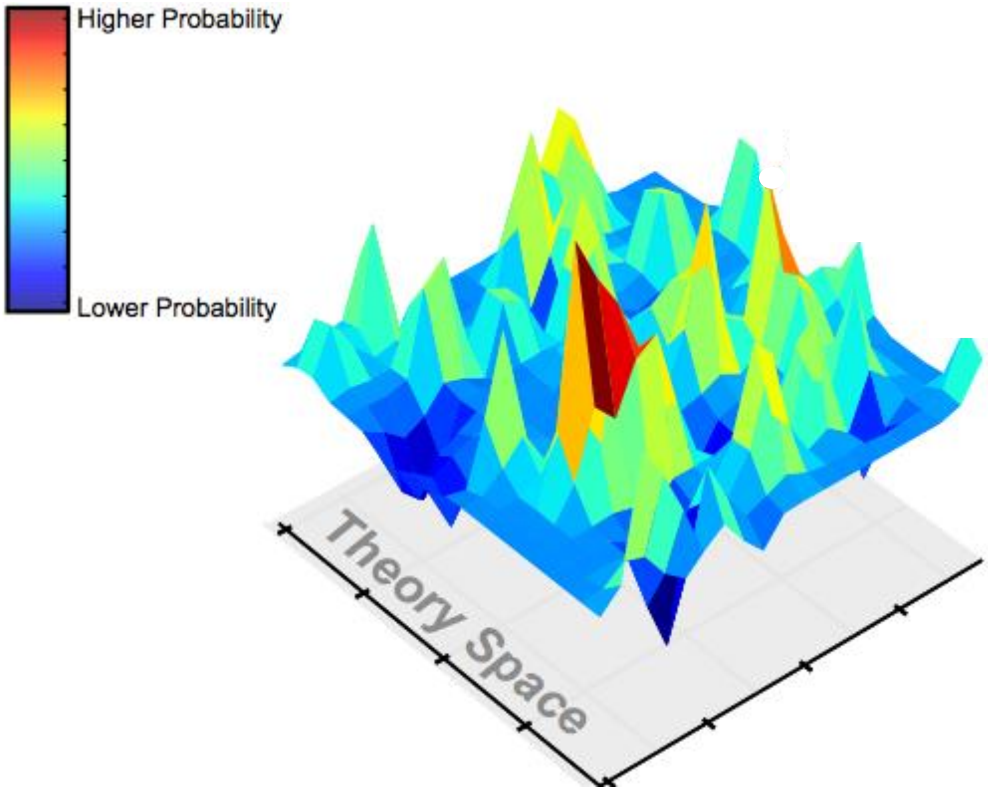
Finding a Good Theory – Ideal Level



Prior (grammar)

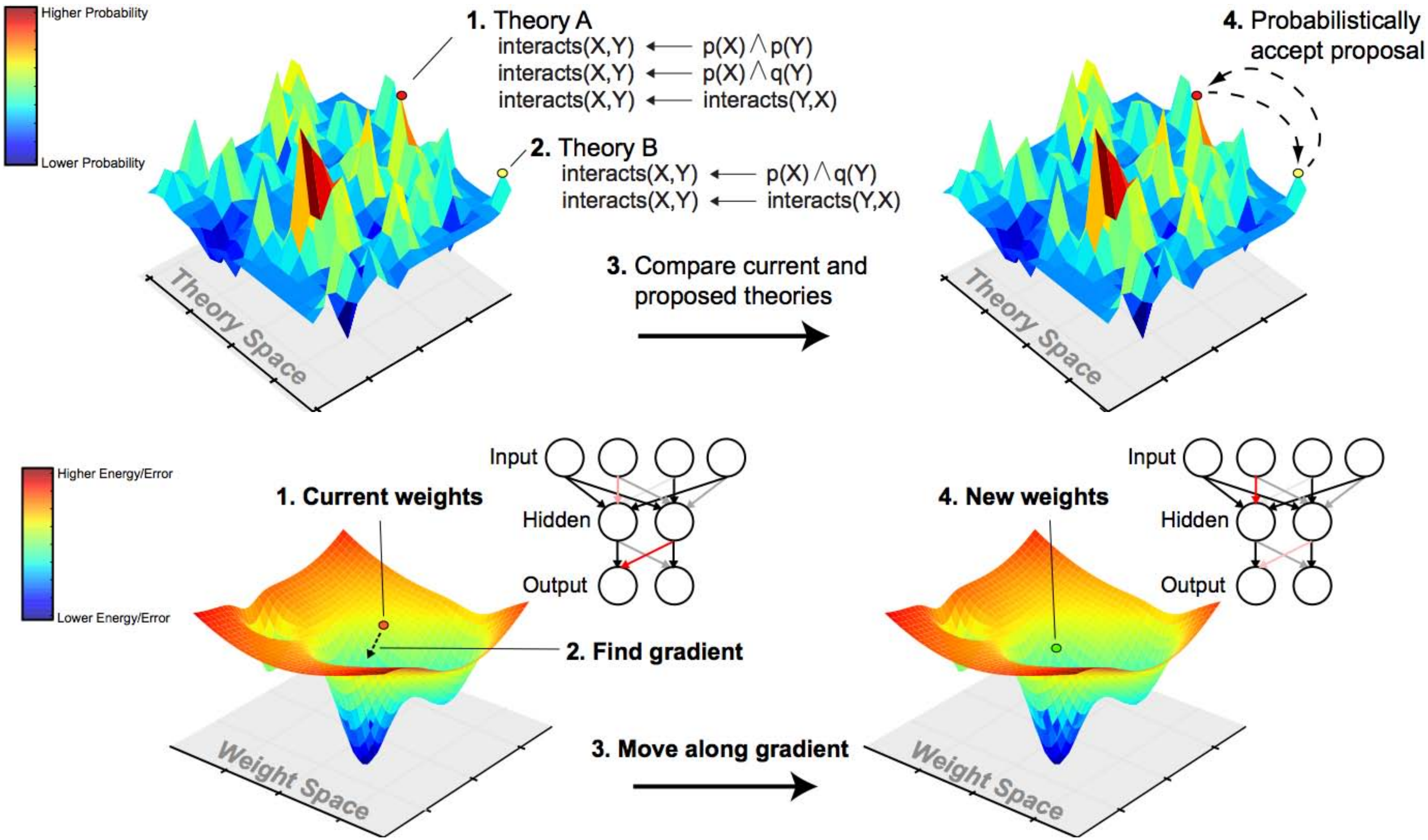
Data comes in

The Problem of Search



Source: T. Ullman, N. Goodman & J. Tenenbaum. "Theory learning as stochastic search in the language of thought." *Cognitive Development* 27 no. 4 (2012): 455-480.

Stochastic Search

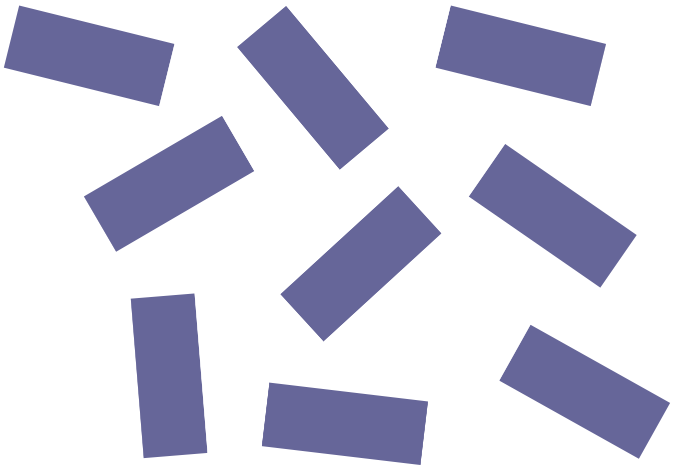


Source: T. Ullman, N. Goodman & J. Tenenbaum. "Theory learning as stochastic search in the language of thought." *Cognitive Development* 27 no. 4 (2012): 455-480.

Proposing Alternative Theories

Observed predicate: interacts(X,Y)

$$\text{Rule 1: } \text{interacts}(X,Y) \leftarrow p(X) \wedge p(Y)$$



Metropolis-Hastings algorithm proposes alternative theories by changing current theory (new rules, predicates, etc)

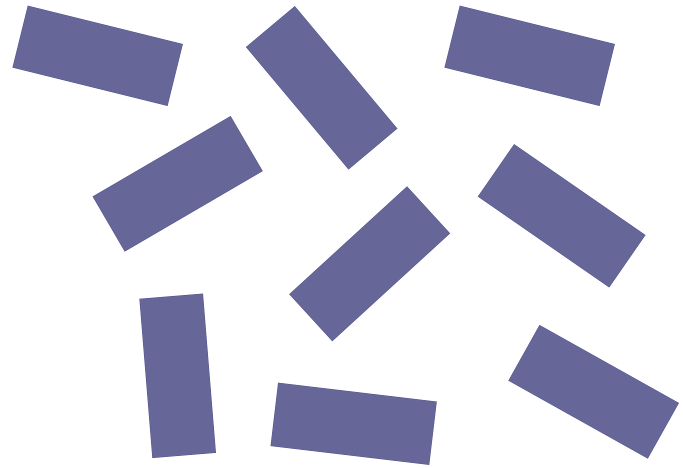
Accept or reject new theory with probability depending on score

Proposing Alternative Theories

Observed predicate: interacts(X,Y)

$$\text{Rule 1: } \text{interacts}(X, Y) \quad p(X) \wedge p(Y)$$

$$\text{Rule 2: } \text{interacts}(X, Y) \leftarrow q(X) \wedge q(Y)$$



Metropolis-Hastings algorithm proposes alternative theories by changing current theory (new rules, predicates, etc)

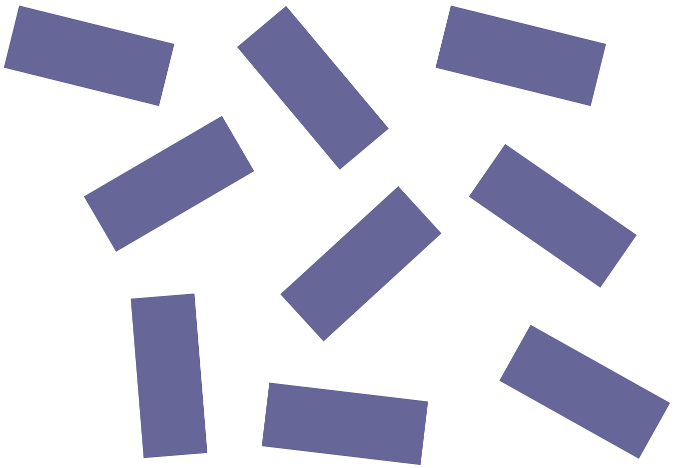
Accept or reject new theory with probability depending on score

Proposing Alternative Theories

Observed predicate: interacts(X,Y)

$$\text{Rule 1: interacts}(X,Y) \leftarrow p(X) \wedge p(Y)$$

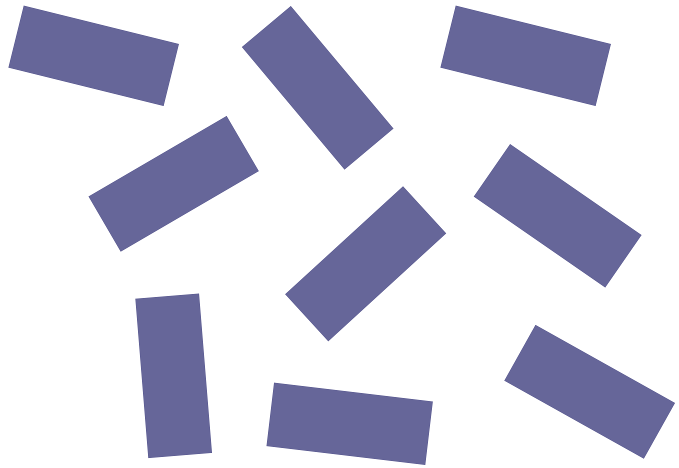
$$\text{Rule 2: interacts}(X,Y) \leftarrow p(X) \wedge q(Y)$$



