Human-Computer Architecture for Scientific Discovery

Zoran Popović
Center for Game Science
Department of CSE
University of Washington

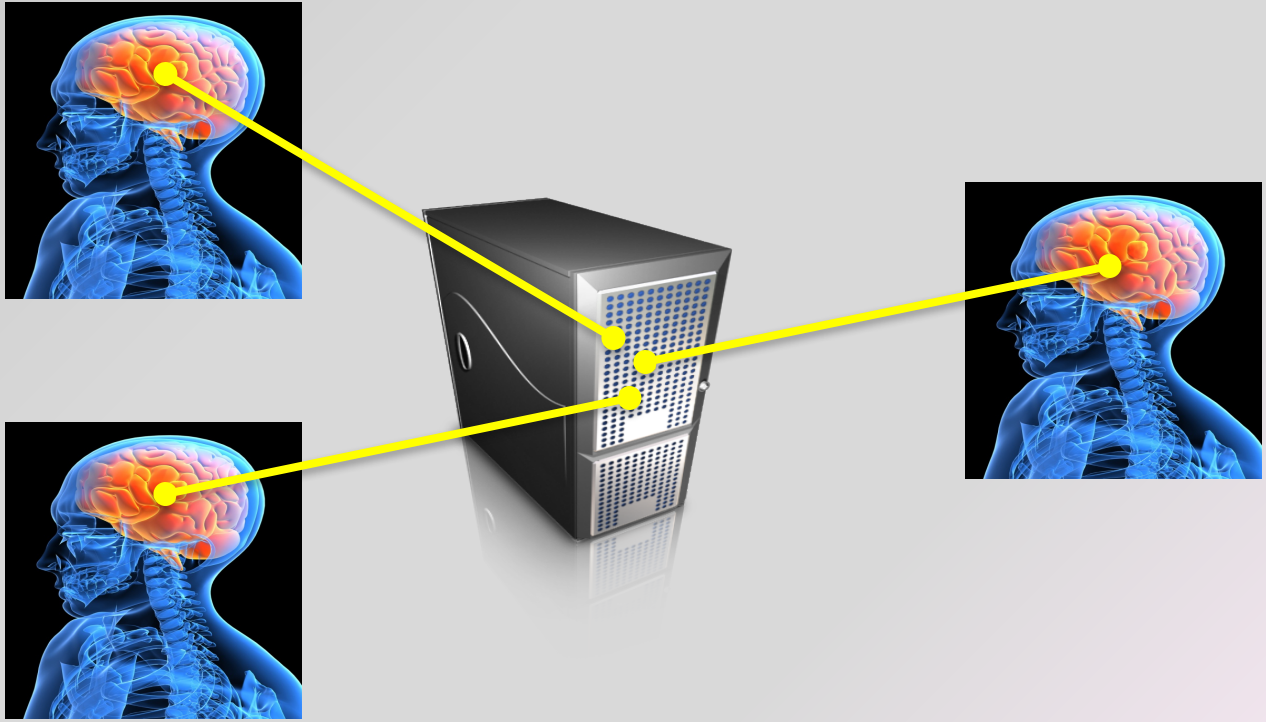
The AI Science Challenge

Solve problems that

- All expert people
- all massive computer architectures

cannot solve by themselves

Symbiotic Human-Computer Computing Architecture



Solving Hard Problems with Human-Computer Symbiosis

Coadaptation (Optimization):

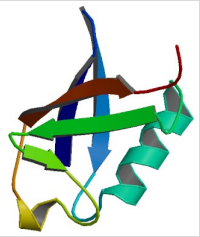
1. People → Experts
2. Interaction → Optimal problem solving tools

Games are an ideal vehicle of coadaptation

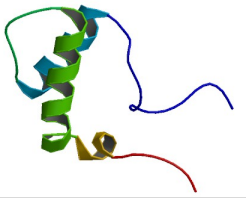
Signs of Promise

Proteins

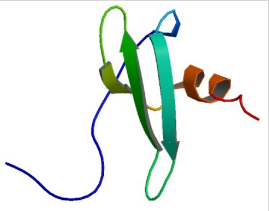
MQIFVKTLTGKTILEVEPSDTIE…



MGKYDKQIDLSTVDLKKLRVKEL…



KPVSLSYRCPCRFFESHVARANV…



⋮

Sequence

⋮

3D Structure

Proteins

MKTLYDLPIVLRNLPEDLVLEI

?

