Growing domain-specific languages alongside neural program synthesizers via wake-sleep program learning

Kevin Ellis
Collaborators: Catherine Wong, Maxwell Nye, Mathias Sablé-Meyer, Lucas Morales, Armando Solar-Lezama, Joshua B. Tenenbaum

2020

neurosym webinar

The premise of program induction

1. Represent knowledge as programs: as symbolic code

The premise of program induction

1. Represent knowledge as programs: as symbolic code

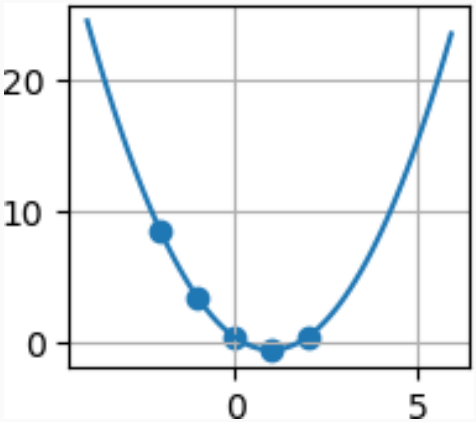
2. Learning=adding to that body of knowledge=
   making new programs=program synthesis

Why program induction?

Why program induction?

strong generalization

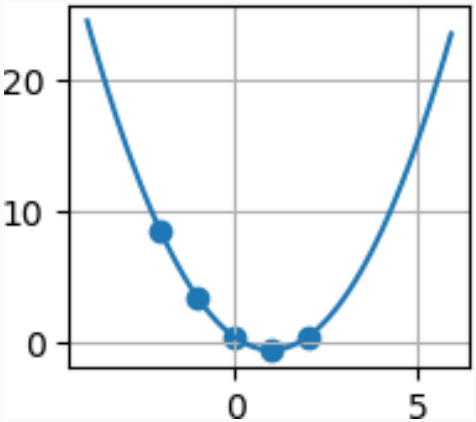
+data efficiency




```
f(x)=(x-1)**2 - 0.5
```

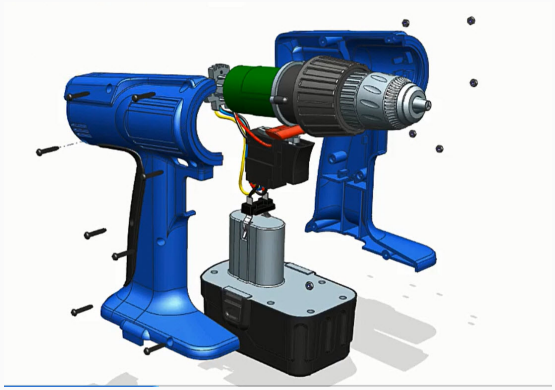
Why program induction?

strong generalization
+data efficiency

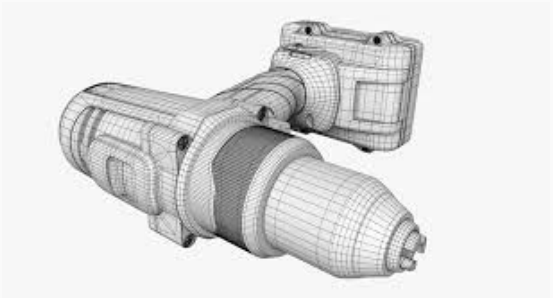



```
f(x)=(x-1)**2 - 0.5
```

interpretability

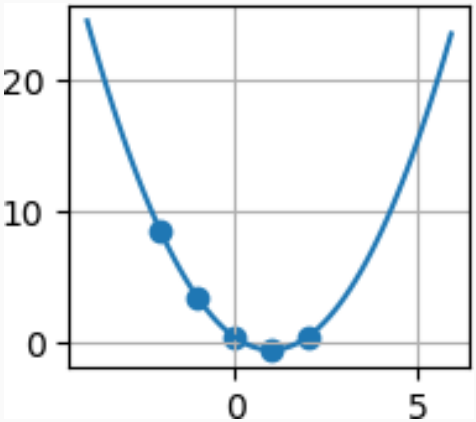


vs



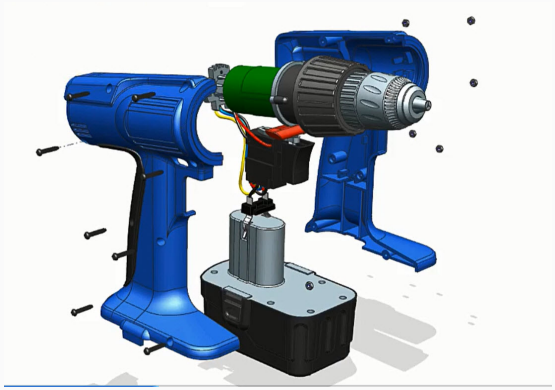
Why program induction?

strong generalization
+data efficiency

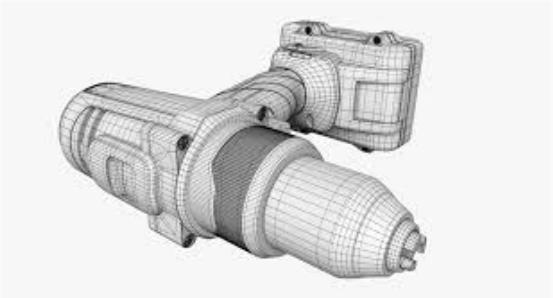



```
f(x)=(x-1)**2 - 0.5
```

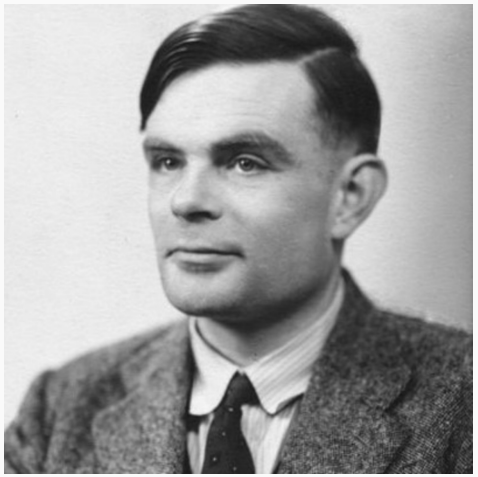
interpretability



vs



universal expressivity



FlashFill (Gulwani 2012)

EXAMPLE 3 (Directory Name Extraction). *Consider the following example taken from an excel online help forum.*

Input v_1	*Output*
Company\Code\index.html	*Company\Code*
Company\Docs\Spec\specs.doc	*Company\Docs\Spec*

String Program:

$\texttt{SubStr}(v_1, \texttt{CPos}(0), \texttt{Pos}(\textit{SlashTok}, \epsilon, -1))$

FlashFill (Gulwani 2012)

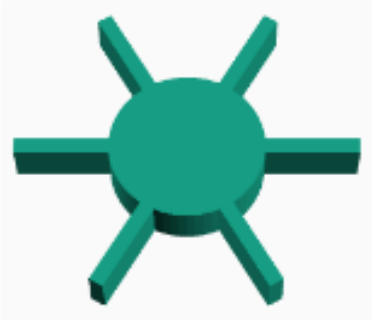
EXAMPLE 3 (Directory Name Extraction). *Consider the following example taken from an excel online help forum.*

Input v_1	*Output*
Company\Code\index.html	*Company\Code*
Company\Docs\Spec\specs.doc	*Company\Docs\Spec*

String Program:

$\texttt{SubStr}(v_1, \texttt{CPos}(0), \texttt{Pos}(\mathit{SlashTok}, \epsilon, -1))$

Szalinski (Nandi 2020)



(a) CAD model of ship's wheel

```
(Union
 (Cylinder [1, 5, 5])
 (Fold Union
  (Tabulate (i 6)
   (Rotate [0, 0, 60i]
    (Translate [1, −0.5, 0]
     (Cuboid [10, 1, 1])))))))
```

(b) Caddy program

FlashFill (Gulwani 2012)

EXAMPLE 3 (Directory Name Extraction). *Consider the following example taken from an excel online help forum.*

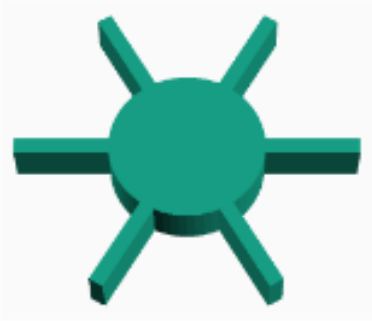
Input v_1	*Output*
Company\Code\index.html	*Company\Code*
Company\Docs\Spec\specs.doc	*Company\Docs\Spec*

String Program:

$\texttt{SubStr}(v_1, \texttt{CPos}(0), \texttt{Pos}(\textit{SlashTok}, \epsilon, -1))$

$$
\begin{array}{rcl}
\text{String expr } P & := & \texttt{Switch}((\mathrm{b}_1, \mathrm{e}_1), \cdot\cdot, (\mathrm{b}_n, \mathrm{e}_n)) \\
\text{Bool b} & := & \mathrm{d}_1 \vee \cdot\cdot \vee \mathrm{d}_n \\
\text{Conjunct d} & := & \pi_1 \wedge \cdot\cdot \wedge \pi_n \\
\text{Predicate } \pi & := & \texttt{Match}(v_i, \mathrm{r}, k) \mid \neg\, \texttt{Match}(v_i, \mathrm{r}, k) \\
\text{Trace expr e} & := & \texttt{Concatenate}(\mathrm{f}_1, \cdot\cdot, \mathrm{f}_n) \\
\text{Atomic expr f} & := & \texttt{SubStr}(v_i, \mathrm{p}_1, \mathrm{p}_2) \\
 & \mid & \texttt{ConstStr}(s) \\
 & \mid & \texttt{Loop}(\lambda w : \mathrm{e}) \\
\text{Position p} & := & \texttt{CPos}(k) \mid \texttt{Pos}(\mathrm{r}_1, \mathrm{r}_2, \mathrm{c}) \\
\text{Integer expr c} & := & k \mid k_1 w + k_2 \\
\text{Regular Expression r} & := & \texttt{TokenSeq}(\mathrm{T}_1, \cdot\cdot, \mathrm{T}_m) \\
\text{Token T} & := & C+ \mid [\neg C]+ \\
 & \mid & \text{SpecialToken}
\end{array}
$$

Szalinski (Nandi 2020)



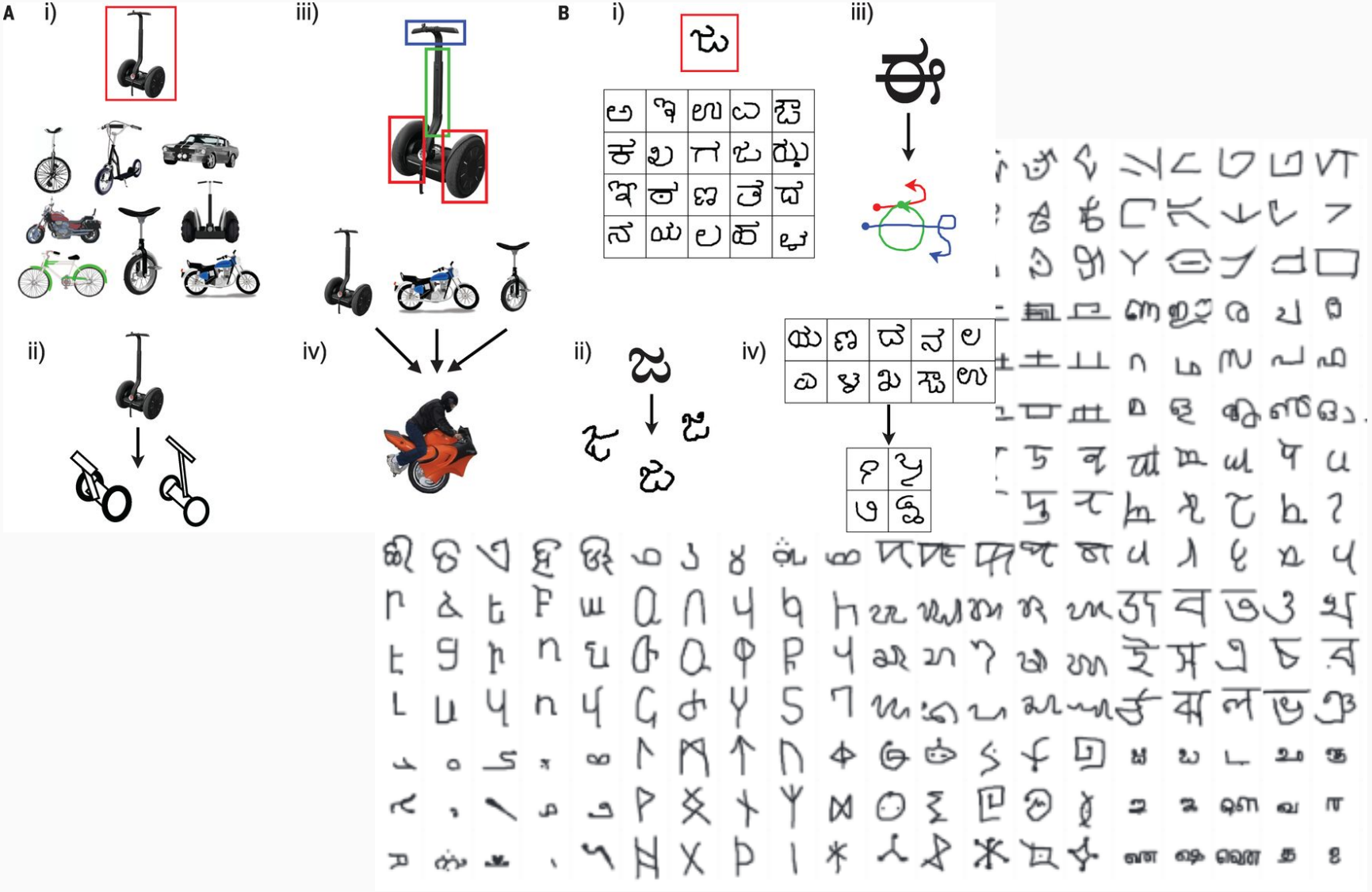
(a) CAD model of ship's wheel

```
(Union
 (Cylinder [1, 5, 5])
 (Fold Union
  (Tabulate (i 6)
   (Rotate [0, 0, 60i]
    (Translate [1, −0.5, 0]
     (Cuboid [10, 1, 1])))))))
```

(b) Caddy program

op	::=	+ | - | × | /
num	::=	$\mathbb{R}$ | ⟨*var*⟩ | ⟨*num*⟩ ⟨*op*⟩ ⟨*num*⟩
vec2	::=	[⟨*num*⟩, ⟨*num*⟩]
vec3	::=	[⟨*num*⟩, ⟨*num*⟩, ⟨*num*⟩]
affine	::=	Translate | Rotate | Scale | **TranslateSpherical**
binop	::=	Union | Difference | Intersection
cad	::=	(Cuboid ⟨*vec3*⟩) | (Sphere ⟨*num*⟩)
	|	(Cylinder ⟨*vec2*⟩) | (HexPrism ⟨*vec2*⟩) | …
	|	(⟨*affine*⟩ ⟨*vec3*⟩ ⟨*cad*⟩)
	|	(⟨*binop*⟩ ⟨*cad*⟩ ⟨*cad*⟩)
	|	**(Fold** ⟨*binop*⟩ ⟨*cad-list*⟩**)**
cad-list	::=	(List ⟨*cad*⟩+)
	|	(Concat ⟨*cad-list*⟩+)
	|	(Tabulate (⟨*var*⟩ $\mathbb{Z}^+$)+ ⟨*cad*⟩)
	|	(Map2 ⟨*affine*⟩ ⟨*vec3-list*⟩ ⟨*cad-list*⟩)
vec3-list	::=	(List ⟨*vec3*⟩+)
	|	(Concat ⟨*vec3-list*⟩+)
	|	(Tabulate (⟨*var*⟩ $\mathbb{Z}^+$)+ ⟨*vec3*⟩)

Program induction for learning and perception



Lake et al 2015

Program Induction and learning to learn

learning a DSL

learning to synthesize

synergy between DSL+learned synthesizer

Learning to write code

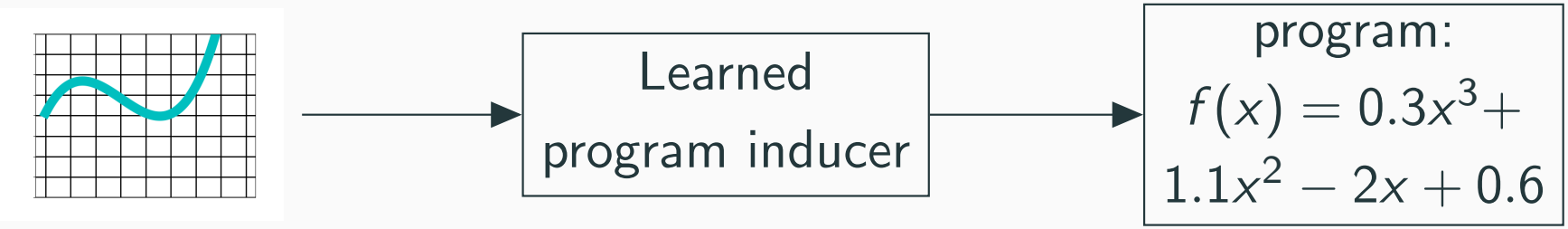
Goal: acquire domain-specific knowledge needed to induce a class of programs

- Library of abstractions (domain specific language)
- Inference strategy (synthesis algorithm)

Learning to write code

Goal: acquire domain-specific knowledge needed to induce a class of programs

- Library of abstractions (domain specific language)
- Inference strategy (synthesis algorithm)



Concepts: x^3, $\alpha x + \beta$, etc

Inference strategy: neurosymbolic search for programs

Library learning

Initial Primitives

Sample Problem: `sort list`

[9 2 7 1] → [1 2 7 9]
[3 8 9 4 2] → [2 3 4 8 9]
[6 2 2 3 8 5] → [2 2 3 5 6 8]
...

Ellis, Morales, Sable-Meyer, Solar-Lezama, Tenenbaum. NeurIPS 2018.

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Library learning

Initial Primitives

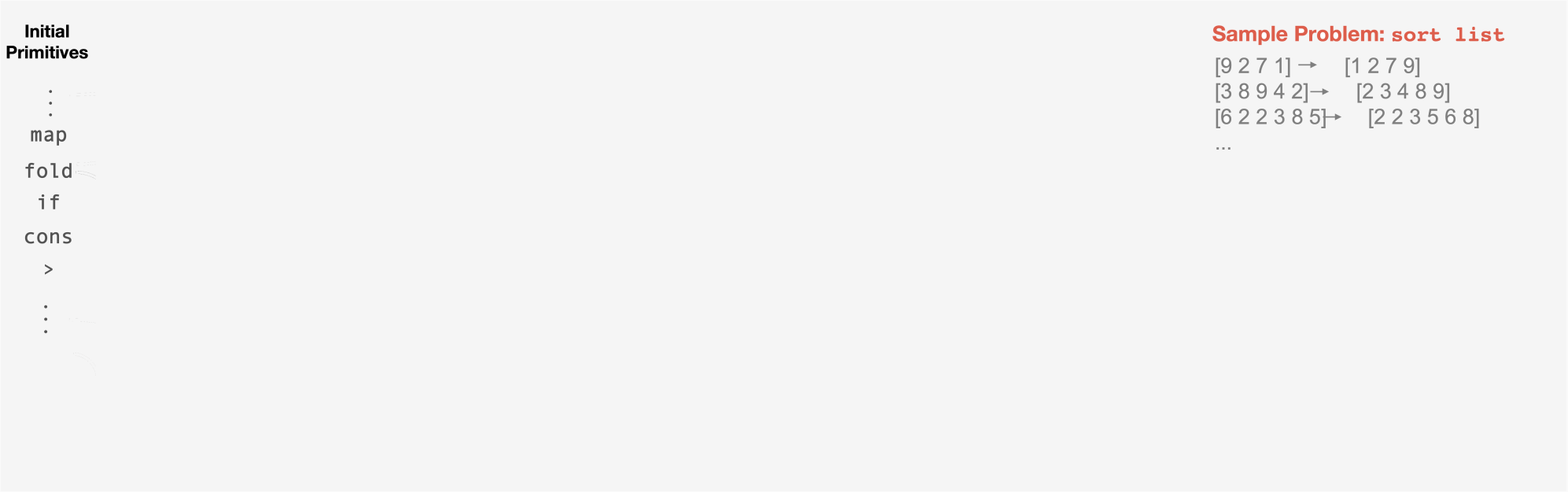
```
  :
 map
fold
 if
cons
  >
  :
```

Sample Problem: `sort list`

[9 2 7 1] → [1 2 7 9]
[3 8 9 4 2] → [2 3 4 8 9]
[6 2 2 3 8 5] → [2 2 3 5 6 8]
...

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Library learning



Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

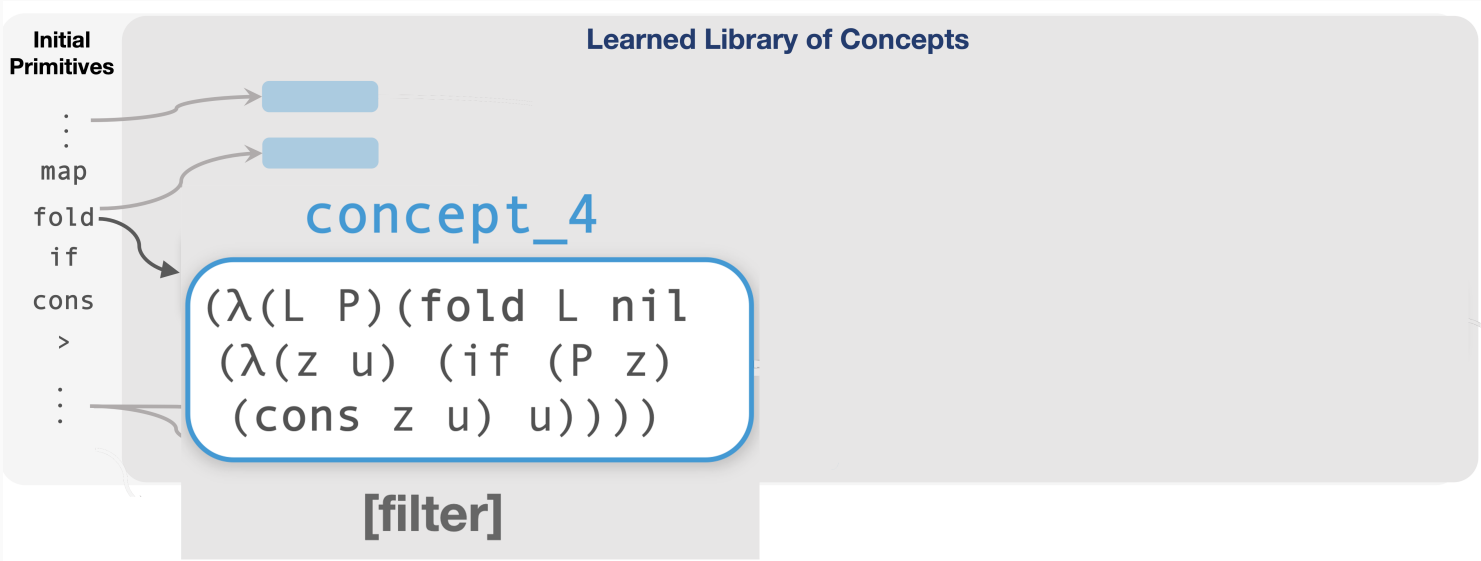
Library learning

Sample Problem: `sort list`

[9 2 7 1] → [1 2 7 9]
[3 8 9 4 2] → [2 3 4 8 9]
[6 2 2 3 8 5] → [2 2 3 5 6 8]
...

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Library learning



Sample Problem: `sort list`

[9 2 7 1] → [1 2 7 9]
[3 8 9 4 2] → [2 3 4 8 9]
[6 2 2 3 8 5] → [2 2 3 5 6 8]
...

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

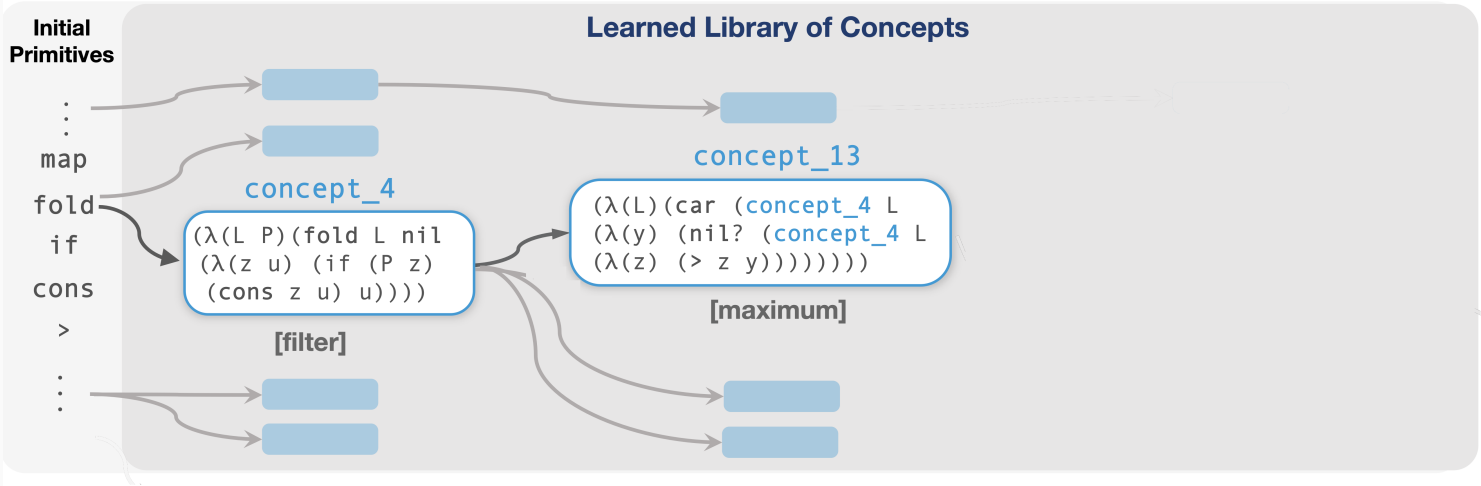
Library learning

Sample Problem: `sort list`

[9 2 7 1] → [1 2 7 9]
[3 8 9 4 2] → [2 3 4 8 9]
[6 2 2 3 8 5] → [2 2 3 5 6 8]
...

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Library learning

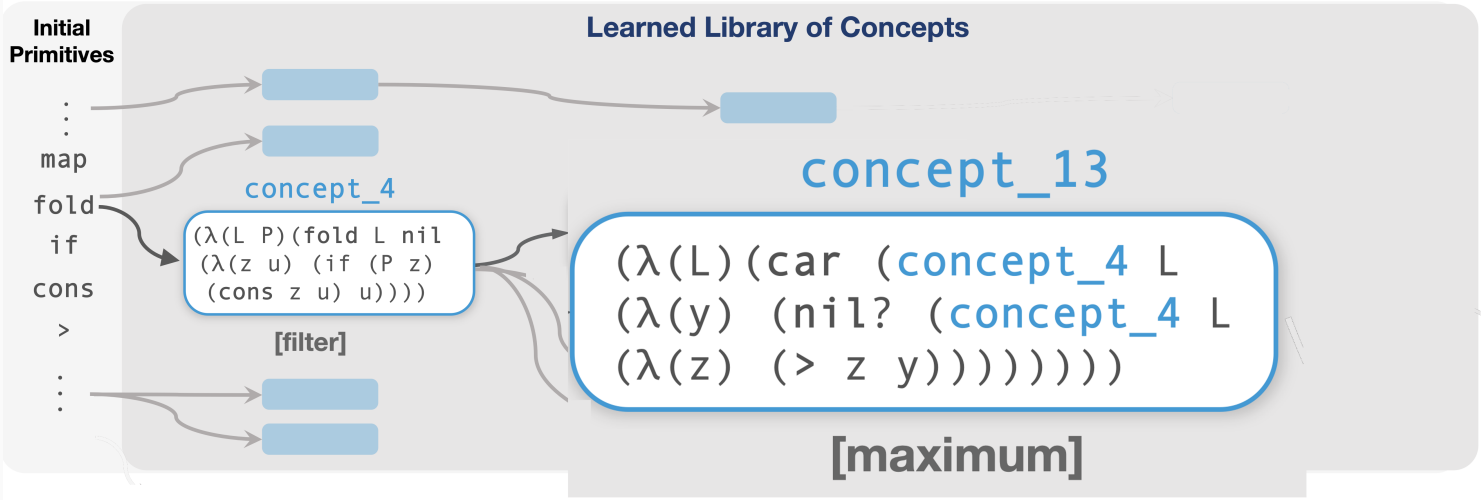


Sample Problem: `sort list`

[9 2 7 1] → [1 2 7 9]
[3 8 9 4 2] → [2 3 4 8 9]
[6 2 2 3 8 5] → [2 2 3 5 6 8]
...

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Library learning

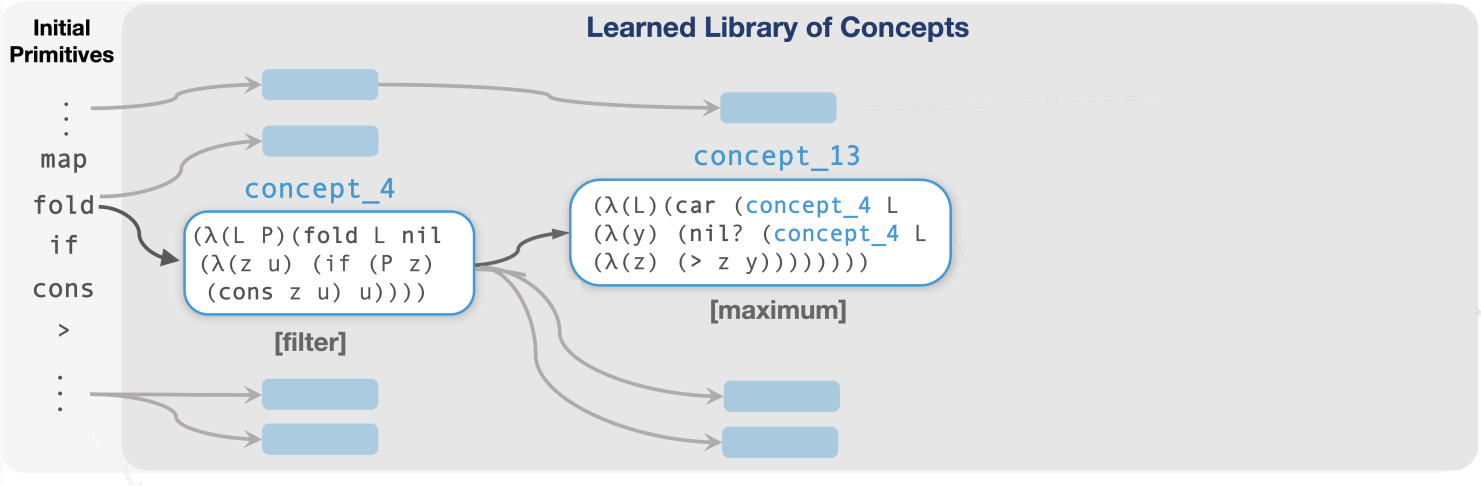


Sample Problem: `sort list`

[9 2 7 1] → [1 2 7 9]
[3 8 9 4 2] → [2 3 4 8 9]
[6 2 2 3 8 5] → [2 2 3 5 6 8]
...

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Library learning

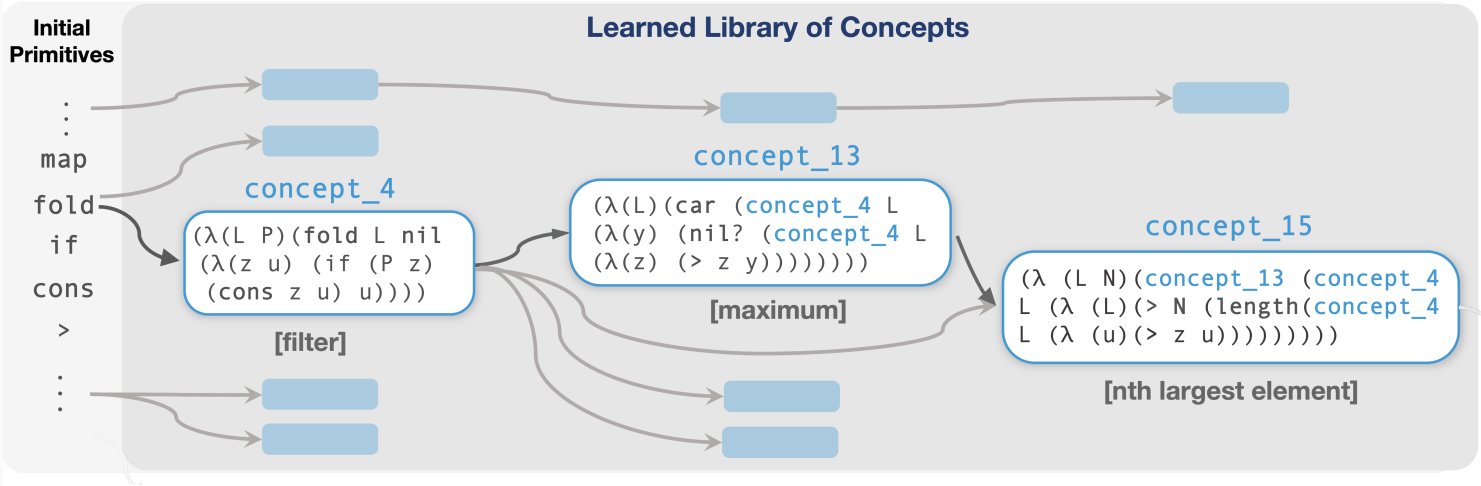


Sample Problem: `sort list`

[9 2 7 1] → [1 2 7 9]
[3 8 9 4 2] → [2 3 4 8 9]
[6 2 2 3 8 5] → [2 2 3 5 6 8]
...

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Library learning

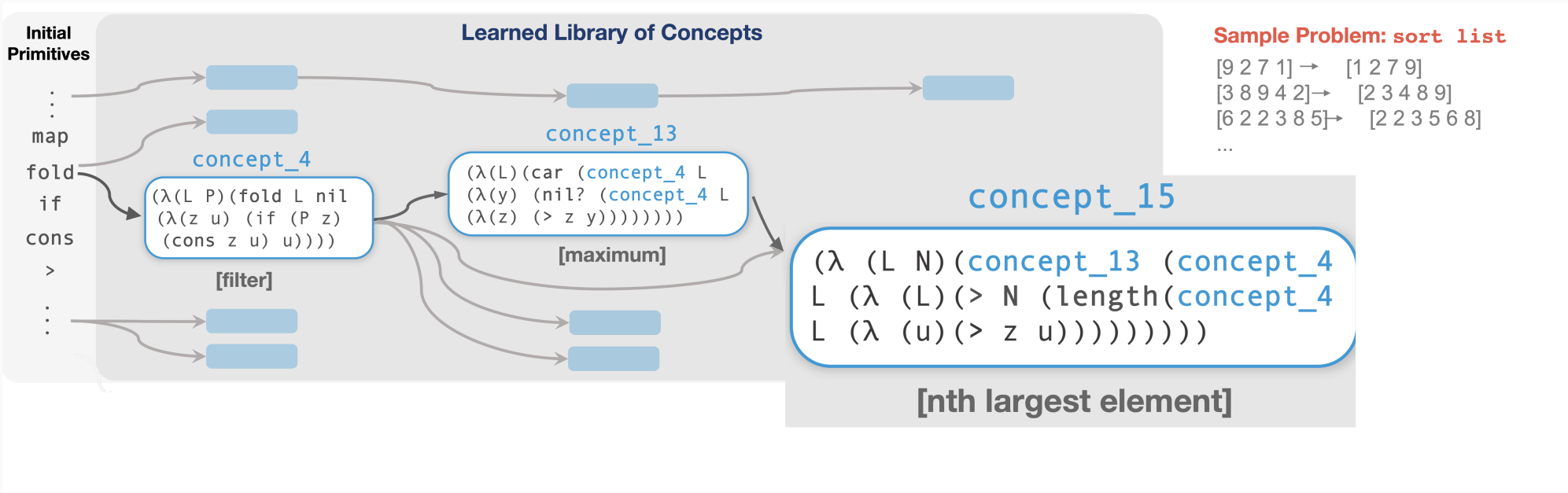


Sample Problem: `sort list`

[9 2 7 1] → [1 2 7 9]
[3 8 9 4 2] → [2 3 4 8 9]
[6 2 2 3 8 5] → [2 2 3 5 6 8]
...

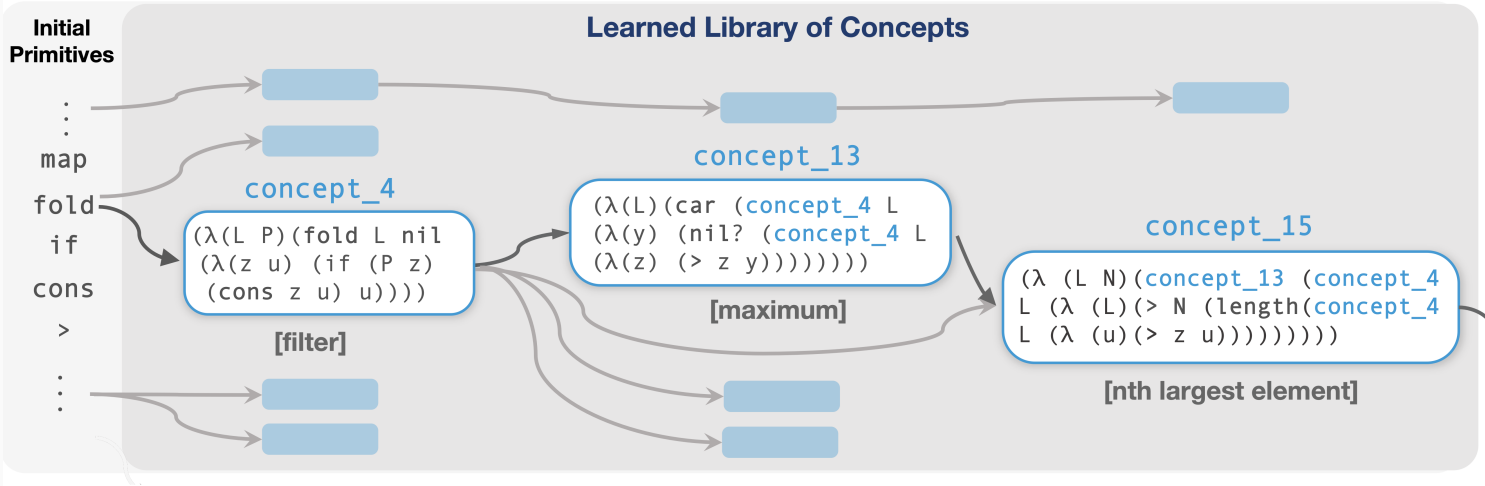
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Library learning



Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Library learning



Sample Problem: `sort list`

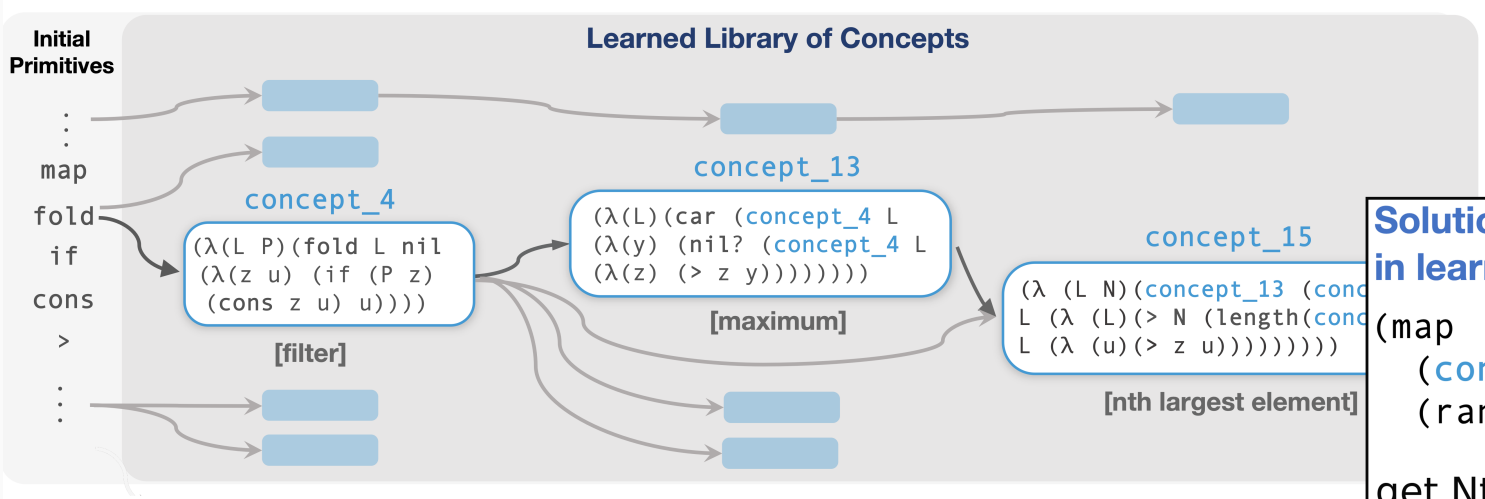
[9 2 7 1] → [1 2 7 9]
[3 8 9 4 2] → [2 3 4 8 9]
[6 2 2 3 8 5] → [2 2 3 5 6 8]
...