Metropolis-Hastings algorithm proposes alternative theories by changing current theory (new rules, predicates, etc)

Accept or reject new theory with probability depending on score

Proposing Alternative Theories



Observed predicate: interacts(X,Y)

$$\text{Rule 1: } \text{interacts}(X,Y) \leftarrow p(X) \wedge p(Y)$$

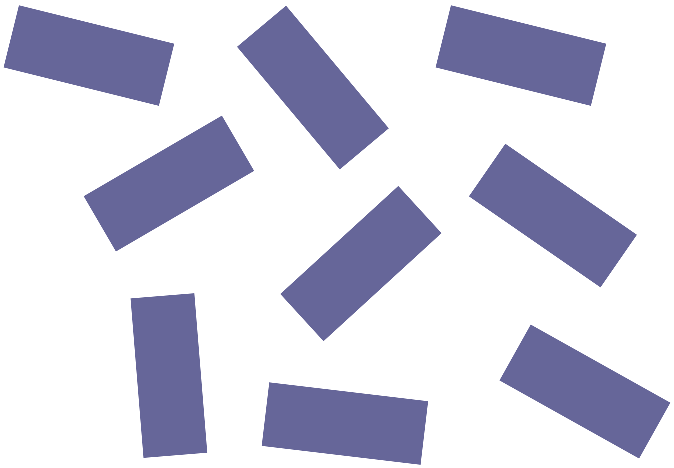
$$\text{Rule 2: } \text{interacts}(X,Y) \leftarrow p(X) \wedge q(Y)$$

$$\text{Rule 3: } \text{interacts}(X,Y) \leftarrow \text{interacts}(Y,X)$$

Metropolis-Hastings algorithm proposes alternative theories by changing current theory (new rules, predicates, etc)

Accept or reject new theory with probability depending on score

Proposing Alternative Theories



Observed predicate: interacts(X,Y)

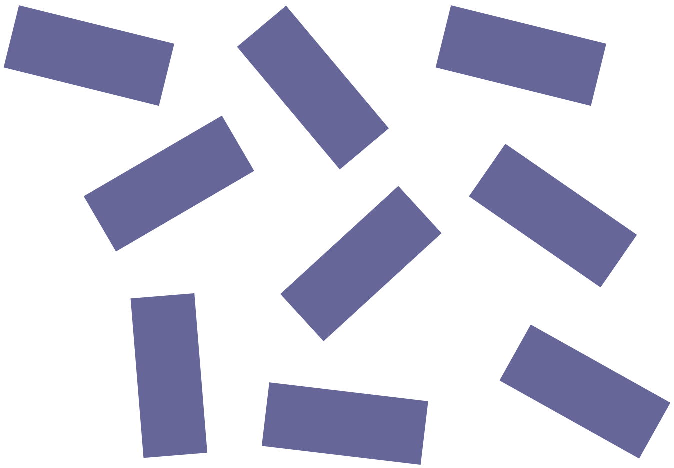
$$\text{Rule 2}: \text{interacts}(X, Y) \leftarrow p(X) \wedge q(Y)$$

$$\text{Rule 3}: \text{interacts}(X, Y) \leftarrow \text{interacts}(Y, X)$$

Metropolis-Hastings algorithm proposes alternative theories by changing current theory (new rules, predicates, etc)

Accept or reject new theory with probability depending on score

Proposing Alternative Theories



Observed predicate: interacts(X,Y)

$$\text{Rule 1: interacts}(X,Y) \leftarrow p(X) \wedge p(Y)$$

$$\text{Rule 2: interacts}(X,Y) \leftarrow p(X) \wedge q(Y)$$

$$\text{Rule 3: interacts}(X,Y) \leftarrow \text{interacts}(Y,X)$$

Metropolis-Hastings algorithm proposes alternative theories by changing current theory (new rules, predicates, etc)

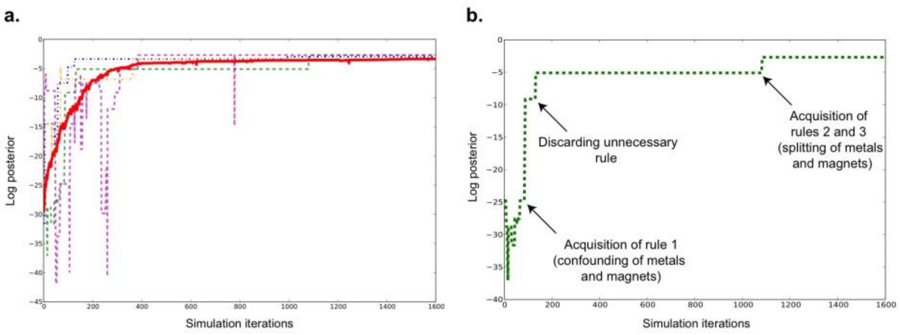
Accept or reject new theory with probability depending on score

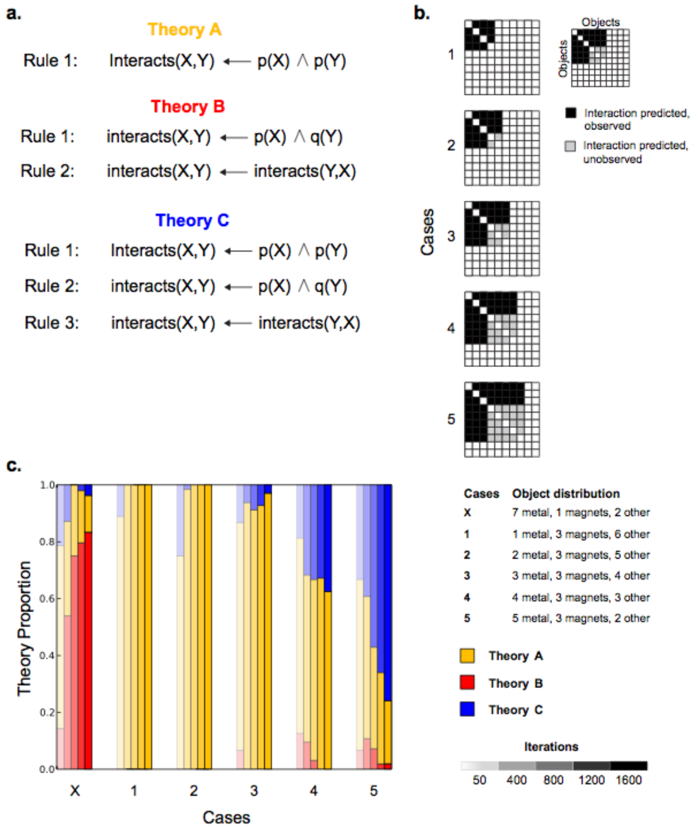
Stochastic Search and Children

Rule 1: interacts(X,Y) ← p(X) ∧ p(Y)

Rule 2: interacts(X,Y) ← p(X) ∧ q(Y)

Rule 3: interacts(X,Y) ← interacts(Y,X)





Ullman, Goodman & Tenenbaum, 2012
Denison, Bonawitz, Gopnik and Griffiths 2013

Mid-Summary

Theories are useful

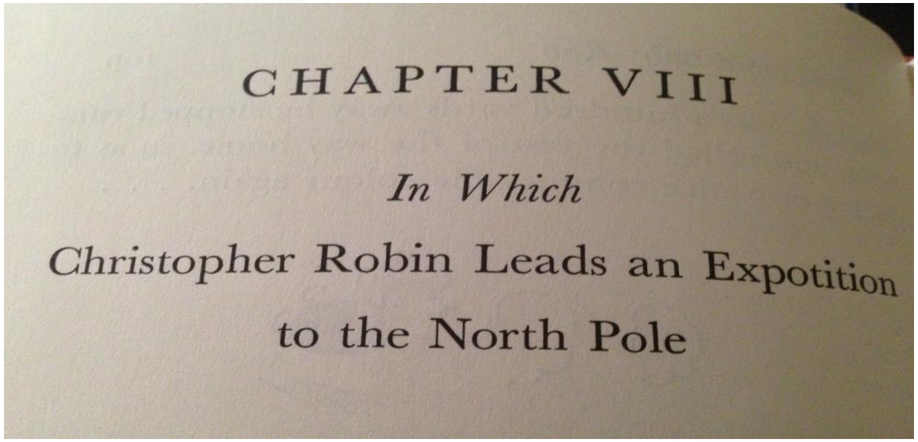
Rich, structured theories define a rich landscape

Algorithmic solution: stochastic search in rich landscape

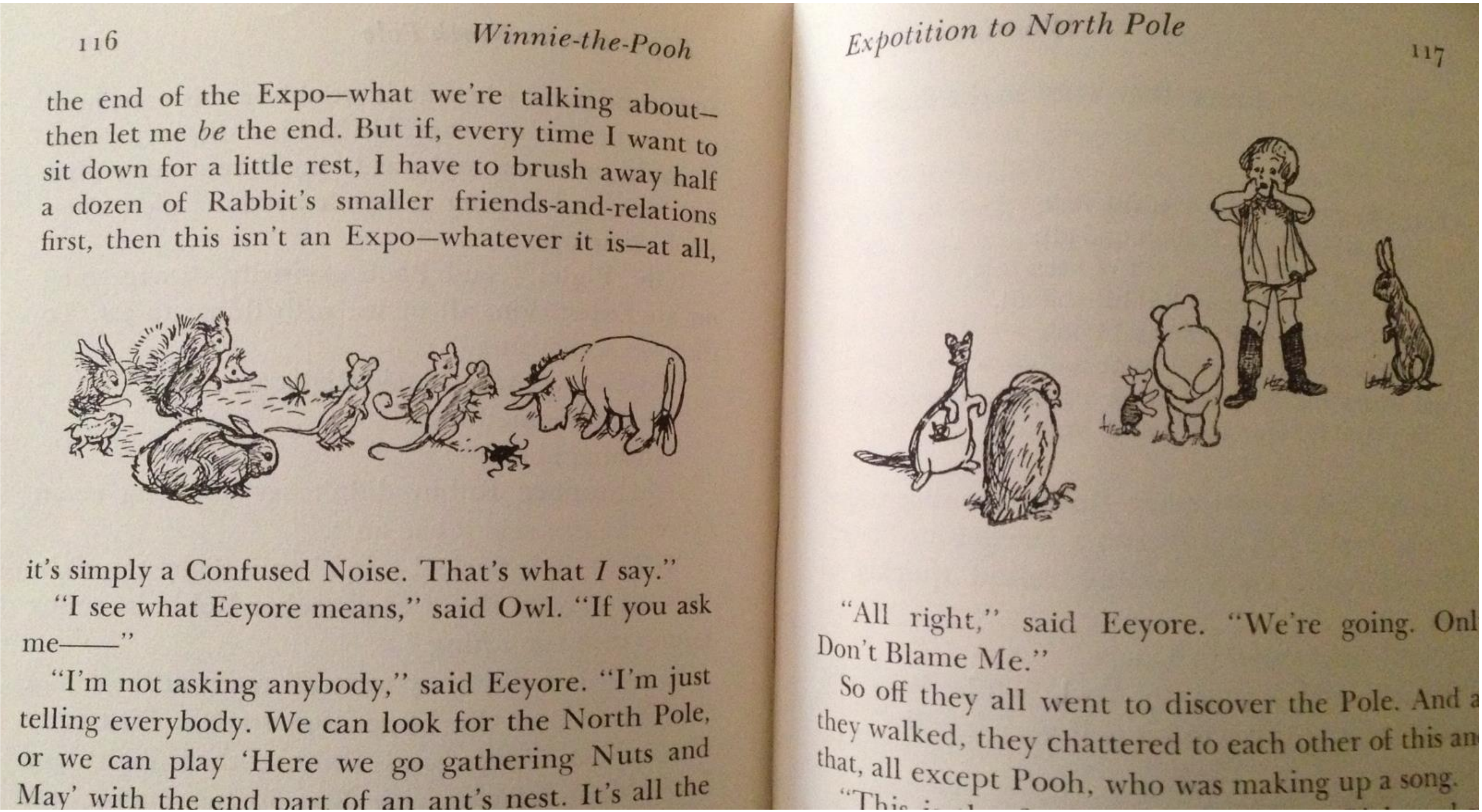
Application to children?