SMSDTAERVKKIVVEHLGVDAI

?

GPLGSDQYIVVNGAPVIPSAKV

?

⋮ ⋮

Sequence

EASY

3D Structure

HARD

Proteins

MKTLYDLPIVLRNLPEDLVLEE

?

SMSDTAERVKKIVVEHLGVDAD

?

GPLGSDQYIVVNGAPVIPSAKV

?

⋮

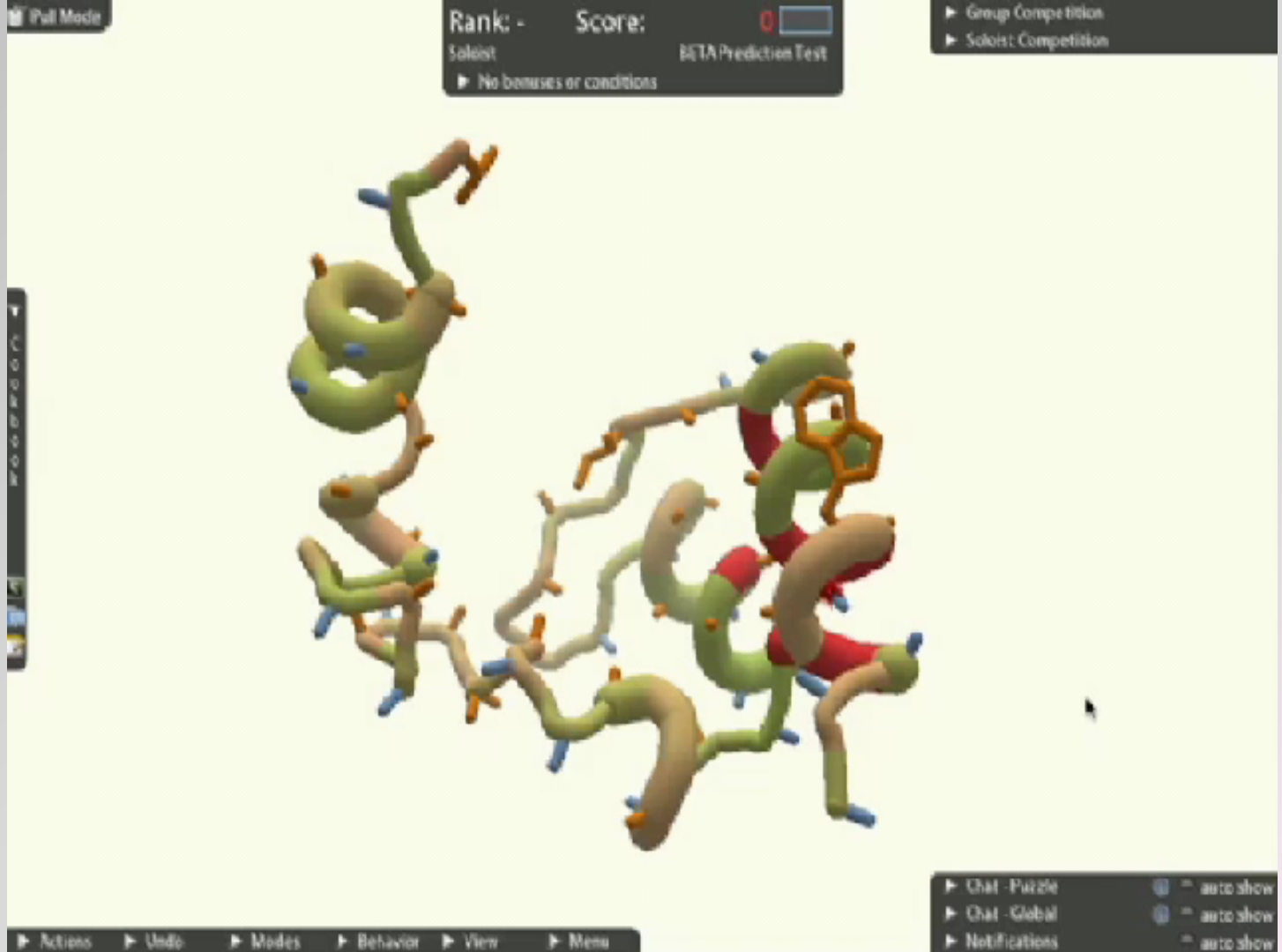
Sequence

EASY

⋮

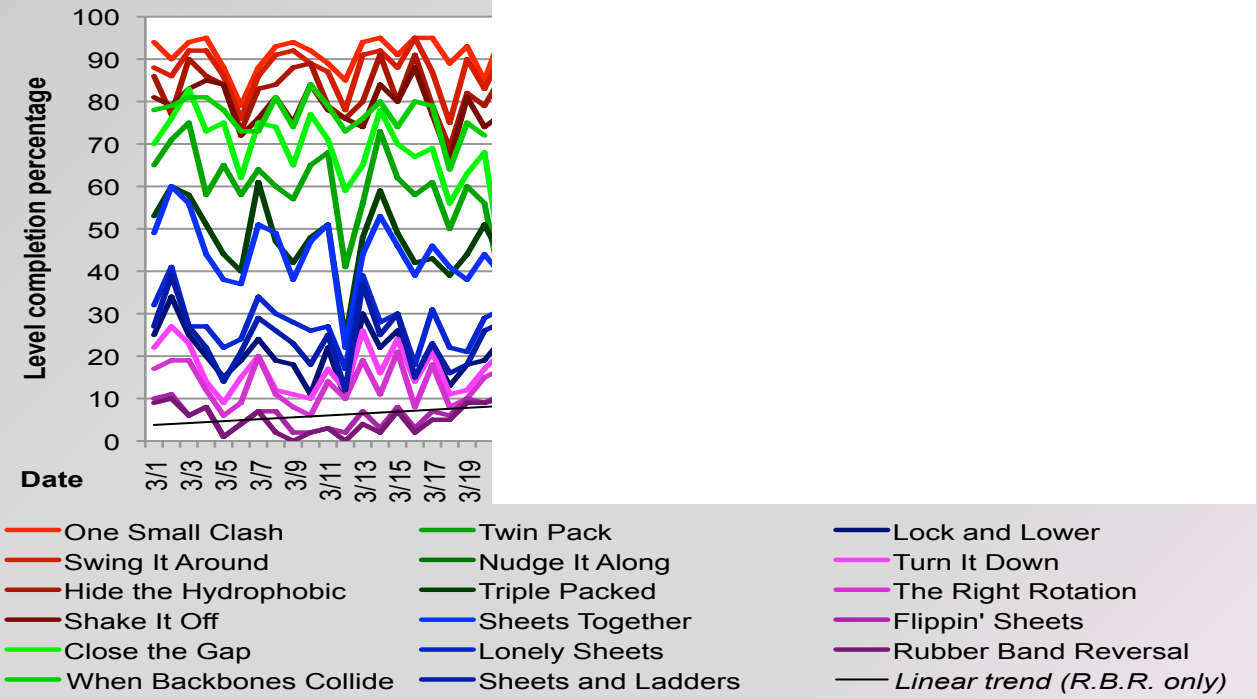
3D Structure

HARD


Pull Mode
Rank: -
Score:
0
Soloist
BETA Prediction Test
No bonuses or conditions
Group Competition
Soloist Competition
Chat - Puzzle
auto show
Chat - Global
auto show
Notifications
auto show
Actions
Undo
Modes
Behavior
View
Menu