Solution to `sort list` discovered in learned language:

```
(map  (λ (n)
  (concept_15 L (+ 1 n)))
  (range (length L)))
```

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Library learning



Sample Problem: `sort list`

[9 2 7 1] → [1 2 7 9]
[3 8 9 4 2] → [2 3 4 8 9]
[6 2 2 3 8 5] → [2 2 3 5 6 8]
...

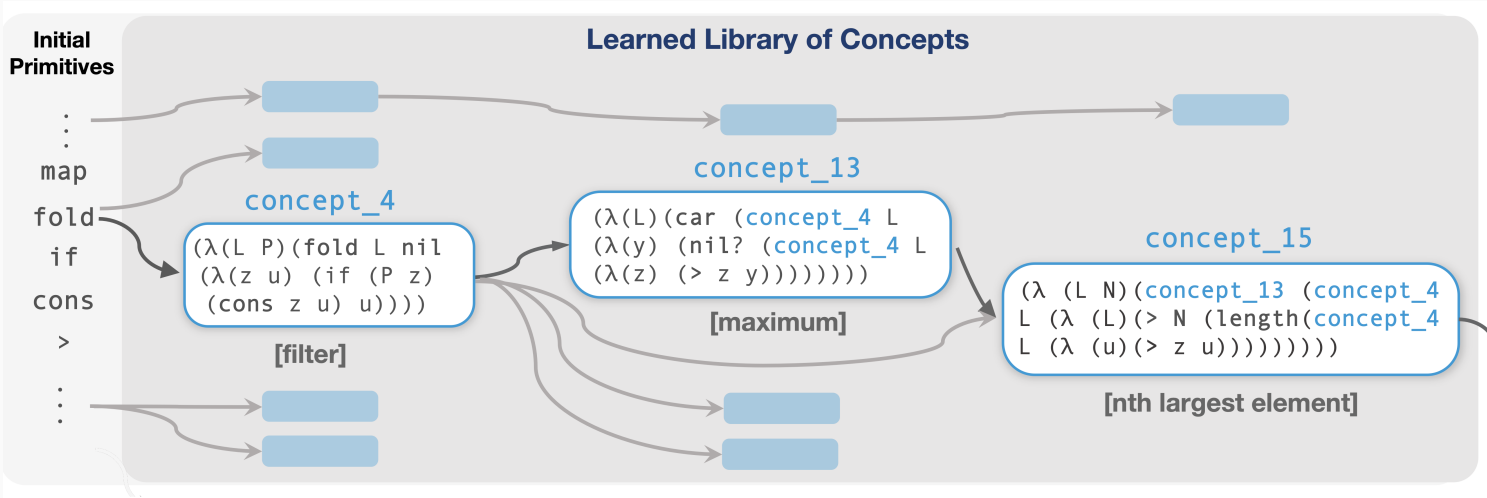
Solution to `sort list` discovered in learned language:

```
(map  (λ (n)
  (concept_15 L (+ 1 n)))
  (range (length L)))
```

get Nth largest element, where N is 1, 2, 3, ...

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Library learning



Sample Problem: `sort list`

[9 2 7 1] → [1 2 7 9]
[3 8 9 4 2] → [2 3 4 8 9]
[6 2 2 3 8 5] → [2 2 3 5 6 8]
...

Solution to `sort list` discovered in learned language:

```
(map  (λ (n)
  (concept_15 L (+ 1 n)))
  (range (length L)))
```

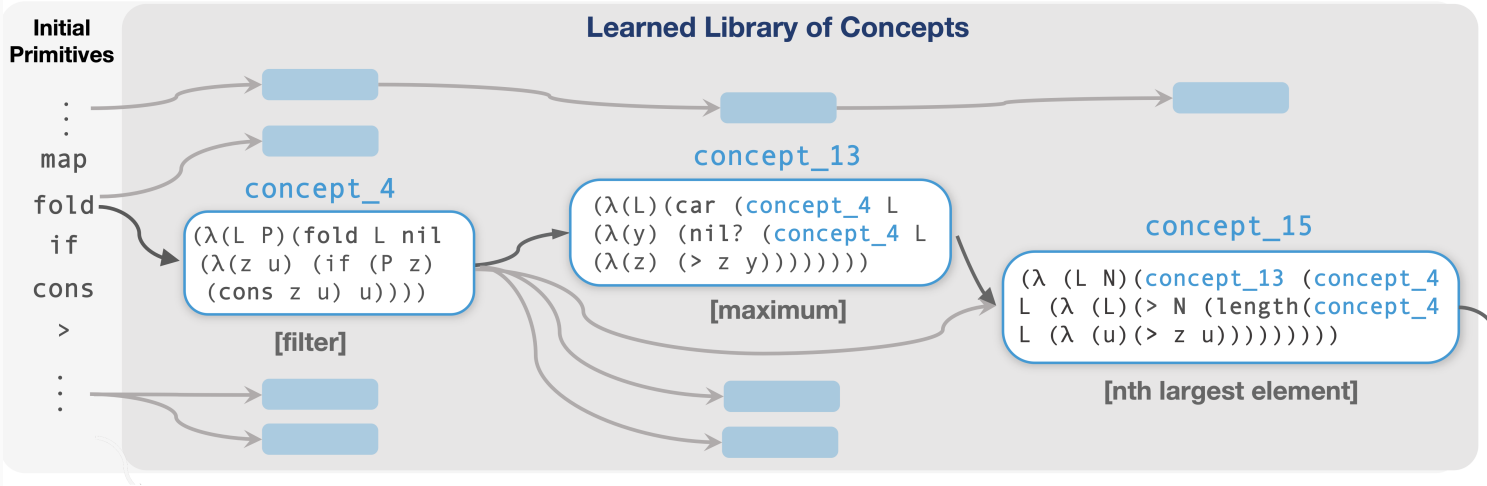
get Nth largest element, where N is 1, 2, 3, ...

Solution rewritten in initial primitives:

```
(lambda (x) (map (lambda (y) (car (fold (fold x nil (lambda (z u) (if (gt?  (+ y 1) (length
(fold x nil (lambda (v w) (if (gt?  z v) (cons v w) w))))) (cons z u) u))) nil (lambda (a b) (if
(nil?  (fold (fold x nil (lambda (c d) (if (gt?  (+ y 1) (length (fold x nil (lambda (e f) (if
(gt?  c e) (cons e f) f))))) (cons c d) d))) nil (lambda (g h) (if (gt?  g a) (cons g h) h))))
(cons a b) b)))) (range (length x))))
```

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Library learning



Sample Problem: `sort list`

[9 2 7 1] → [1 2 7 9]
[3 8 9 4 2] → [2 3 4 8 9]
[6 2 2 3 8 5] → [2 2 3 5 6 8]
...

Solution to `sort list` discovered in learned language:

```
(map  (λ (n)
  (concept_15 L (+ 1 n)))
  (range (length L)))
```

get Nth largest element, where N is 1, 2, 3, ...

Solution rewritten in initial primitives:

```
(lambda (x) (map (lambda (y) (car (fold (fold x nil (lambda (z u) (if (gt?  (+ y 1) (length
(fold x nil (lambda (v w) (if (gt?  z v) (cons v w) w))))) (cons z u) u))) nil (lambda (a b) (if
(nil?  (fold (fold x nil (lambda (c d) (if (gt?  (+ y 1) (length (fold x nil (lambda (e f) (if
(gt?  c e) (cons e f) f))))) (cons c d) d))) nil (lambda (g h) (if (gt?  g a) (cons g h) h))))
(cons a b) b)))) (range (length x))))
```

induced sort program found in $\leq$ 10min. Brute-force search without learned library would take $\approx 10^{73}$ years

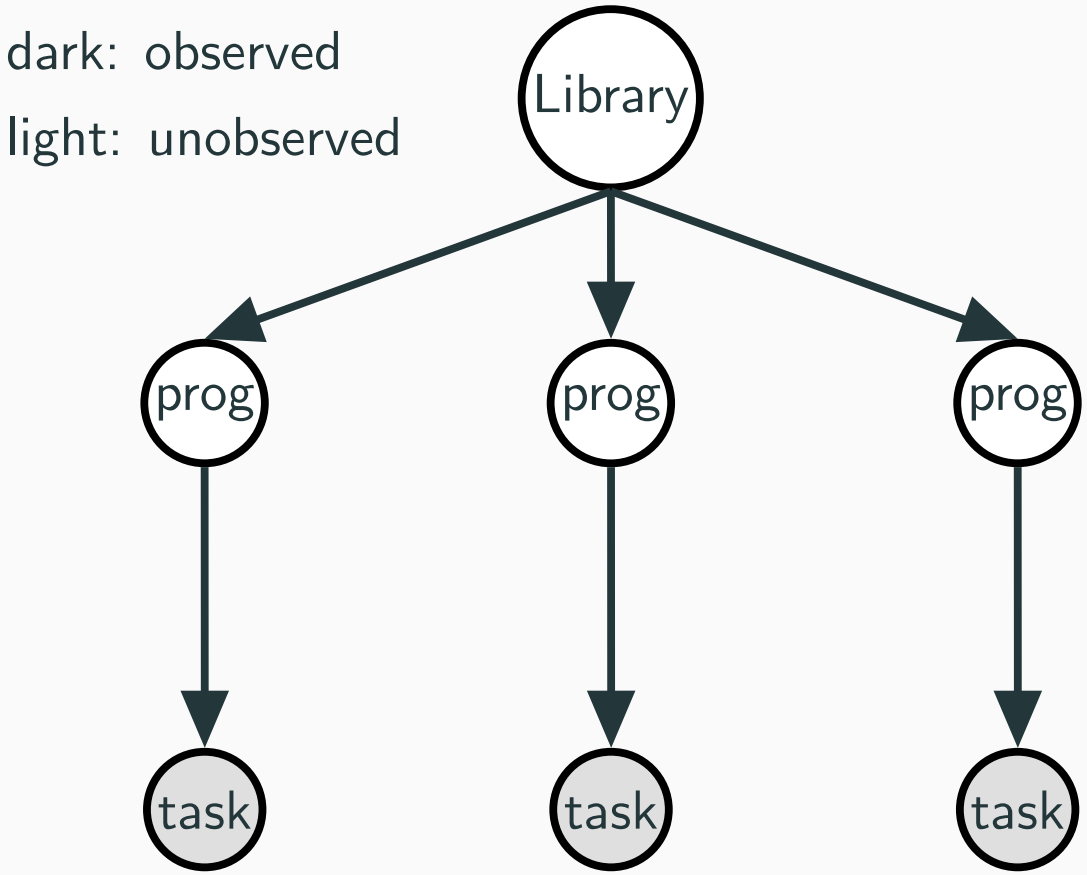
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

DreamCoder

- **Wake:** Solve problems by writing programs
- **Sleep:** Improve library and neural recognition model:
  - **Abstraction sleep:** Improve library
  - **Dream sleep:** Improve neural recognition model

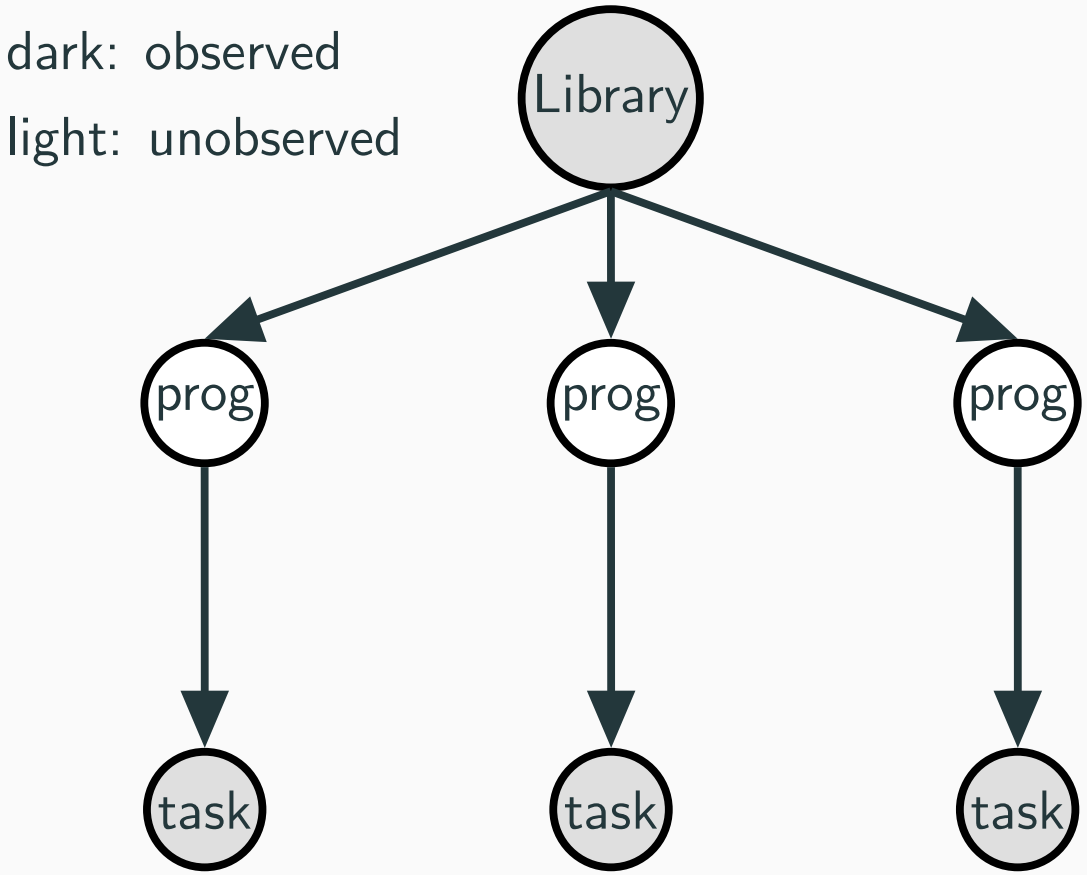
cf. Helmholtz machine, wake/sleep neural network training algorithms

Library learning as Bayesian inference



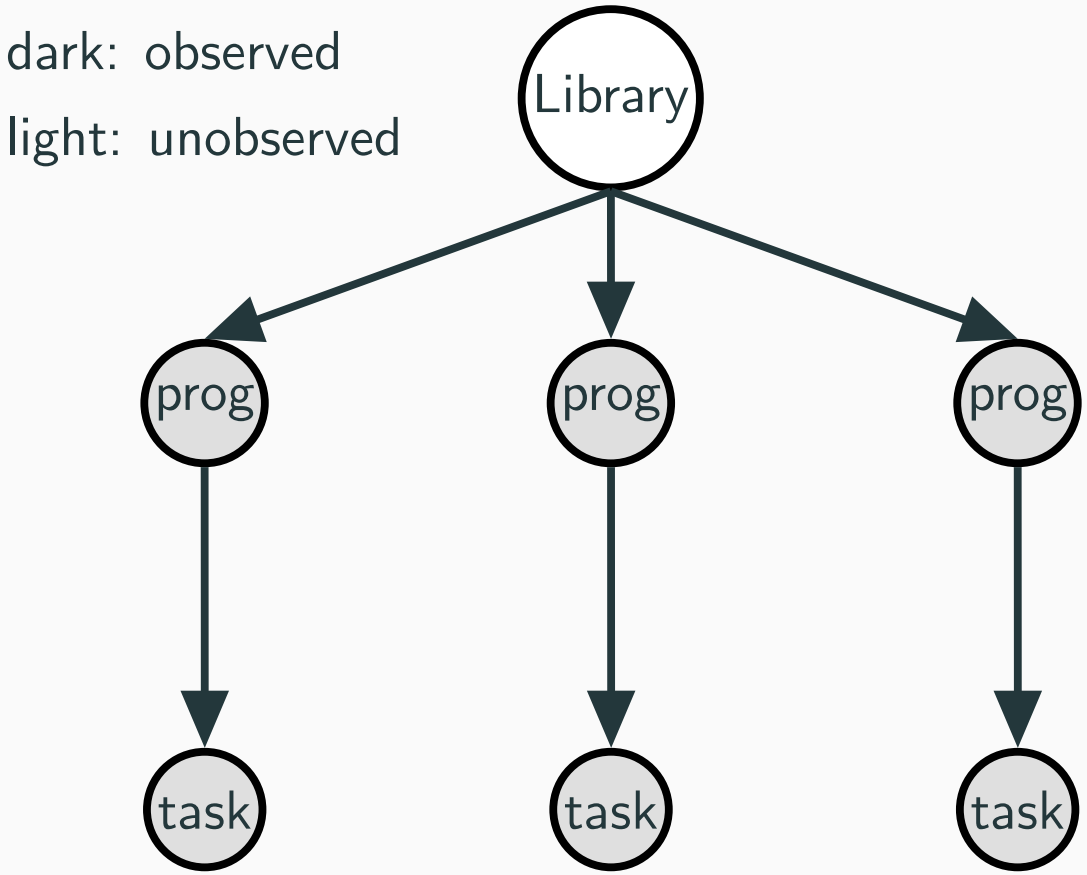
[Dechter et al, 2013] [Liang et al, 2010] [Lake et al, 2015]

Library learning as Bayesian inference



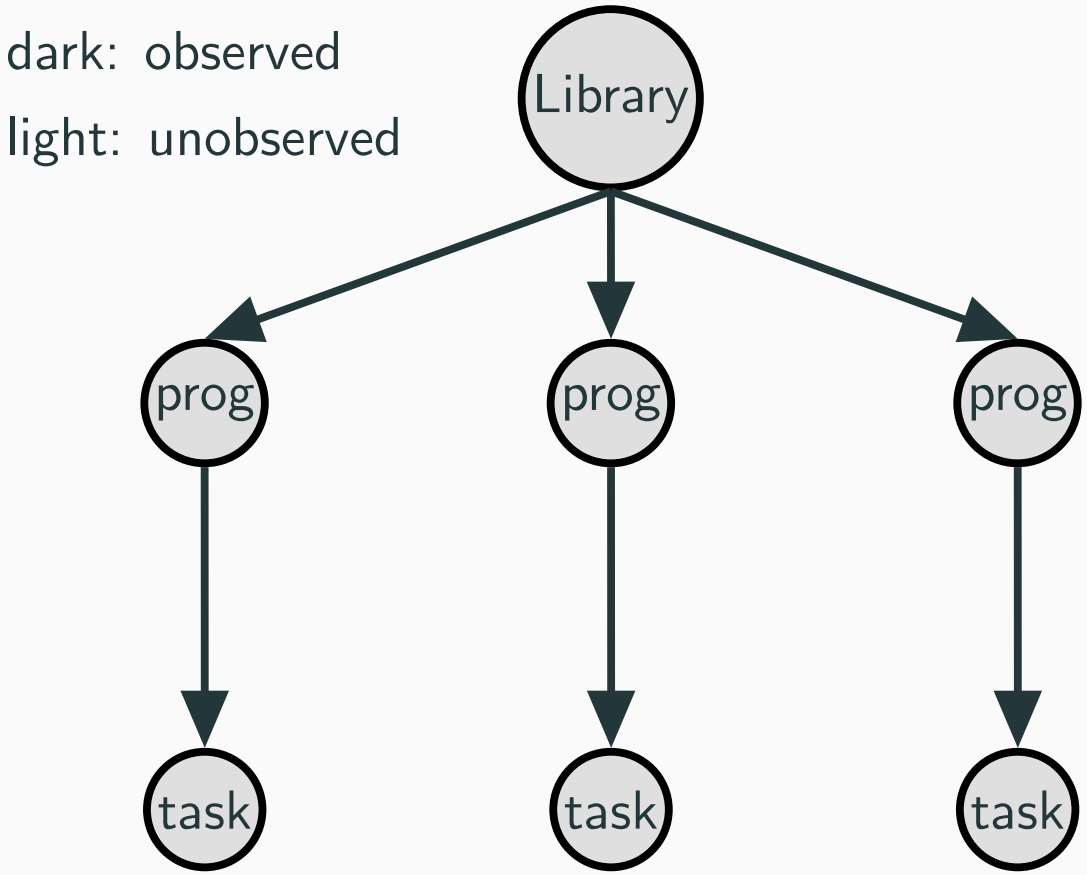
[Dechter et al, 2013] [Liang et al, 2010] [Lake et al, 2015]

Library learning as Bayesian inference



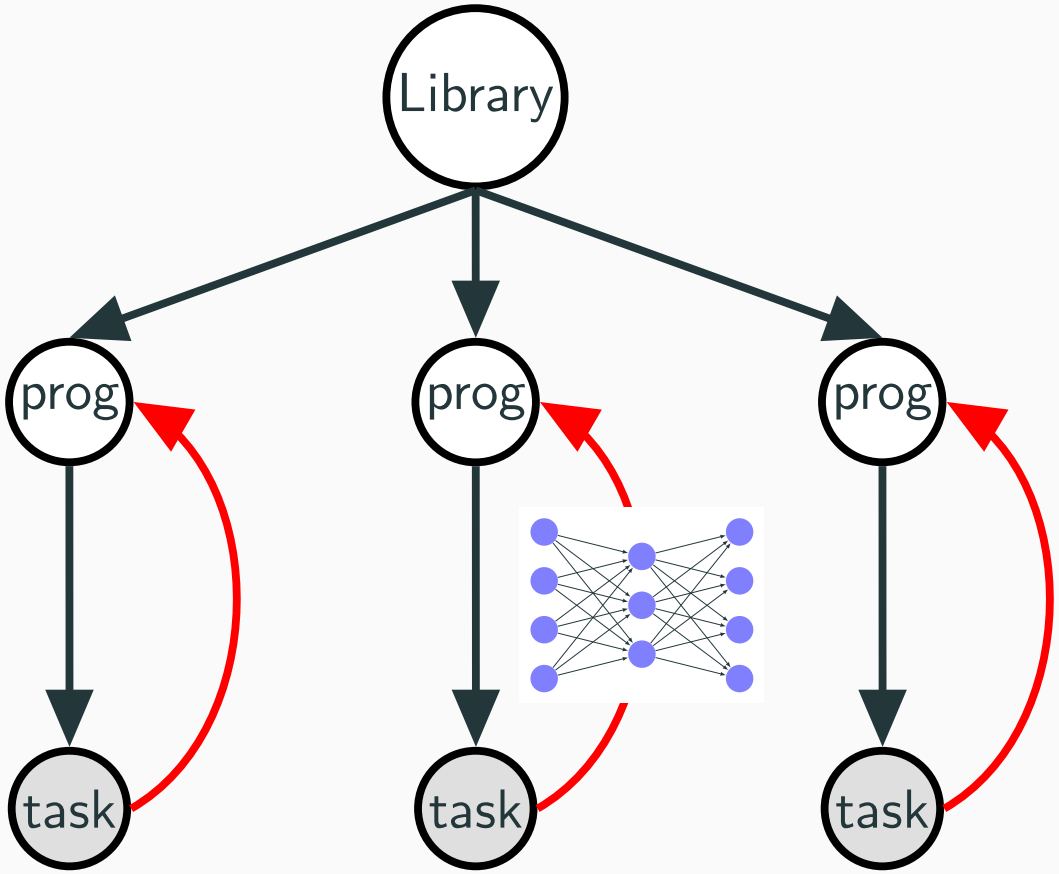
[Dechter et al, 2013] [Liang et al, 2010] [Lake et al, 2015]

Library learning as Bayesian inference



[Dechter et al, 2013] [Liang et al, 2010] [Lake et al, 2015]

Library learning as neurally-guided Bayesian inference

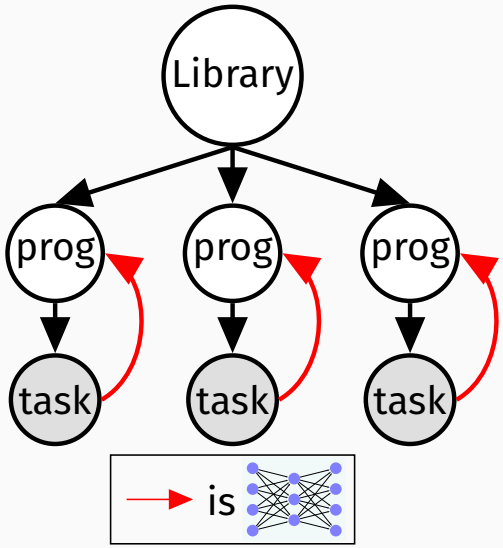


library learning via program analysis +

new neural inference network for program synthesis +

better program representation (Lisp+polymorphic types [Milner 1978])

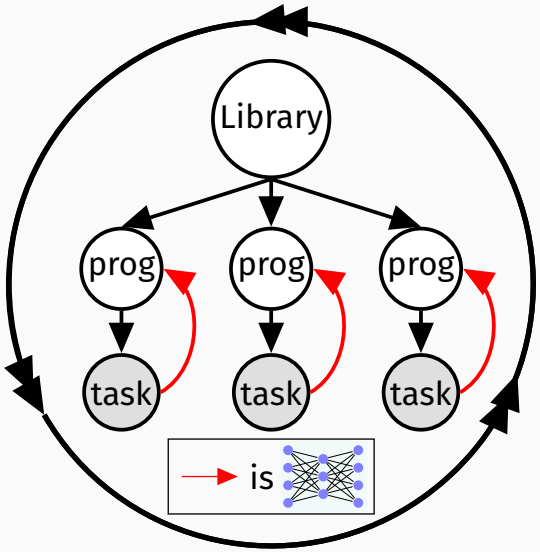
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.


Library
prog
prog
prog
task
task
task
is

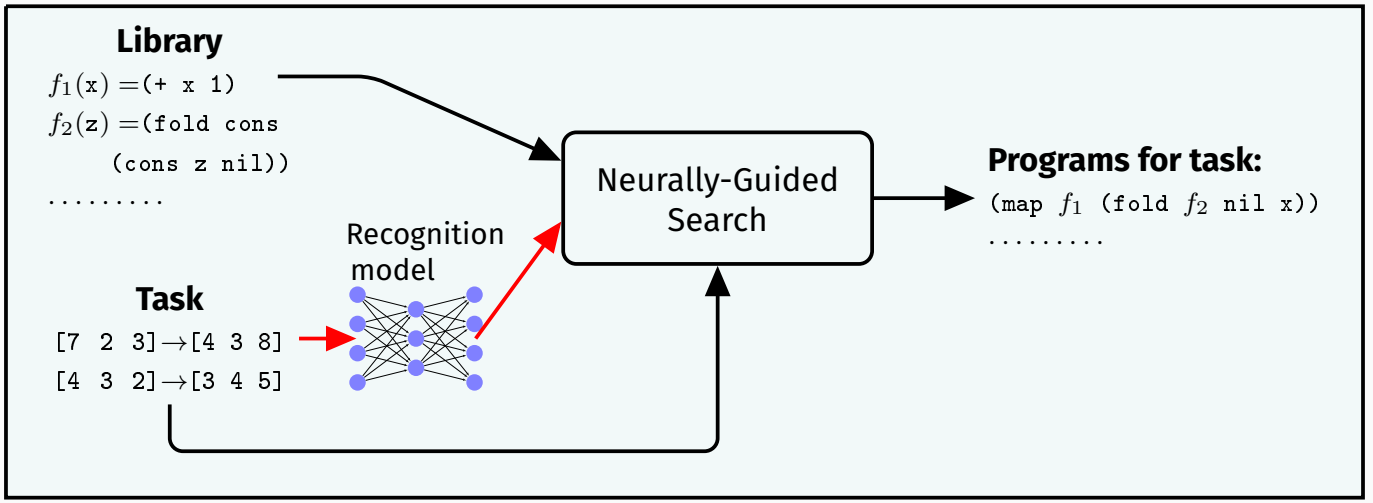
Wake

Sleep: Abstraction

Sleep: Dreaming

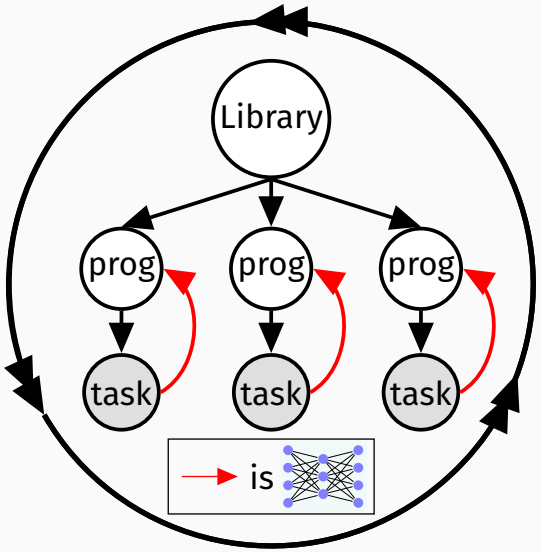


WAKE

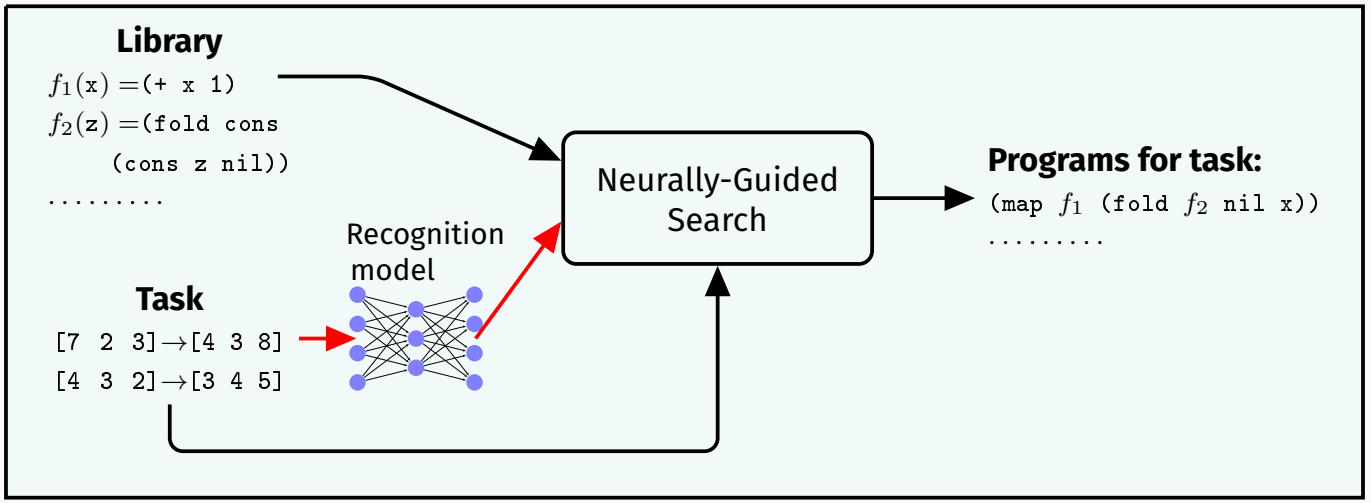


SLEEP: ABSTRACTION

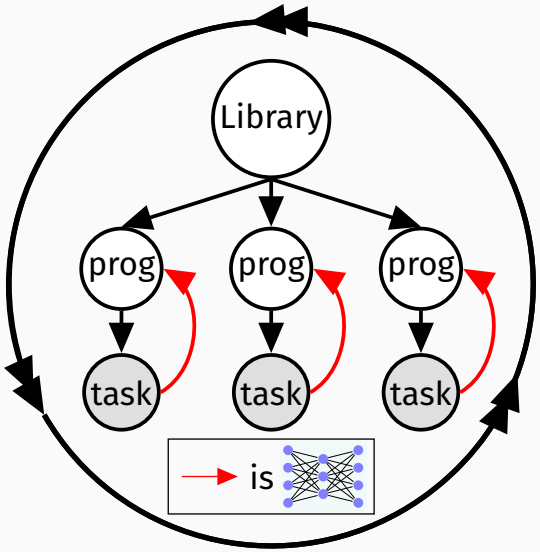
SLEEP: DREAMING



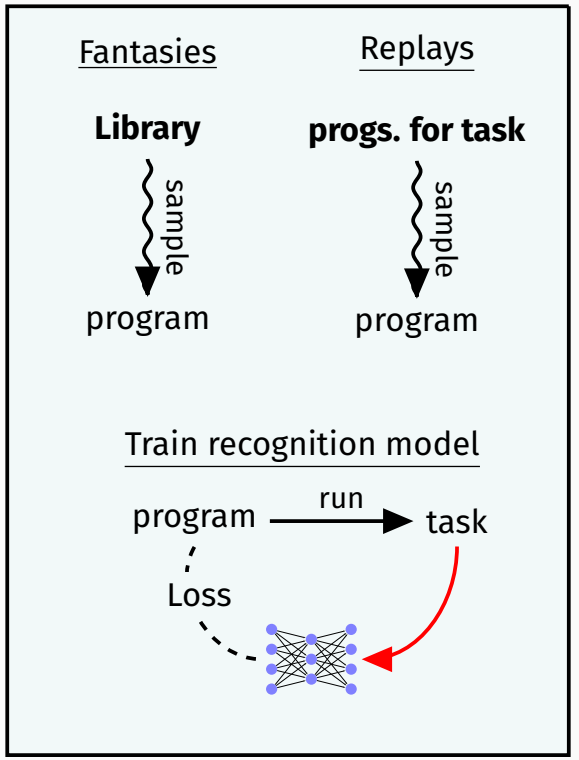
Wake

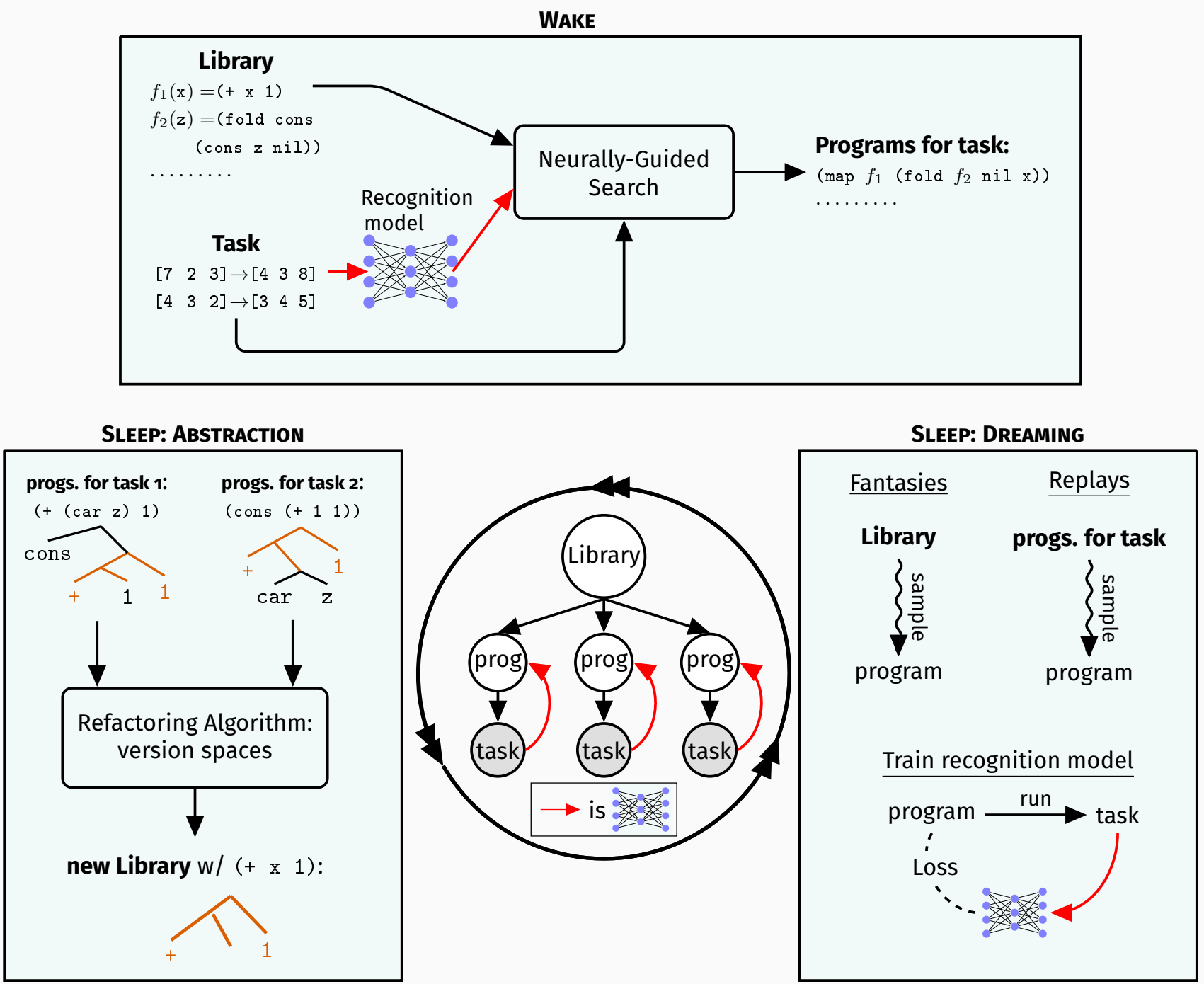


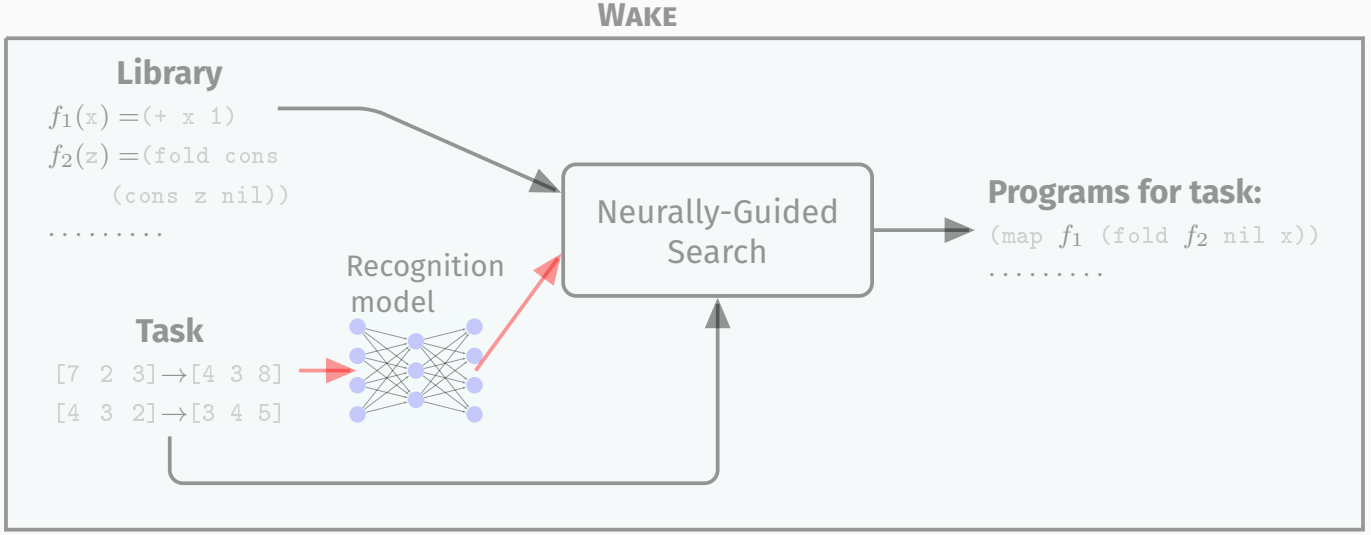
Sleep: Abstraction

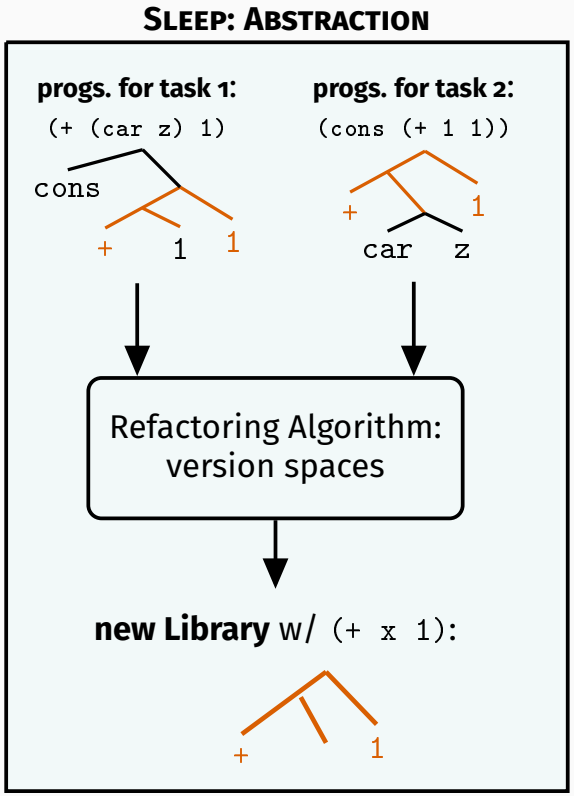


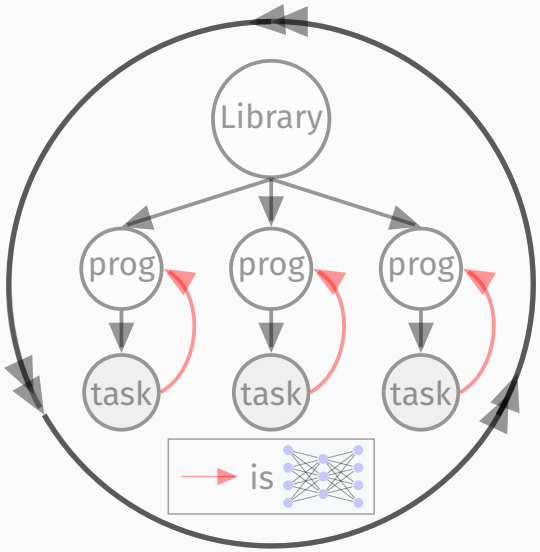
Sleep: Dreaming

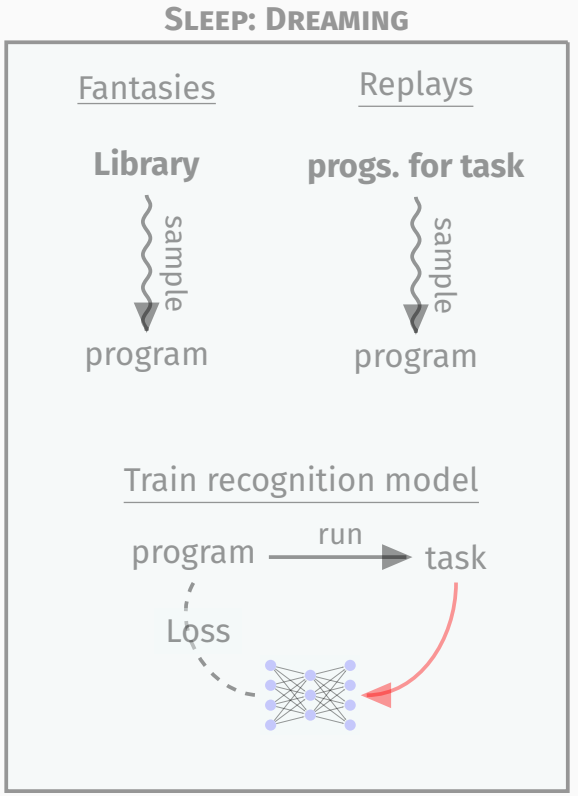



WAKE
Library
f_1(x) =(+ x 1)
f_2(z) =(fold cons
(cons z nil))
.....
Task
[7 2 3]→[4 3 8]
[4 3 2]→[3 4 5]
Recognition model
Neurally-Guided Search
Programs for task:
(map f_1 (fold f_2 nil x))
.....
SLEEP: ABSTRACTION
progs. for task 1:
(+ (car z) 1)
cons
+
1
1
progs. for task 2:
(cons (+ 1 1))
+
1
car
z
Refactoring Algorithm: version spaces
new Library w/ (+ x 1):
+
1
Library
prog
prog
prog
task
task
task
is
SLEEP: DREAMING
Fantasies
Replays
Library
progs. for task
sample
sample
program
program
Train recognition model
program
run
task
Loss


WAKE
Library
$f_1(x) = (+\ x\ 1)$
$f_2(z) = (\text{fold cons } (\text{cons } z \text{ nil}))$
.....
Task
[7 2 3]→[4 3 8]
[4 3 2]→[3 4 5]
Recognition model
Neurally-Guided Search
Programs for task:
(map f_1 (fold f_2 nil x))
.....