Handoff to Laura

In Which, following an elegant exposition of a formal model, attendant experiments and quantitative data, Laura proceeds to wave her hands around …



What’s wrong with stochastic search?



116 *Winnie-the-Pooh*

the end of the Expo—what we’re talking about—then let me *be* the end. But if, every time I want to sit down for a little rest, I have to brush away half a dozen of Rabbit’s smaller friends-and-relations first, then this isn’t an Expo—whatever it is—at all, it’s simply a Confused Noise. That’s what *I* say.”

“I see what Eeyore means,” said Owl. “If you ask me——”

“I’m not asking anybody,” said Eeyore. “I’m just telling everybody. We can look for the North Pole, or we can play ‘Here we go gathering Nuts and May’ with the end part of an ant’s nest. It’s all the

Expotition to North Pole 117

“All right,” said Eeyore. “We’re going. Onl Don’t Blame Me.”

So off they all went to discover the Pole. And a they walked, they chattered to each other of this an that, all except Pooh, who was making up a song.

Issues with Stochastic Search

Problem 1: Even with prior knowledge, templates, and a bias towards simplicity, the search space is infinite.

"I know why you have to turn off your cell phone when you get on the airplane"

"Oh yeah? Why?"

Issues with Stochastic Search

Problem 1: Even with prior knowledge, templates, and a bias towards simplicity, the search space is infinite.

"Because when the plane takes off it's too noisy to hear. "

"Oh yeah? Why?"

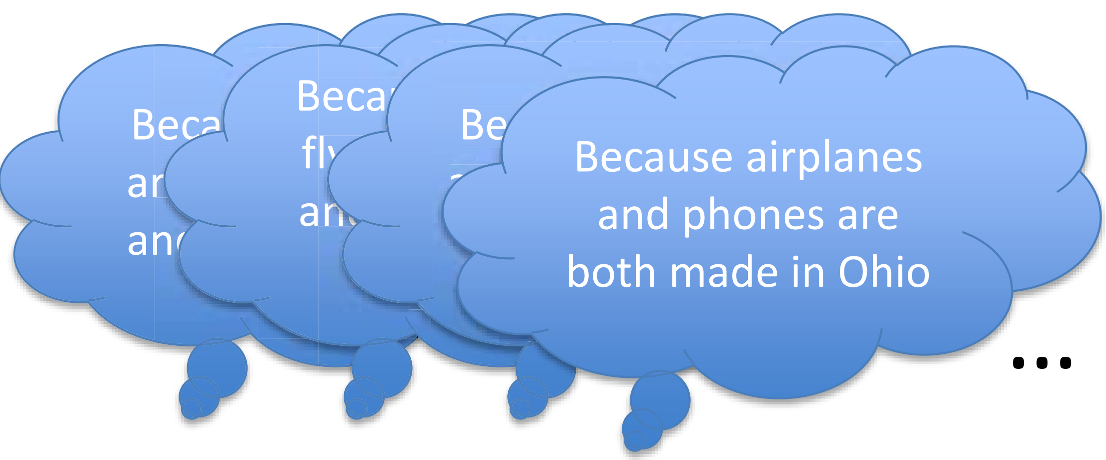
Because airplanes are made of metal and so are phones

Issues with Stochastic Search

Problem 1: Even with prior knowledge, templates, and a bias towards simplicity, the search space is infinite.

There are innumerable logical, constitutive, causal, and relational hypotheses consistent with the grammar of our intuitive theories How do we rapidly converge on ones that actually might explain the data?

“Oh yeah? Why?”



…

Issues with Stochastic Search

Problem 1: Even with prior knowledge, templates, and a bias towards simplicity, the search space is infinite.

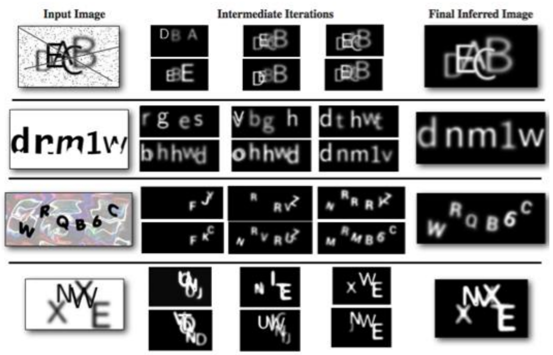
Not just toy problem. Modeling even relatively simple, well-understood problems takes long time.

Winnie-the-Pooh film image removed due to copyright restrictions.

Iterations spent searching in hopeless places

Approximate Bayesian Image Interpretation using Generative Probabilistic Graphics Programs

Vikash K. Mansinghka* [1,2], Tejas D. Kulkarni* [1,2], Yura N. Perov[1,2,3], and Joshua B. Tenenbaum[1,2]



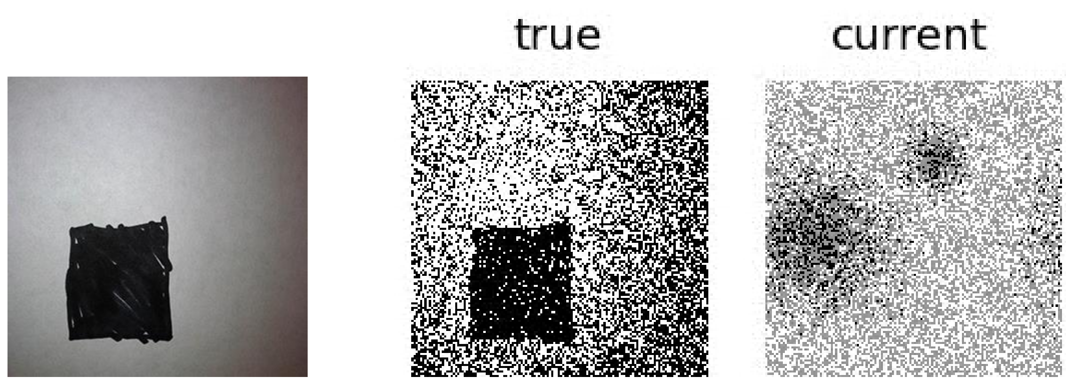
Issues with Stochastic Search

Problem 1: Even with prior knowledge, templates, and a bias towards simplicity, the search space is infinite.

Not just toy problem. Modeling even relatively simple, well-understood problems takes long time.

Winnie-the-Pooh film image removed due to copyright restrictions.

Iterations spent searching in hopeless places



Issues with Stochastic Search

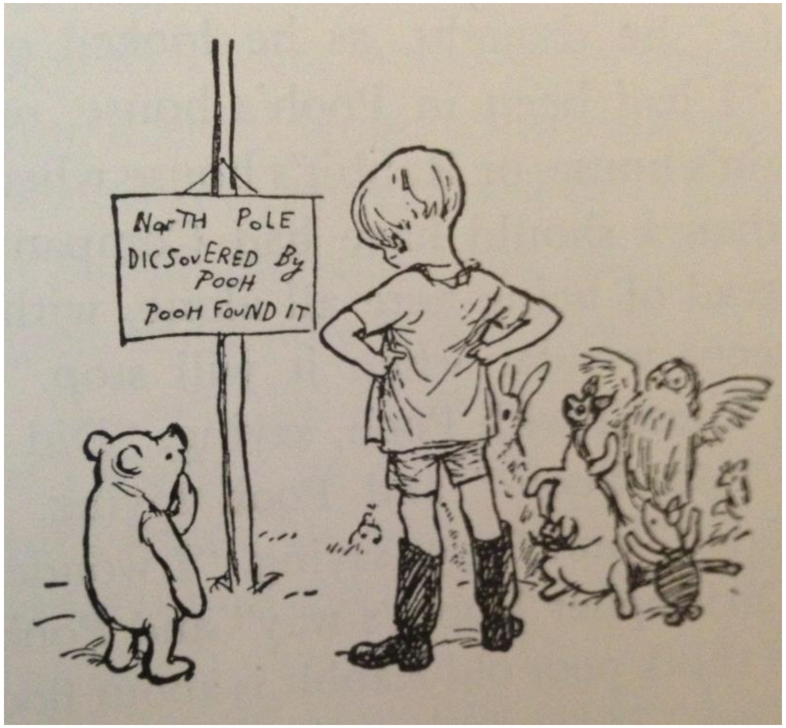
Problem 2: Stochastic search does not make use of knowledge and abilities we seem to have.

- We know a lot about our problems, well before we can solve them.
- Abstract representation of what the solution might look like could help guide searching the space.

Issues with Stochastic Search

Problem 2: Stochastic search does not make use of knowledge and abilities we seem to have.

“Sure to be a pole,” said Rabbit, “because of calling it a pole, and if it’s a pole, well, I should think it would be sticking in the ground, shouldn’t you, because there’d be nowhere else to stick it.”



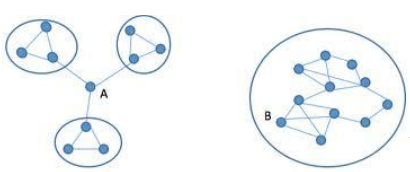
Issues with Stochastic Search

Problem 2: Stochastic search does not make use of knowledge and abilities we seem to have.

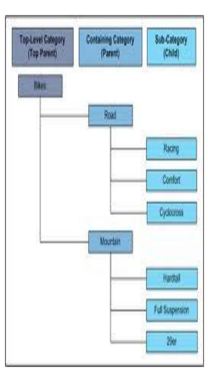
Form of the problem as **input** to algorithm should increase the probability that it proposes useful ideas

Consider the information contained in question words

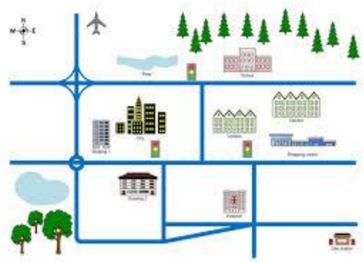
Who?



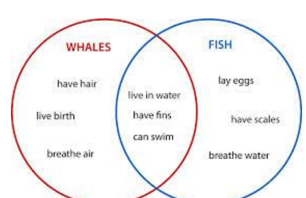
What?



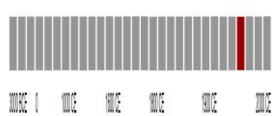
Where?



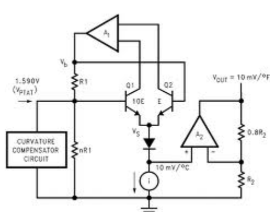
Which?



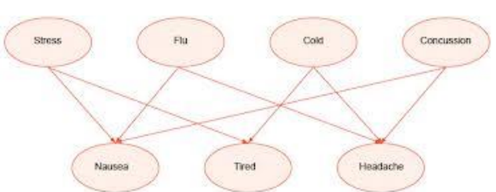
When?