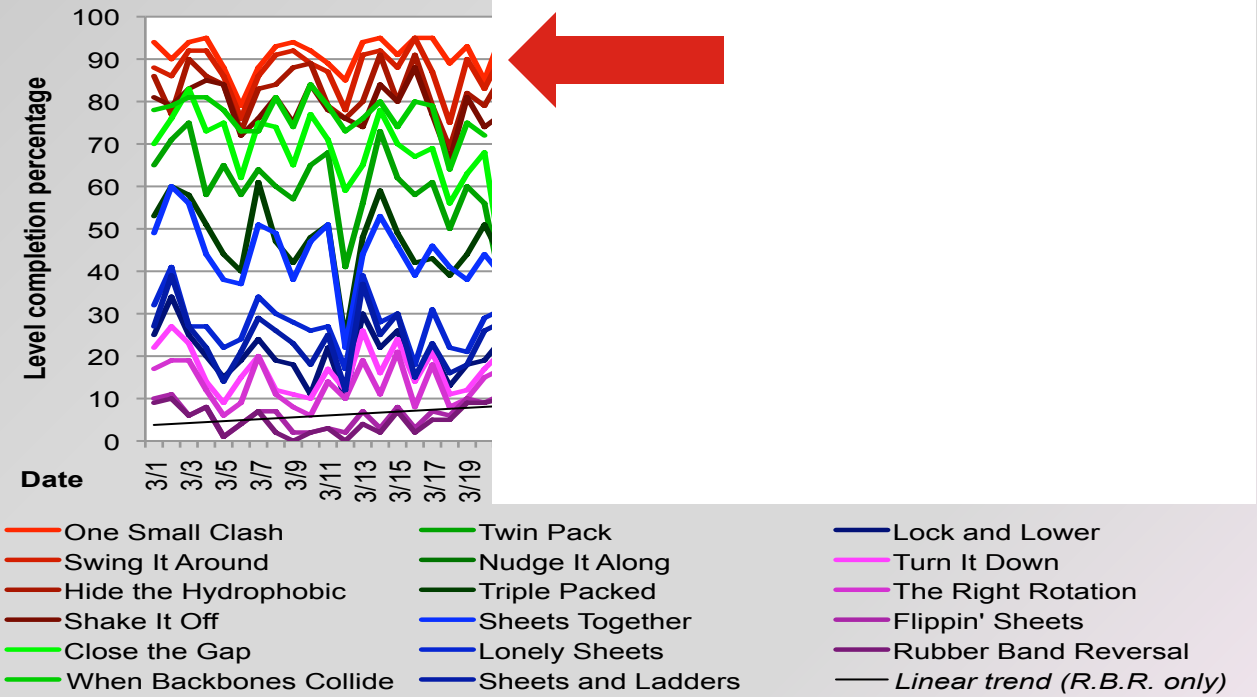
Training

refinement



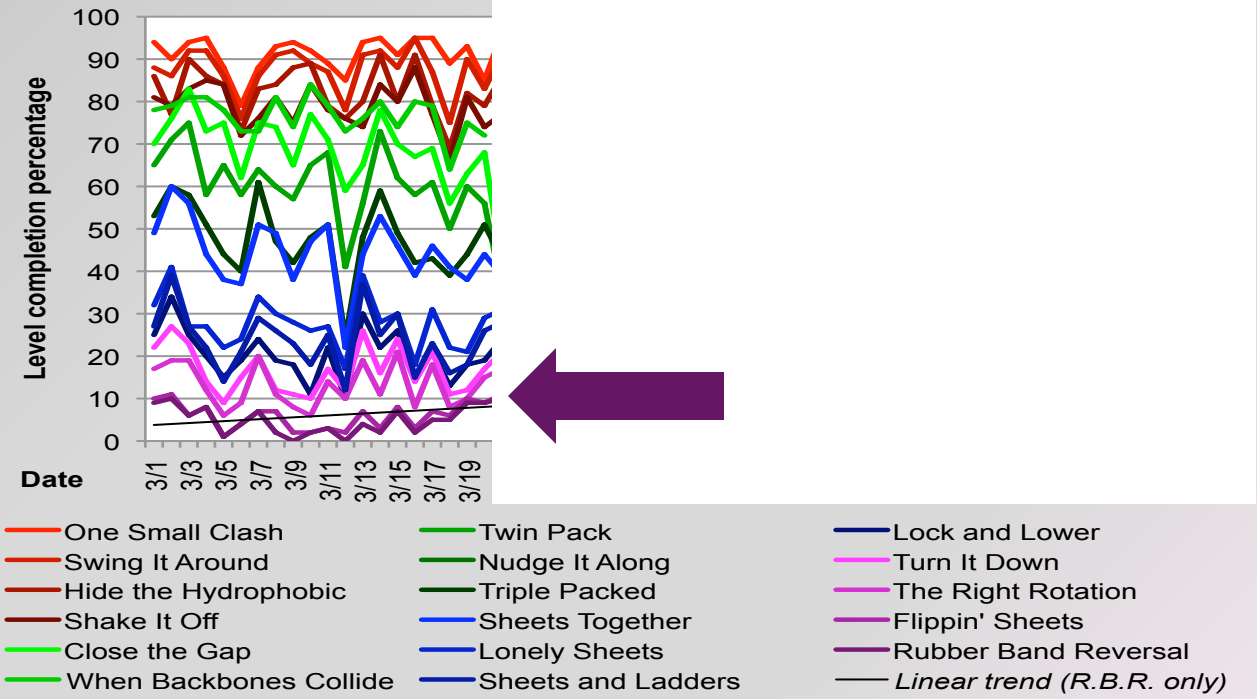