SLEEP: ABSTRACTION
progs. for task 1:
(+ (car z) 1)
cons
+
1
1
progs. for task 2:
(cons (+ 1 1))
+
1
car
z
Refactoring Algorithm: version spaces
new Library w/ (+ x 1):
+
1


Library
prog
prog
prog
task
task
task
is


SLEEP: DREAMING
Fantasies
Replays
Library
progs. for task
sample
sample
program
program
Train recognition model
program
run
task
Loss

Program Induction and learning to learn

learning a DSL

learning to synthesize

synergy between DSL+learned synthesizer

Abstraction Sleep: Growing the library via refactoring

$5 + 5$

Abstraction Sleep: Growing the library via refactoring

$5 + 5$

```
(+ 5 5)
```

Abstraction Sleep: Growing the library via refactoring

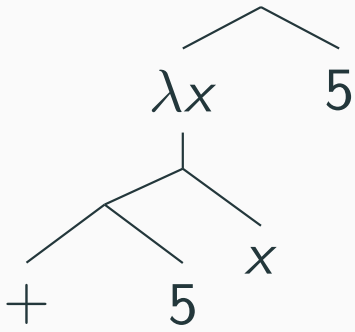
$5 + 5$

```
(+ 5 5)
```

Abstraction Sleep: Growing the library via refactoring

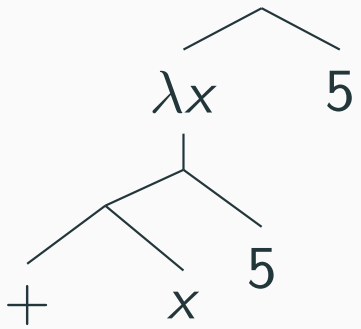
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

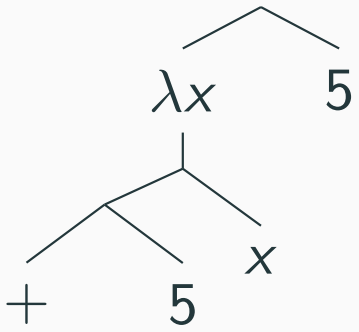
Abstraction Sleep: Growing the library via refactoring



Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

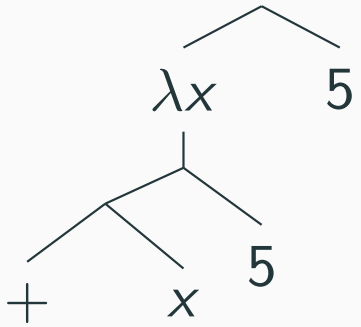
Abstraction Sleep: Growing the library via refactoring

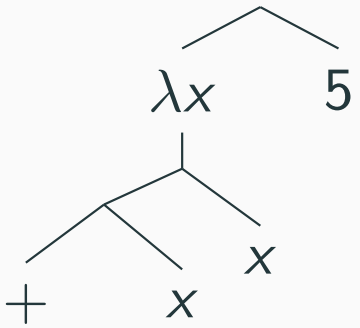


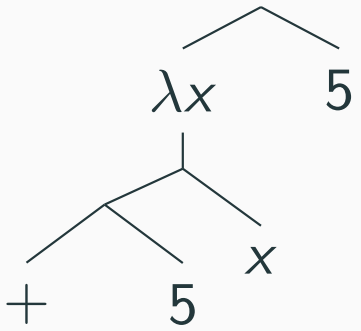


Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Abstraction Sleep: Growing the library via refactoring

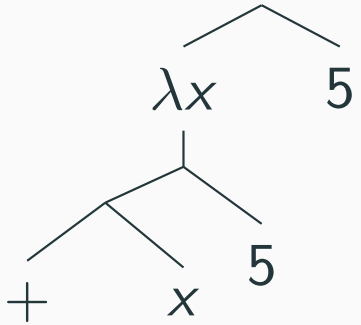


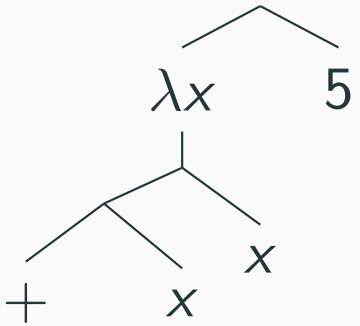


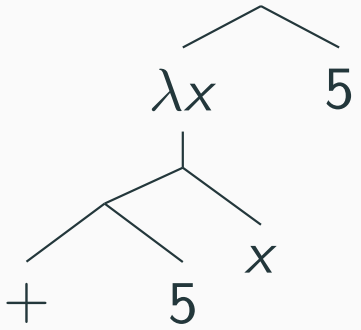


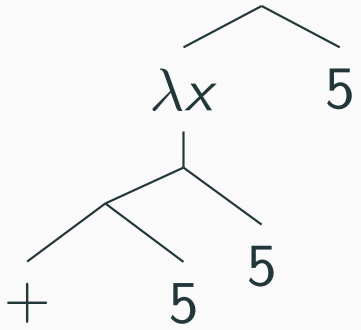
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Abstraction Sleep: Growing the library via refactoring



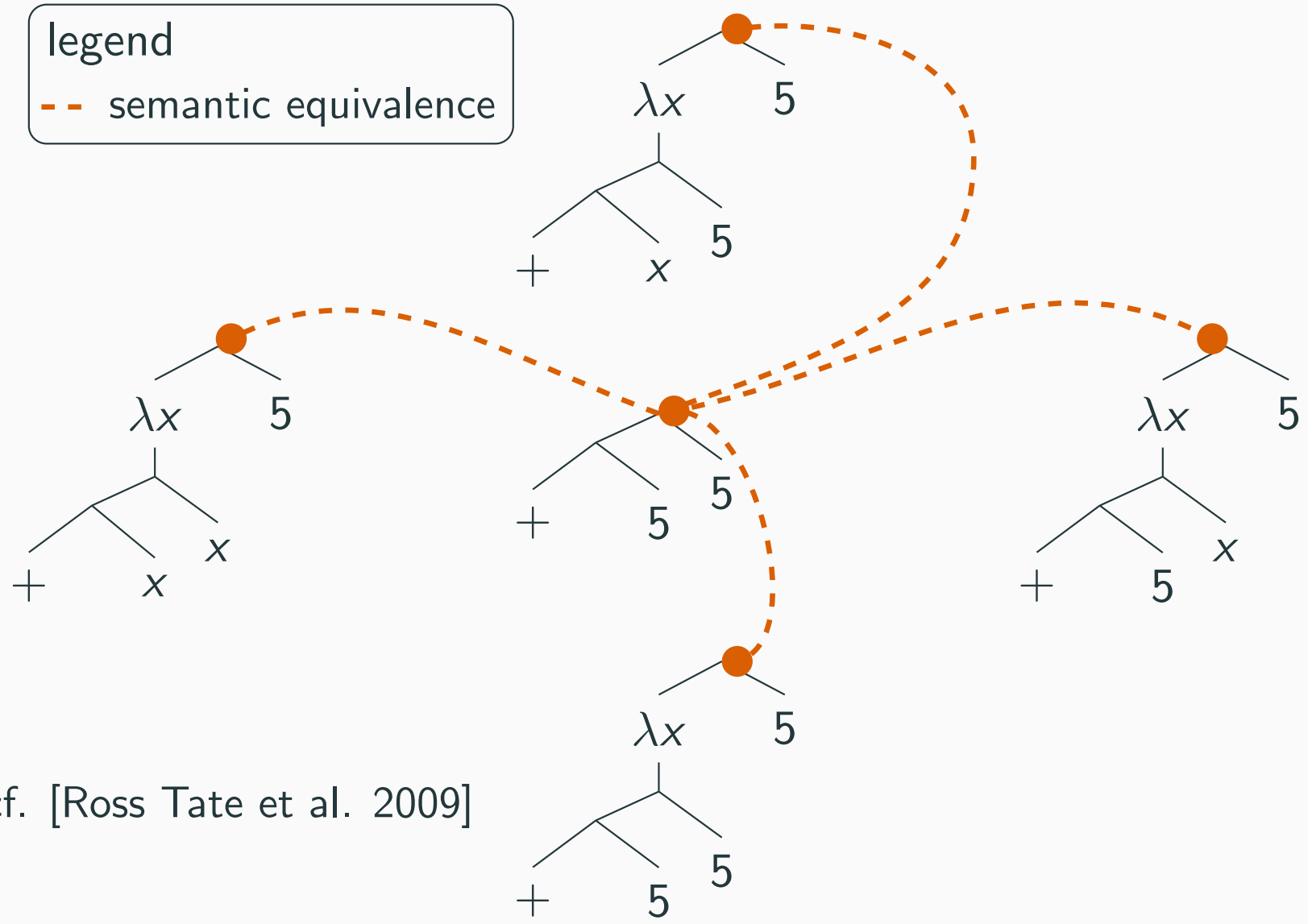






Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Abstraction Sleep: Growing the library via refactoring

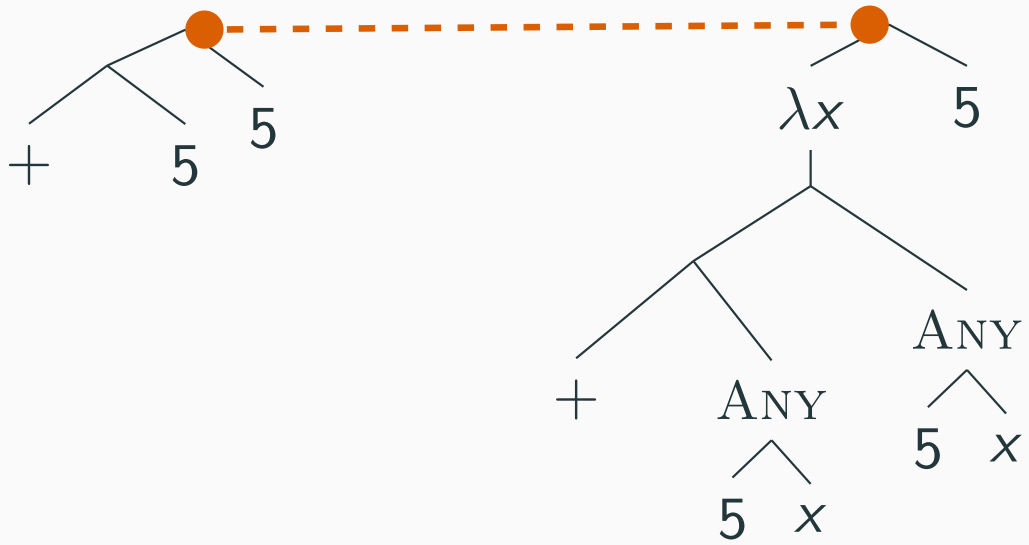


cf. [Ross Tate et al. 2009]

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Abstraction Sleep: Growing the library via refactoring

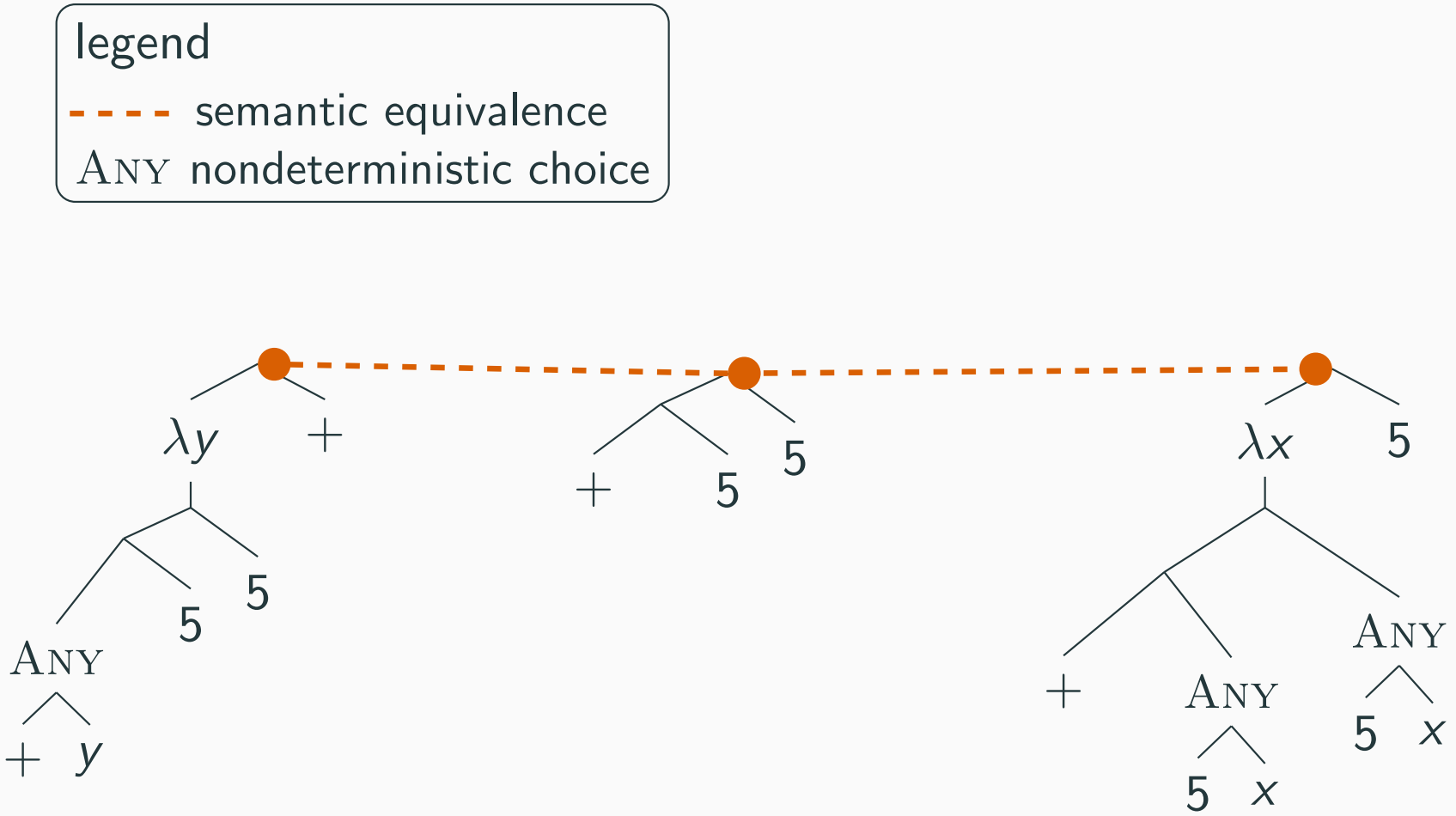
legend
- - - - semantic equivalence
ANY nondeterministic choice



cf. [Gulwani 2012]

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

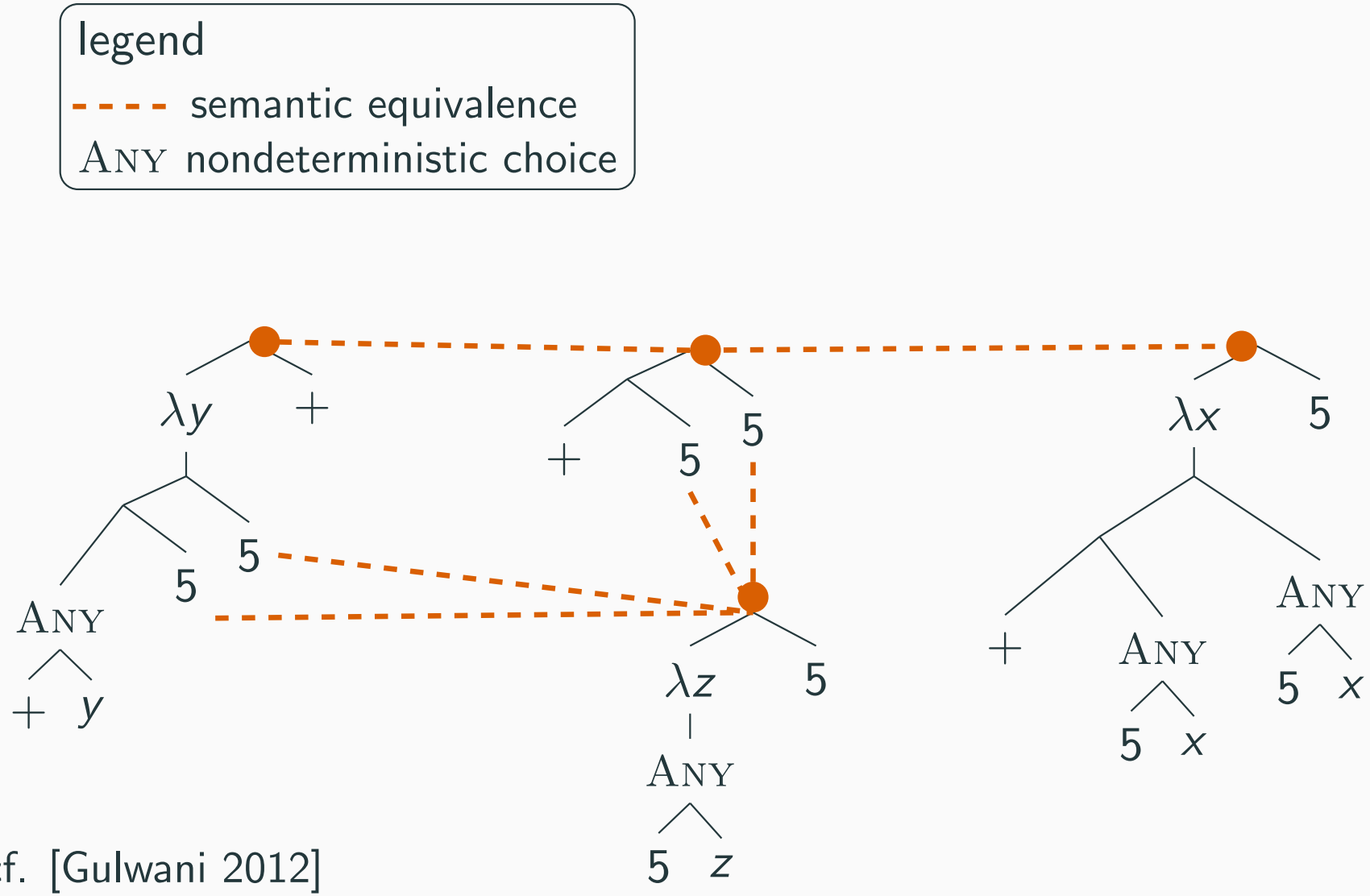
Abstraction Sleep: Growing the library via refactoring



cf. [Gulwani 2012]

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

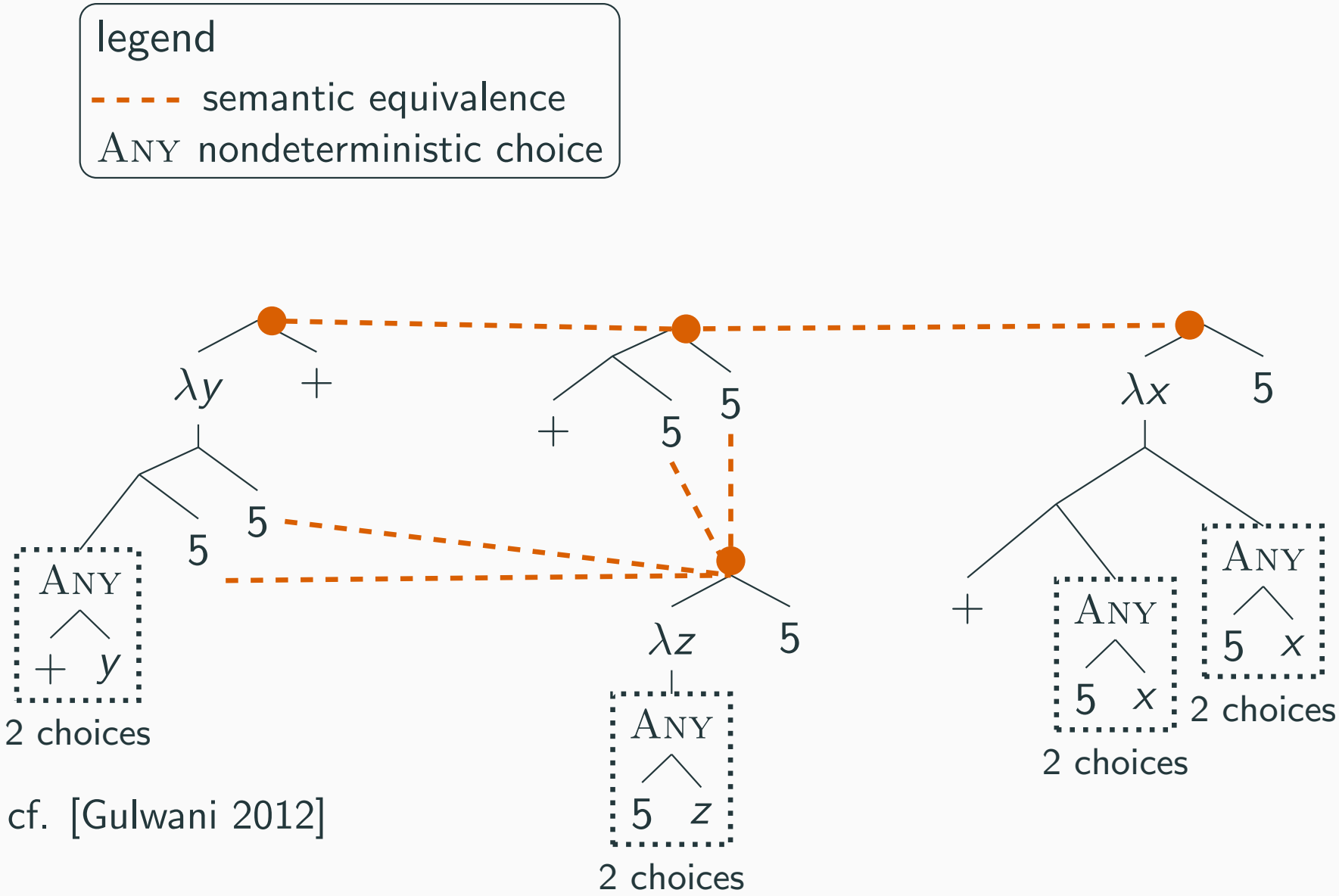
Abstraction Sleep: Growing the library via refactoring



cf. [Gulwani 2012]

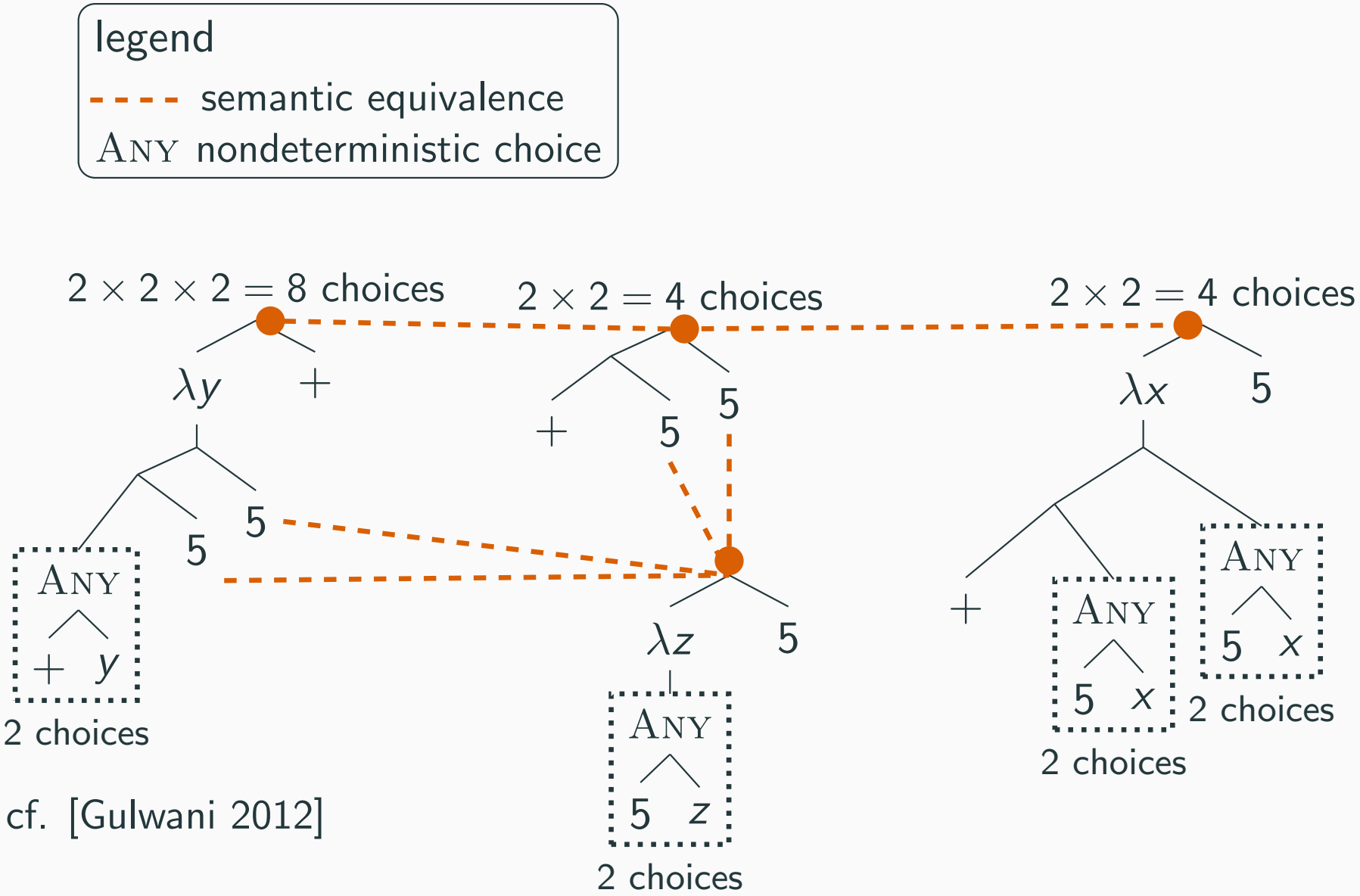
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Abstraction Sleep: Growing the library via refactoring



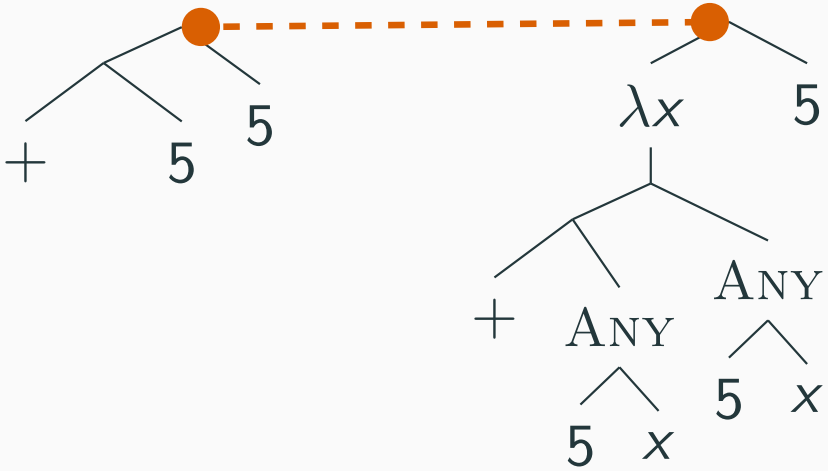
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

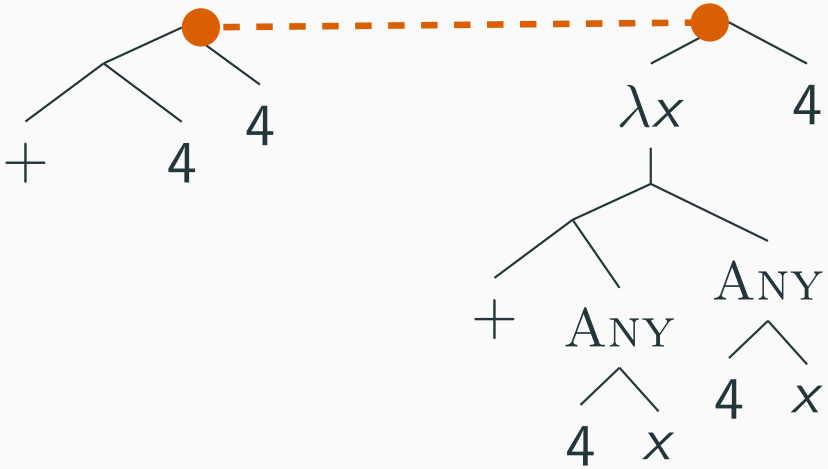
Abstraction Sleep: Growing the library via refactoring



cf. [Gulwani 2012]

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.



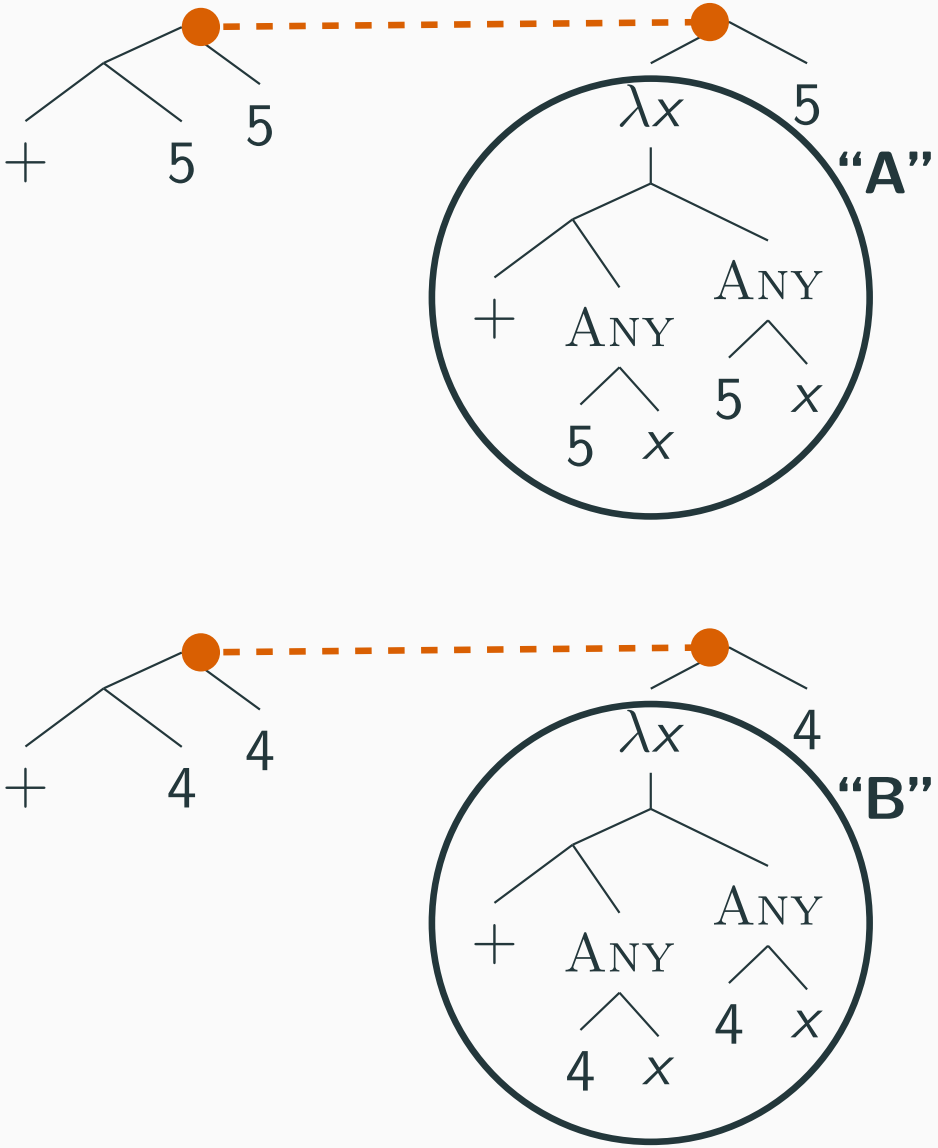


legend

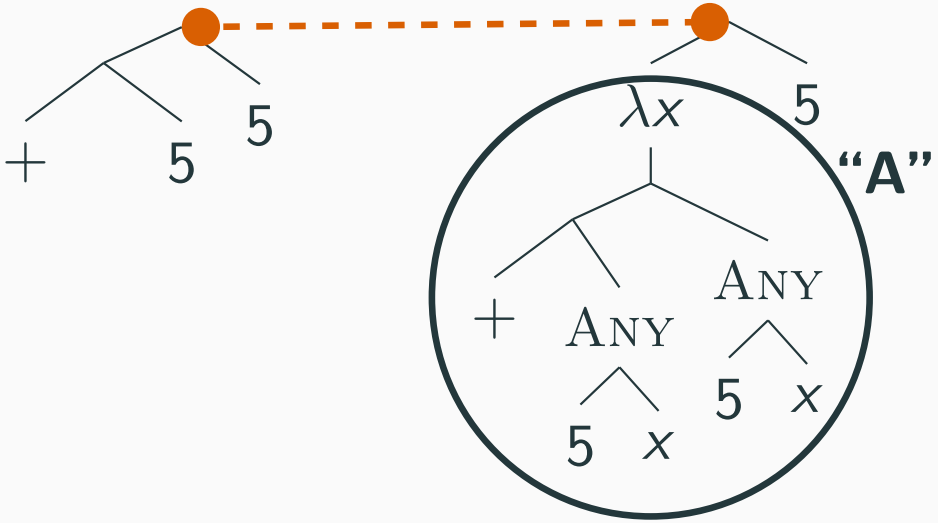
semantic equivalence

ANY nondeterministic choice

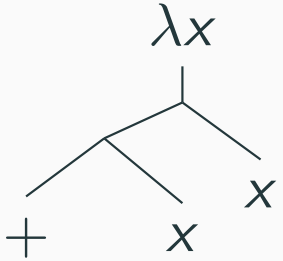
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

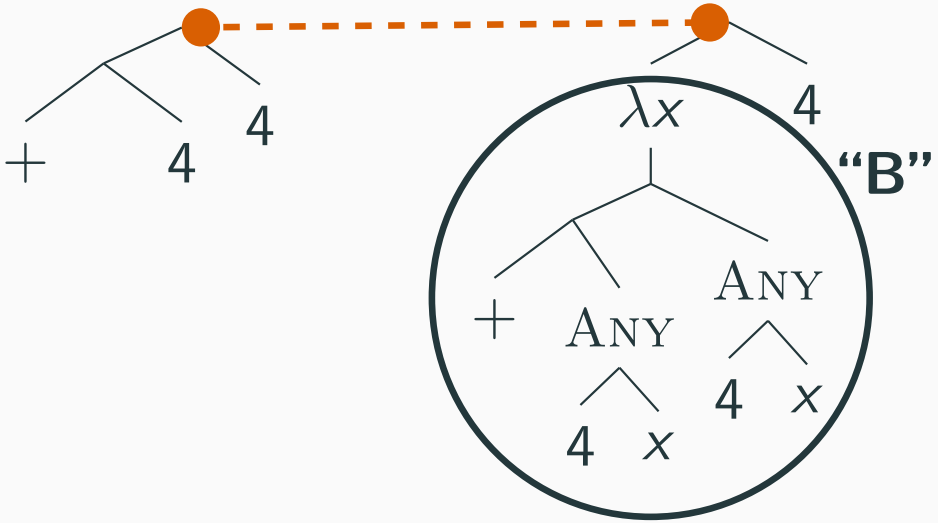


Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.



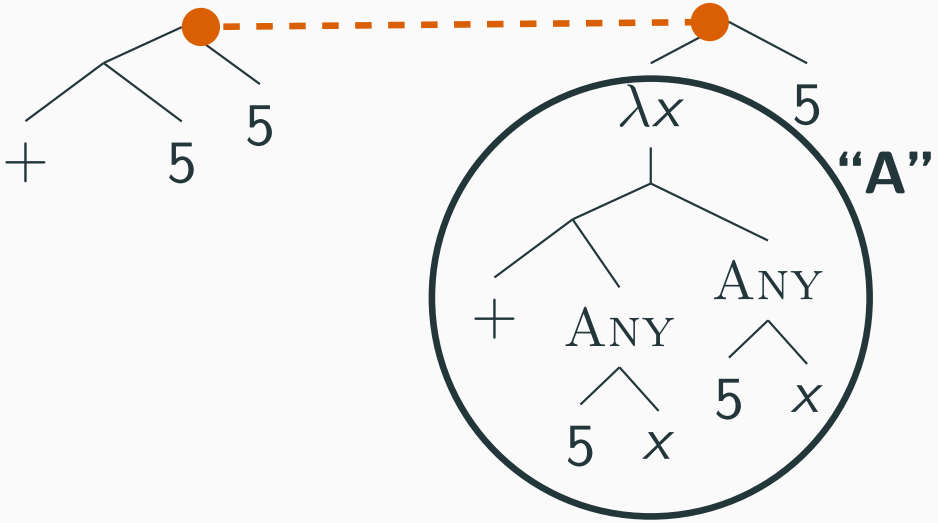
A intersect B:



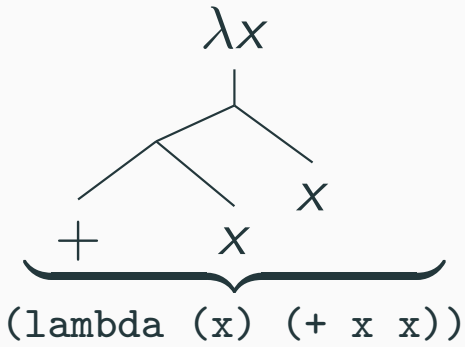


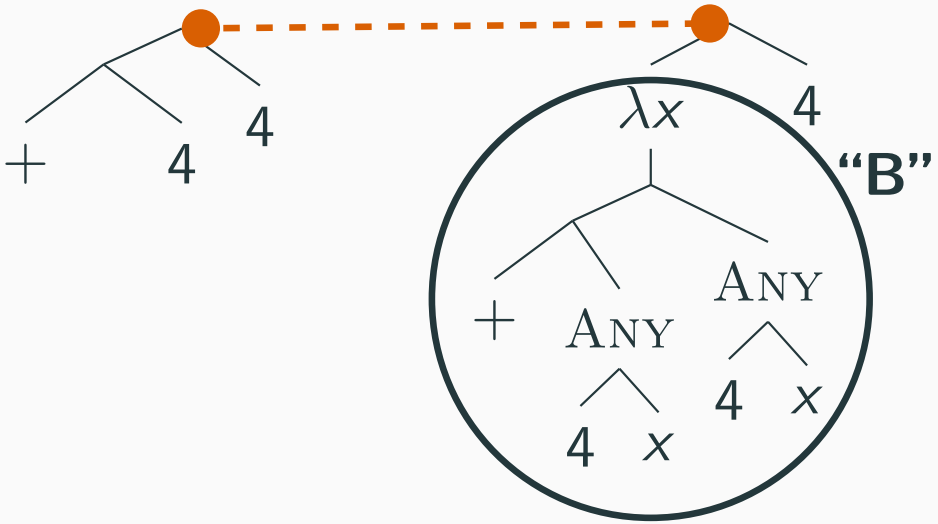
legend
- - - - semantic equivalence
ANY nondeterministic choice

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.