How?



Why?

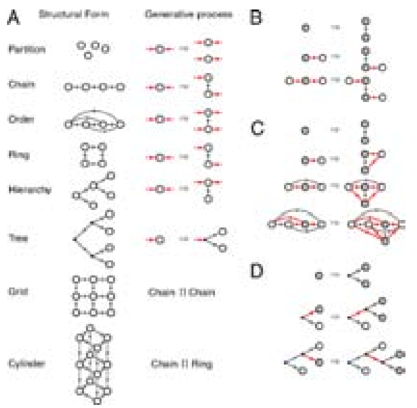


Issues with Stochastic Search

Problem 2: Stochastic search does not make use of knowledge and abilities we seem to have.

Models use abstract form to evaluate hypotheses (Kemp & Tenenbaum 2008)

BUT representation of the problem could also constrain space



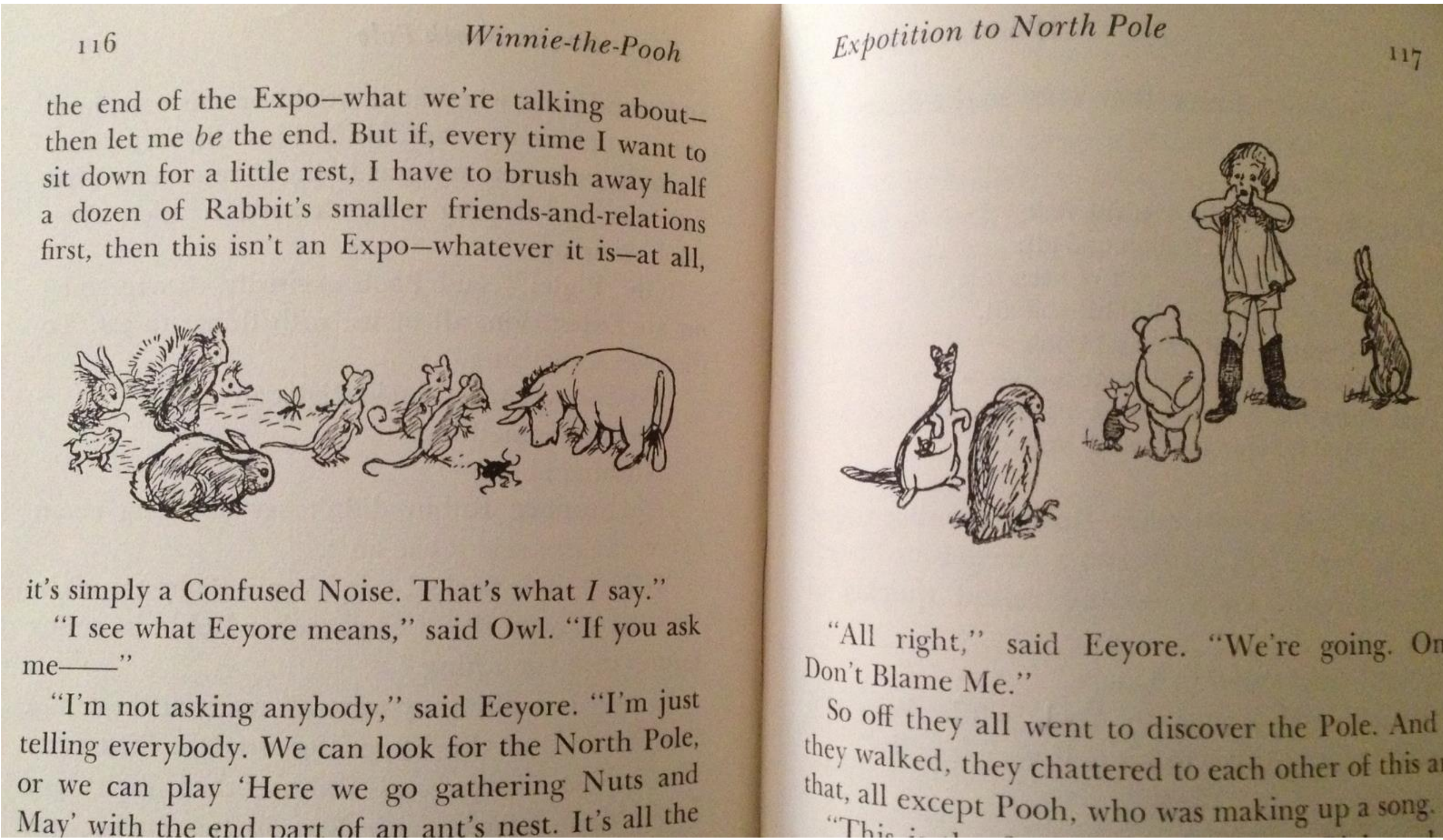
Learners have rich constraints far beyond question words.

Kinds of problems & criteria for solving them derive from multiple sources:

- The kinds of problems we want to solve (e.g., navigation, explanation, etc.)
- Broader epistemic ends (persuading, instructing, deceiving, etc.)
- Non-epistemic ends (impressing, soothing, entertaining, etc.)

Goals are innumerable, ways to achieve goals are limited

Proposal: Goal-oriented hypothesis generation


116 Winnie-the-Pooh

the end of the Expo—what we're talking about—then let me *be* the end. But if, every time I want to sit down for a little rest, I have to brush away half a dozen of Rabbit's smaller friends-and-relations first, then this isn't an Expo—whatever it is—at all, it's simply a Confused Noise. That's what *I* say."

"I see what Eeyore means," said Owl. "If you ask me——"

"I'm not asking anybody," said Eeyore. "I'm just telling everybody. We can look for the North Pole, or we can play 'Here we go gathering Nuts and May' with the end part of an ant's nest. It's all the

Expotition to North Pole 117

"All right," said Eeyore. "We're going. Onl Don't Blame Me."

So off they all went to discover the Pole. And a they walked, they chattered to each other of this an that, all except Pooh, who was making up a song.

Proposal: Goal-oriented hypothesis generation

Goal-oriented hypothesis generation

When we do not have an abstract representation of what might count as a solution to a problem we resort to very inefficient and often ineffective searches.

- Indeed, what it might mean for us to think that a problem is “tractable” might be to recognize that we don’t know the answer
- but we at least have a precise enough representation of the problem to guide the search.

Goal-oriented hypothesis generation

Representing what “counts” as a solution to a problem might explain….

- Sense of “being on the right track”

- “Great idea!”, even when we know it is wrong

Can constrain proposals based on how well….

- They fit prior knowledge & data [“TRUTH”]
- They solve problems if they were true [“TRUTHINESS”]

What does it mean to think of a new idea?

- Generating new ideas is not about radical concept/theory change
- It is the problem of ordinary, everyday, productive thinking
- Can reliably make up new – relevant – answers to any *ad hoc* question. Answers may be trivial and may be false, but they are…
  - Genuinely new (didn't have them until we thought of them)
  - Genuinely made up (didn't learn them from new evidence/testimony)
  - Answers to the question (not non-sequiters)
- Only possible if we can use the form of the question to guide search

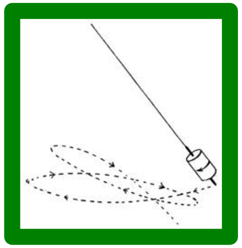
What does it mean to think of a new idea?

What's a good name for a new theater company?

How do they get the stripes on peppermints?

Fresh ink

Asaccharolyticus



Goal-oriented hypothesis generation

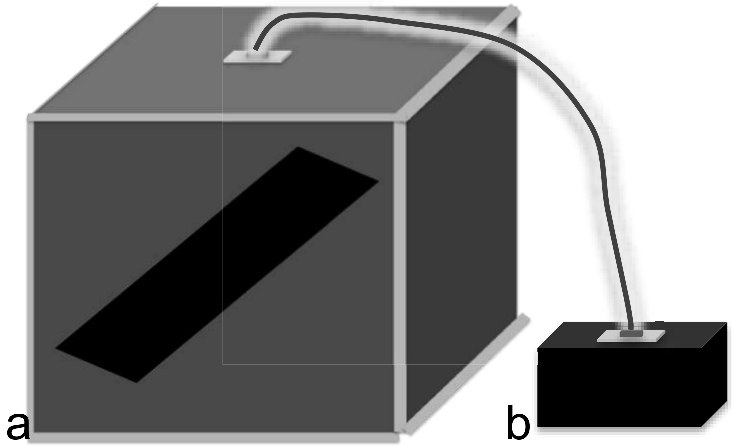
Is there any evidence that information contained only in the abstract form of the problem can help learners converge on solutions? (“Look Ma. No data.”)

Rachel Magid

Mark Sheskin

Goal-oriented hypothesis generation

Is there any evidence that information contained only in the abstract form of the problem can help learners converge on solutions? ("Look Ma. No data.")



Two visual effects

Continuous: ball flowing up and down.
Discrete: ball appearing at the bottom, disappearing, and then appearing at top

Two auditory effects

Continuous: low tone (225 Hz) gradually rising in pitch to high tone (900 Hz) and back
Discrete: low tone (225 Hz) alternating with high tone (900 Hz)

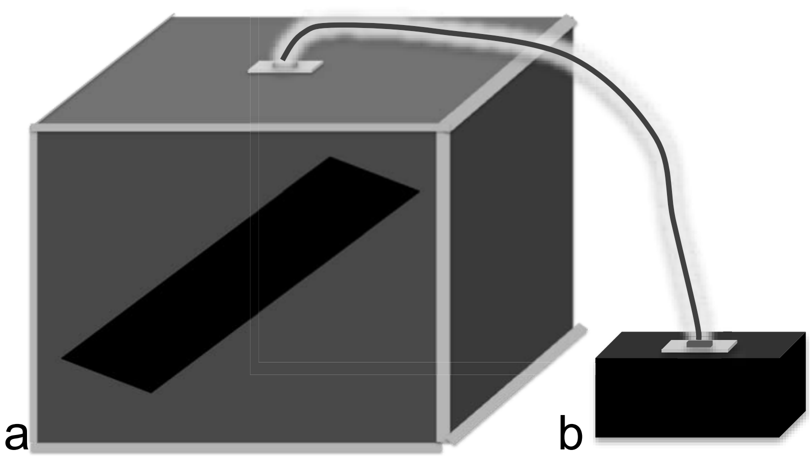
Experiment 1

- Do you see the ball? It's going low, high, low. I'm using one of these parts to make the ball go low, high, low.
- Do you see the ball? It's going higher and lower. I'm using one of these parts to make the ball go higher and lower.

" Which part made the ball go_______?"

Half the children asked about continuous visual and discrete auditory

Half asked about discrete visual and continuous auditory



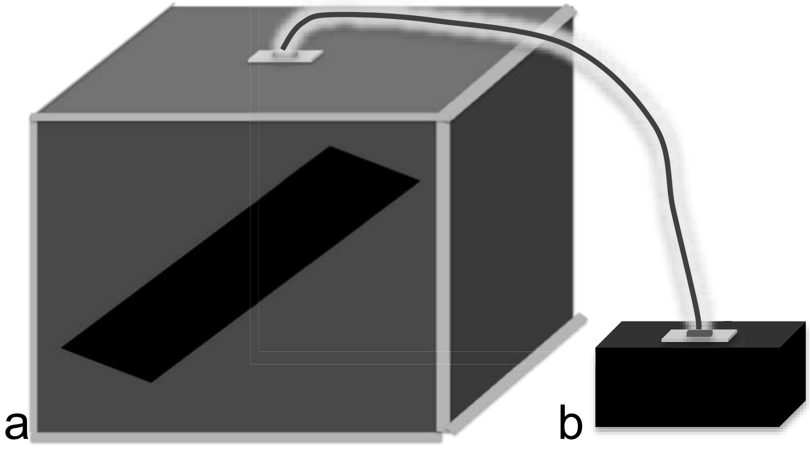
Experiment 2

- Do you see the ball? It's gazzing. I'm using one of these parts to make the machine gazz.
- Do you see the ball? It's blicking. I'm using one of these parts to make the machine blick.