Training

refinement



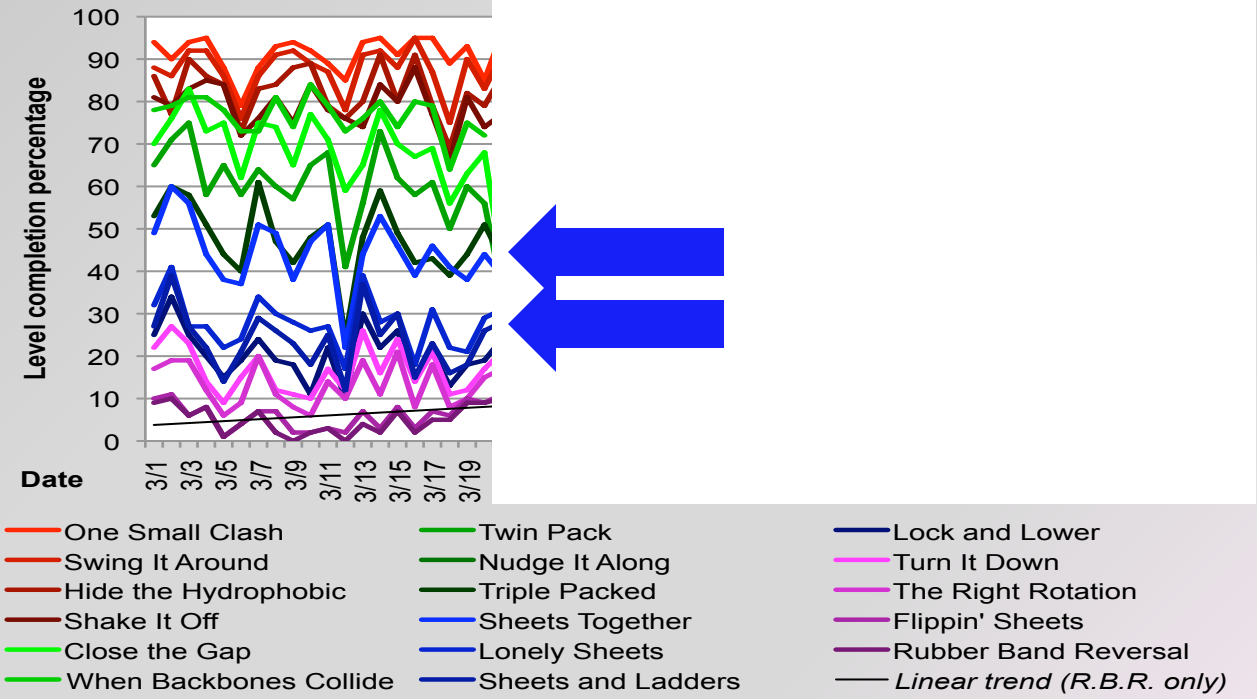
Training