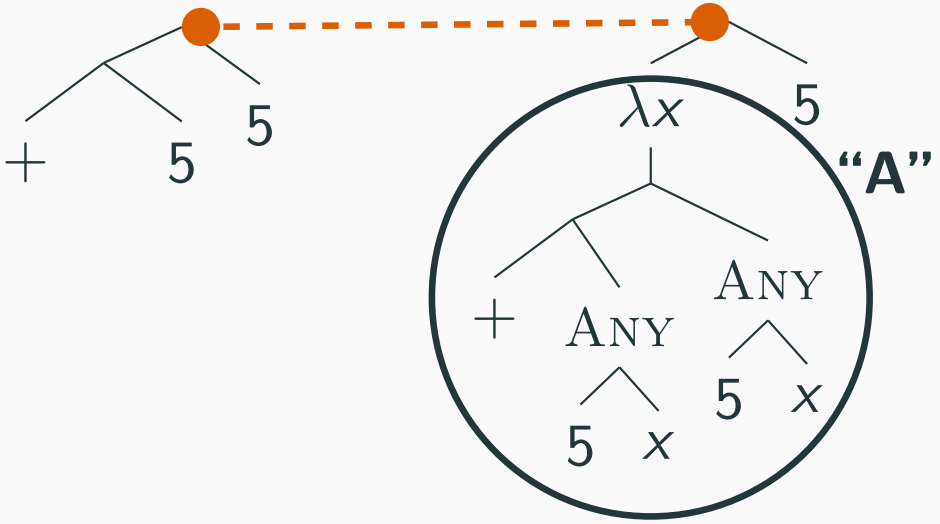
A intersect B:



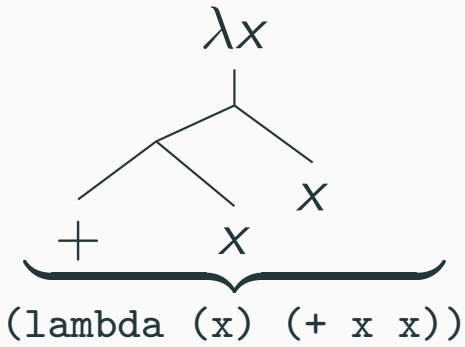


legend
- - - - semantic equivalence
ANY nondeterministic choice

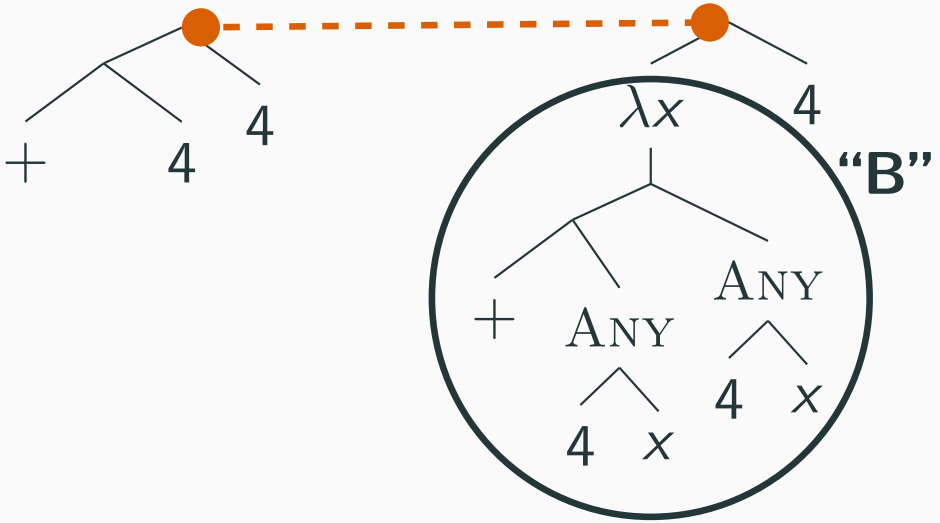
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.



A intersect B:

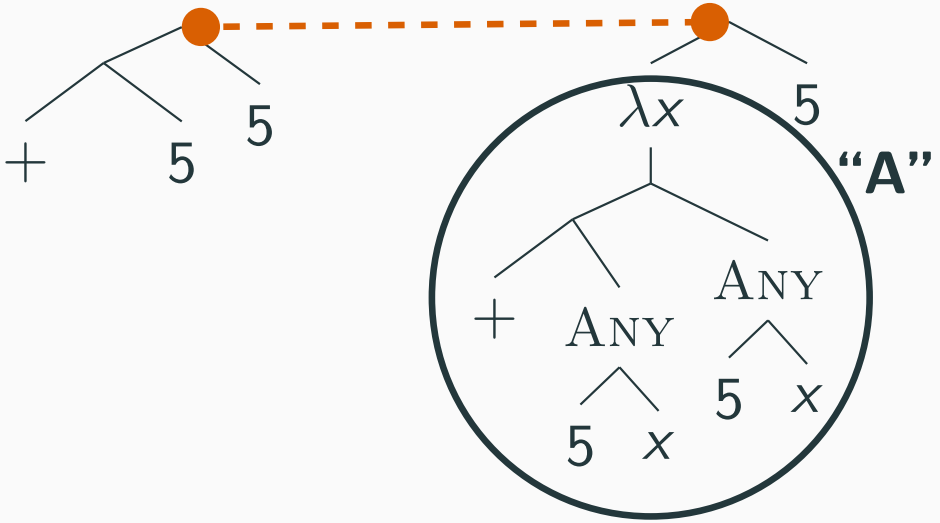


= `double`

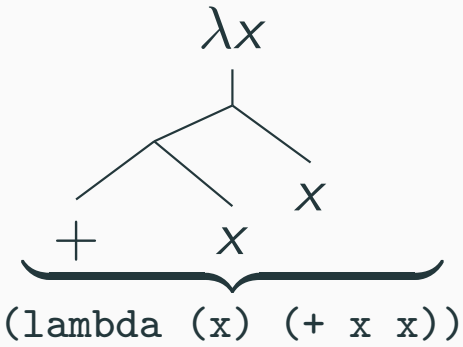


legend
- - - - semantic equivalence
ANY nondeterministic choice

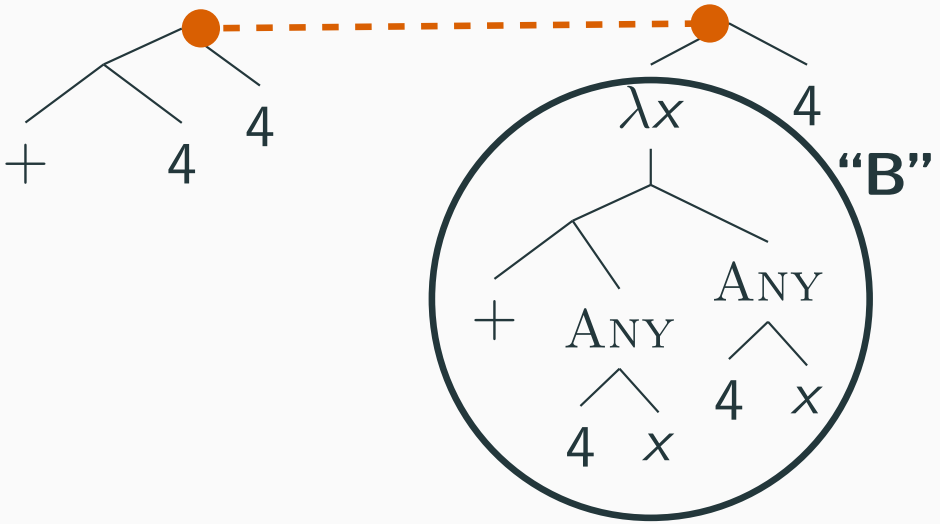
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.



A intersect B:



= double

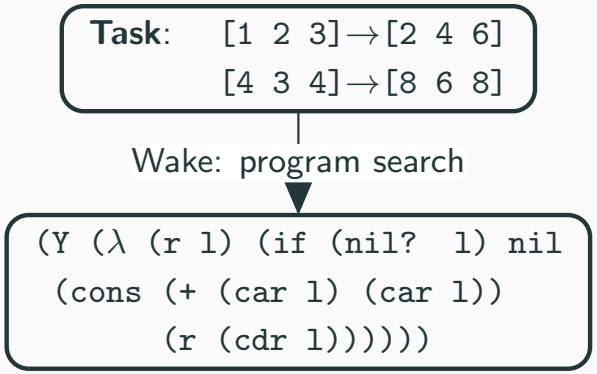


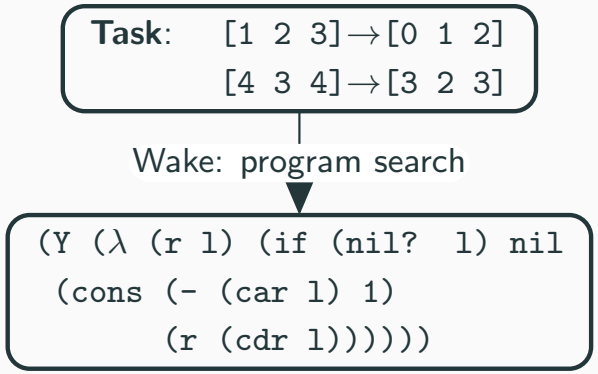
w/o double	w/ double
(+ 5 5)	(double 5)
(+ 4 4)	(double 4)
(+ 3 3)	(double 3)
...	

legend
- - - - semantic equivalence
ANY nondeterministic choice

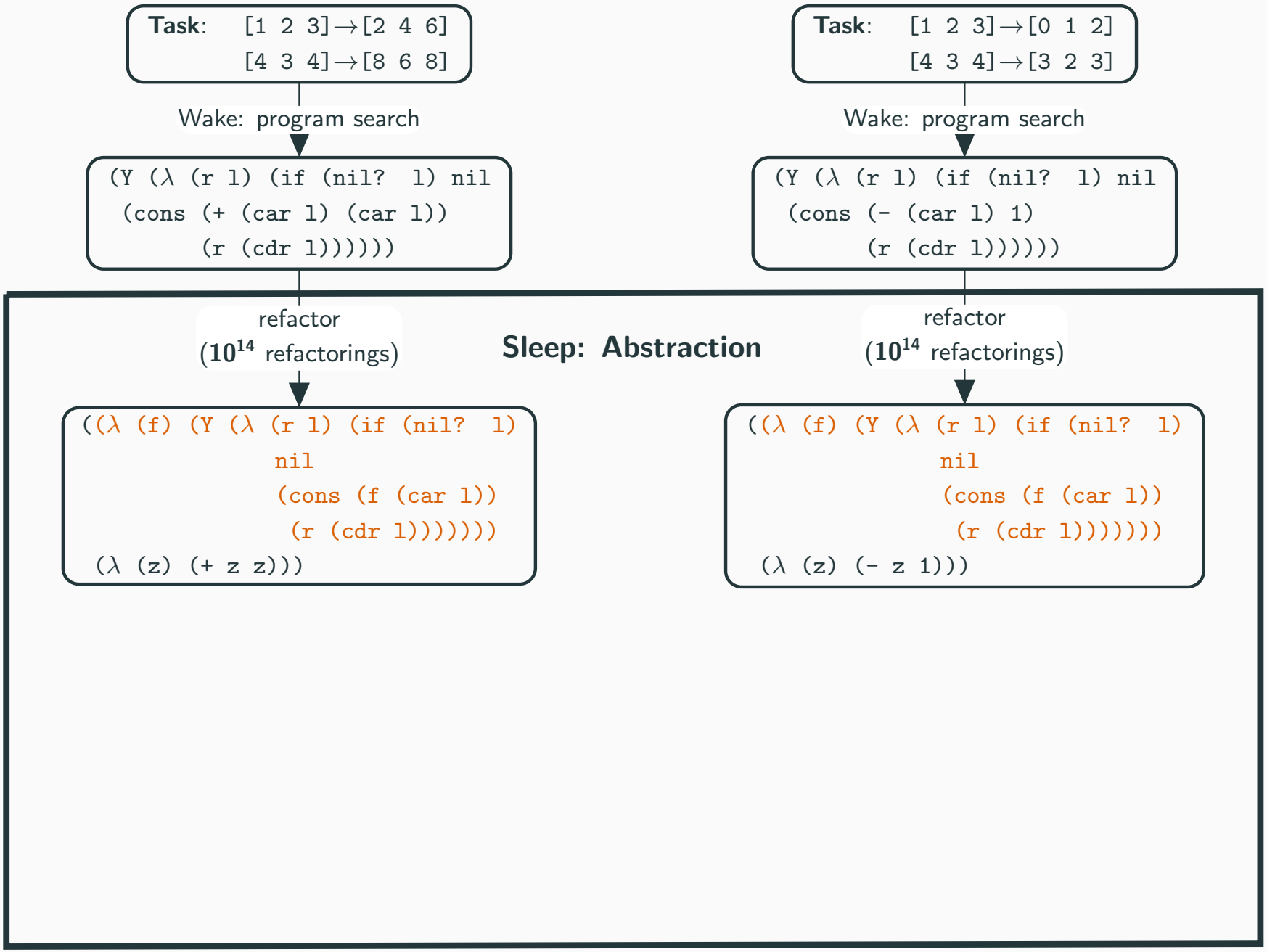
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Abstraction Sleep: Growing the library via refactoring

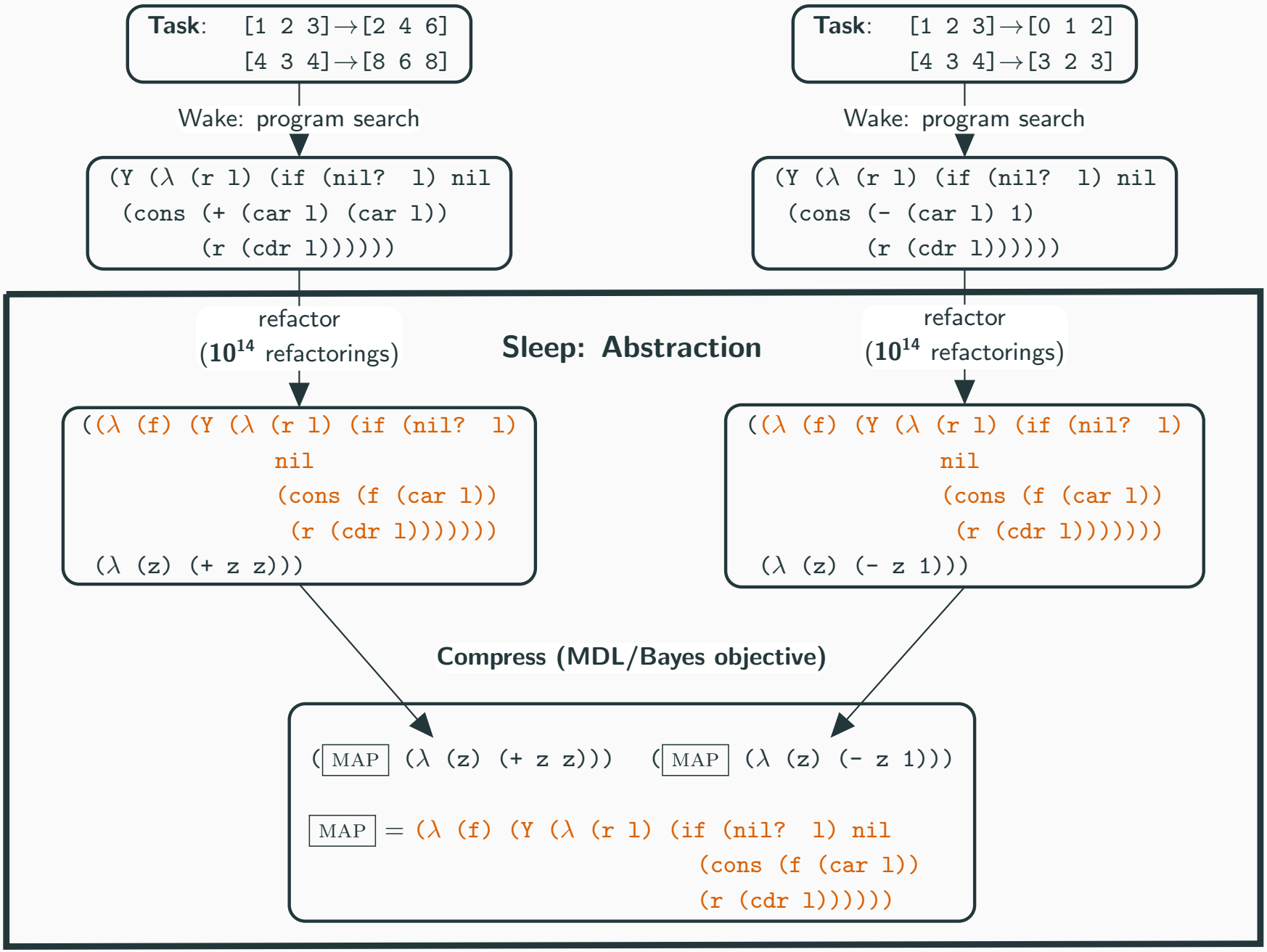




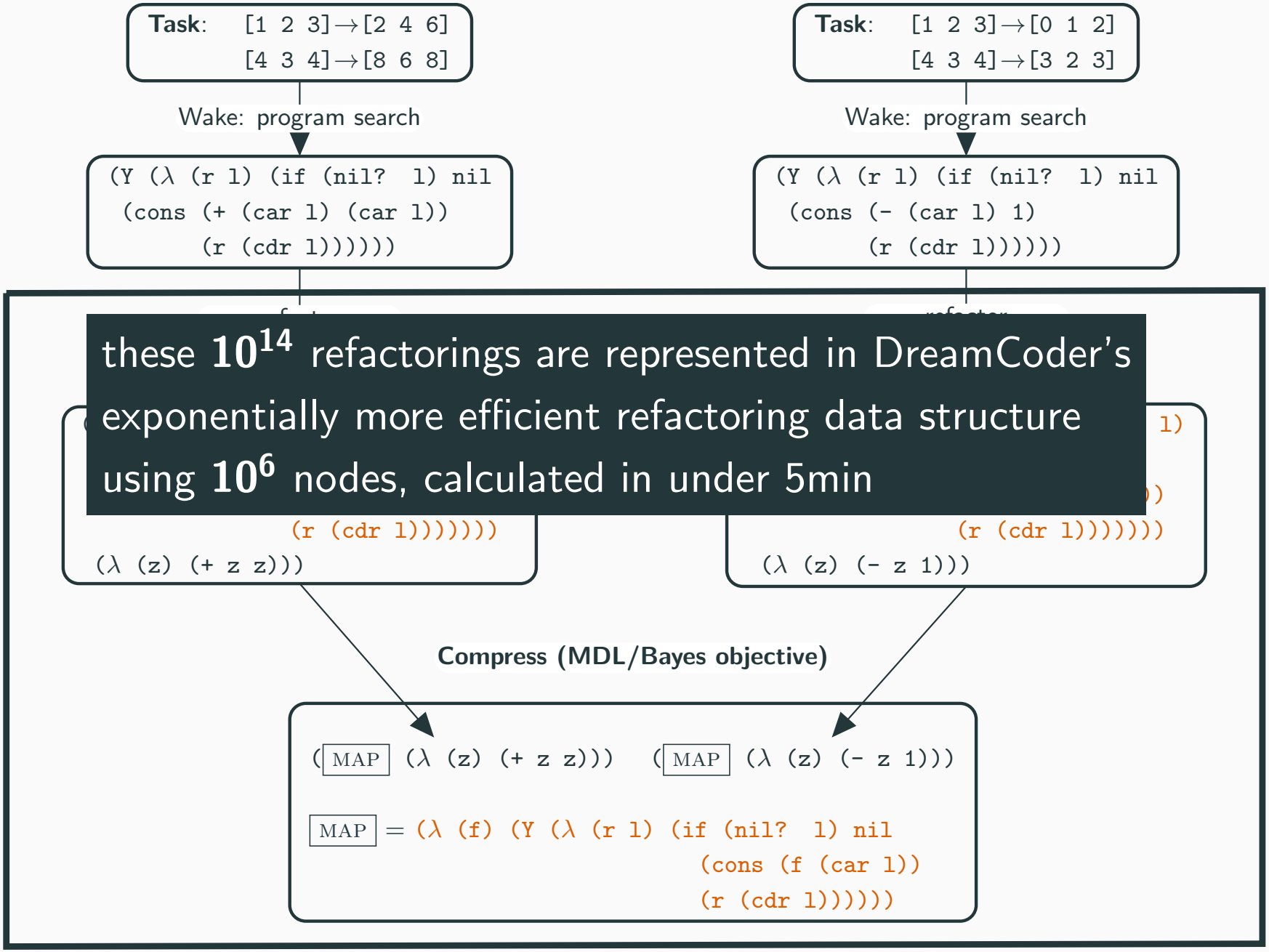
Abstraction Sleep: Growing the library via refactoring



Abstraction Sleep: Growing the library via refactoring



Abstraction Sleep: Growing the library via refactoring

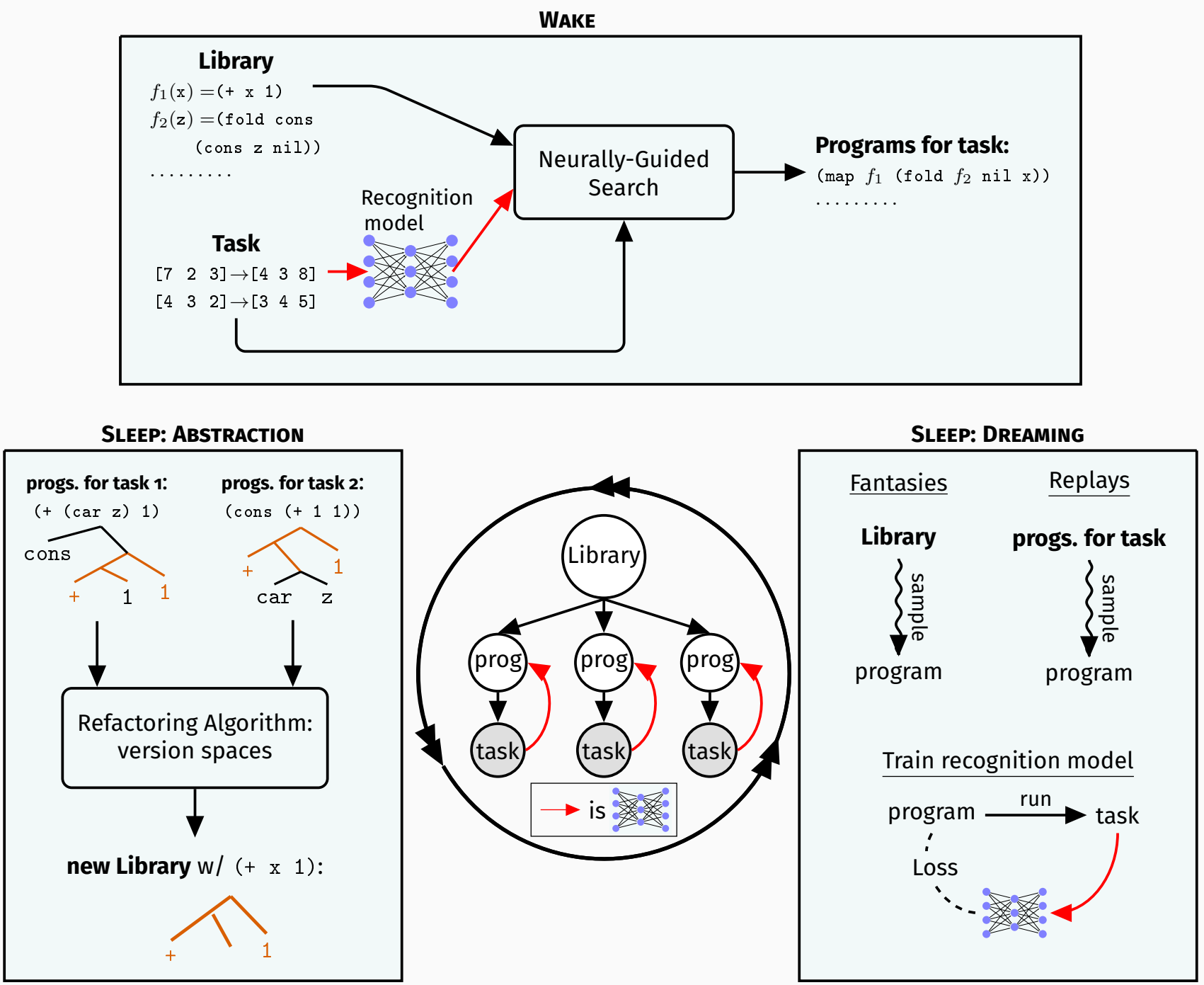


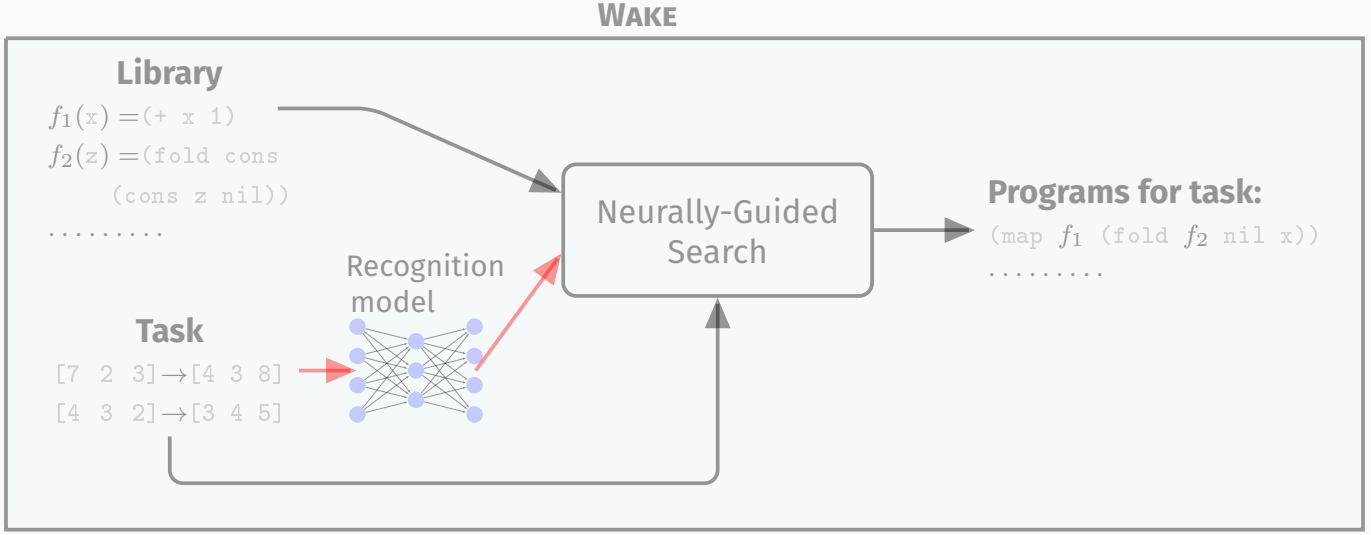
Program Induction and learning to learn

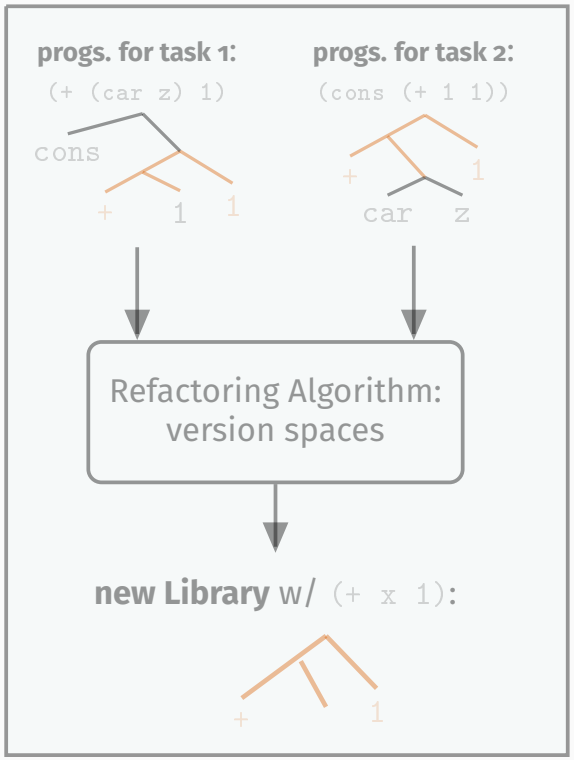
learning a DSL

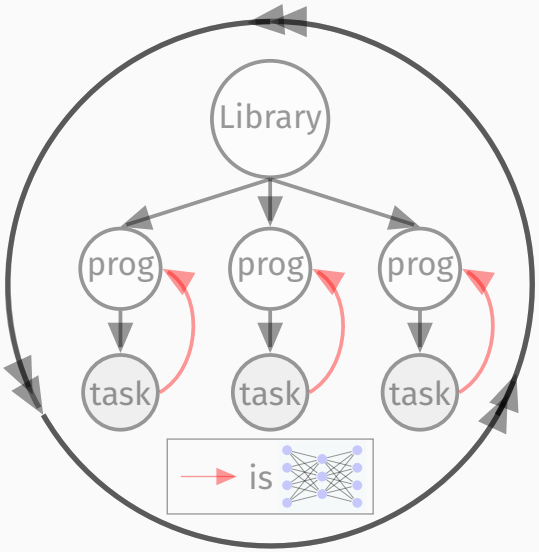
learning to synthesize

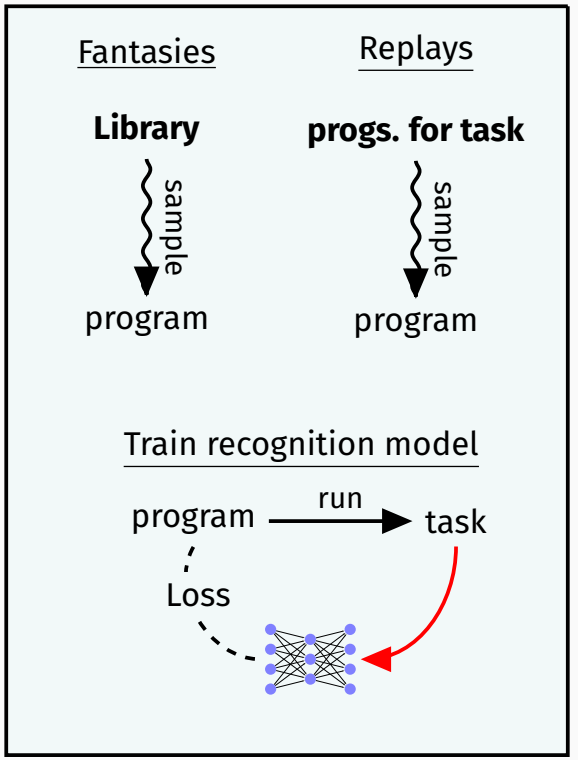
synergy between DSL+learned synthesizer


WAKE
Library
f_1(x) =(+ x 1)
f_2(z) =(fold cons
(cons z nil))
.....
Task
[7 2 3]→[4 3 8]
[4 3 2]→[3 4 5]
Recognition model
Neurally-Guided Search
Programs for task:
(map f_1 (fold f_2 nil x))
.....
SLEEP: ABSTRACTION
progs. for task 1:
(+ (car z) 1)
progs. for task 2:
(cons (+ 1 1))
cons
+
1
1
+
1
car
z
Refactoring Algorithm: version spaces
new Library w/ (+ x 1):
+
1
Library
prog
prog
prog
task
task
task
is
SLEEP: DREAMING
Fantasies
Replays
Library
progs. for task
sample
sample
program
program
Train recognition model
program
run
task
Loss

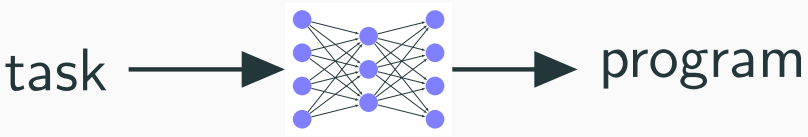

WAKE
Library
$f_1(x) = (+\ x\ 1)$
$f_2(z) = (\text{fold cons } (\text{cons } z \text{ nil}))$
.....
Recognition model
Task
[7 2 3]→[4 3 8]
[4 3 2]→[3 4 5]
Neurally-Guided Search
Programs for task:
(map f_1 (fold f_2 nil x))
.....


SLEEP: ABSTRACTION
progs. for task 1:
(+ (car z) 1)
progs. for task 2:
(cons (+ 1 1))
cons
+
1
1
+
1
car
z
Refactoring Algorithm: version spaces
new Library w/ (+ x 1):
+
1

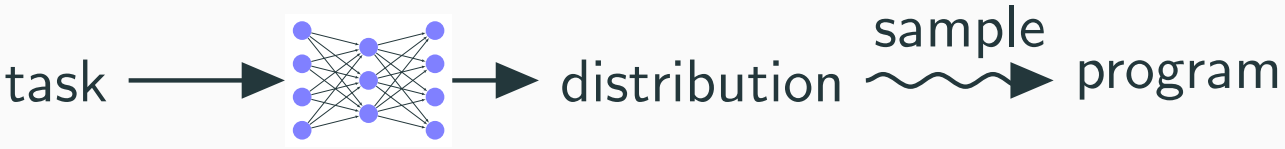

Library
prog
prog
prog
task
task
task
is


SLEEP: DREAMING
Fantasies
Replays
Library
progs. for task
sample
sample
program
program
Train recognition model
program
run
task
Loss

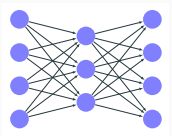
Neural recognition model guides search



Neural recognition model guides search



Neural recognition model guides search

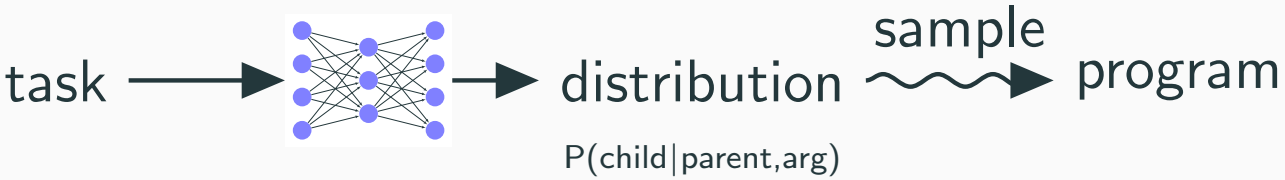


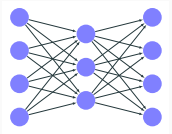
is a...

recurrent network (Devlin et al 2017)

unigram model (Menon et al 2013; Balog et al 2016)

Neural recognition model guides search

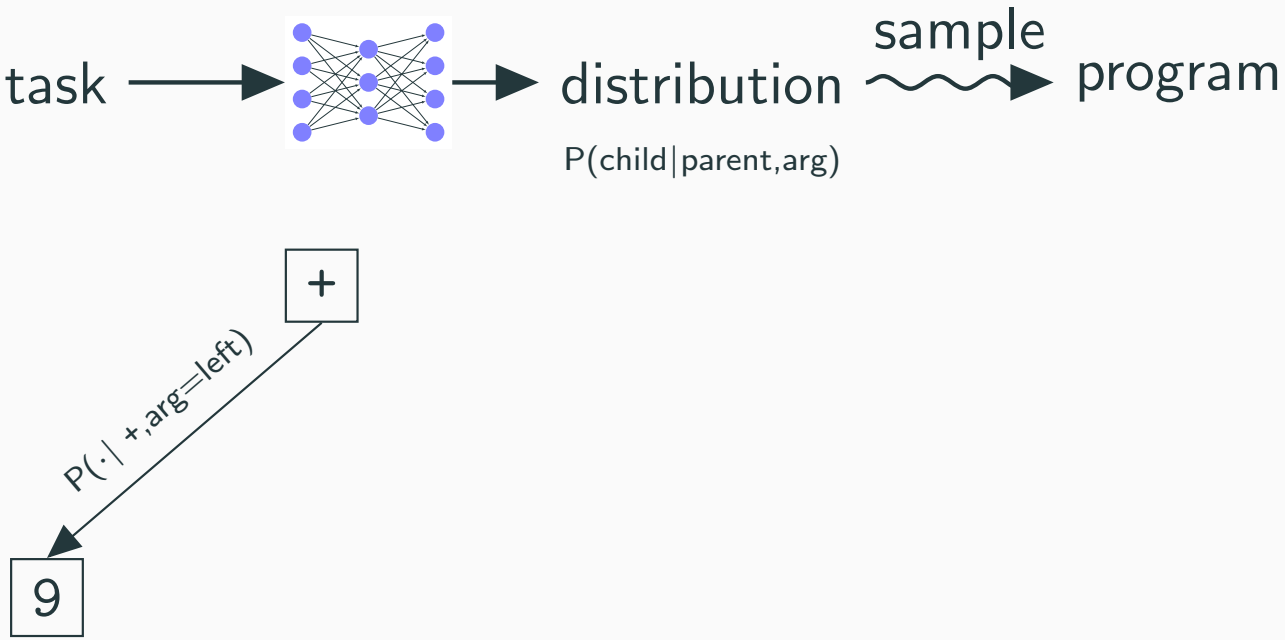




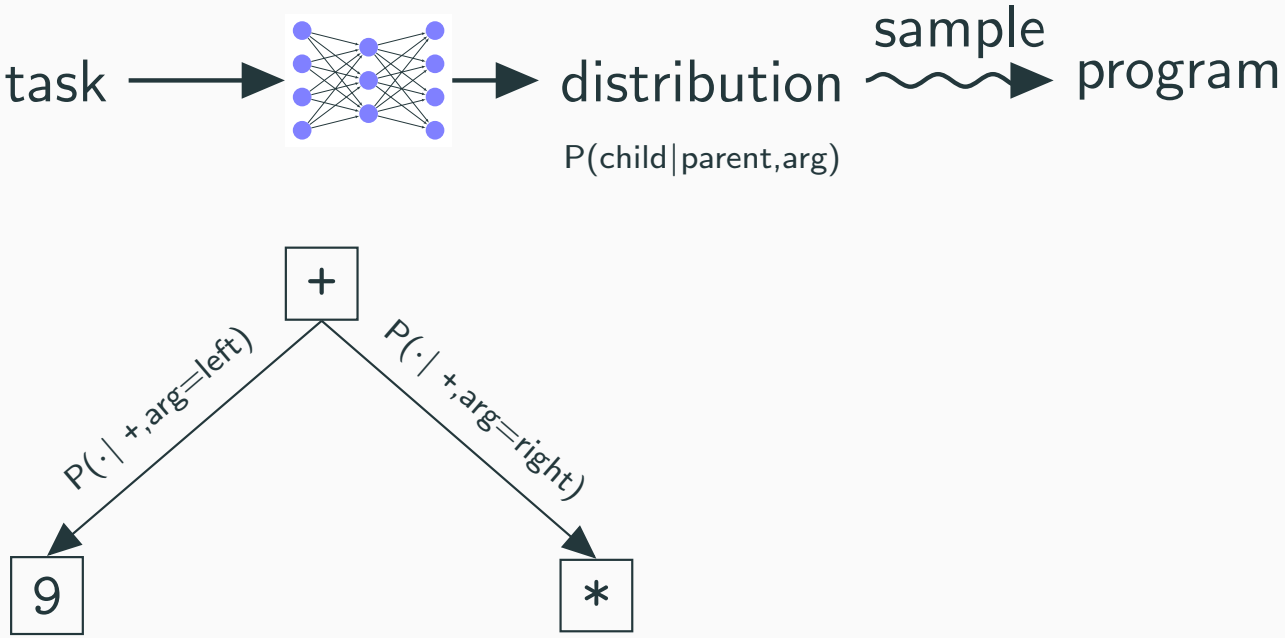
is a **"bigram" model over syntax trees**