" Which part made the machine _______?"

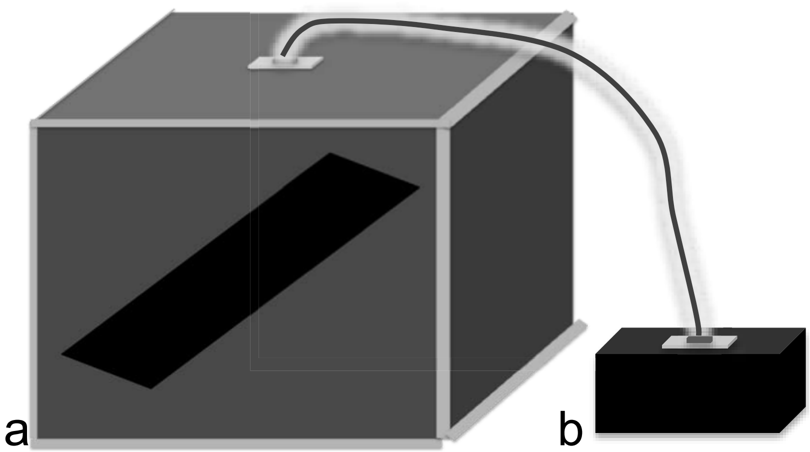
Half the children asked about continuous visual and discrete auditory

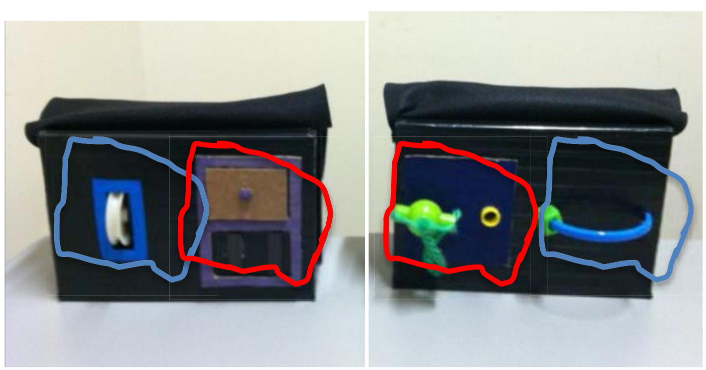
Half asked about discrete visual and continuous auditory



Goal-oriented hypothesis generation

- No fact of the matter. No covariation evidence.





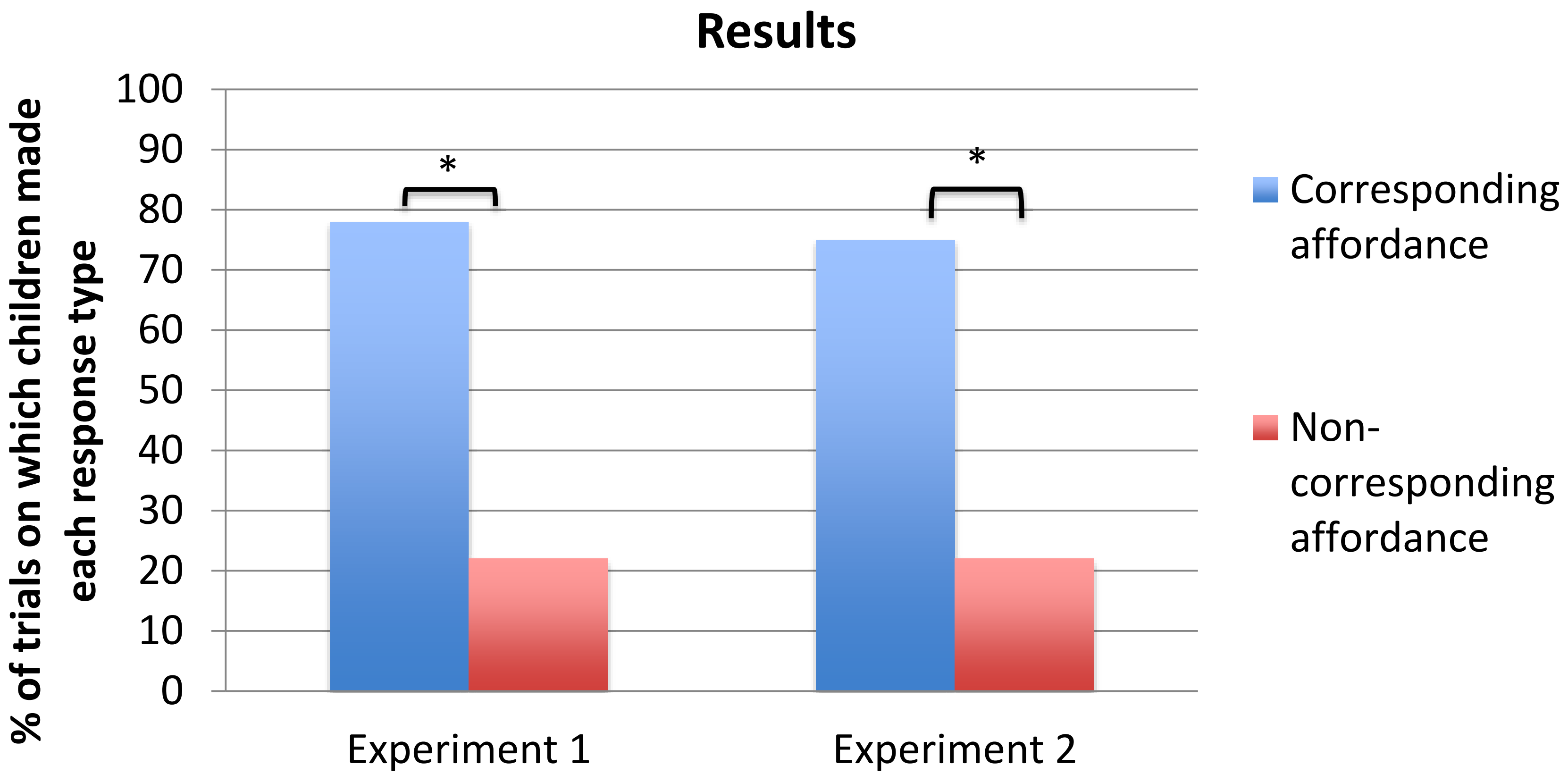
Two visual stimuli

Continuous: ball flowing up and down.
Discrete: ball appearing at the bottom, disappearing, and then appearing at top

Two auditory stimuli

Continuous: low tone (225 Hz) gradually rising in pitch to high tone (900 Hz) and back
Discrete: low tone (225 Hz) alternating with high tone (900 Hz)

Experiments 1 and 2

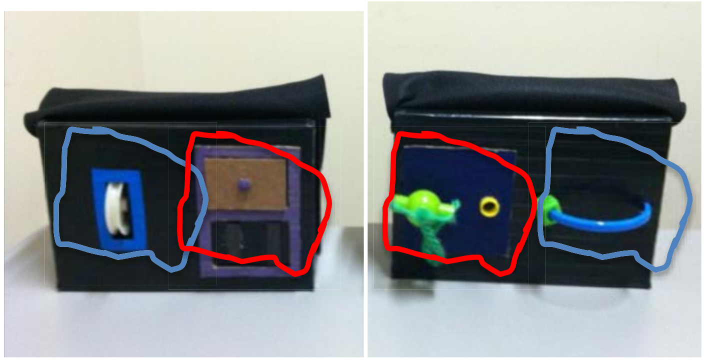


Four-six year-olds.
Mean: 62 months.
N = 16/Experiment

Magid, Sheskin, & Schulz, in press, *Cognitive Development*

Goal-oriented hypothesis generation

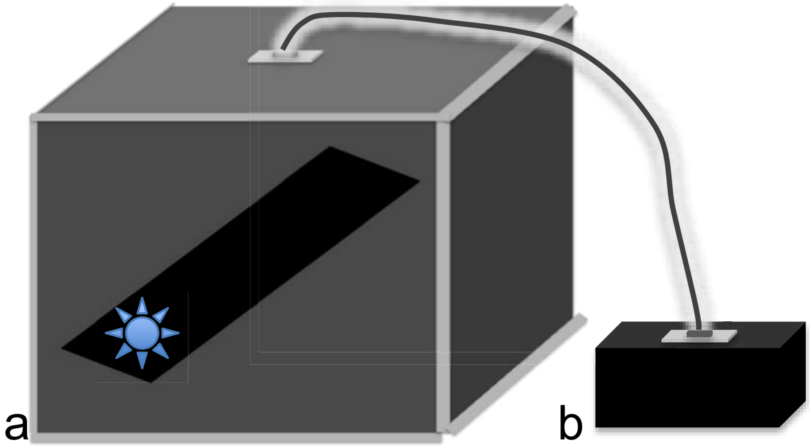
- So what? We said there was no fact of the matter and no covariation evidence.
- If children don't know the answer and there's no way to find out, maybe they just use cross-modal mapping to map from the affordance to the stimuli.
- We wanted to know if they were actually using the form of the problem to constrain the solution.
- If so, they should give different answers given different problems.

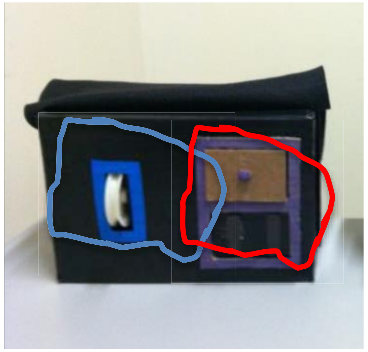


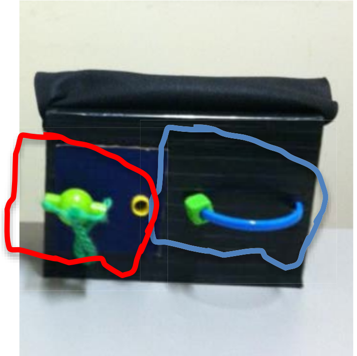
Goal-oriented hypothesis generation

- Do you see (hear) the machine? It's gazzing (flurping). I'm using one of these parts to make the machine gazz (flurp).
- Do you see (hear) the machine. It's blicking (daxing). I'm using one of these parts to make the machine blick (dax).