refinement



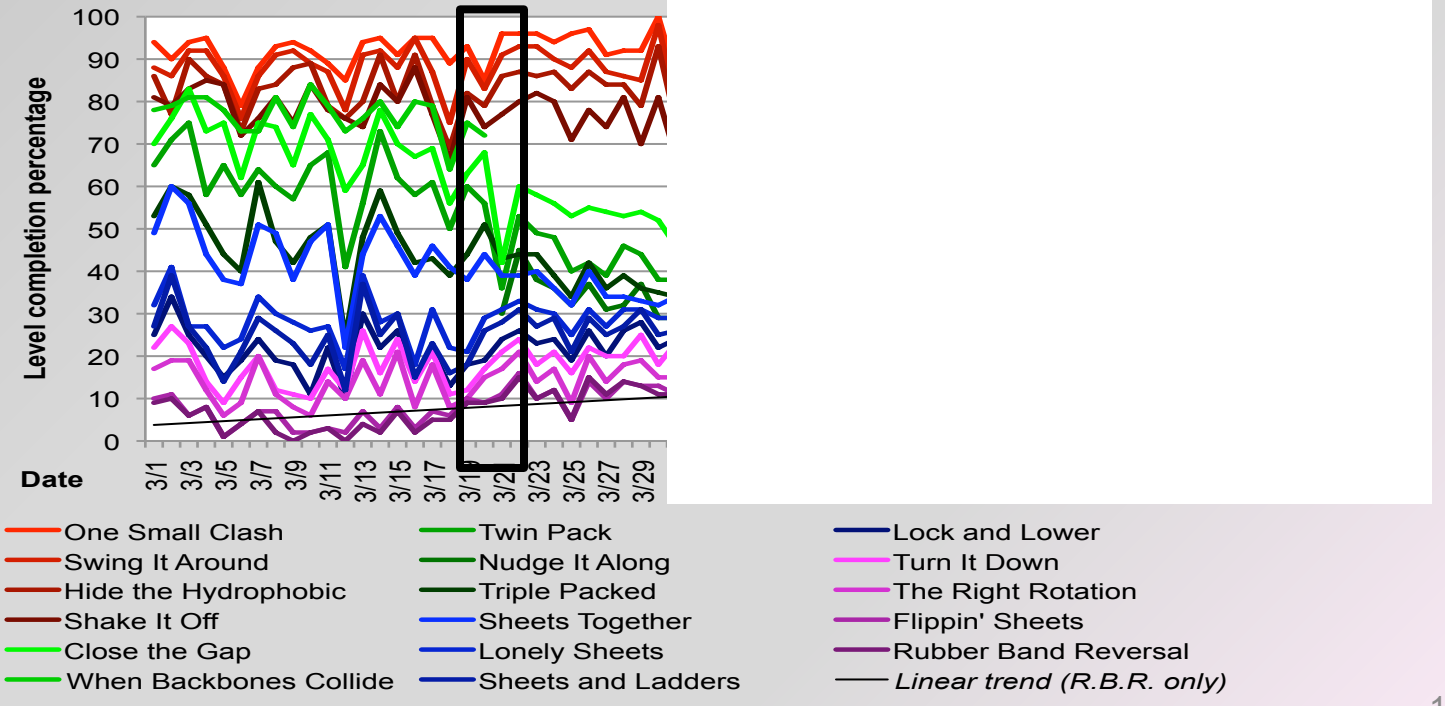
Training

refinement