Neural recognition model guides search

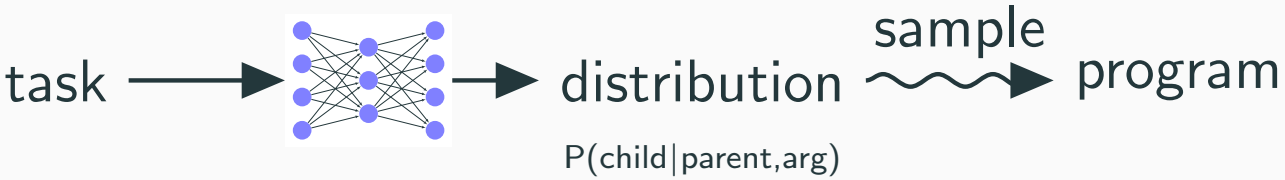
Neural recognition model guides search

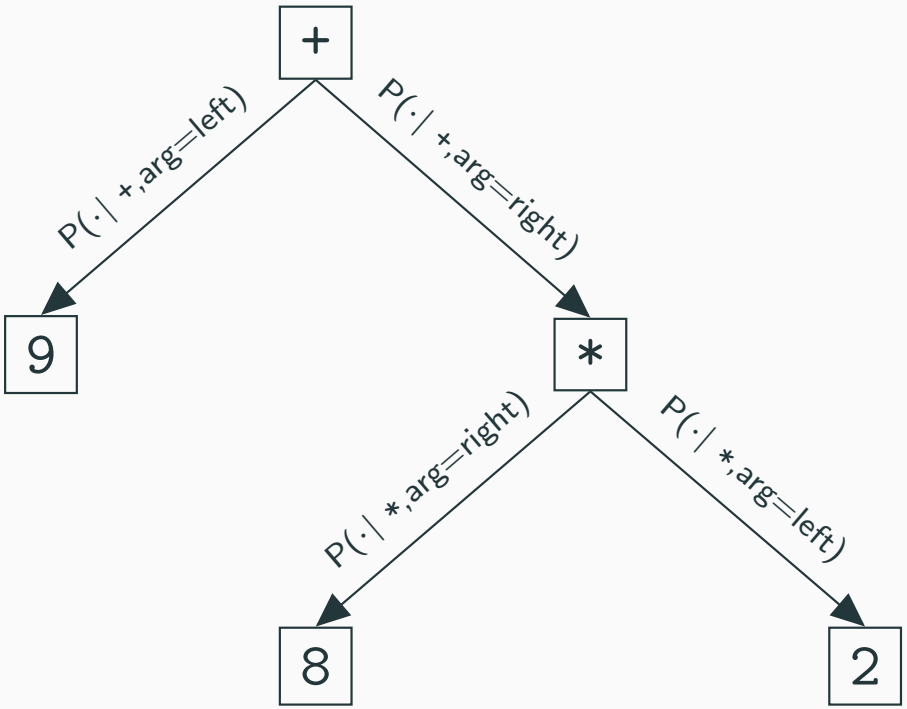


Neural recognition model guides search

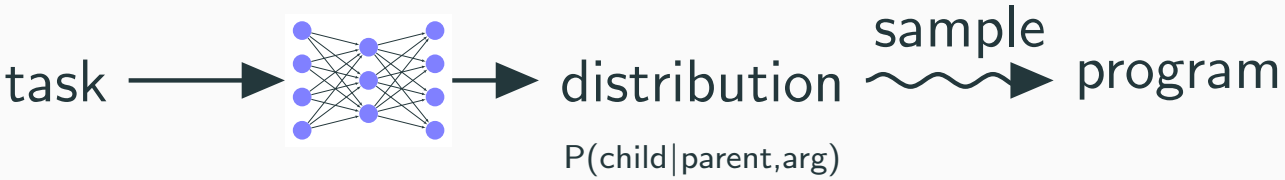


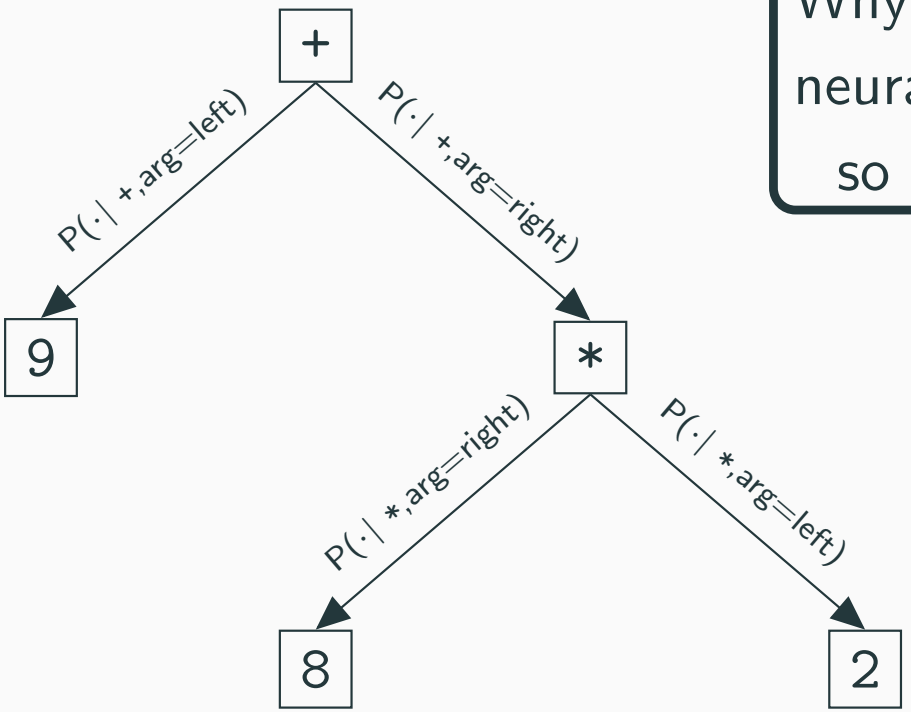
Neural recognition model guides search





Neural recognition model guides search



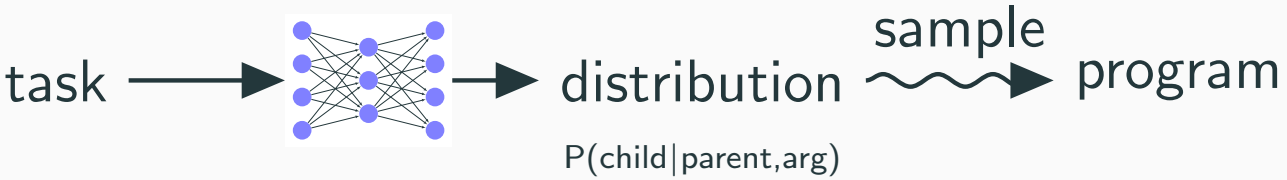


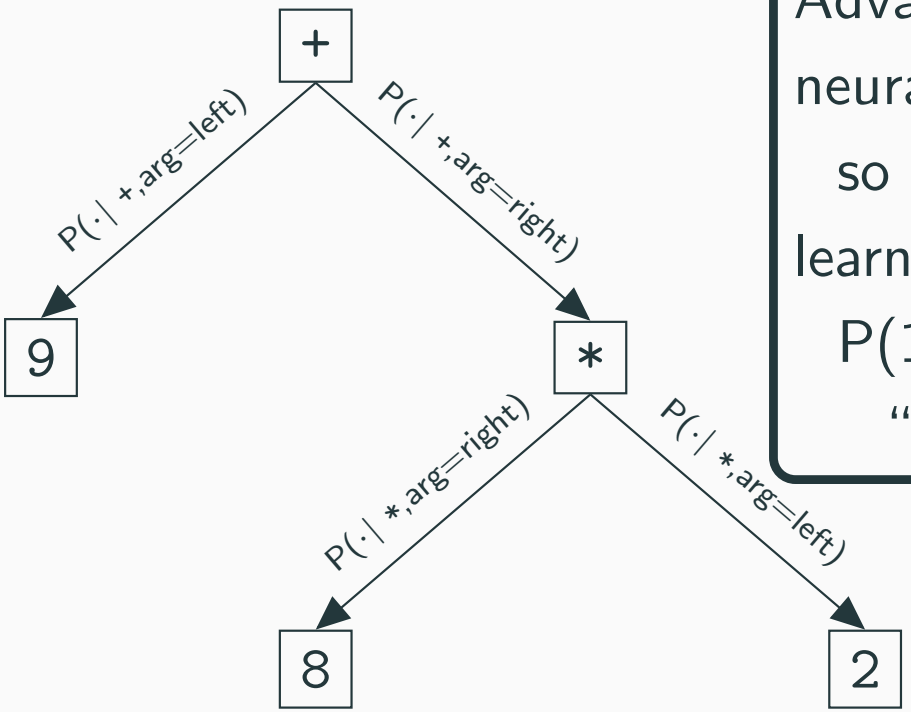
Why we do this:

neural net runs once per task,

so CPU bottlenecks instead of GPU

Neural recognition model guides search





Advantages:

neural net runs once per task,
so CPU bottlenecks instead of GPU

learns to break syntactic symmetries:
P(1|*,arg=left)=0.0
"do not multiply by one"

Program Induction and learning to learn

learning a DSL

learning to synthesize

synergy between DSL+learned synthesizer

DreamCoder Domains

List Processing

Sum List

```
[1 2 3] → 6
[4 6 8 1] → 17
```

Double

```
[1 2 3] → [2 4 6]
[4 5 1] → [8 10 2]
```

Text Editing

Abbreviate

```
Allen Newell → A.N.
Herb Simon → H.S.
```

Drop Last Three

```
shrdlu → shr
shakey → sha
```

Regexes

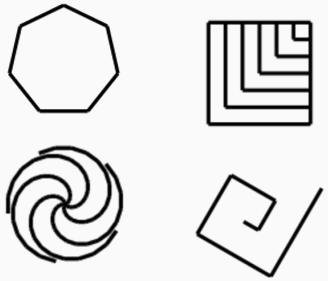
Phone numbers

```
(555) 867-5309
(650) 555-2368
```

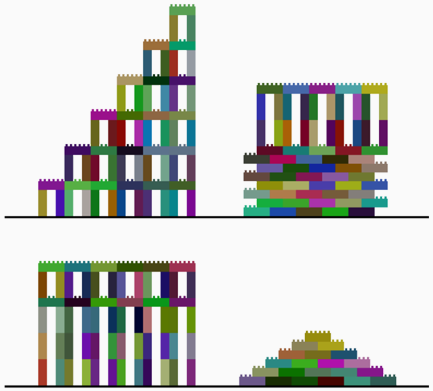
Currency

```
$100.25
$4.50
```

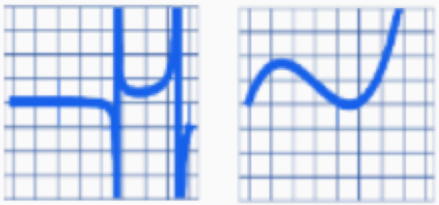
LOGO Graphics



Block Towers



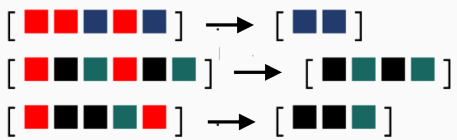
Symbolic Regression



$$y = f(x)$$

Recursive Programming

Filter Red



Physical Laws

$$\vec{a} = \frac{1}{m}\sum_i \vec{F}_i$$

$$\vec{F} \propto \frac{q_1 q_2}{|\vec{r}|^2}\hat{r}$$

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

DreamCoder Domains

List Processing

Sum List

```
[1 2 3] → 6
[4 6 8 1] → 17
```

Double

```
[1 2 3] → [2 4 6]
[4 5 1] → [8 10 2]
```

Text Editing

Abbreviate

```
Allen Newell → A.N.
Herb Simon → H.S.
```

Drop Last Three

```
shrdlu → shr
shakey → sha
```

Regexes

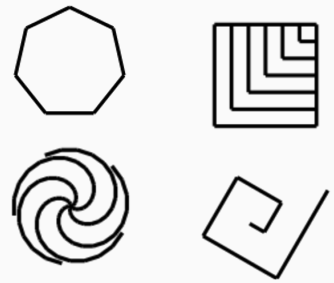
Phone numbers

```
(555) 867-5309
(650) 555-2368
```

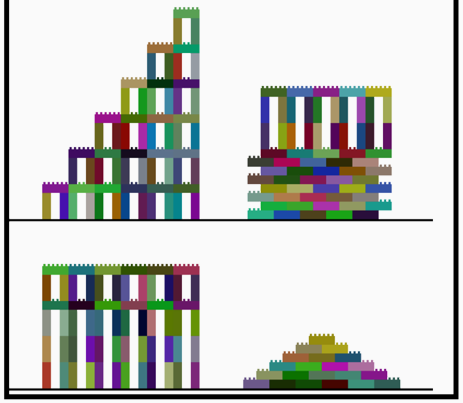
Currency

```
$100.25
$4.50
```

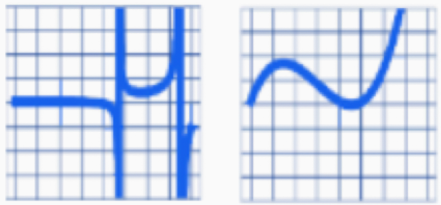
LOGO Graphics



Block Towers



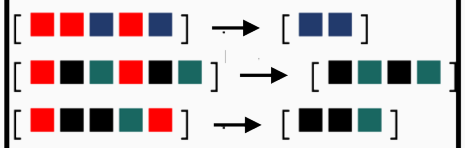
Symbolic Regression



$$y = f(x)$$

Recursive Programming

Filter Red



Physical Laws

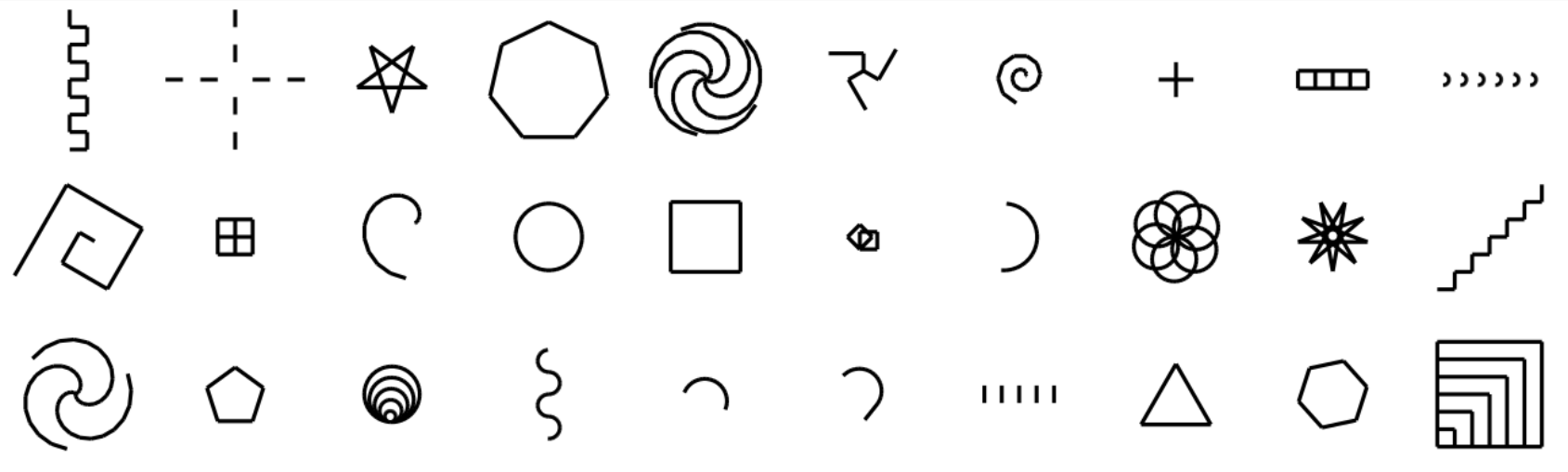
$$\vec{a} = \frac{1}{m}\sum_i \vec{F}_i$$

$$\vec{F} \propto \frac{q_1 q_2}{|\vec{r}|^2}\hat{r}$$

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

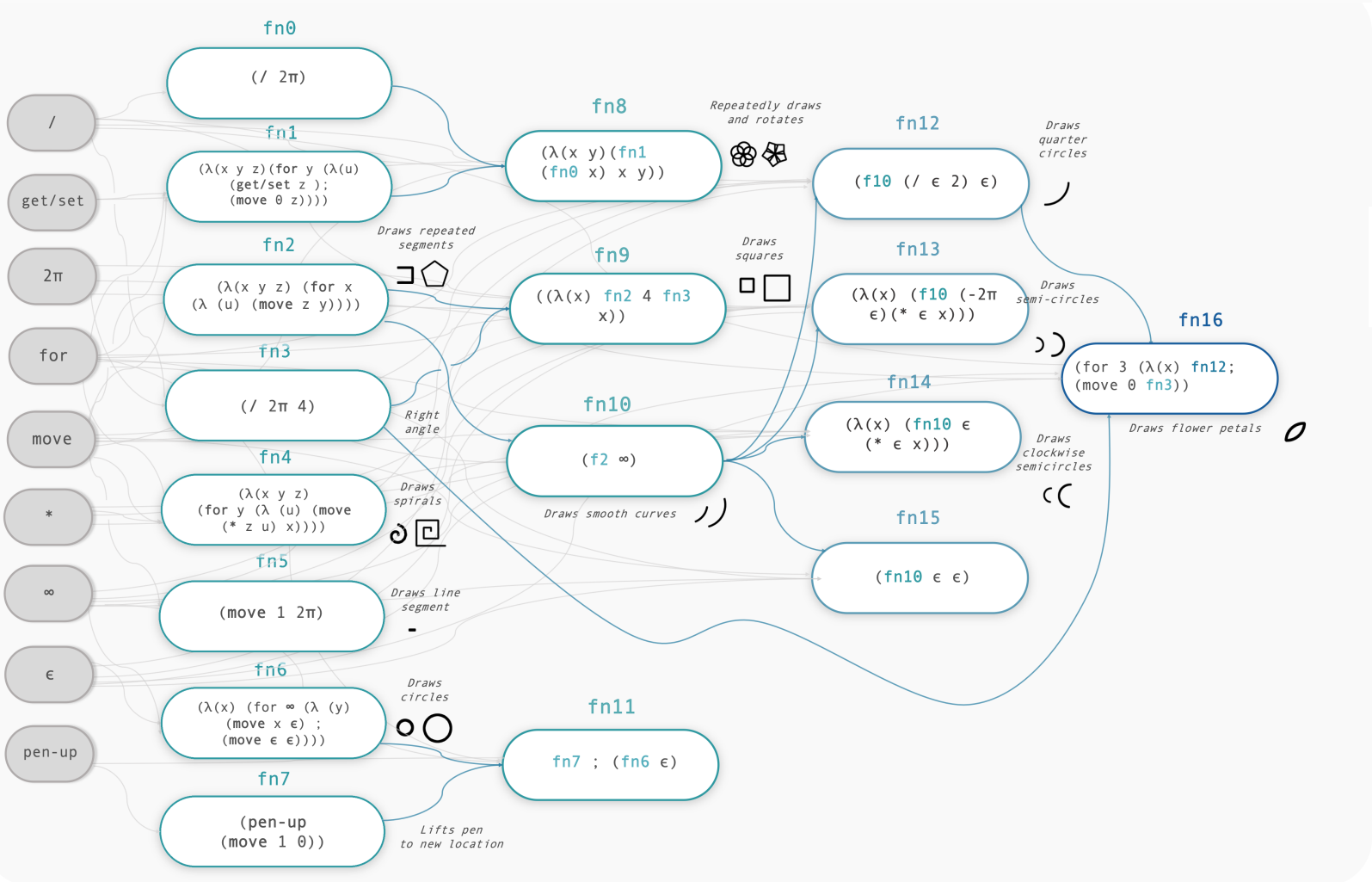
LOGO Turtle Graphics

30 out of 160 tasks

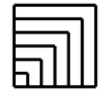


Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

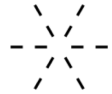
LOGO Turtle Graphics – learning an interpretable library



(fn8 5 (fn4 (* ϵ 2) ∞ ϵ))



(for 7 (λ (x) (fn9 x)))

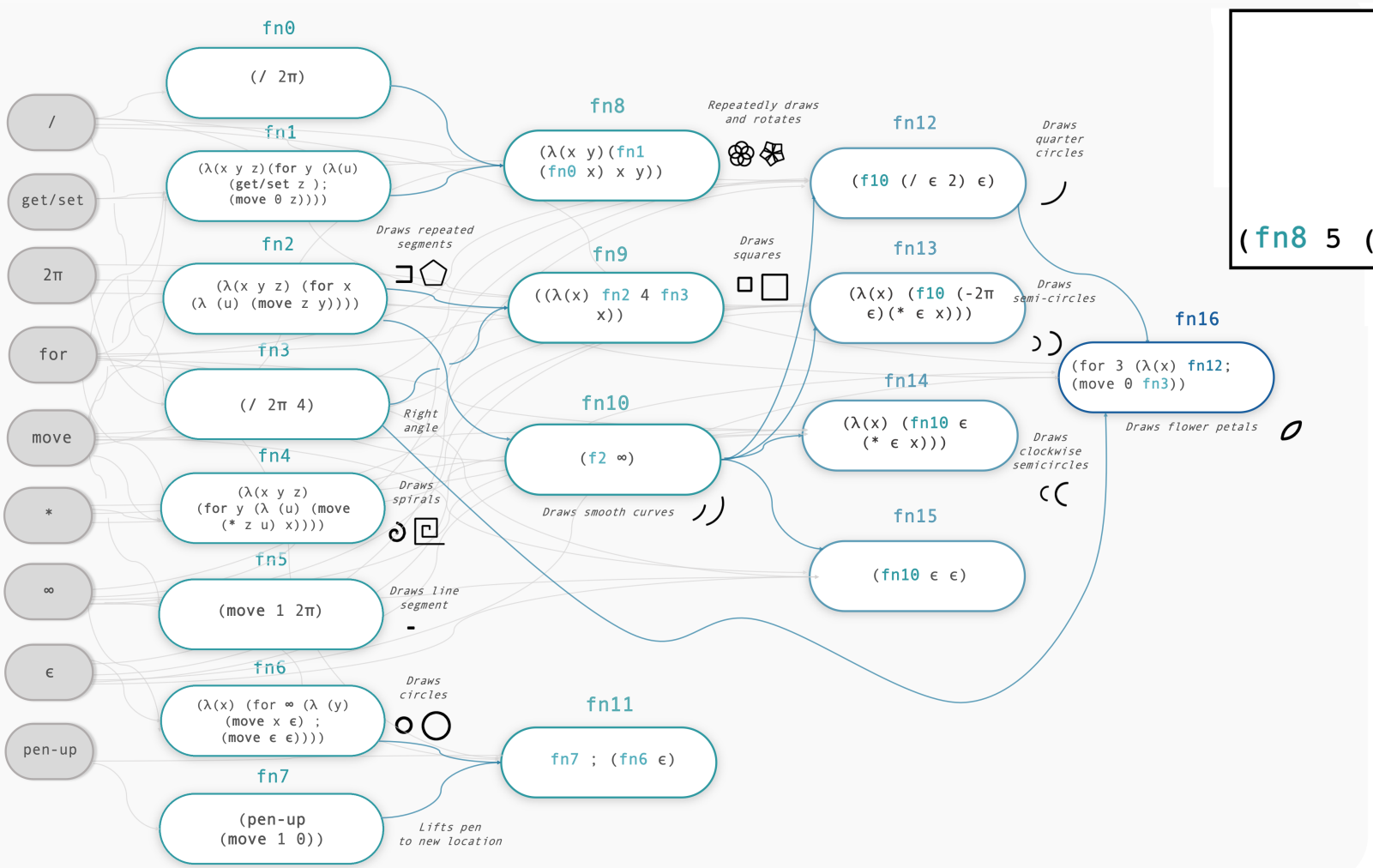


(fn8 6 (fn7 ; fn5 ; fn7 ; fn5))

(move 0 (fn0 7)); fn5 ; (fn13 4)

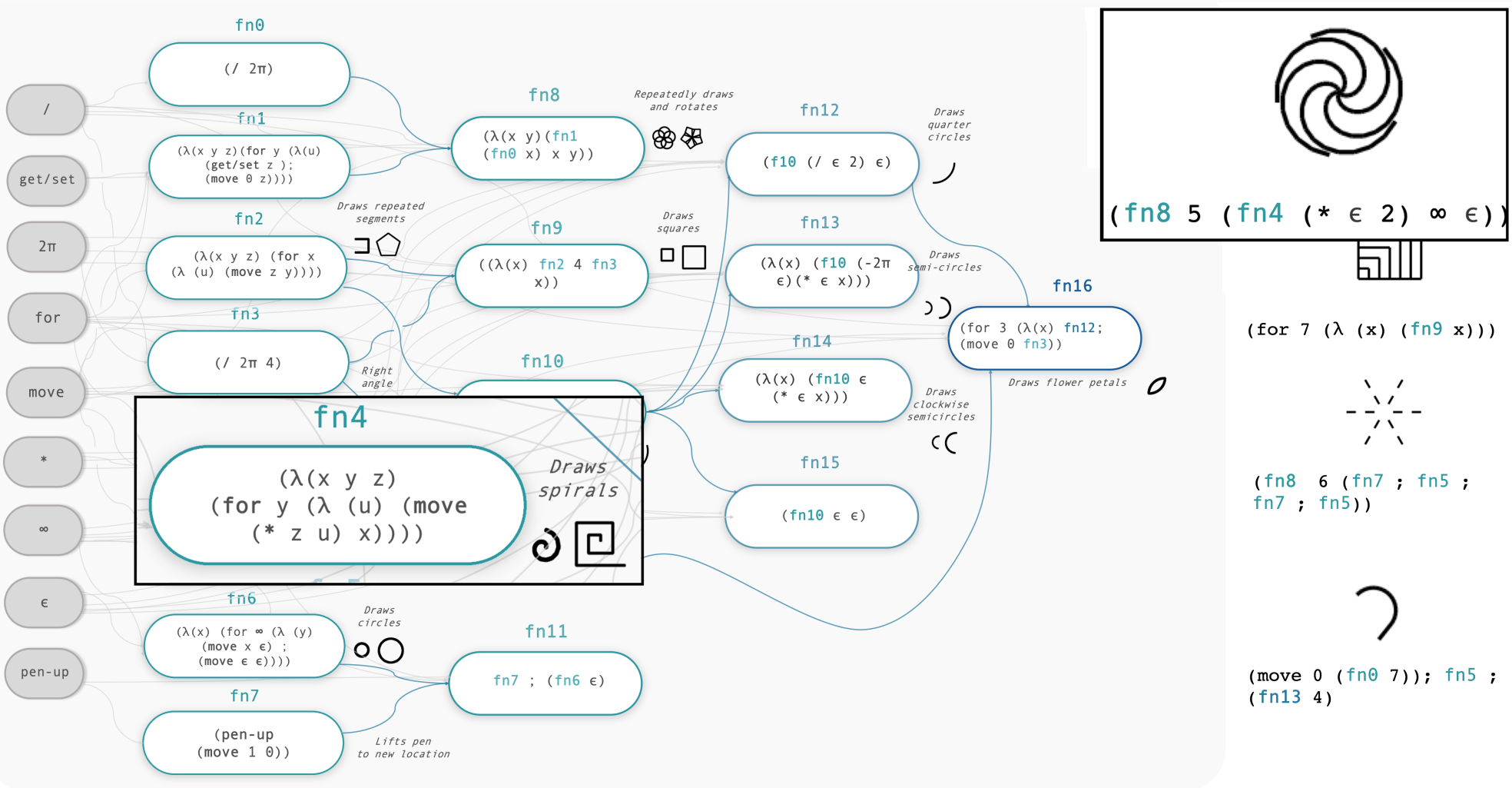
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

LOGO Turtle Graphics – learning an interpretable library



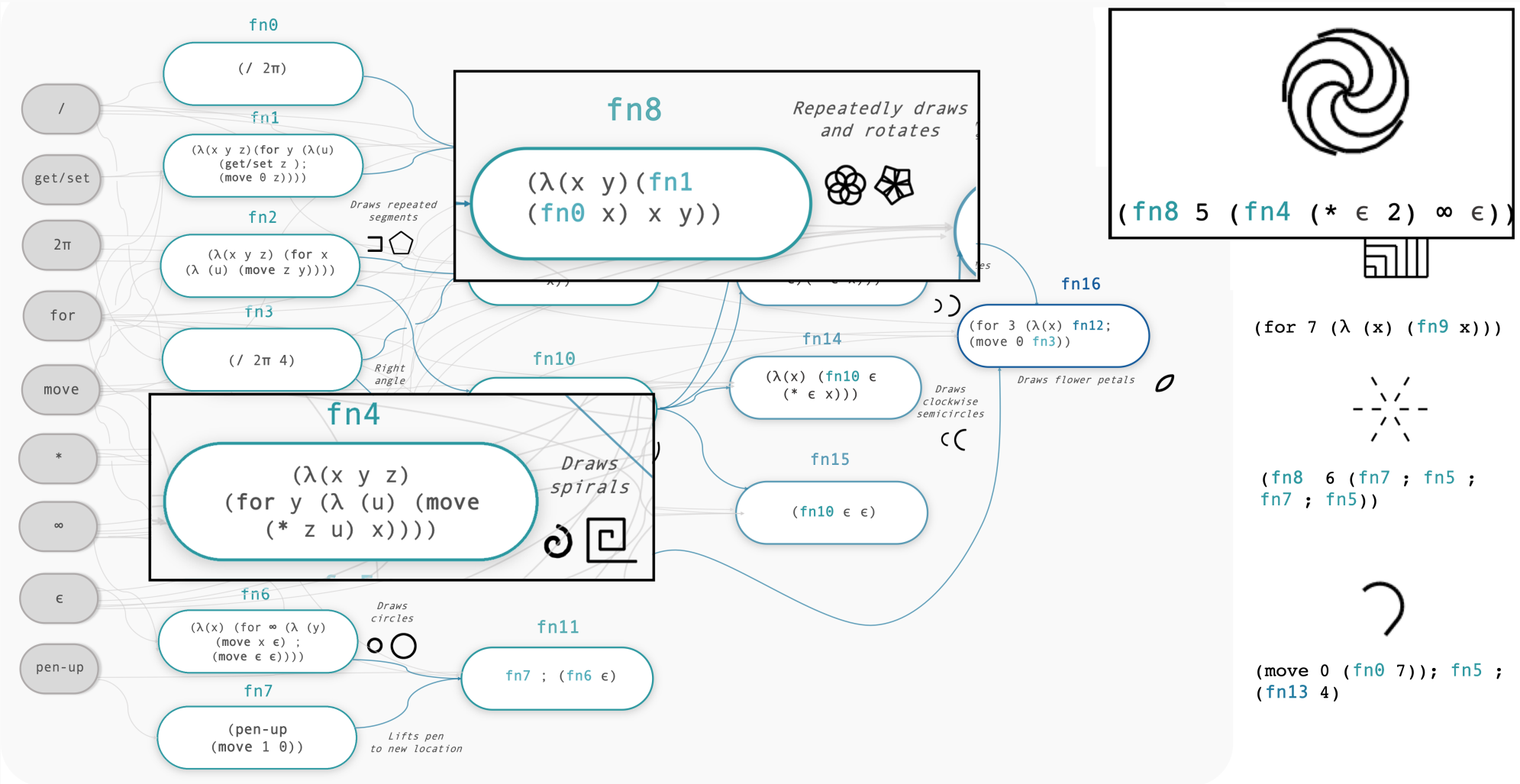
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

LOGO Turtle Graphics – learning an interpretable library



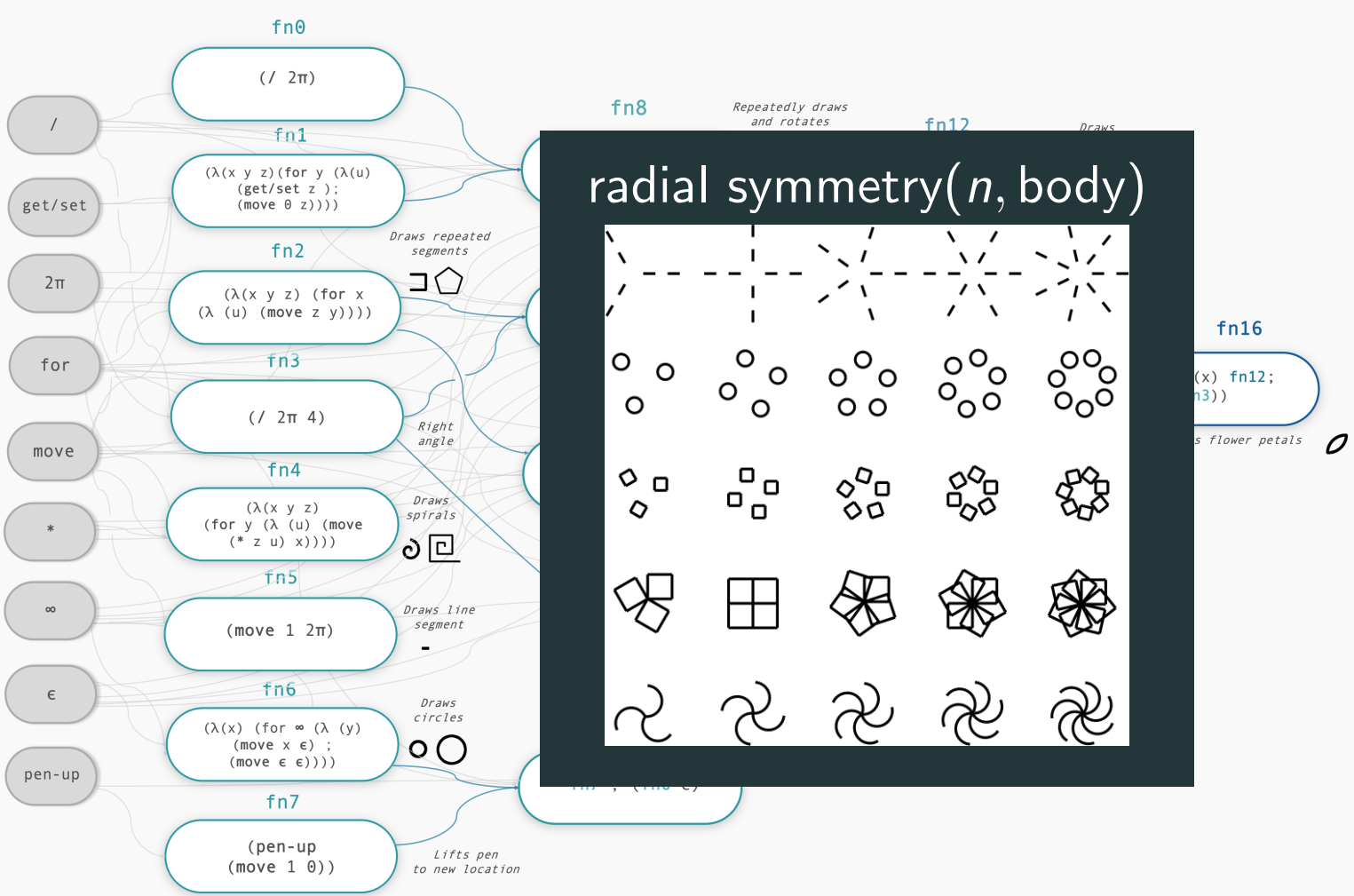
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

LOGO Turtle Graphics – learning an interpretable library



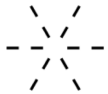
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

LOGO Turtle Graphics – learning an interpretable library



(fn8 5 (fn4 (* ϵ 2) ∞ ϵ))

(for 7 (λ (x) (fn9 x)))

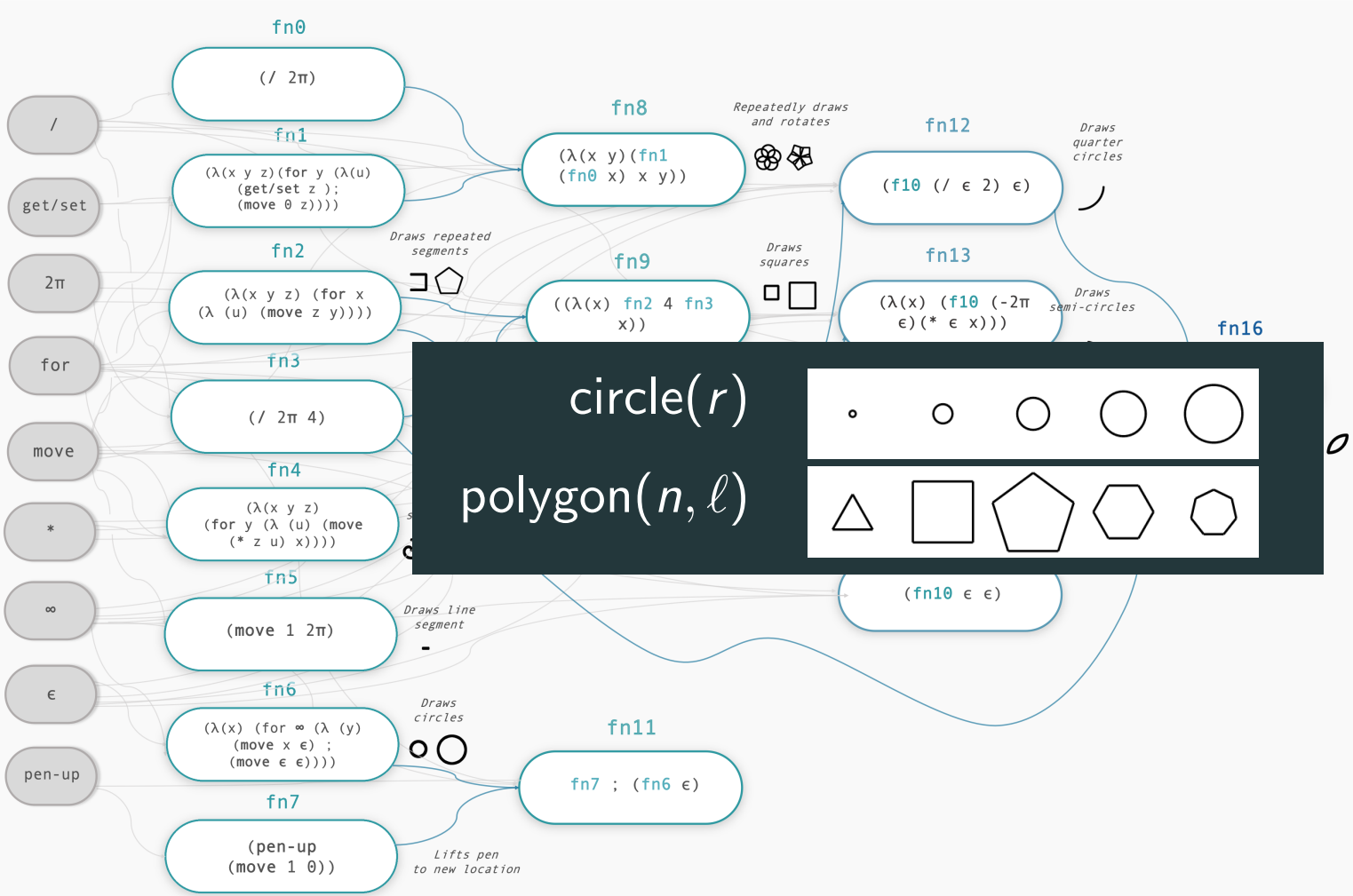


(fn8 6 (fn7 ; fn5 ; fn7 ; fn5))

(move 0 (fn0 7)); fn5 ; (fn13 4)

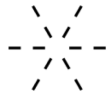
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

LOGO Turtle Graphics – learning an interpretable library



```
(fn8 5 (fn4 (* ϵ 2) ∞ ϵ))
```

```
(for 7 (λ (x) (fn9 x)))
```