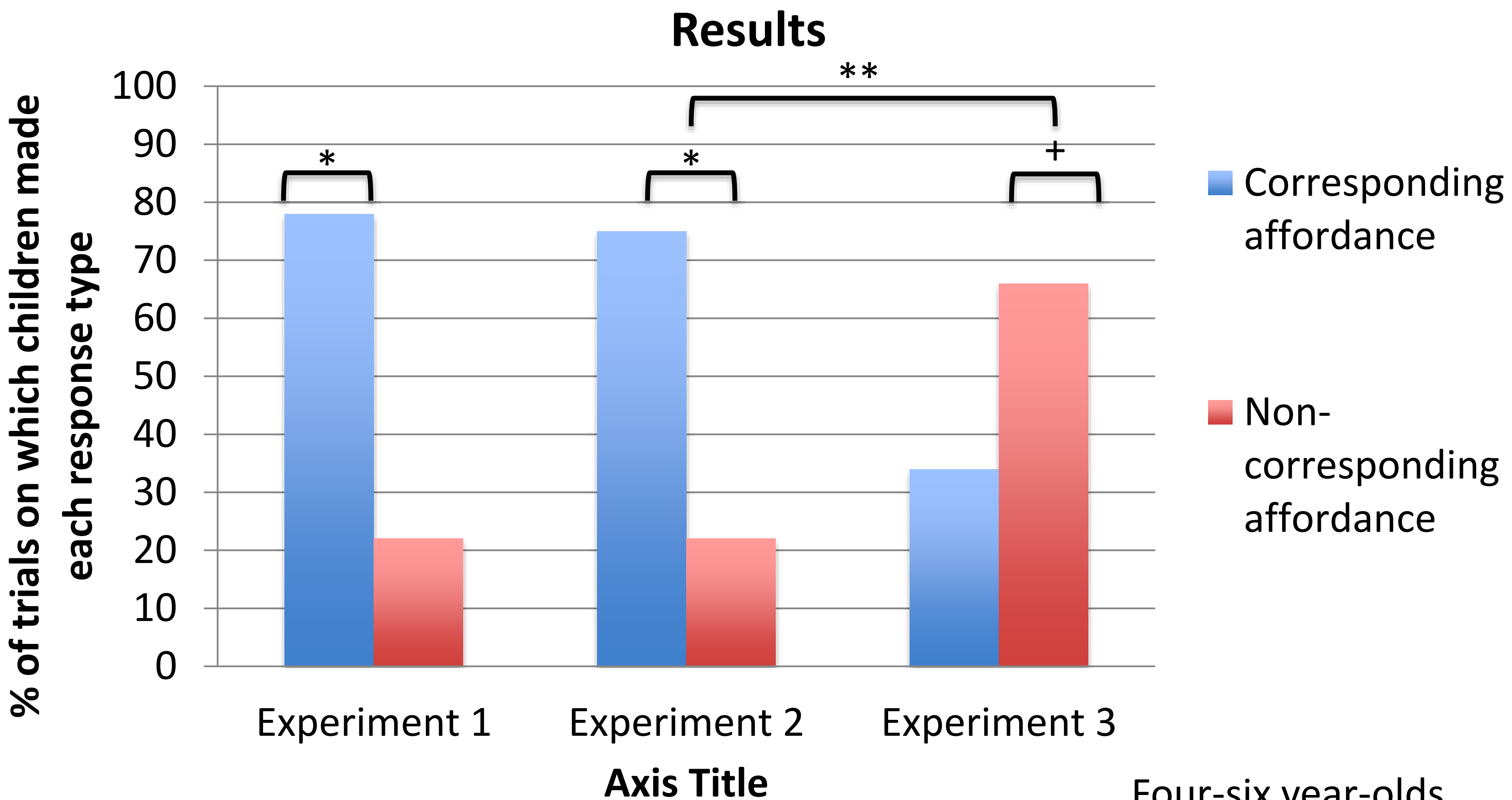
Showed the children the continuous visual stimuli and asked them how to generate the auditory one, and vice versa.







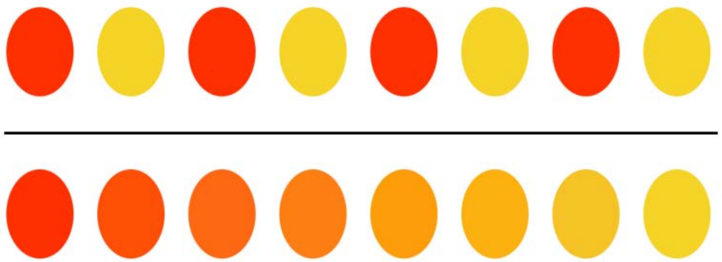
Goal-oriented hypothesis generation

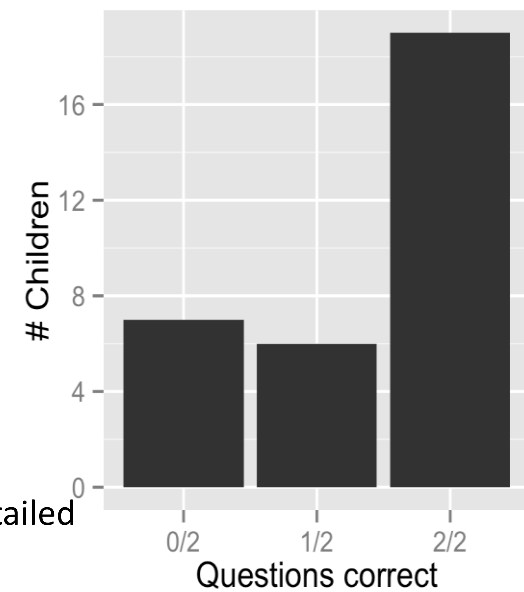


Four-six year-olds.
Mean: 62 months.
N = 16/Experiment

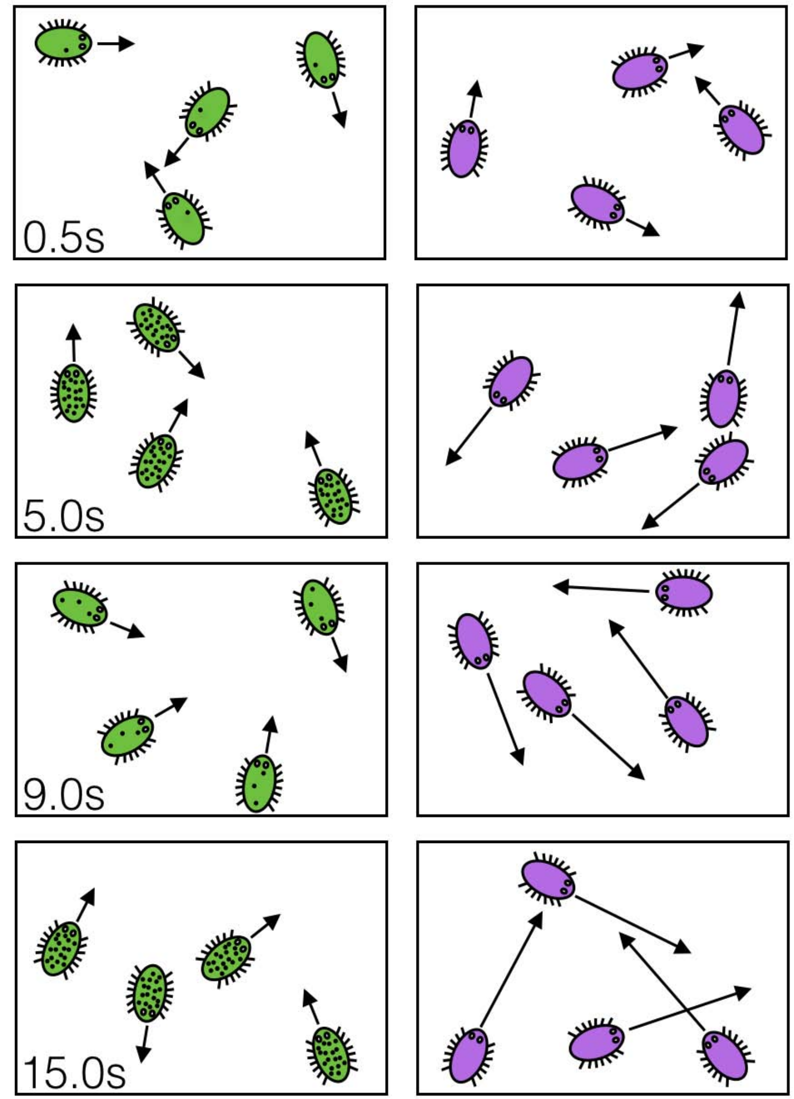
Magid, Sheskin, & Schulz, in press, *Cognitive Development*

Children prefer causal processes that preserve the dynamic form regardless of the lower level features and the absence of covariation evidence (Tsividis, Tenenbaum, & Schulz in prep)





p < .001 by two-tailed binomial test.



Goal-oriented hypothesis generation

Is this just analogical reasoning?

- Funny kind of analogy. Not a mapping between a known problem and a solution to a new problem and new solution.
- Instead a mapping between the form of the problem to the form of the solution.
- Also the argument is that this applies to any possible goal we might have, including cases where it is not obvious that analogical reasoning applies.
- “What’s a good name for a new theater company?”

Goal-oriented hypothesis generation

- Rather, children seem to have **data-independent criteria** for the evaluation of hypotheses -- criteria that extend beyond simplicity or compatibility with prior knowledge.
- Children can consider the extent to which a hypothesis fulfills the abstract goals of a solution to a problem, not just the degree to which a hypothesis fits the data.

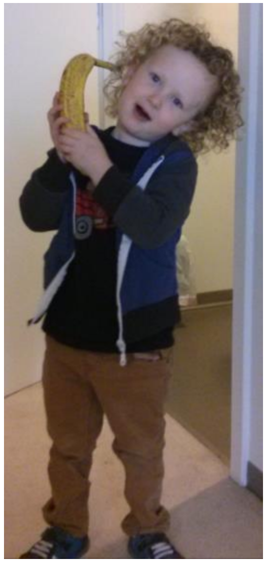
Imagination

A mystery of human cognition:

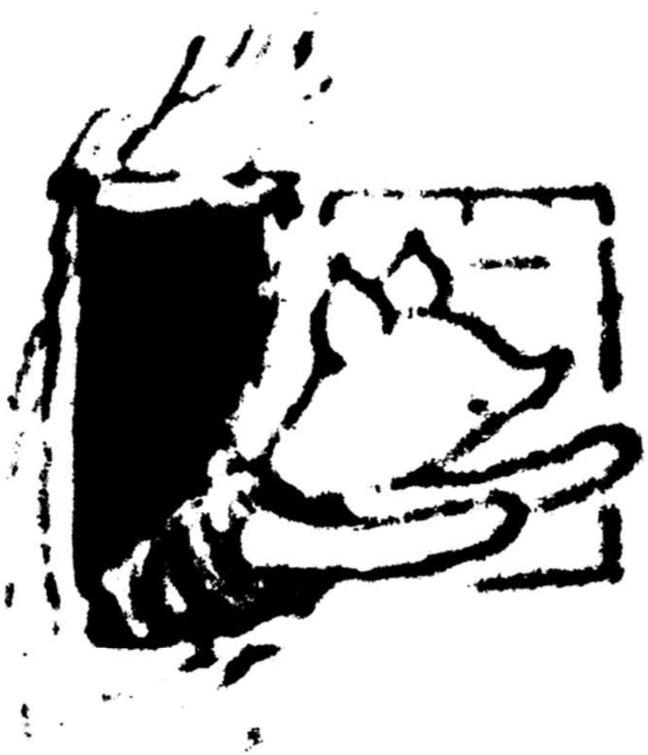
So much time pretending and making up stories

Good stories do not have to be true, BUT

Pose problems, solve problems



Goal-oriented hypothesis generation and imagination



"And that is really the end of the story, and as I am very tired after that last sentence, I think I shall stop there."

Response

"Run away, run away" clip from
Monty Python and the Holy Grail
removed due to copyright restrictions.

Response

Steve Piantadosi

Fast Proposals

Owen Lewis

Good Proposals

"'Tis but a scratch" clip from *Monty Python and the Holy Grail* removed due to copyright restrictions.

Eyal Dechter

Good Primitives

Ad-hoc Spaces

Response I: Fast Proposals

Critique: "Stochastic search does not make use of, or account for, some abilities we know people have"

Rebuttal: You're wrong(?)