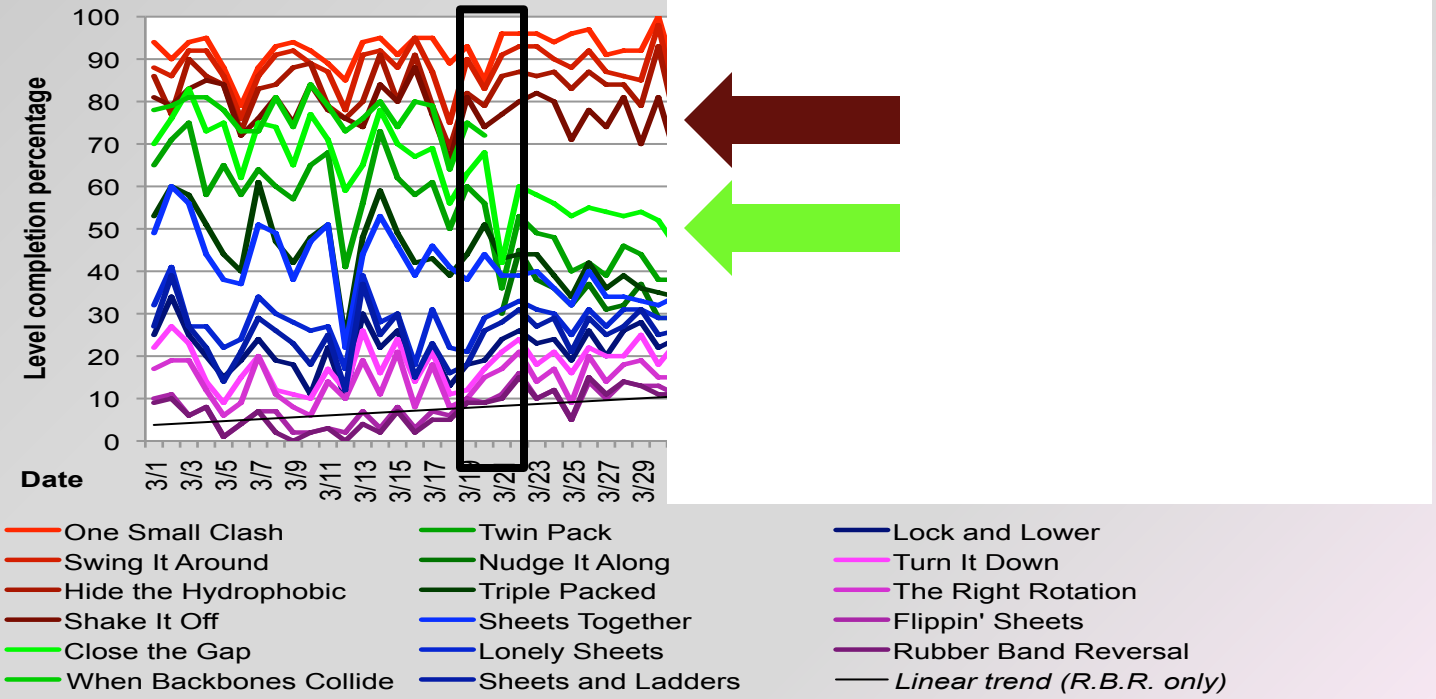
Training

refinement



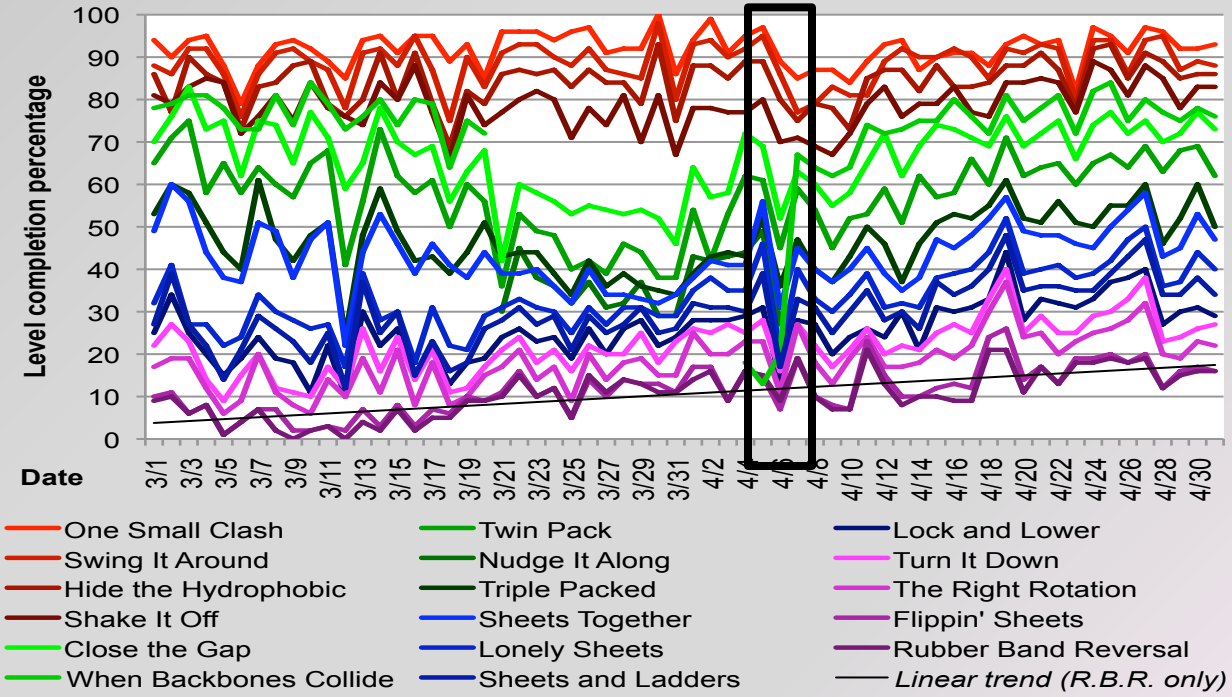