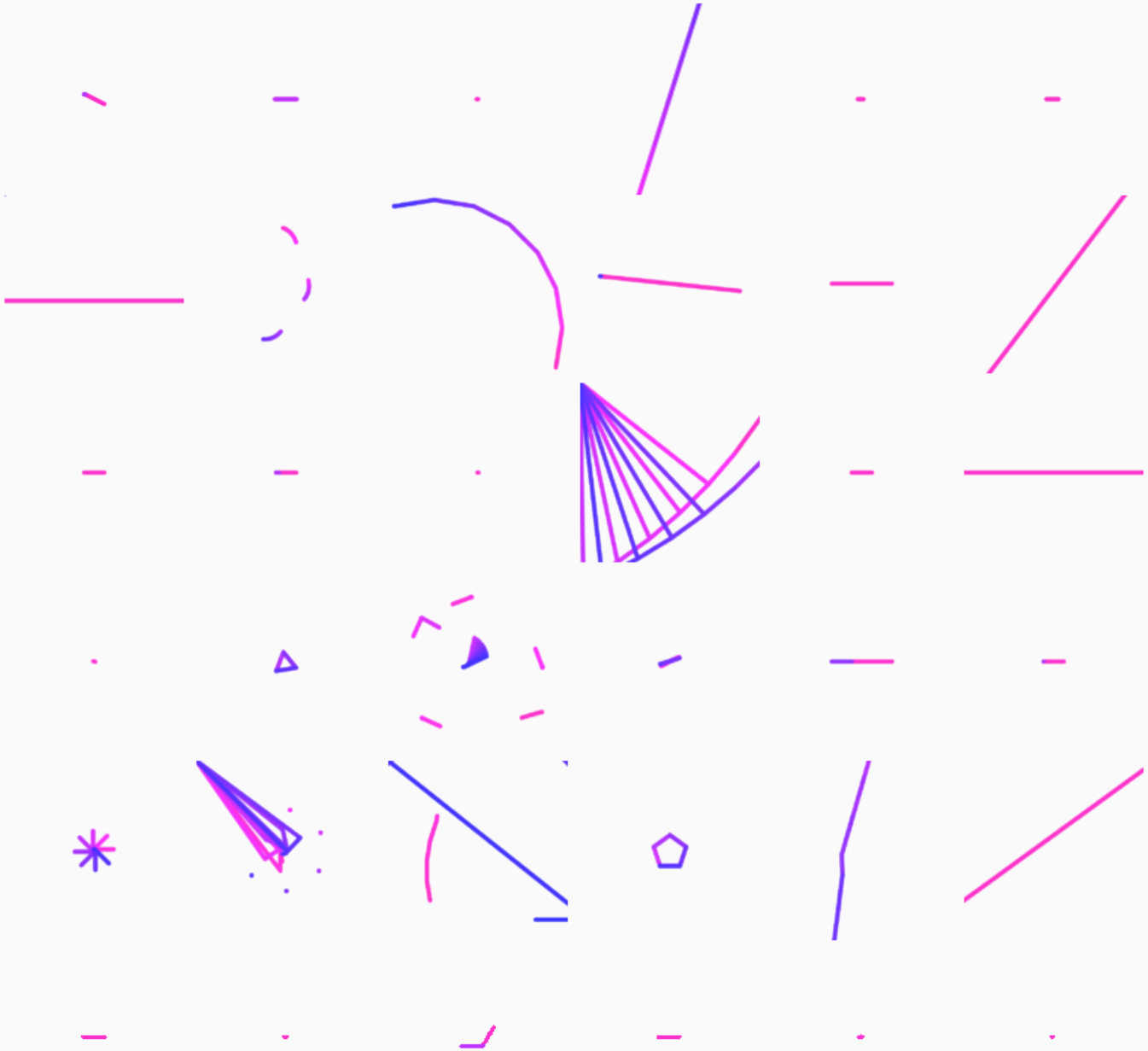


```
(fn8  6 (fn7 ; fn5 ;
fn7 ; fn5))
```

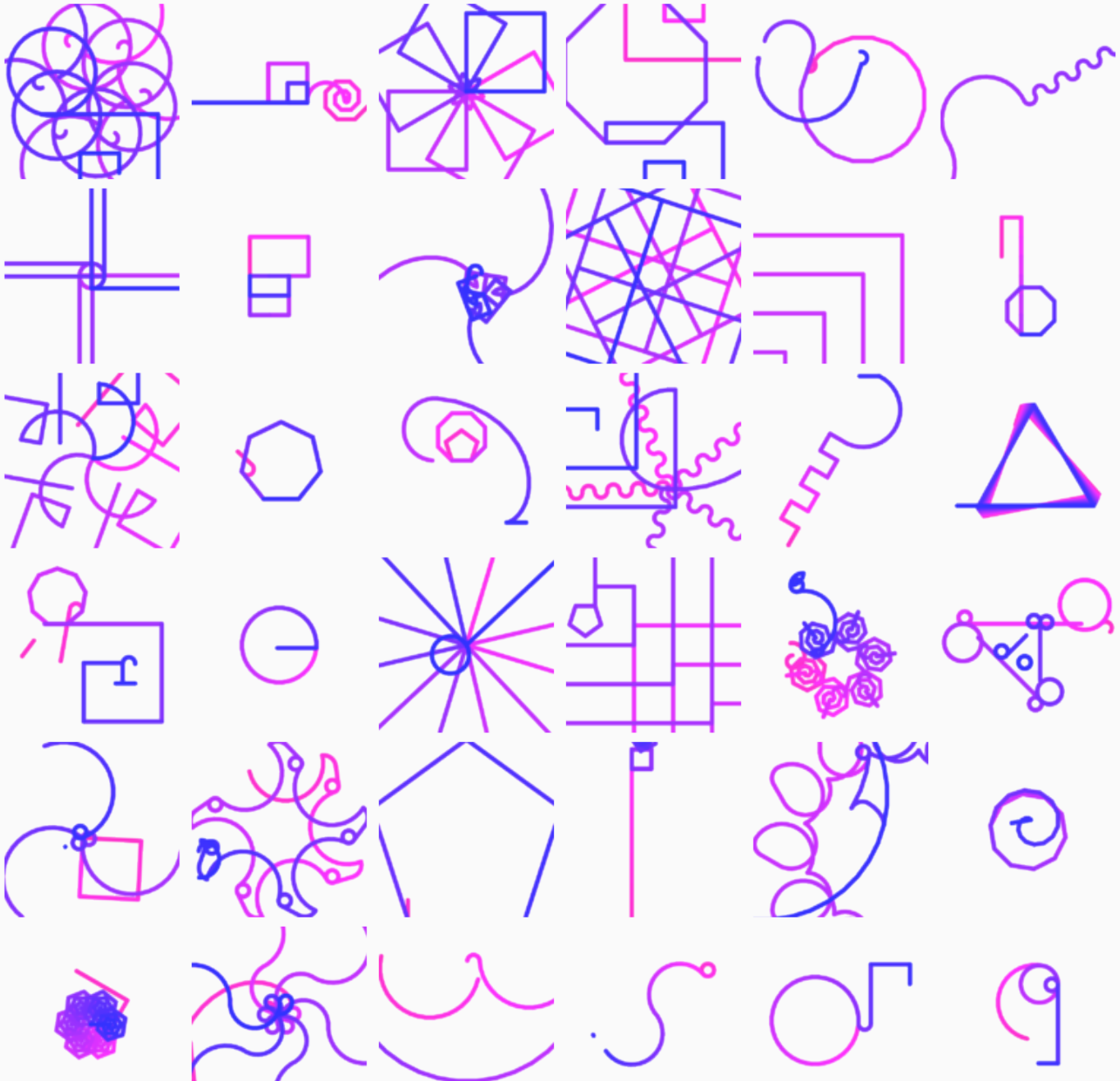
```
(move 0 (fn0 7)); fn5 ;
(fn13 4)
```

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

What does DreamCoder dream of? (before learning)

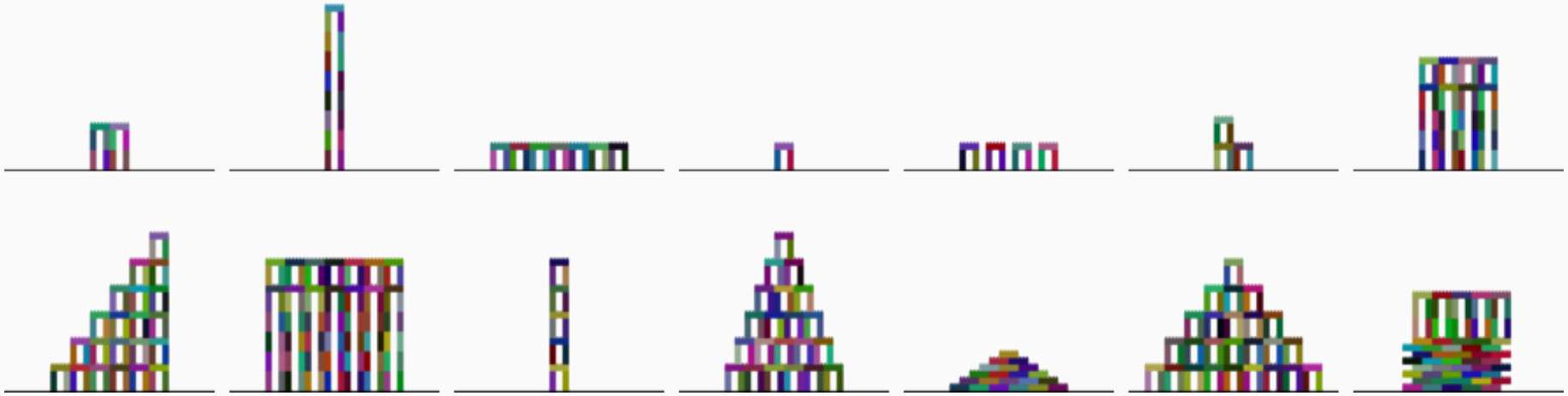


What does DreamCoder dream of? (after learning)



Planning to build towers

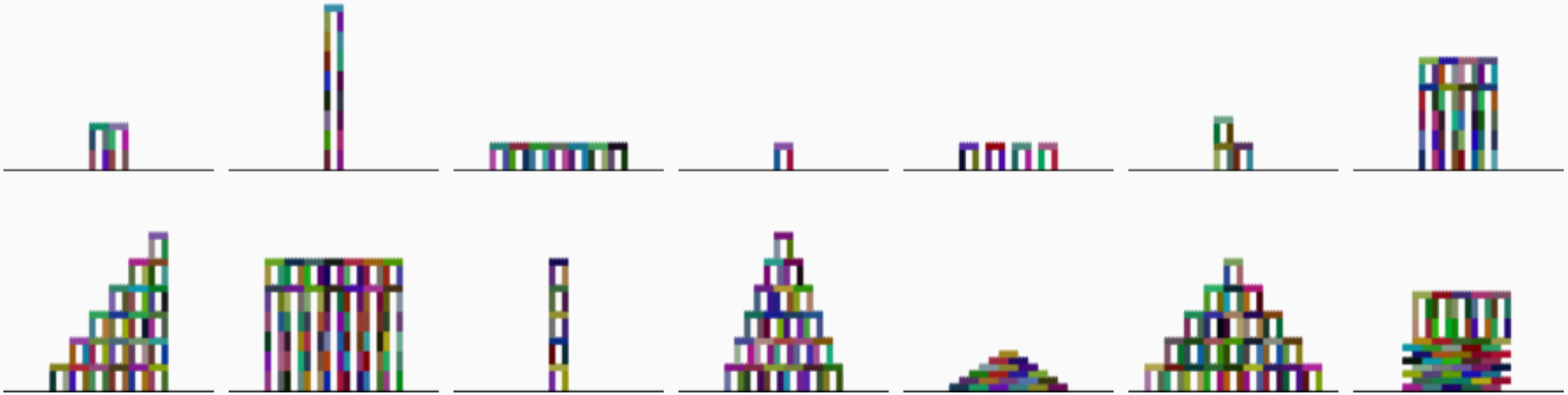
example tasks (112 total)



Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

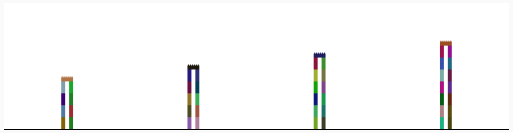
Planning to build towers

example tasks (112 total)

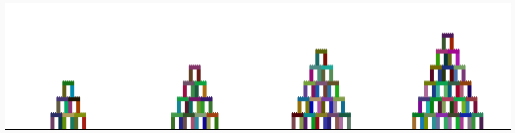


learned library routines ($\approx$ 20 total)

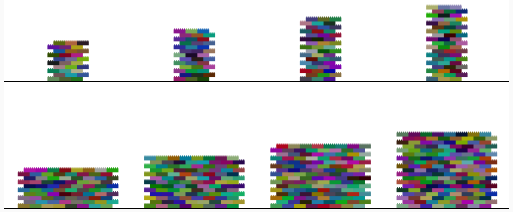
arch(h)



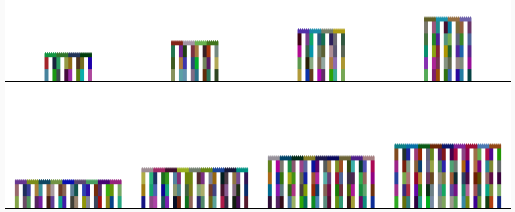
pyramid(h)



wall(w, h)

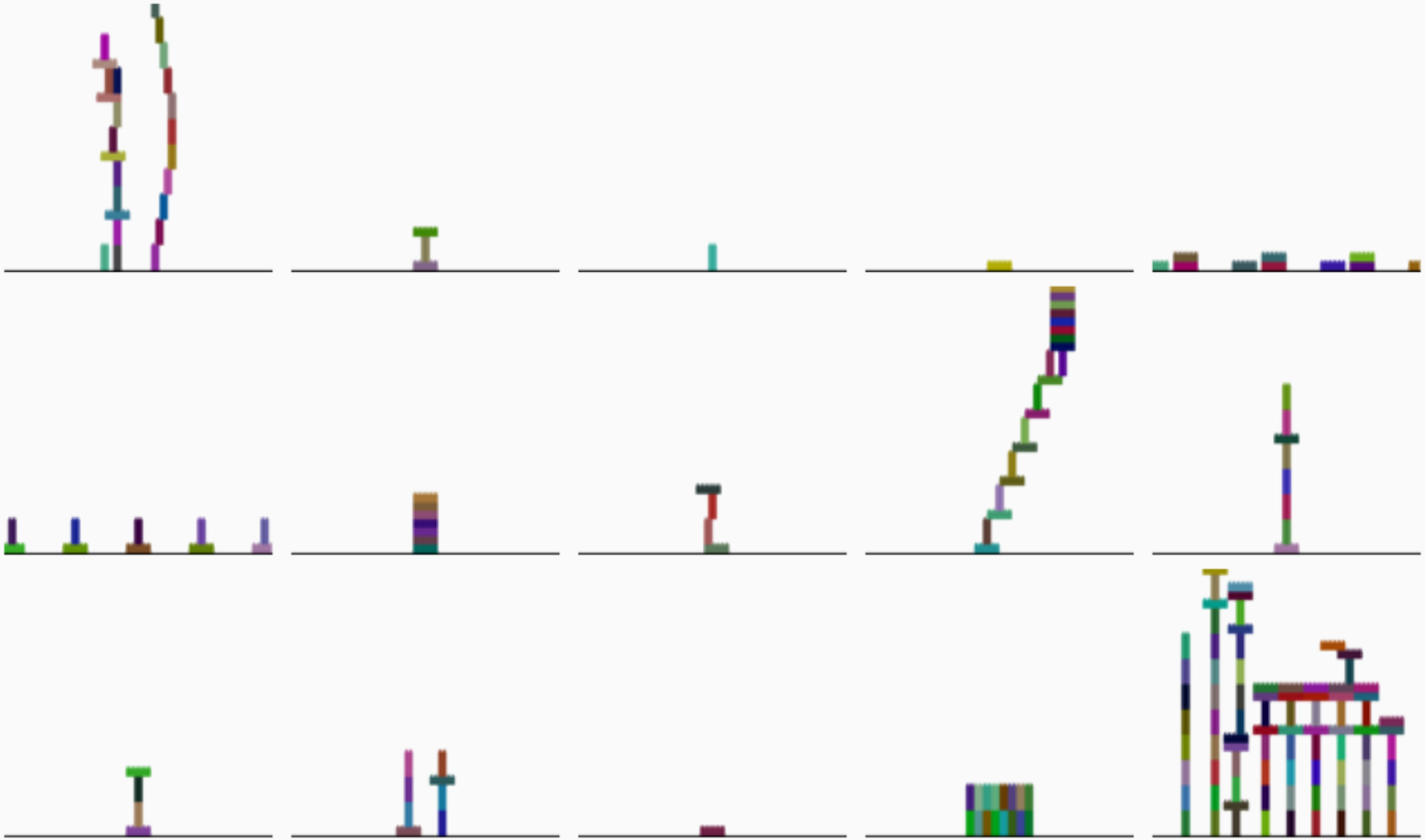


bridge(w, h)

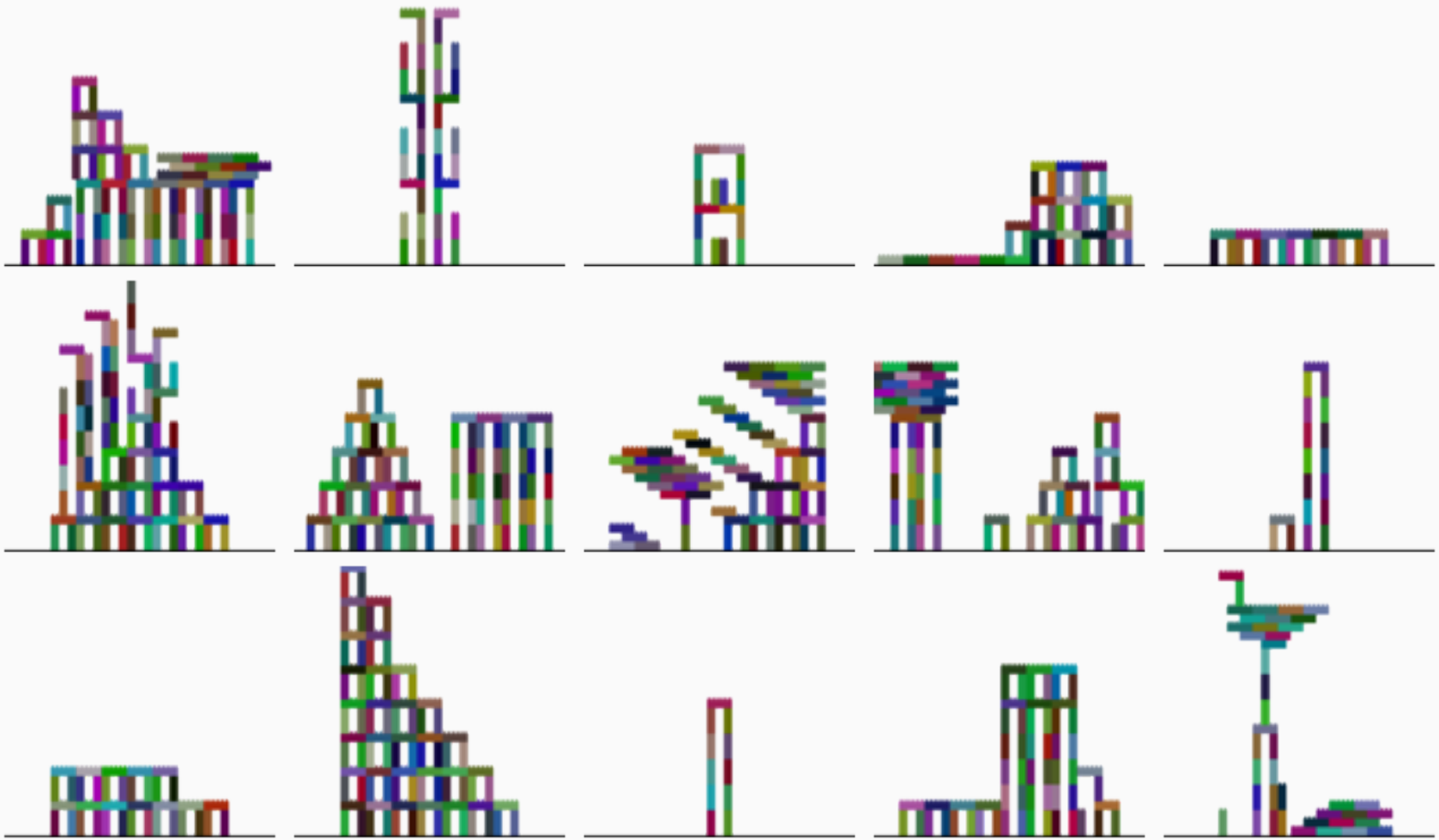


Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

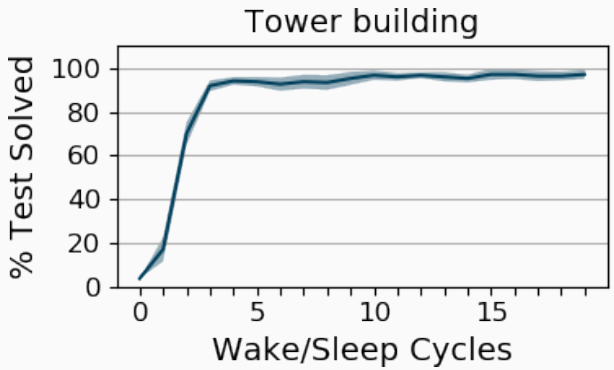
Dreams before learning



Dreams after learning

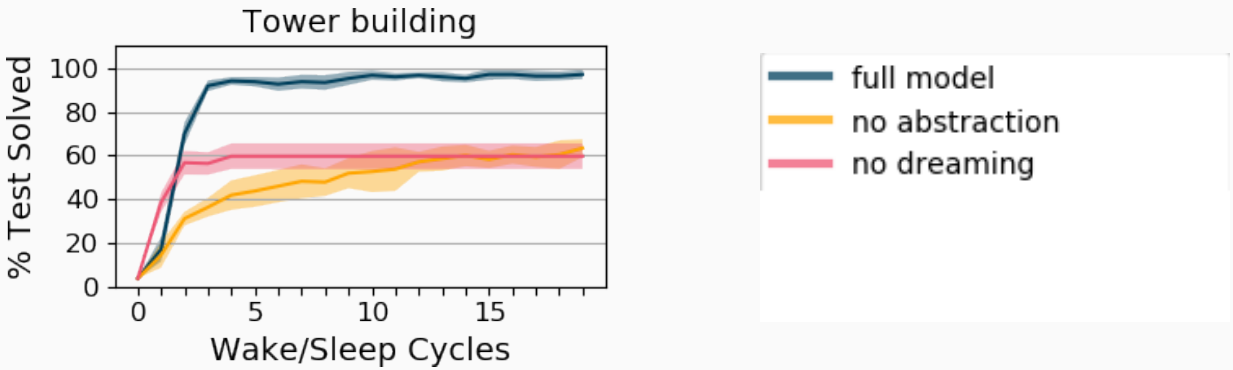


Learning dynamics



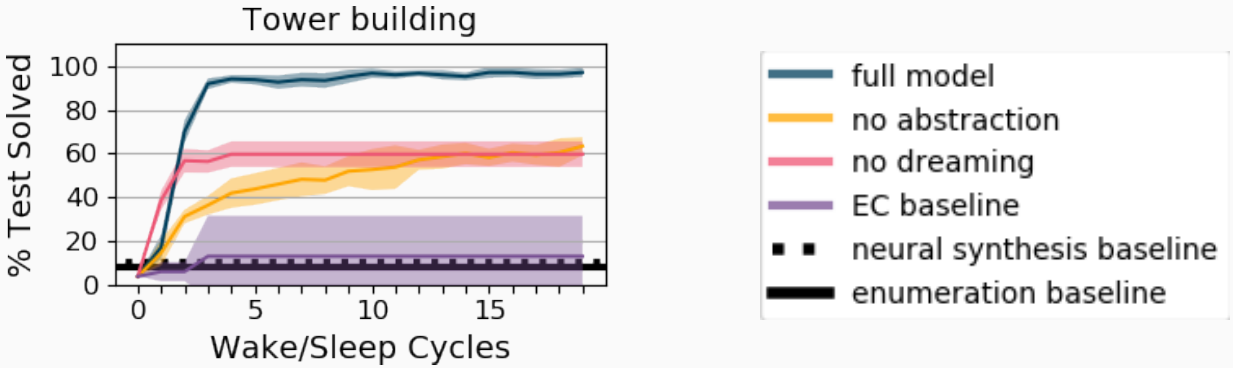
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Learning dynamics



Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

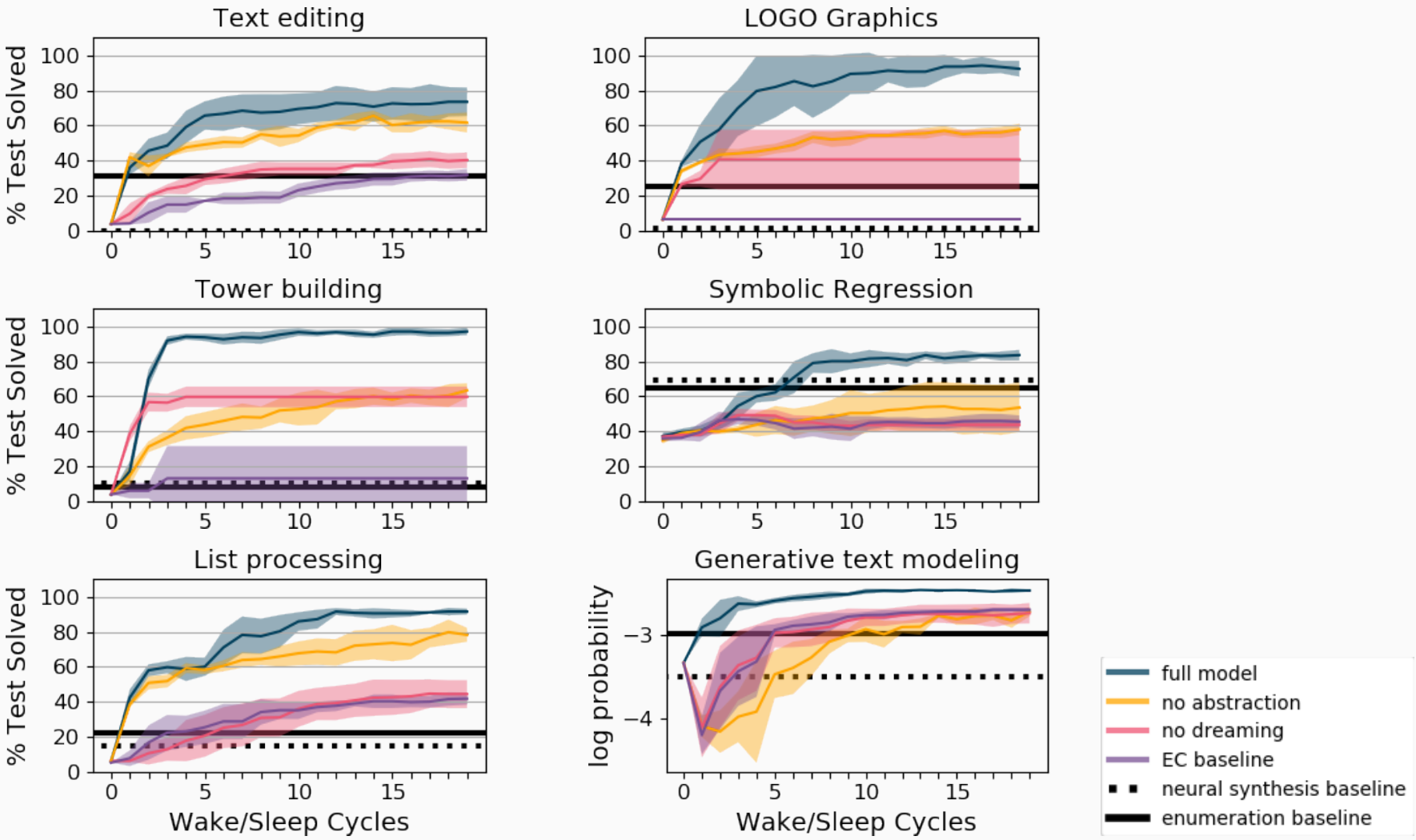
Learning dynamics



baselines: Exploration-Compression, EC [Dechter et al. 2013]
neural program synthesis, RobustFill [Devlin et al. 2017]
24 hours of brute-force enumeration

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Learning dynamics



Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Synergy between dreaming and library learning

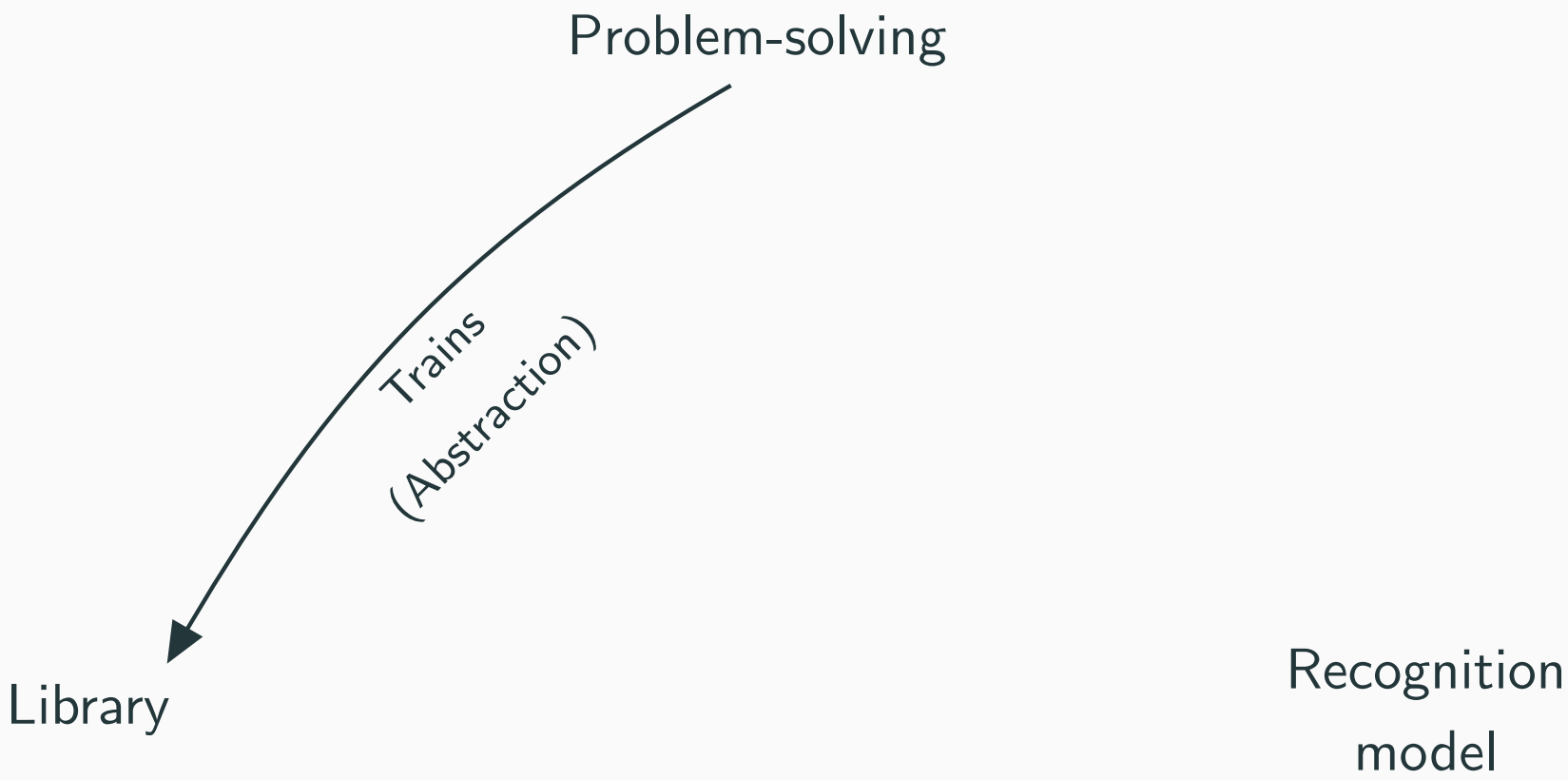
Problem-solving

Library

Recognition model

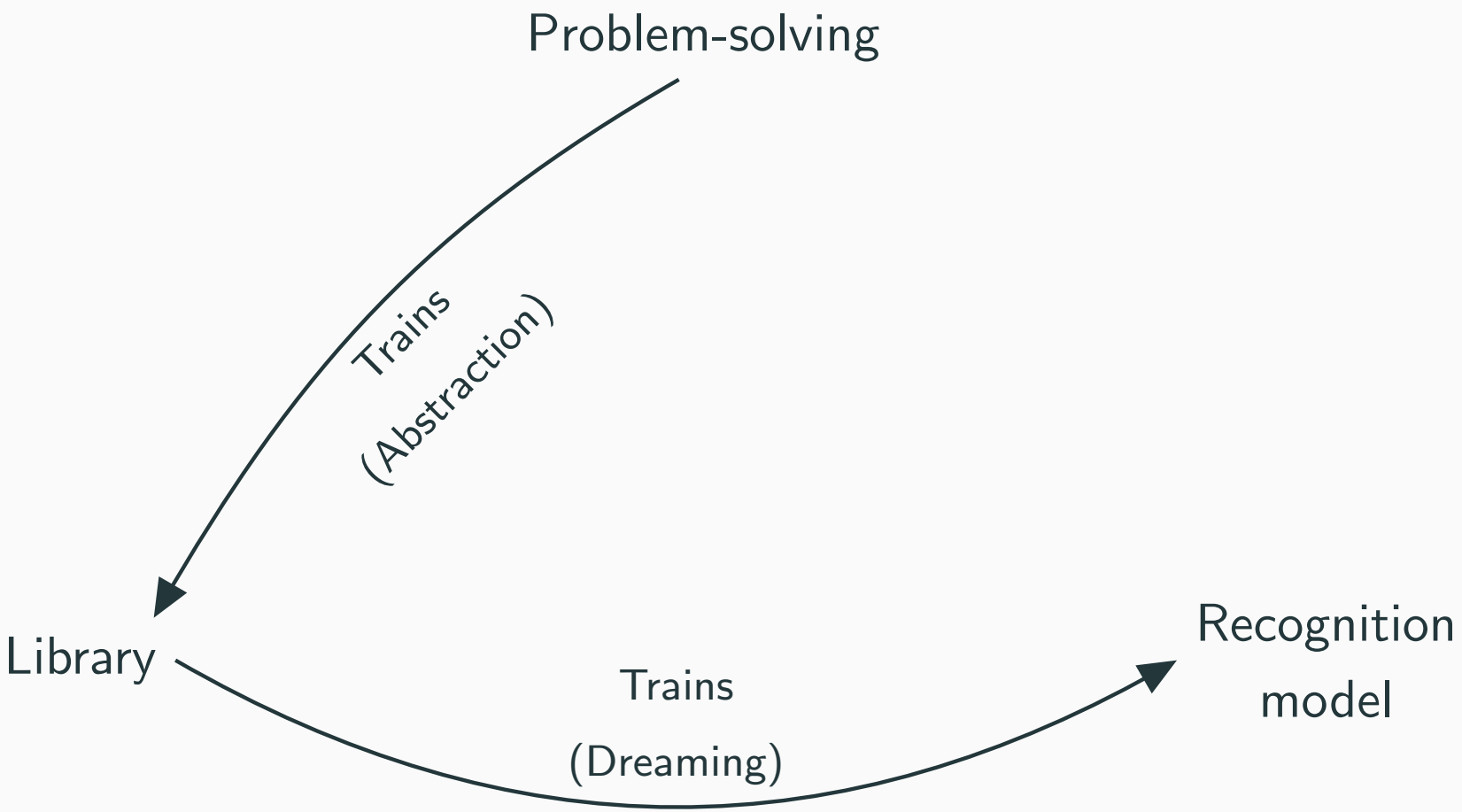
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Synergy between dreaming and library learning



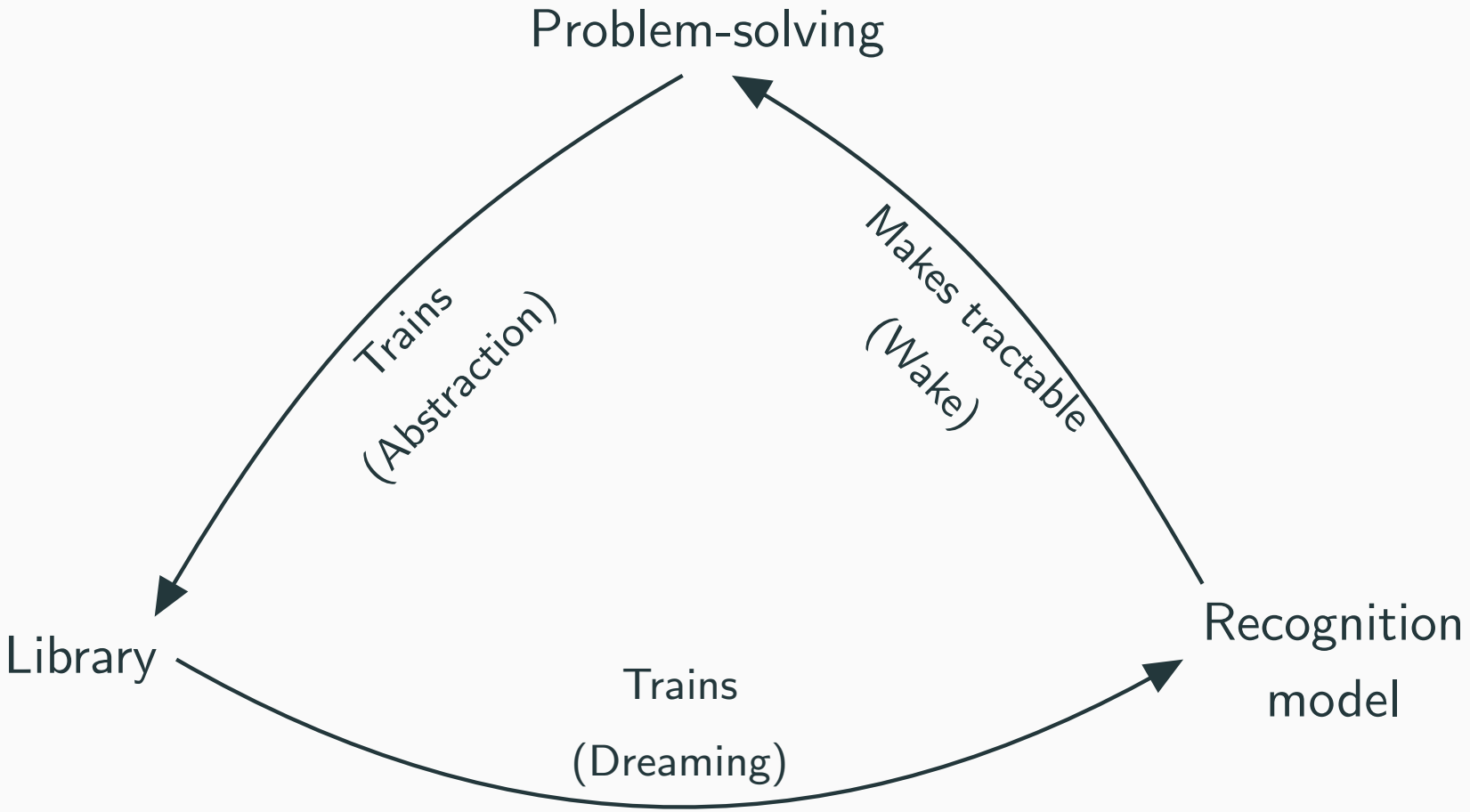
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Synergy between dreaming and library learning



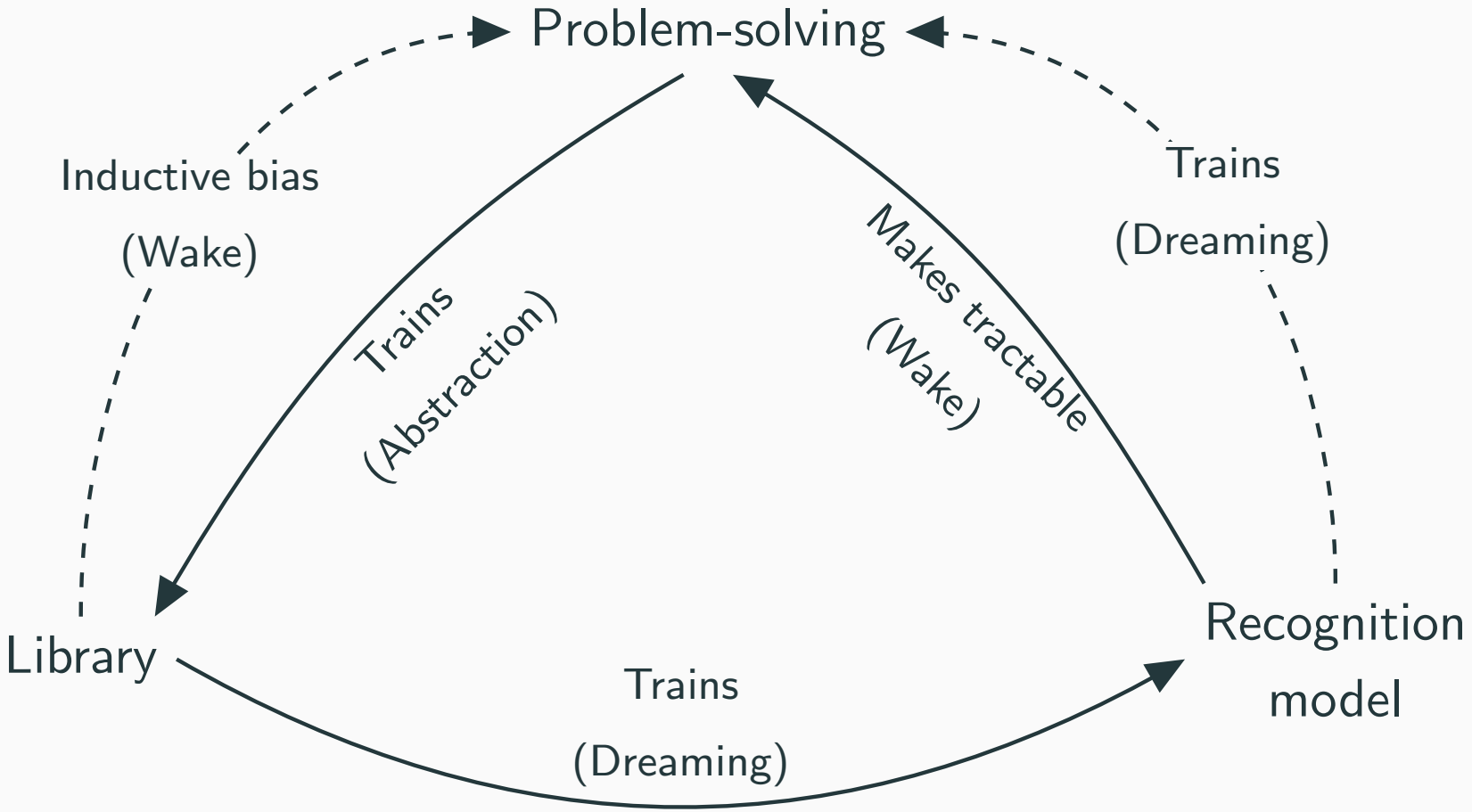
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Synergy between dreaming and library learning