Many hypotheses - only aware of (relatively) good ones

Requires ability to suggest many hypotheses

Steve Piantadosi

Response I: Fast Proposals

Stochastic search algorithms can be parallelized (in some cases)

- Run many “chains” in parallel, not one chain for a long time

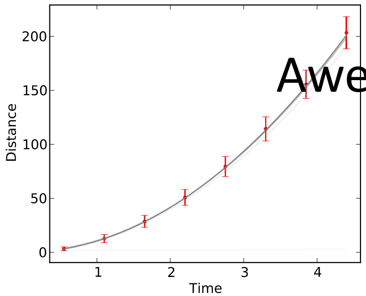
- Take advantage of GPU architecture, not CPU

~30 times faster than CPU

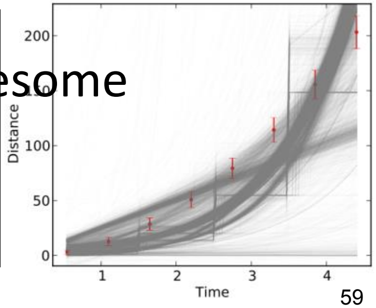
GPU’s are cheap and plentiful, search scales in number of GPU’s.

+

=

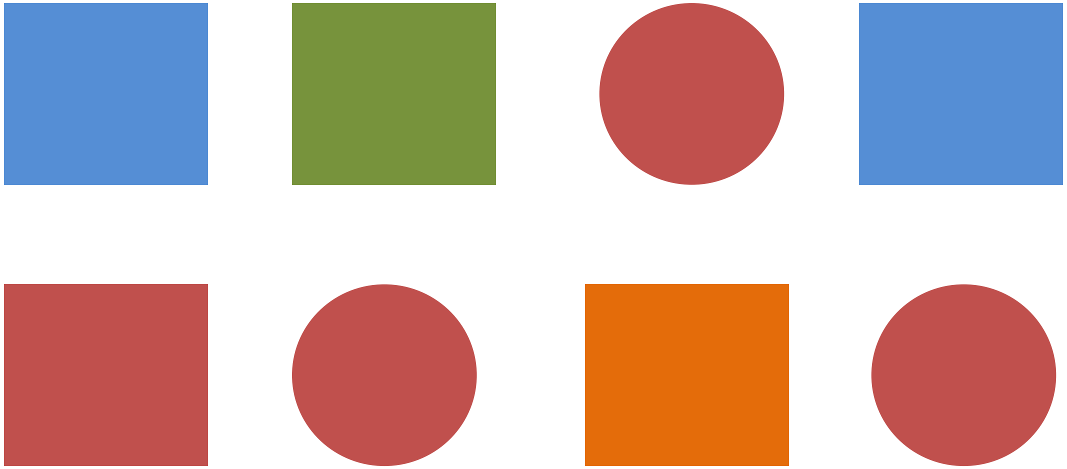


Awesome



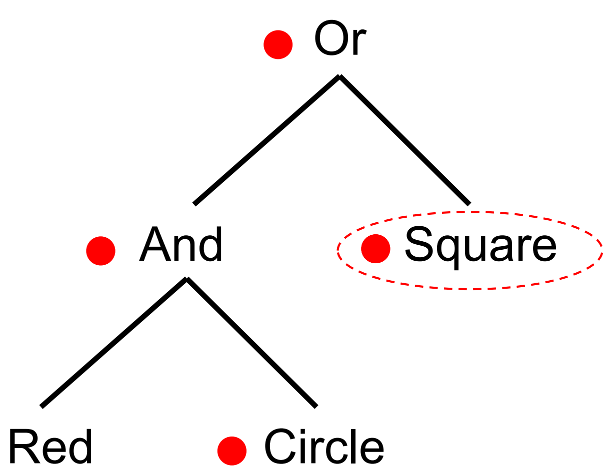
Response 2: Relevant Proposals

Owen Lewis (MIT)

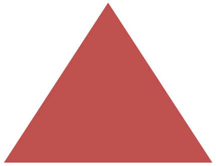


Response 1: Relevant Proposals

Current Hypothesis

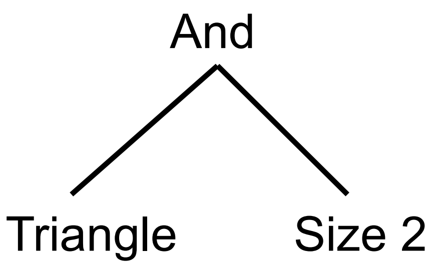


New example



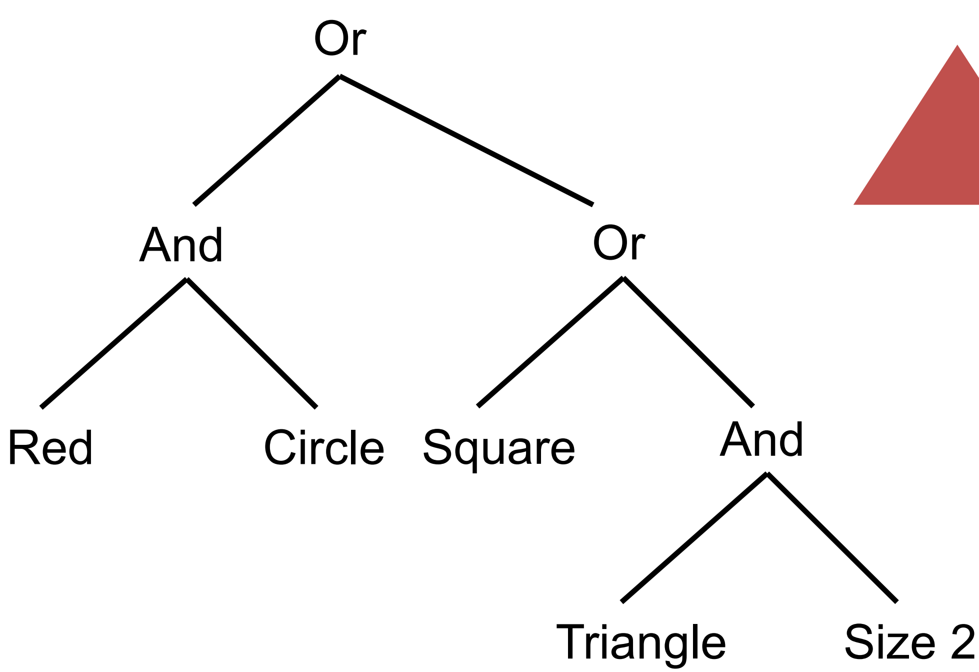
Owen Lewis (MIT)

Stochastic description



Response 2: Relevant Proposals

Current Hypothesis

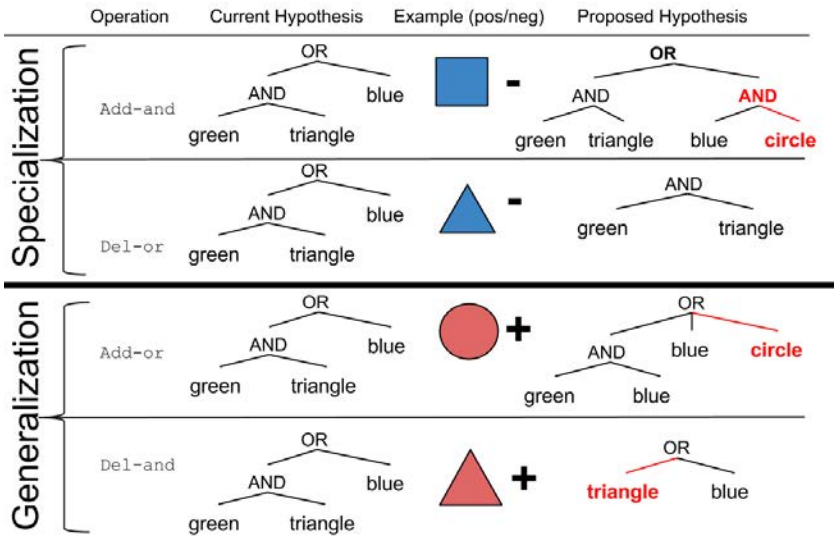


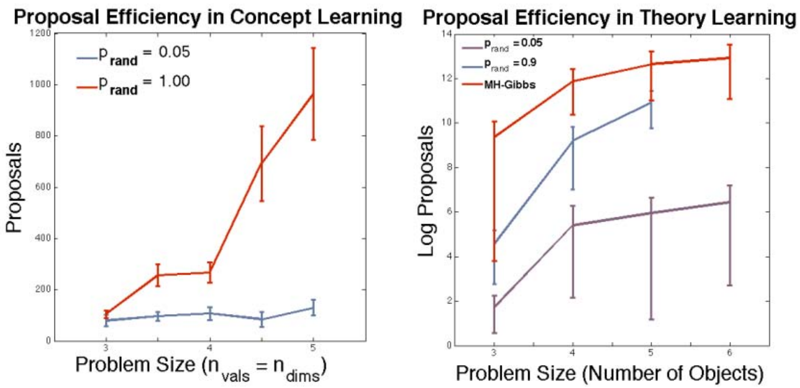
New example

Owen Lewis (MIT)

Bottom line: propose relevant hypotheses

Response 2: Relevant Proposals





Owen Lewis (MIT)

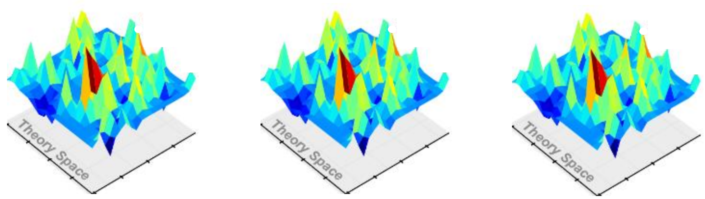
Response 3 – Good Primitives

Reminder: Templates as smart proposals

Template 1:	P(X,Y)	$\leftarrow$	P(X,Z) $\wedge$ P(Z,Y)
Template 2:	P(X,Y)	$\leftarrow$	P(Z,X) $\wedge$ P(Z,Y)
Template 3:	P(X,Y)	$\leftarrow$	P(X,Z) $\wedge$ P(Y,Z)
Template 4:	P(X,Y)	$\leftarrow$	P(Z,X) $\wedge$ P(Y,Z)
Template 5:	P(X,Y)	$\leftarrow$	P(X,Y) $\wedge$ P(X)
Template 6:	P(X,Y)	$\leftarrow$	P(Y,X) $\wedge$ P(X)
Template 7:	P(X,Y)	$\leftarrow$	P(X,Y) $\wedge$ P(Y)
Template 8:	P(X,Y)	$\leftarrow$	P(Y,X) $\wedge$ P(Y)
Template 9:	P(X,Y)	$\leftarrow$	P(X) $\wedge$ P(Y)
Template 10:	P(X,Y)	$\leftarrow$	P(Y,X)
Template 11:	P(X,Y)	$\leftarrow$	P(X,Y)
Template 12:	P(X)	$\leftarrow$	P(X)
Template 13:	P(X)	$\leftarrow$	P(X,Y) $\wedge$ P(X)
Template 14:	P(X)	$\leftarrow$	P(Y,X) $\wedge$ P(X)
Template 15:	P(X)	$\leftarrow$	P(X,Y) $\wedge$ P(Y)
Template 16:	P(X)	$\leftarrow$	P(Y,X) $\wedge$ P(Y)

e.g. Transitivity

Useful when there are multiple problems:



• • •

BUT: Discovery of templates?