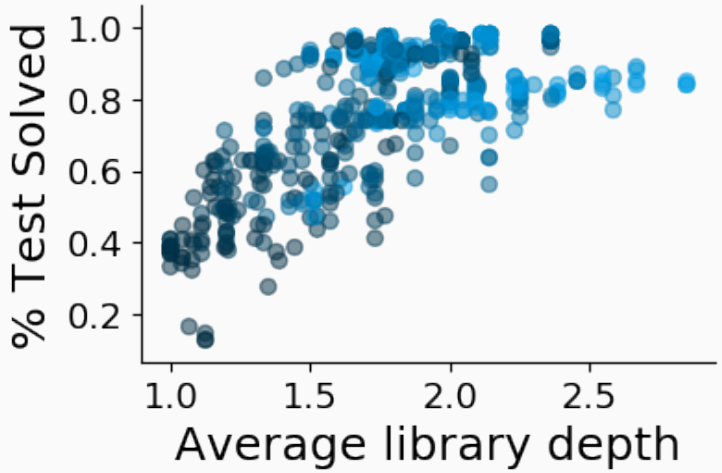
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Synergy between dreaming and library learning



Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Evidence for dreaming bootstrapping better libraries

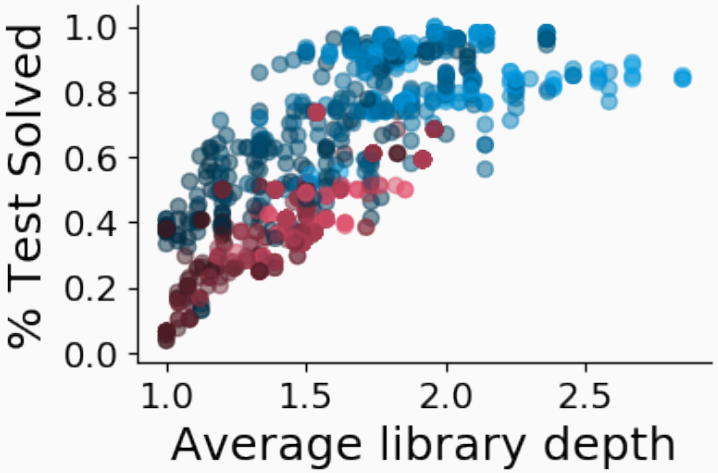


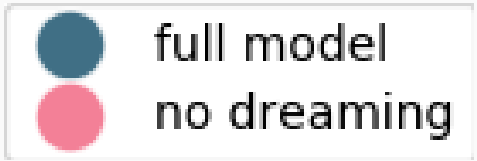
Darker: Early in learning

Brighter: Later in learning

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Evidence for dreaming bootstrapping better libraries



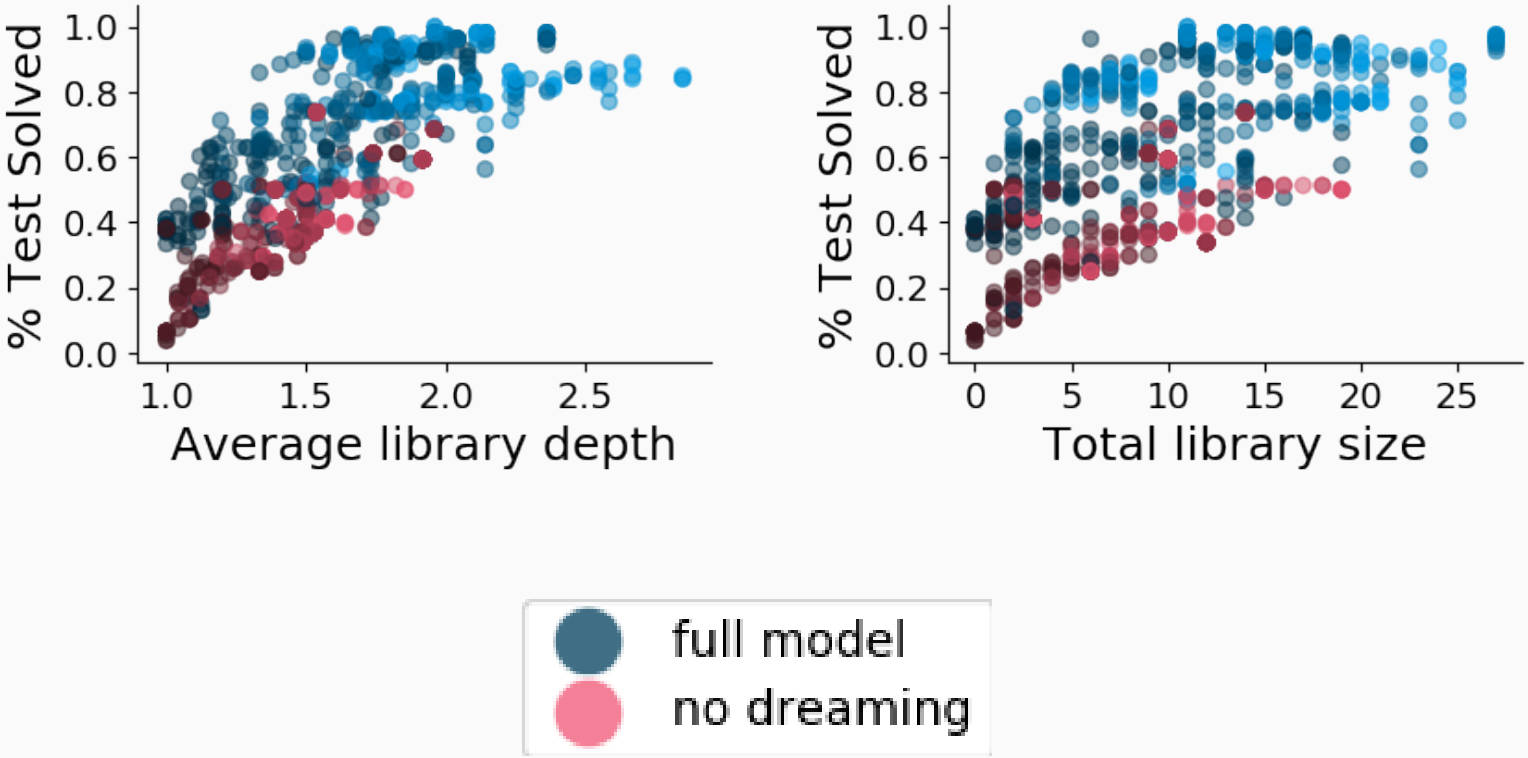


Darker: Early in learning

Brighter: Later in learning

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Evidence for dreaming bootstrapping better libraries



Darker: Early in learning

Brighter: Later in learning

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

From learning libraries, to learning languages

From learning libraries, to learning languages

modern functional programming $\rightarrow$ physics

From learning libraries, to learning languages

1950's Lisp $\rightarrow$ modern functional programming $\rightarrow$ physics

Physics Formula Sheet

Mechanics

$x = x_0 + v_{x0}t + \frac{1}{2}a_x t^2$

$v = v_0 + at$

$v_x^2 - v_{x0}^2 = 2a(x - x_0)$

$\vec{a} = \frac{\Sigma \vec{F}}{m} = \frac{\vec{F}_{net}}{m}$

$|\vec{F}_{friction}| \leq \mu |\vec{F}_{Normal}|$

$\vec{p} = m\vec{v}$

$\Delta \vec{p} = \vec{F}\Delta t$

$KE = \frac{1}{2}mv^2$

$\Delta PE = mg\Delta y$

$\Delta E = W = Fd\cos\theta$

$a_c = \frac{v^2}{r}$

$\theta = \theta_0 + \omega_0 t + \frac{1}{2}\alpha t^2$

$\omega = \omega_0 + \alpha t$

$T = \frac{2\pi}{\omega} = \frac{1}{f}$

$\nu = f\lambda$

$x = A\cos(2\pi f t)$

$\vec{\alpha} = \frac{\Sigma \vec{\tau}}{I} = \frac{\vec{\tau}_{net}}{I}$

$\vec{\tau} = r \times F$

$L = I\,\omega$

$\Delta L = \tau \Delta t$

$|\vec{F}_{spring}| = k|\vec{x}|$

$PE_{spring} = \frac{1}{2}kx^2$

$T_{spring} = 2\pi\sqrt{\frac{m}{k}}$

$T_{pendulum} = 2\pi\sqrt{\frac{\ell}{g}}$

$|\vec{F}_{gravity}| = G\frac{m_1 m_2}{r^2}$

$|\vec{F}_{gravity}| = m\vec{g}$

$PE_{gravity} = -G\frac{m_1 m_2}{r}$

$\rho = \frac{m}{V}$

$KE = \frac{1}{2}I\,\omega^2$

Electricity

$|\vec{F}_E| = k\left|\frac{q_1 q_2}{r^2}\right|$

$\Delta V = IR$

$R = \frac{\rho \ell}{A}$

$I = \frac{\Delta q}{\Delta t}$

$P = I\Delta V$

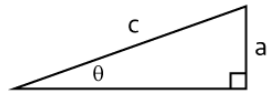
$R_{series} = R_1 + R_2 + \ldots + R_n$

$\frac{1}{R_{Parallel}} = \frac{1}{R_1} + \frac{1}{R_2} + \ldots + \frac{1}{R_n}$

Geometry

Rectangle	$A = bh$	Rectangular Solid	$V = \ell wh$	Triangle	$A = \frac{1}{2}bh$
Circle	$A = \pi r^2$ $C = 2\pi r$	Cylinder	$V = \pi r^2 \ell$ $S = 2\pi r\ell + 2\pi r^2$	Sphere	$V = \frac{4}{3}\pi r^3$ $S = 4\pi r^2$

Trigonometry



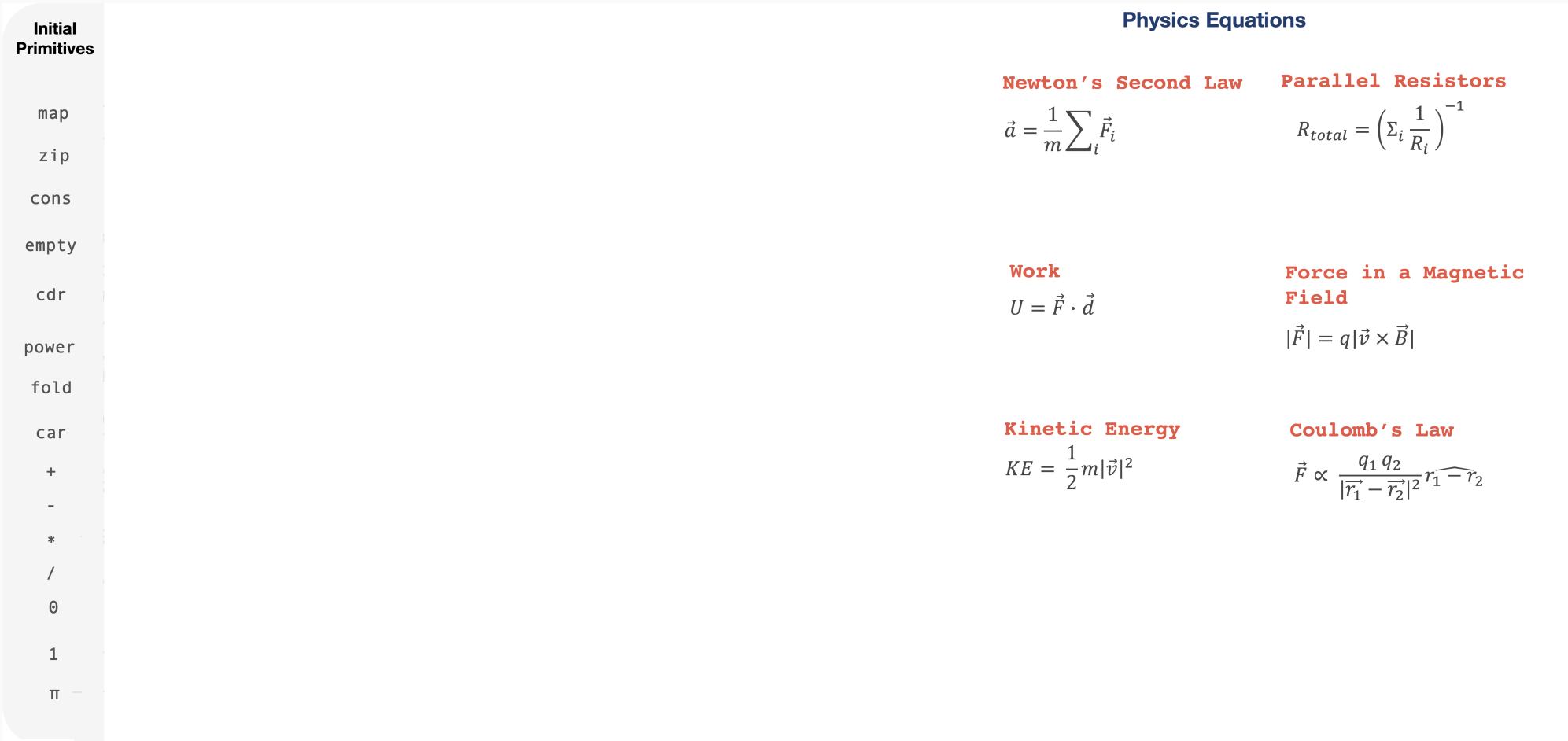
$c^2 = a^2 + b^2$

$\sin\theta = \frac{a}{c}$ $\cos\theta = \frac{b}{c}$ $\tan\theta = \frac{a}{b}$

Variables

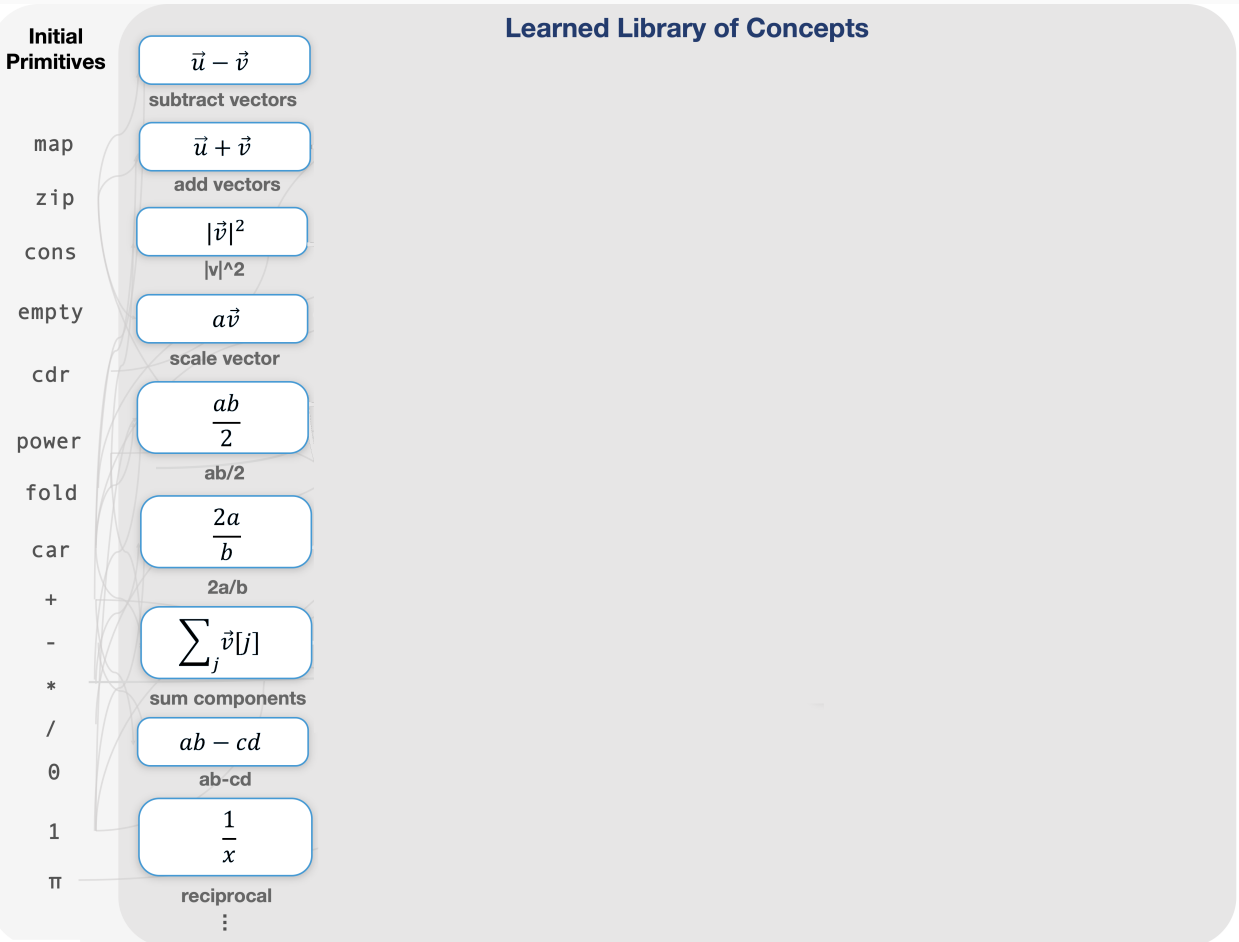
a = acceleration
A = amplitude
A = Area
b = base length
C = circumference
d = distance
E = energy
f = frequency
F = force
h = height
I = current
I = rotational inertia
KE = kinetic energy
k = spring constant
L = angular momentum
ℓ = length
m = mass
P = power
p = momentum
q = charge
r = radius
R = resistance
S = surface area
T = period
t = time
PE = potential energy
V = electric potential
V = volume
v = velocity
w = width
W = work
x = position
y = height
α = angular acceleration
λ = wavelength
μ = coefficient of friction

Growing languages for vector algebra and physics



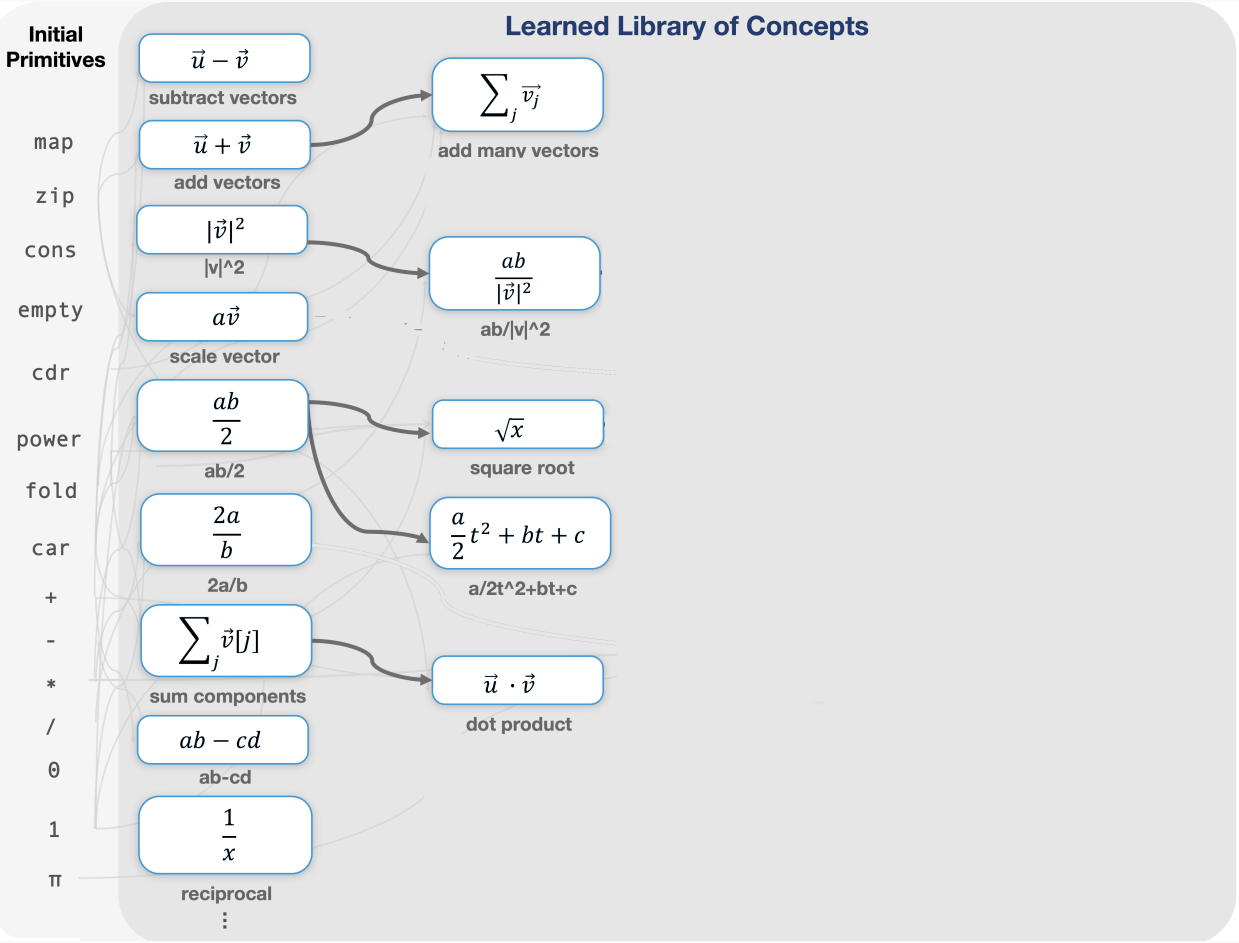
Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Growing languages for vector algebra and physics



Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Growing languages for vector algebra and physics



Physics Equations

Newton's Second Law

$$\vec{a} = \frac{1}{m}\sum_i \vec{F}_i$$

Parallel Resistors

$$R_{total} = \left(\Sigma_i \frac{1}{R_i}\right)^{-1}$$

Work

$$U = \vec{F} \cdot \vec{d}$$

Force in a Magnetic Field

$$|\vec{F}| = q|\vec{v} \times \vec{B}|$$

Kinetic Energy

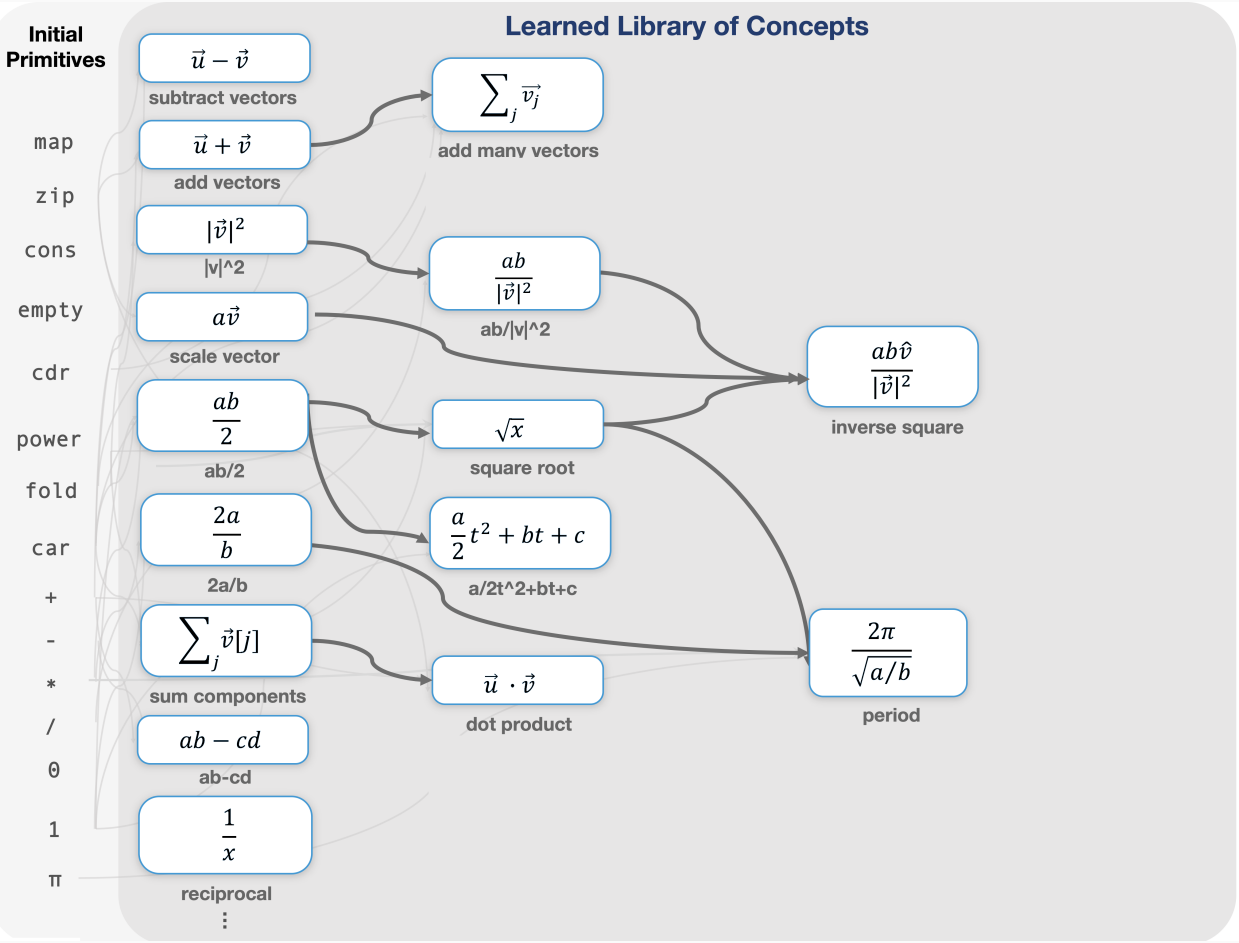
$$KE = \frac{1}{2}m|\vec{v}|^2$$

Coulomb's Law

$$\vec{F} \propto \frac{q_1 q_2}{|\vec{r_1} - \vec{r_2}|^2}\widehat{r_1 - r_2}$$

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Growing languages for vector algebra and physics



Physics Equations

Newton's Second Law

$$\vec{a} = \frac{1}{m}\sum_i \vec{F}_i$$

Parallel Resistors

$$R_{total} = \left(\Sigma_i \frac{1}{R_i}\right)^{-1}$$

Work

$$U = \vec{F} \cdot \vec{d}$$

Force in a Magnetic Field

$$|\vec{F}| = q|\vec{v} \times \vec{B}|$$

Kinetic Energy

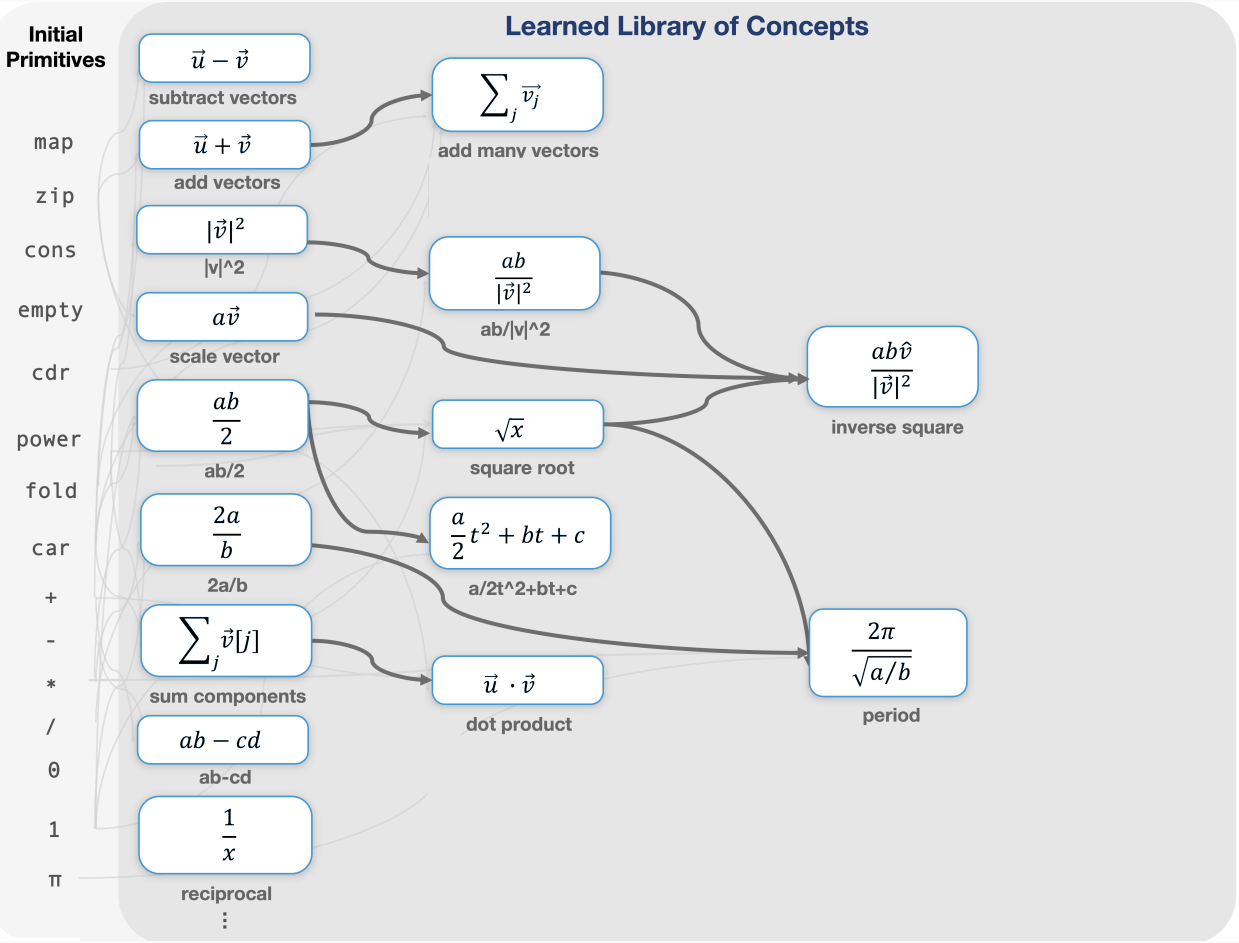
$$KE = \frac{1}{2}m|\vec{v}|^2$$

Coulomb's Law

$$\vec{F} \propto \frac{q_1 q_2}{|\vec{r_1} - \vec{r_2}|^2}\widehat{r_1 - r_2}$$

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Growing languages for vector algebra and physics



Physics Equations

Newton's Second Law

$$\vec{a}=\frac{1}{m}\sum_i \vec{F}_i$$

```
(scale-vector(reciprocal m)
(add-many-vectors Fs))
```

Parallel Resistors

$$R_{total}=\left(\Sigma_i \frac{1}{R_i}\right)^{-1}$$

```
(reciprocal (sum-components
(map (λ(r) (reciprocal r))
Rs)))
```

Work

$$U=\vec{F}\cdot\vec{d}$$

```
(dot-product F d)
```

Force in a Magnetic Field

$$|\vec{F}|=q|\vec{v}\times\vec{B}|$$

```
(* q (ab-cd v_x b_y v_y b_x))
```

Kinetic Energy

$$KE=\frac{1}{2}m|\vec{v}|^2$$

```
(ab/2 m (|v|^2 v))
```

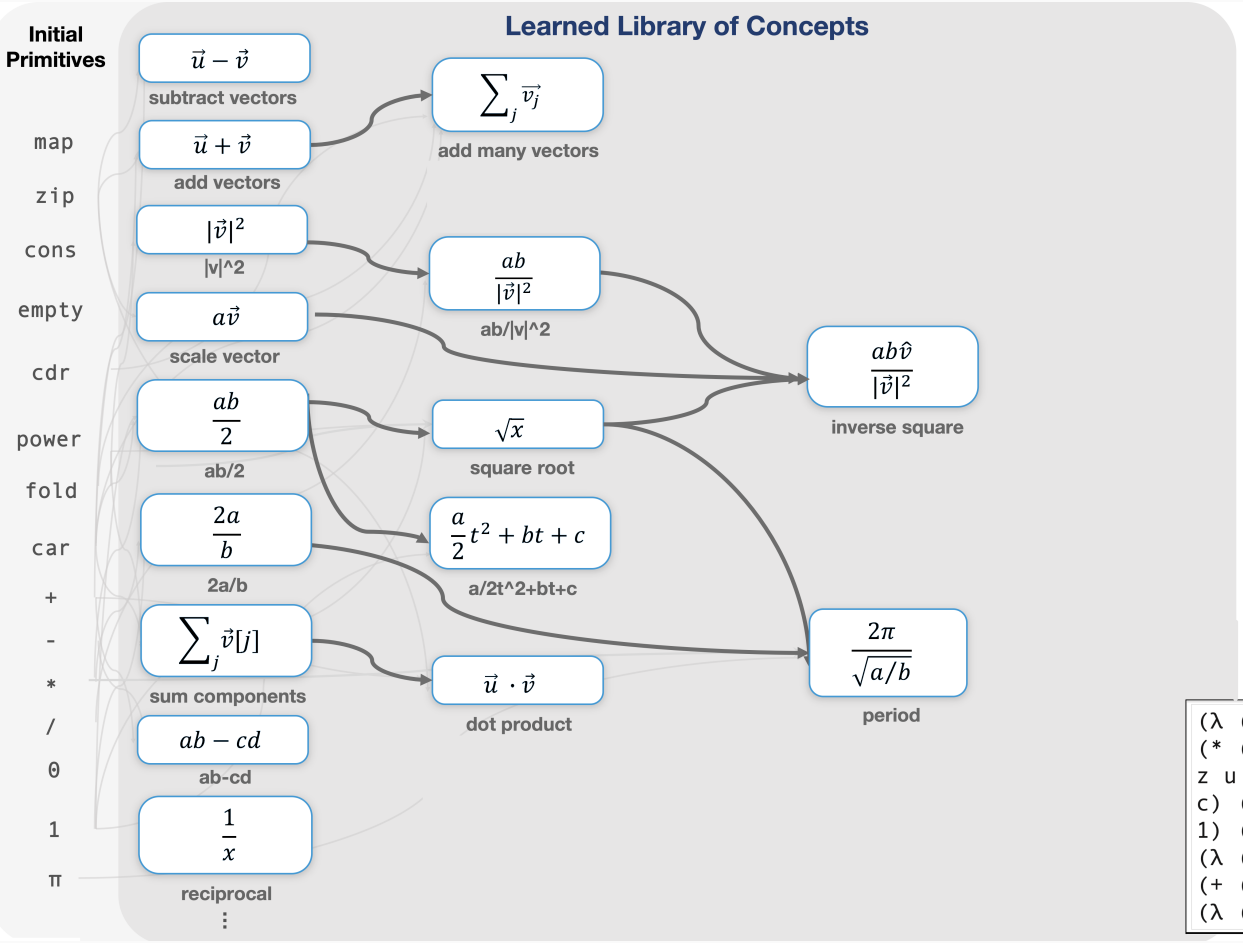
Coulomb's Law

$$\vec{F}\propto\frac{q_1 q_2}{|\vec{r_1}-\vec{r_2}|^2}\widehat{r_1-r_2}$$

```
(inverse-square q_1 q_2
(subtract-vectors r_1 r_2))
```

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Growing languages for vector algebra and physics



Physics Equations

Newton's Second Law

$$\vec{a}=\frac{1}{m}\sum_i \vec{F}_i$$

```
(scale-vector(reciprocal m)
(add-many-vectors Fs))
```

Parallel Resistors

$$R_{total}=\left(\Sigma_i \frac{1}{R_i}\right)^{-1}$$

```
(reciprocal (sum-components
(map (λ(r) (reciprocal r))
Rs)))
```

Work

$$U=\vec{F}\cdot\vec{d}$$

```
(dot-product F d)
```

Force in a Magnetic Field

$$|\vec{F}|=q|\vec{v}\times\vec{B}|$$

```
(* q (ab-cd v_x b_y v_y b_x))
```

Kinetic Energy

$$KE=\frac{1}{2}m|\vec{v}|^2$$

```
(ab/2 m (|v|^2 v))
```

Coulomb's Law

$$\vec{F}\propto\frac{q_1 q_2}{|\vec{r_1}-\vec{r_2}|^2}\widehat{r_1-r_2}$$

```
(inverse-square q_1 q_2
(subtract-vectors r_1 r_2))
```

```
(λ (x y z u) (map (λ (v) (* (/
(* (power (/ (* x x) (fold (zip
z u (λ (w a) (- w a))) 0 (λ (b
c) (+ (* b b)    c)))) (/ (* 1
1) (+ 1 1))) y) (fold (zip z u
(λ (d e)  (- d e))) 0 (λ (f g)
(+ (* f f) g)))) v)) (zip z u
(λ (h i) (- h i)))))
```

Solution to `Coulomb's Law` if expressed in initial primitives

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Growing a language for recursive programming

Initial Primitives

- Y combinator
- cons
- car
- cdr
- nil
- if
- nil?
- +
- -
- 0
- 1
- =

Recursive Programming Algorithms

Stutter

[■■] → [■■■■]
[■■■] → [■■■■■■]

Take every other

[■■■■] → [■■]
[■■■■■■] → [■■■]

List lengths

[[■■■], [■]] → [3 1]
[[■■], [], [■]] → [2 0 1]

List differences

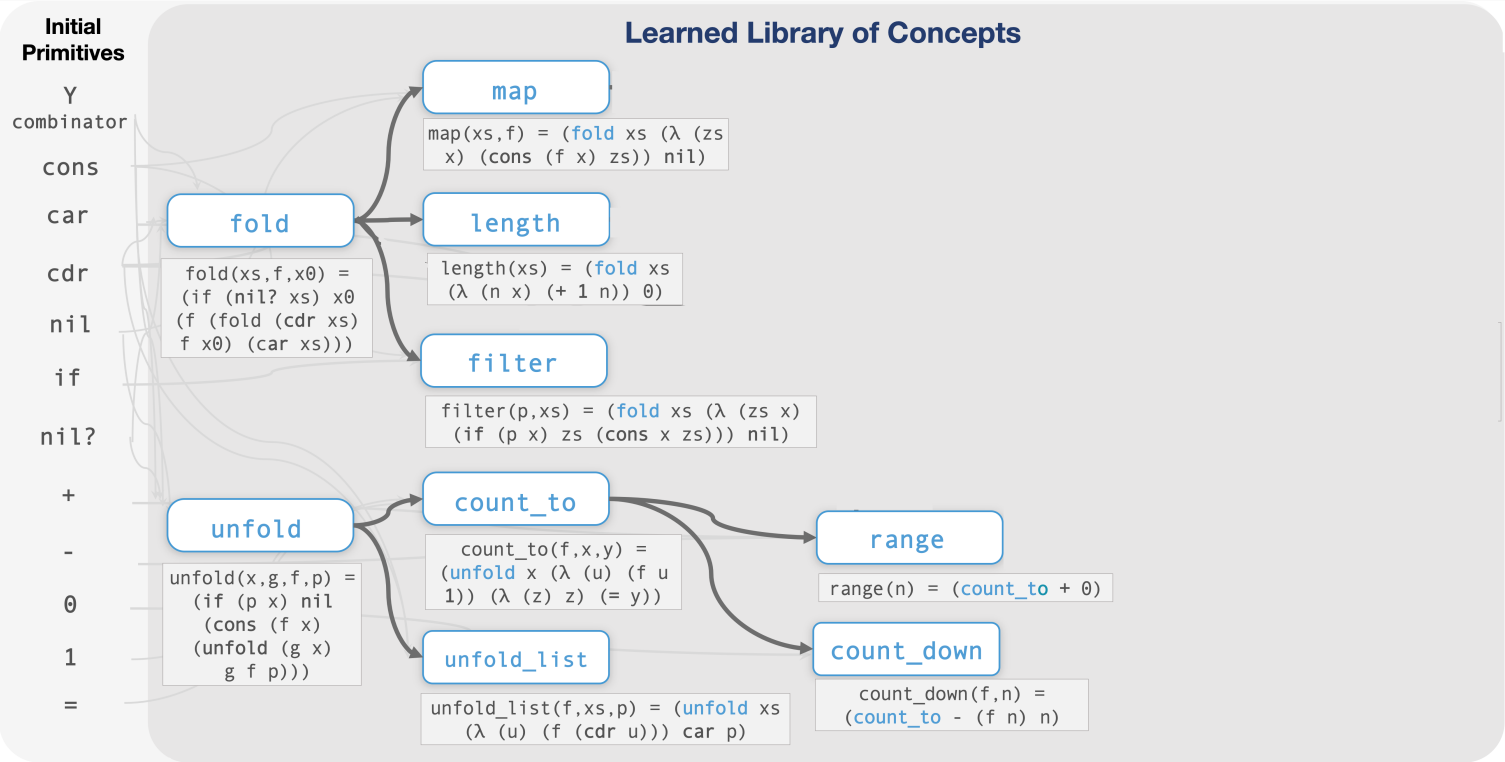
[1 8 2], [0 5 1] → [1 3 1]
[2 3 6], [1 2 4] → [1 1 2]

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Growing a language for recursive programming

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Growing a language for recursive programming



Recursive Programming Algorithms

Stutter

[■■] → [■■■■]
[■■■] → [■■■■■■]

```
(fold A (λ (u v) (cons
v (cons v u))) nil)
```

Take every other

[■■■■] → [■■]
[■■■■■■] → [■■■]

```
(unfold_list cdr A nil?)
```

List lengths

[[■■■], [■]] → [3 1]
[[■■], [], [■]] → [2 0 1]

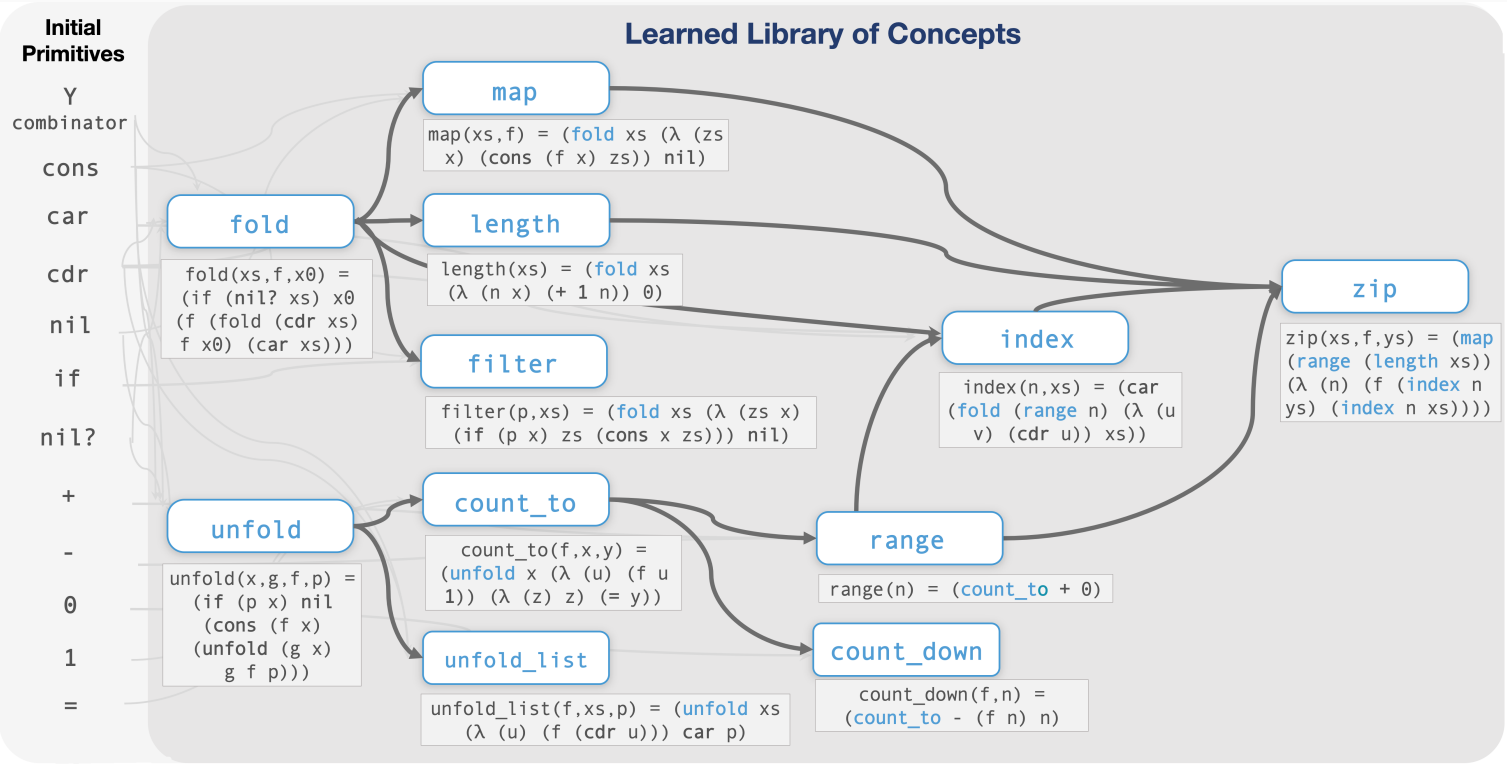
```
(map A length)
```

List differences

[1 8 2], [0 5 1] → [1 3 1]
[2 3 6], [1 2 4] → [1 1 2]

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Growing a language for recursive programming



Recursive Programming Algorithms

Stutter

[■■] → [■■■■]
[■■■] → [■■■■■■]

```
(fold A (λ (u v) (cons
v (cons v u))) nil)
```

Take every other

[■■■■] → [■■]
[■■■■■■] → [■■■]

```
(unfold_list cdr A nil?)
```

List lengths

[[■■■], [■]] → [3 1]
[[■■], [], [■]] → [2 0 1]

```
(map A length)
```

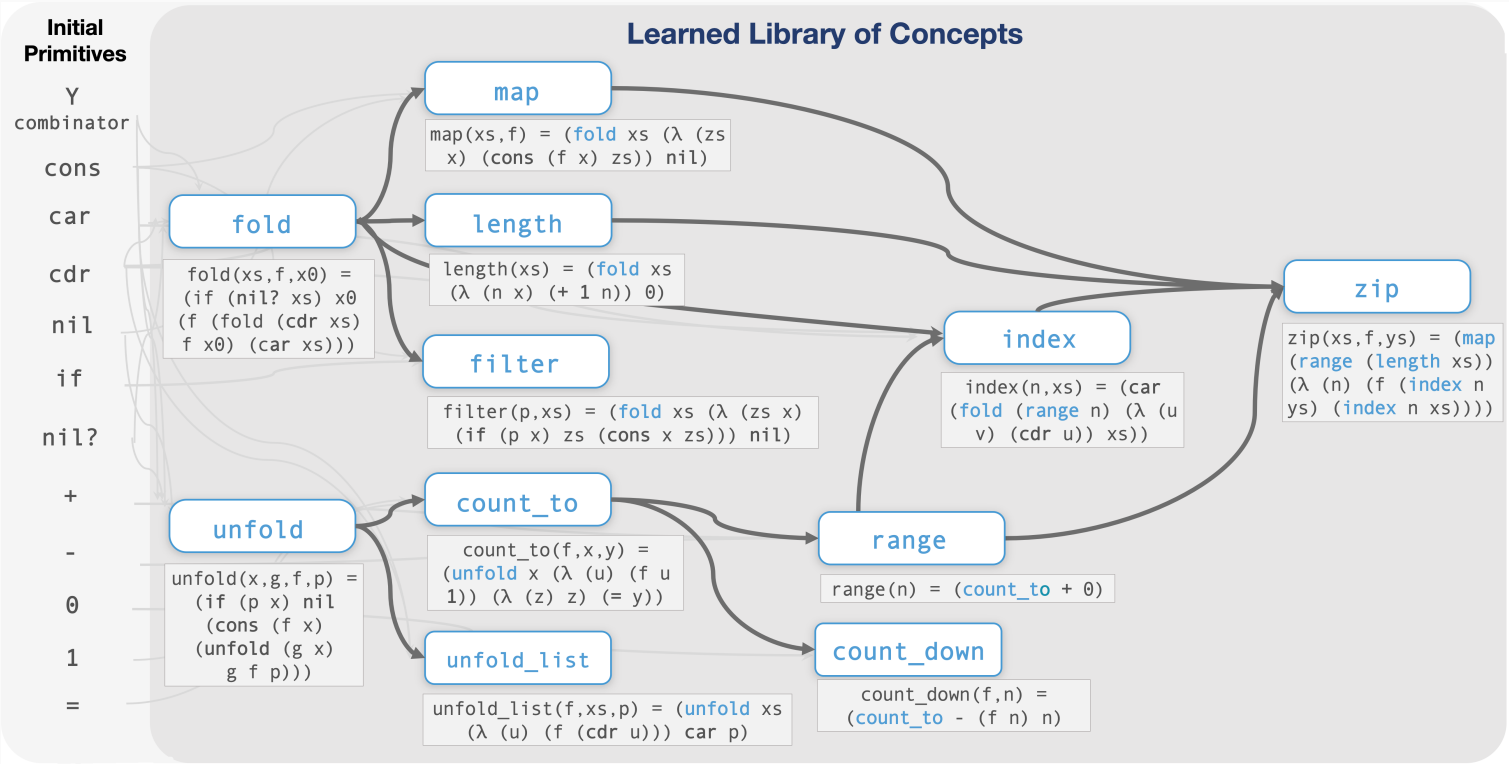
List differences

[1 8 2], [0 5 1] → [1 3 1]
[2 3 6], [1 2 4] → [1 1 2]

```
(zip A - B )
```

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Growing a language for recursive programming



Recursive Programming Algorithms

Stutter

[■■] → [■■■■]
[■■■] → [■■■■■■]

```
(fold A (λ (u v) (cons
v (cons v u))) nil)
```

Take every other

[■■■■] → [■■]
[■■■■■■] → [■■■]

```
(unfold_list cdr A nil?)
```

List lengths

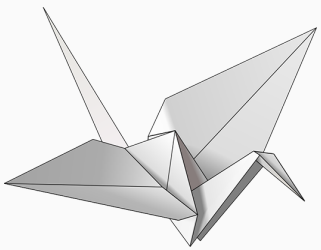
[[■■■], [■]] → [3 1]
[[■■], [], [■]] → [2 0 1]

```
(map A length)
```

List differences

[1 8 2], [0 5 1] → [1 3 1]
[2 3 6], [1 2 4] → [1 1 2]

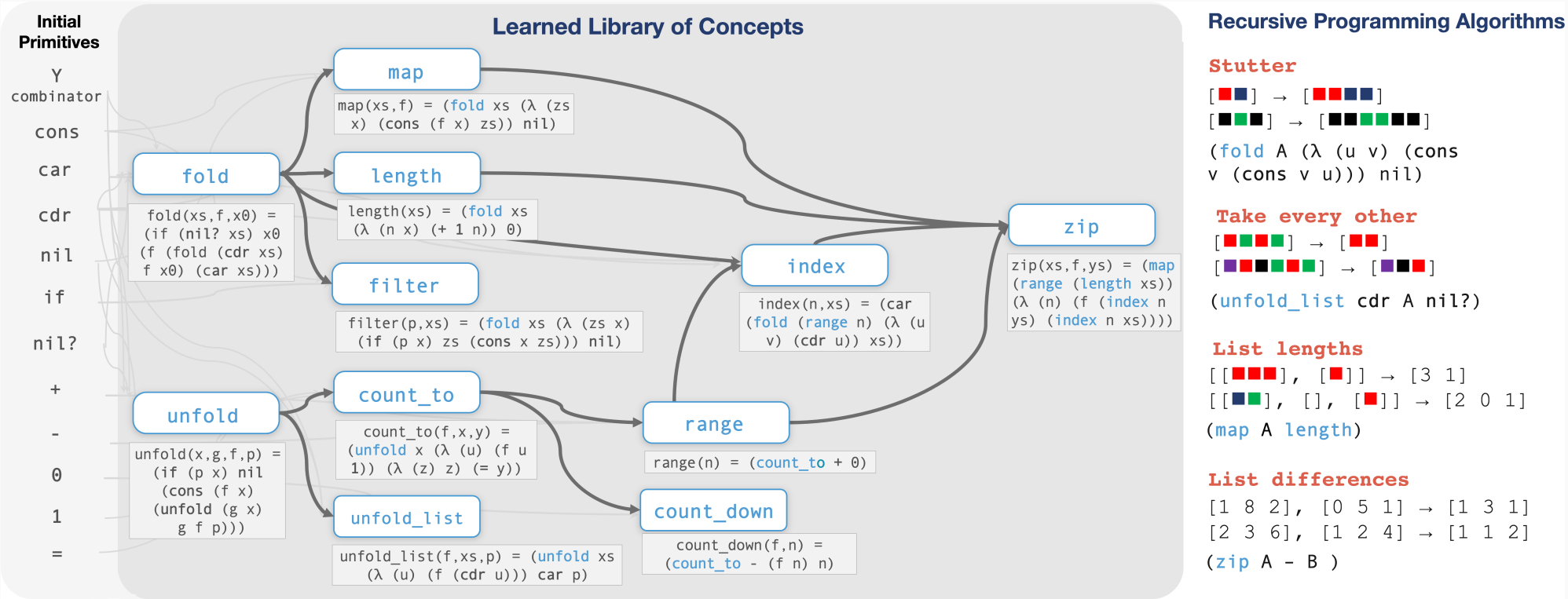
```
(zip A - B )
```



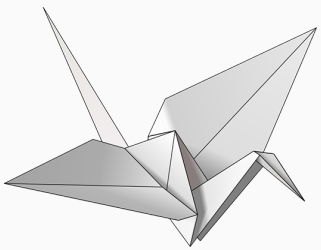
Origami Programming: Jeremy Gibbons, 2003

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Growing a language for recursive programming



1 year of compute. 5 days on 64 CPUs.



Origami Programming: Jeremy Gibbons, 2003

Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

Lessons

Symbols aren't necessarily interpretable. Flexibly grow the language based on experience to make it more powerful *and* more human understandable

Learning-from-scratch is possible in principle. Don't do it. But program induction makes it convenient to build in what we know how to build in, and then learn and adapt on top of that

Program Induction and learning to learn

learning a DSL

learning to synthesize

synergy between DSL+learned synthesizer

Some unresolved questions

Not everything is crisp and symbolic. How do we learn DSLs for neurosymbolic hybrid programs?
(see Memoised Wake-Sleep, Hewitt et al 2020)

Some unresolved questions

Not everything is crisp and symbolic. How do we learn DSLs for neurosymbolic hybrid programs?
(see Memoised Wake-Sleep, Hewitt et al 2020)

Search is still hard
(see "REPL": Ellis, Nye, Pu, Sosa et al 2019)

Some unresolved questions

Not everything is crisp and symbolic. How do we learn DSLs for neurosymbolic hybrid programs?
(see Memoised Wake-Sleep, Hewitt et al 2020)

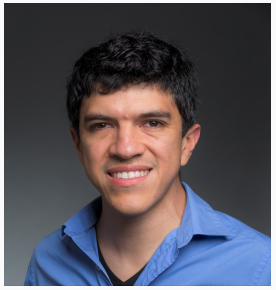
Search is still hard
(see "REPL": Ellis, Nye, Pu, Sosa et al 2019)

Library≠Language:
How do we learn data structures, or discover new types, and thereby *actually* learn DSLs?

Collaborators

Josh Tenenbaum

Armando Solar-Lezama



Max Nye

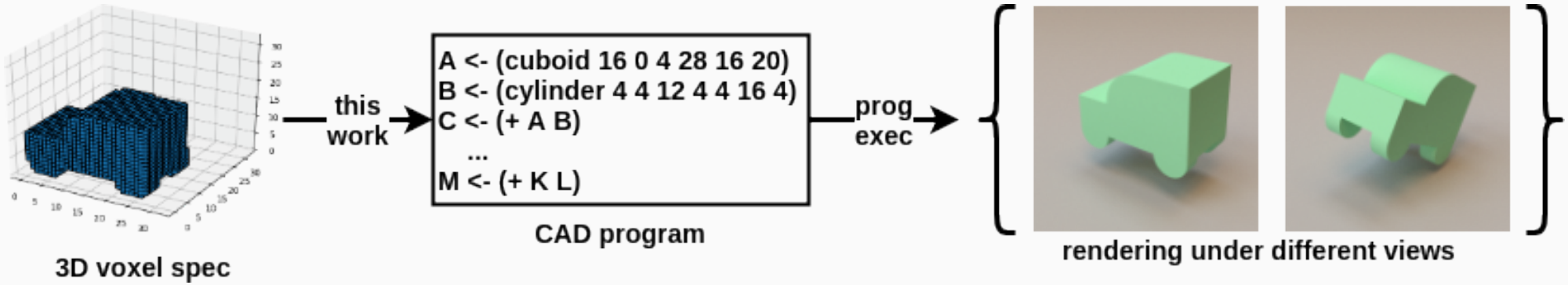
Cathy Wong

Mathias Sable-Meyer

Lucas Morales

thank you

3D program induction



Challenge: combinatorial search!

Branching factor: > 1.3 million per line of code, ≈ 20 lines of code

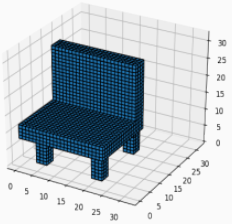
search space size: $(1.3 \text{ million})^{20} \approx 10^{122}$ programs

Ellis*, Nye*, Pu*, Sosa*, Tenenbaum, Solar-Lezama. NeurIPS 2019.

*equal contribution

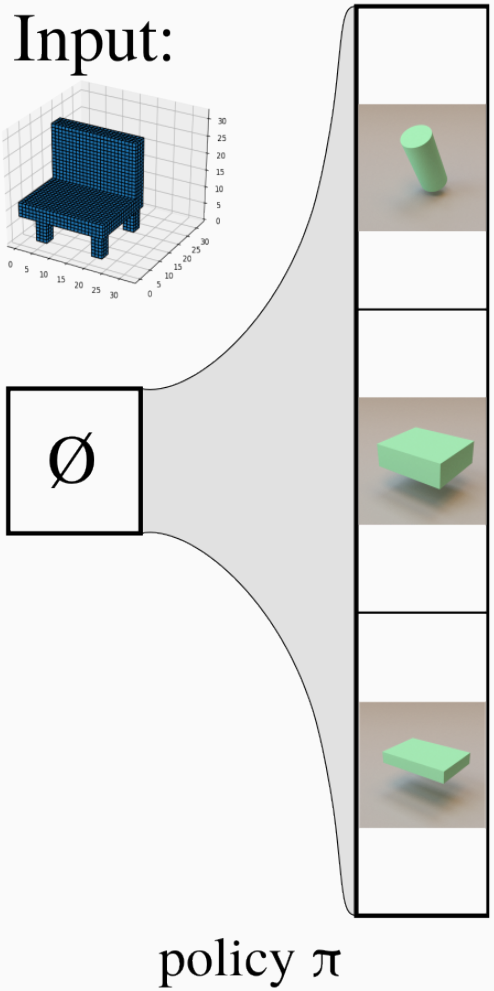
Solution: stochastic **tree search** + learn **policy** that writes code + learn **value** function that assesses execution of program so far; analogous to **AlphaGo** [Silver et al. 2016]

Input:

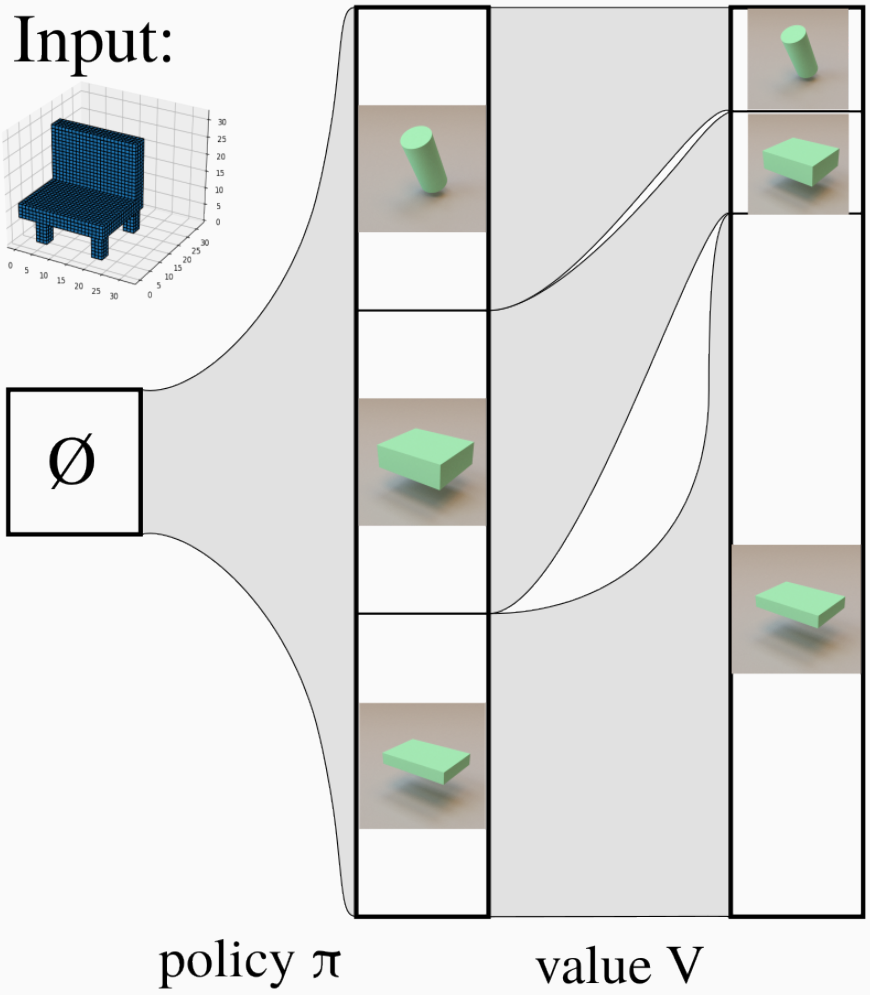


Ø

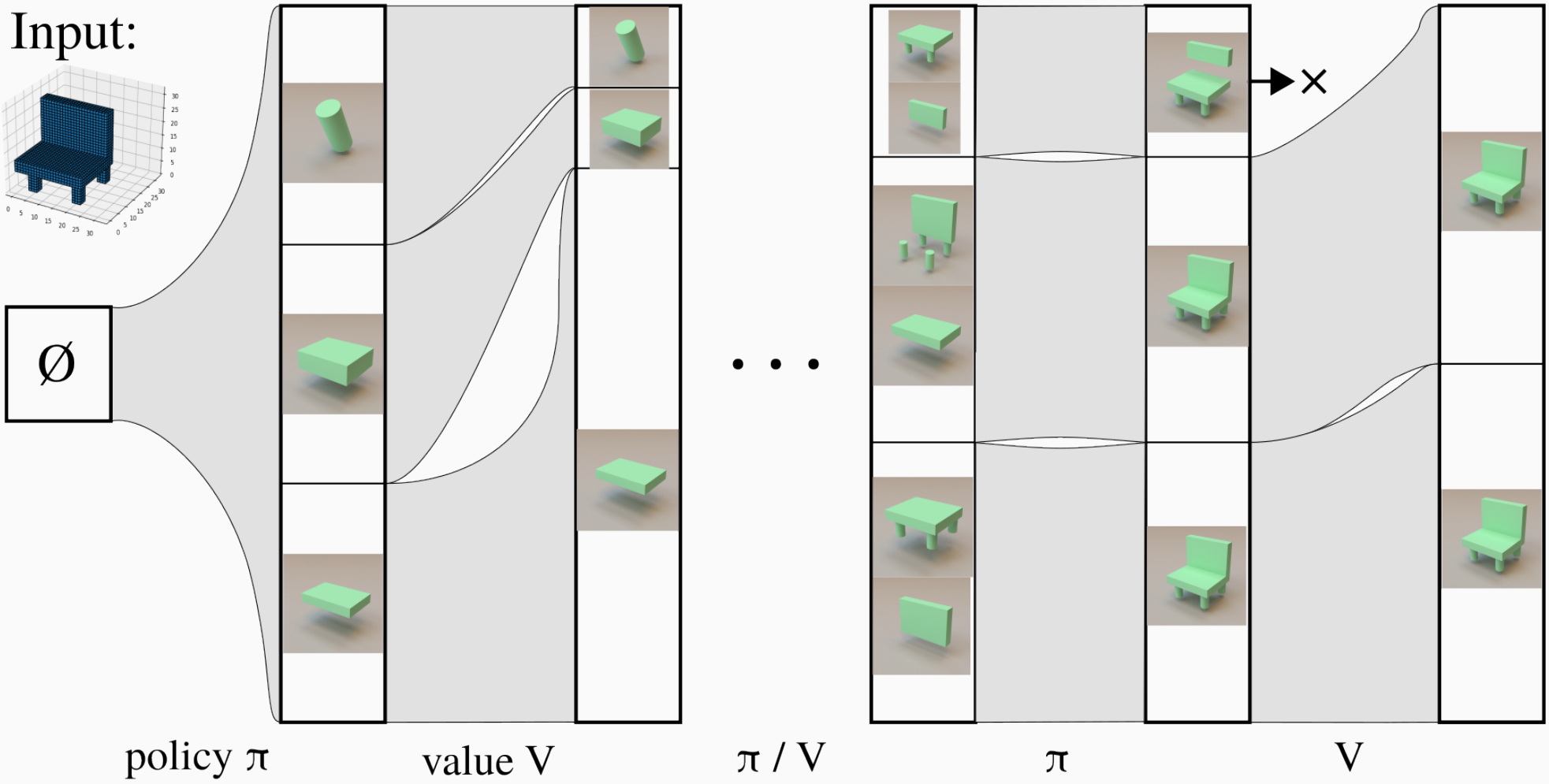
Solution: stochastic **tree search** + learn **policy** that writes code + learn **value** function that assesses execution of program so far; analogous to **AlphaGo** [Silver et al. 2016]



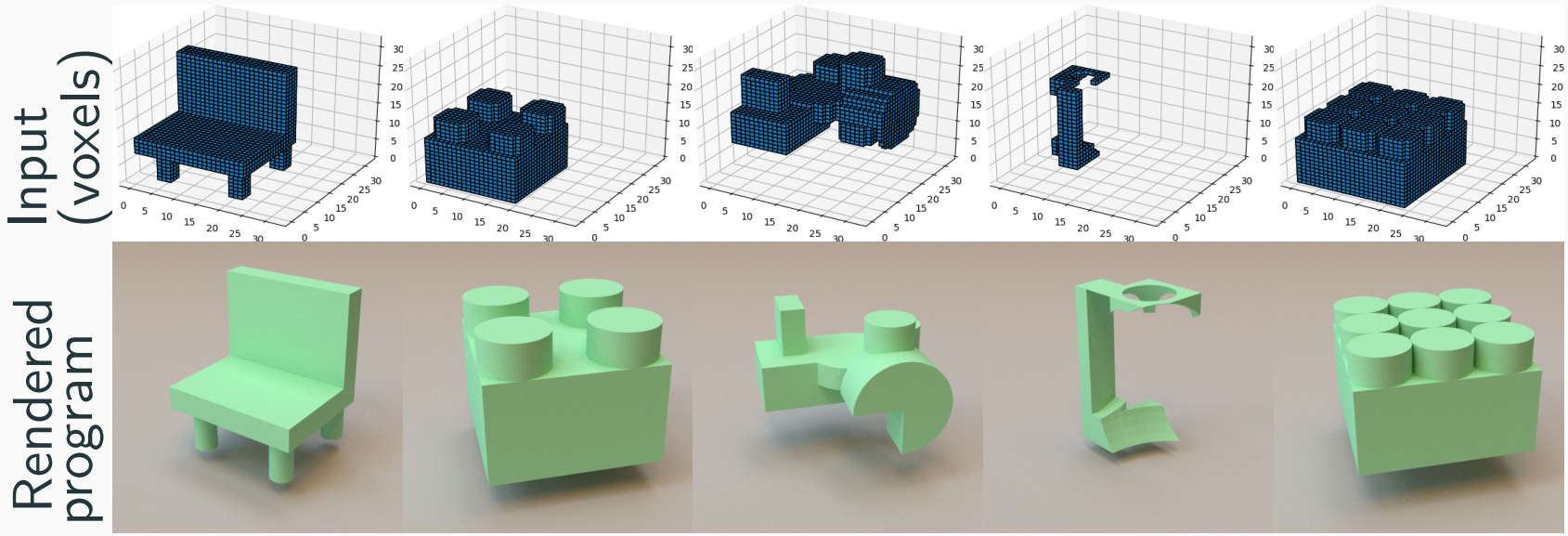
Solution: stochastic **tree search** + learn **policy** that writes code + learn **value** function that assesses execution of program so far; analogous to **AlphaGo** [Silver et al. 2016]



Solution: stochastic **tree search** + learn **policy** that writes code + learn **value** function that assesses execution of program so far; analogous to **AlphaGo** [Silver et al. 2016]



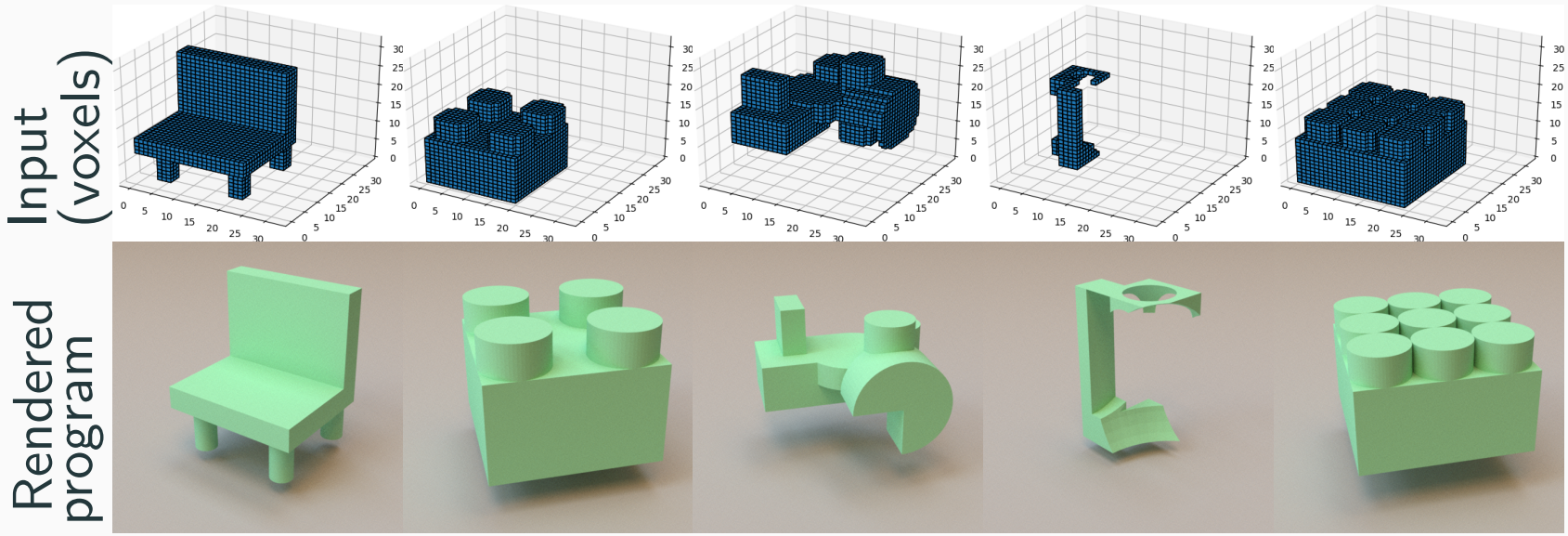
3D program induction



Ellis*, Nye*, Pu*, Sosa*, Tenenbaum, Solar-Lezama. NeurIPS 2019.

*equal contribution

3D program induction

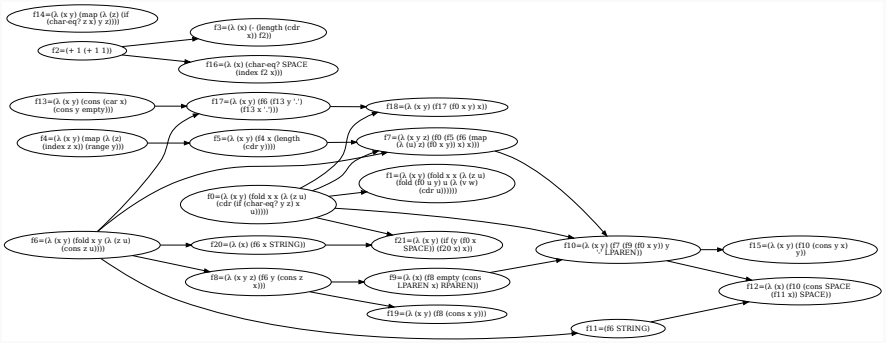


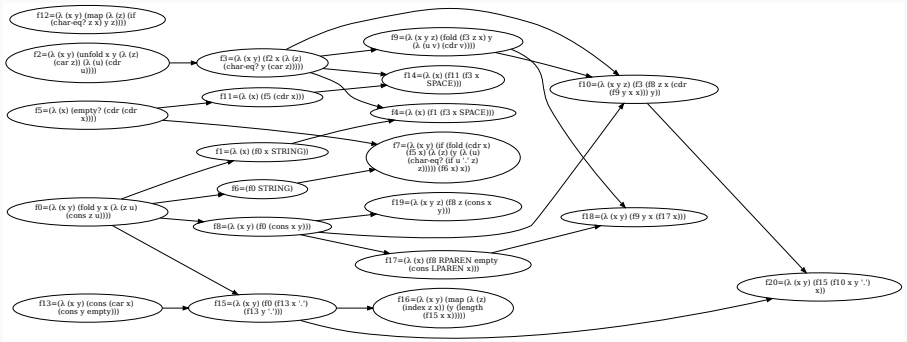
same architecture learns to synthesize text editing programs (FlashFill, Gulwani 2012)

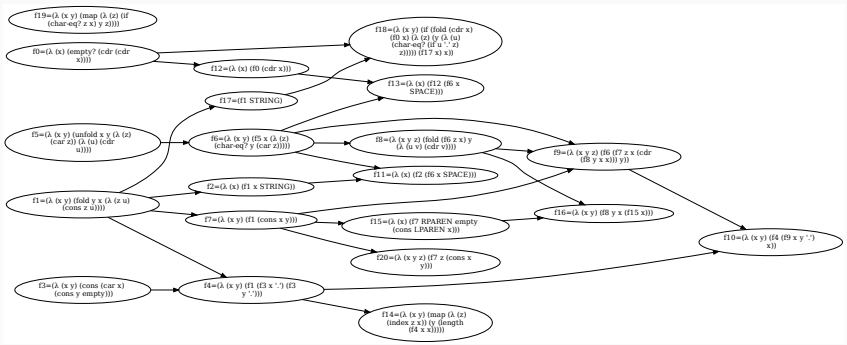
Ellis*, Nye*, Pu*, Sosa*, Tenenbaum, Solar-Lezama. NeurIPS 2019.
*equal contribution

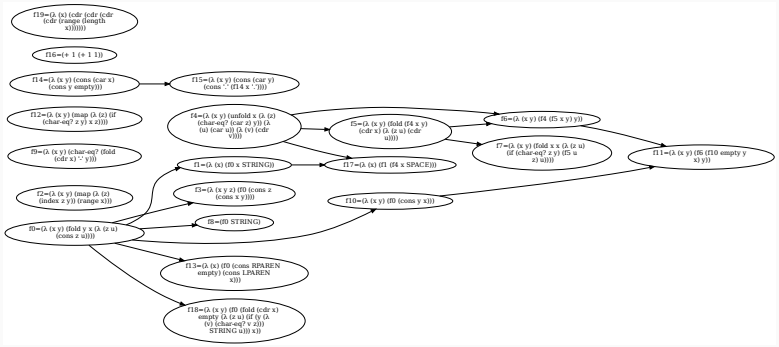
Library structure: Text Editing

DreamCoder learns libraries for FlashFill-style text editing [Gulwani 2012]





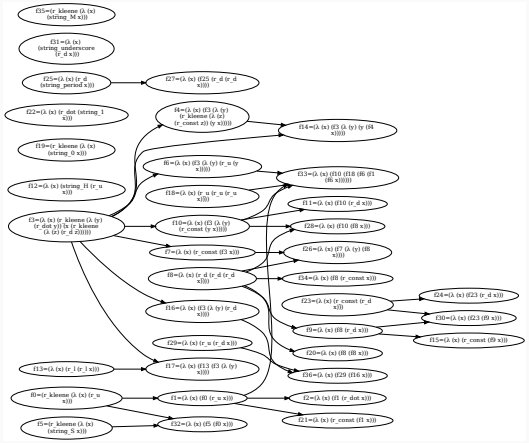


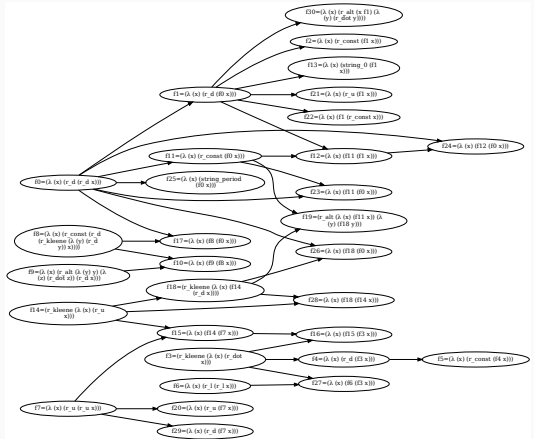


Ellis, Wong, Nye, ..., Solar-Lezama, Tenenbaum. arxiv 2020.

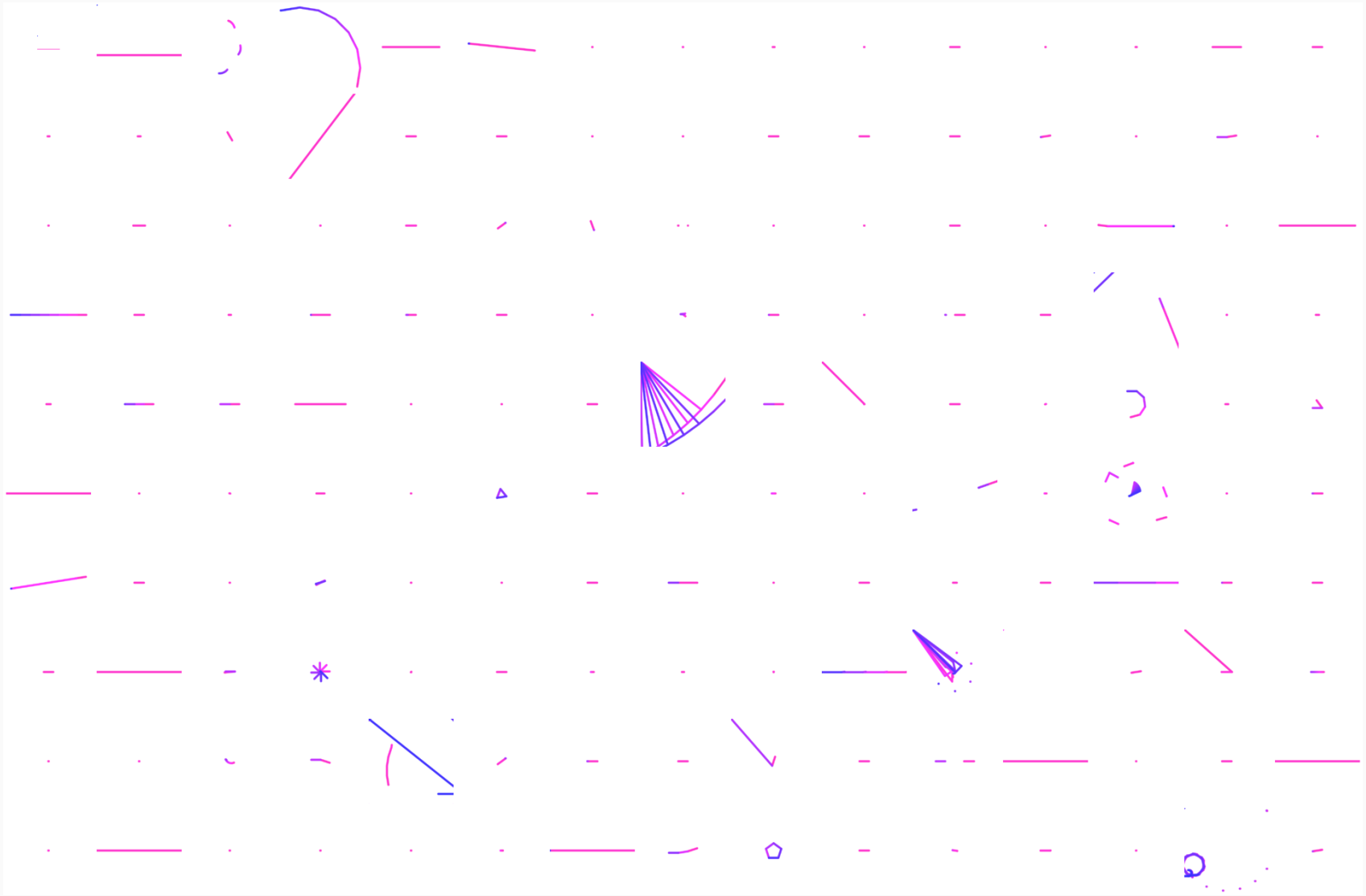
Library structure: Generating Text

Libraries for probabilistic generative models over text:
data from crawling web for CSV files





150 random dreams before learning



150 random dreams after learning