Response 3 – Good Primitives

Exploration Compression Algorithm

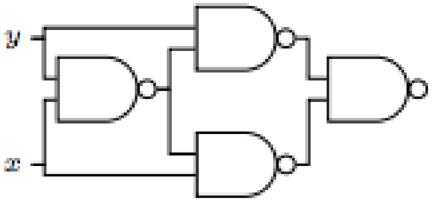
(Dechter, Malmaud, Adams & Tenenbaum 2013)

Eyal Dechter (MIT)

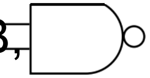
Takes stochastic grammar over programs & primitives:

Generates function library

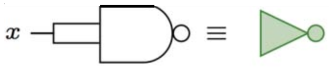
Library 'encapsulates' useful concepts



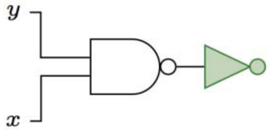
Example: Boolean circuits

Primitives: {I, S, C, B,  }

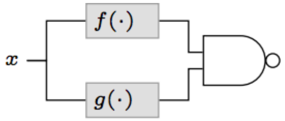
Learned concepts:



[NOT]



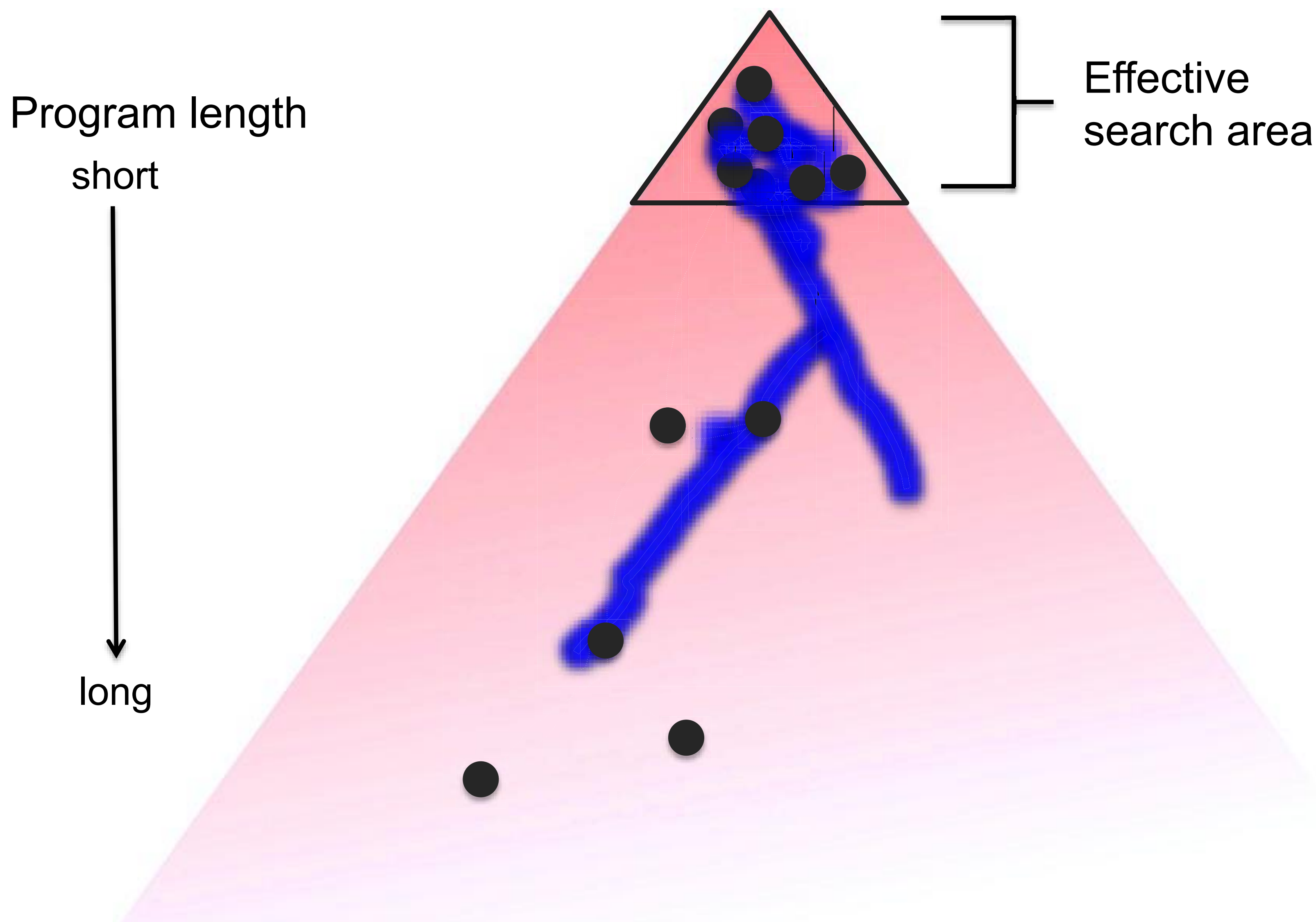
[AND]



[E2]

Bottom line: learn and re-use good 'chunks'

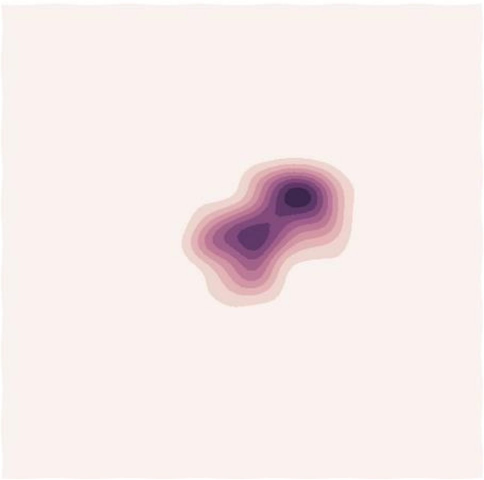
Response 2 – Good Primitives

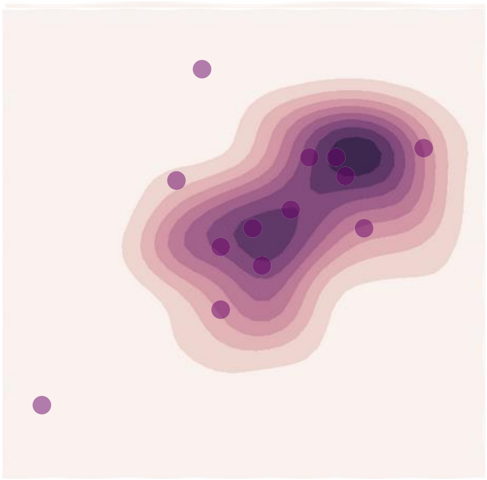


Response 4 – Relevant Spaces

Construct relevant spaces on the fly

“Good name for new romantic drama”





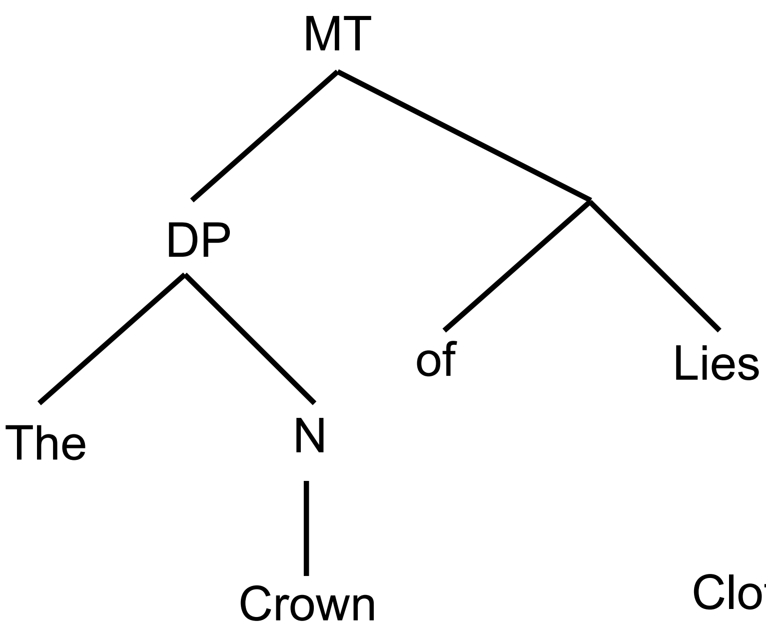
“Give me a paper title for SRCD”

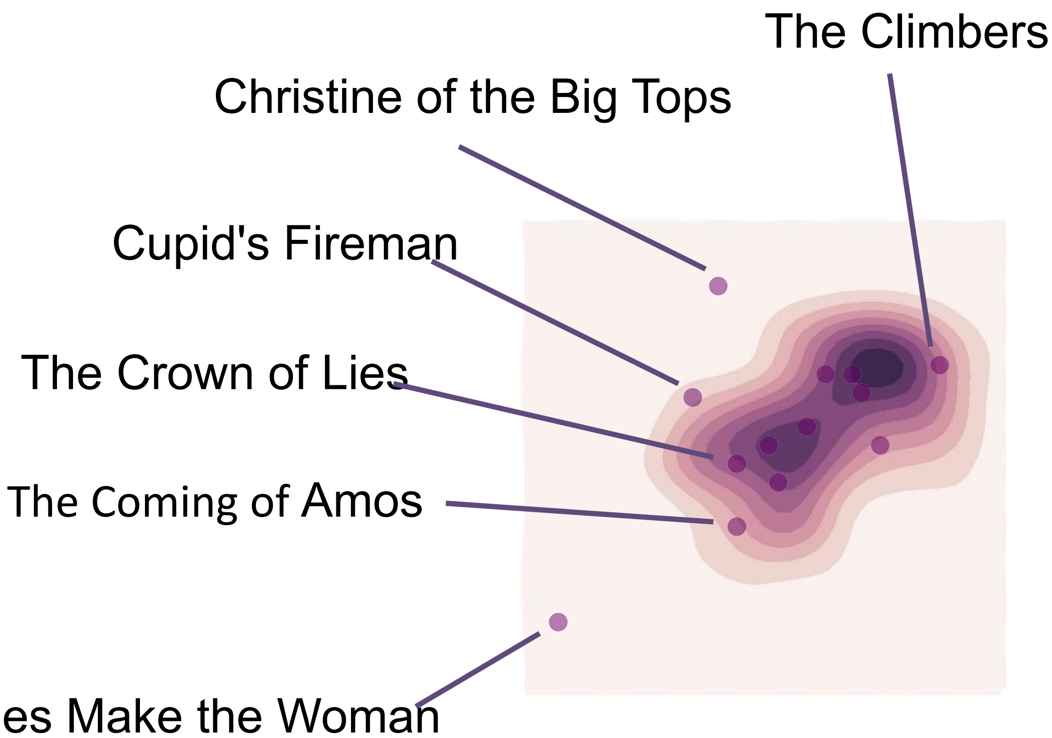
Max Siegel (MIT)

Response 4 – Relevant Spaces

"Good name for new romantic drama"

Mini-Grammar





Response 4 – Relevant Spaces

“Good name for new romantic drama”

Those we Meet Again

Essence of Time

Value of Love

Love Lightly

Girls in Ships

Hunchback of Monte Cristo

Land of Roses

Get it Did

Endless Love

Legend of Paris

Belle of a Lesser God

“Good name for new action movie”

Swordsman in China III

Hu: the Annihilation

The Oversight

Eagle Shooting Heroes

Jack Death

Tomb Raider: the Raging God of Violence

The Chase

The Hit

Legend of Legend

The Edge

Among Heroes

Summary

Still a long way to go to model children,
meet Laura's critique

Hard to say what is hard (early days)

People in development should (continue) to care
about search algorithms, to everyone's benefit

Goal-oriented hypothesis generation and imagination

Very cool. Error-driven proposals. But still driven by the data. We seem to treat the problem itself as part of the "data".

Also very cool. Explains how you develop new representational resources. But not all learning problems can be solved just by changing the representational format

Might be true. "That's what an expedition means. A long line of everybody."
But … not as good a story.

Questions?