Training

refinement



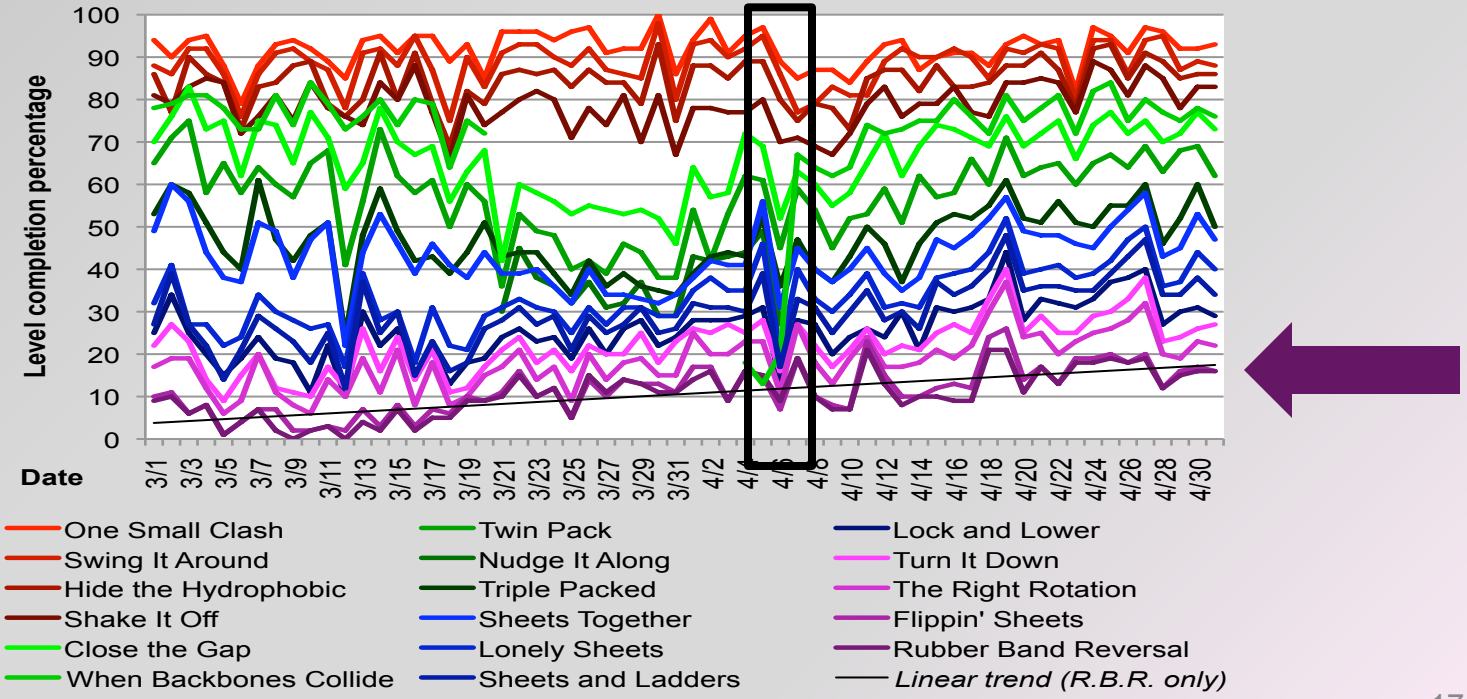
Training

refinement

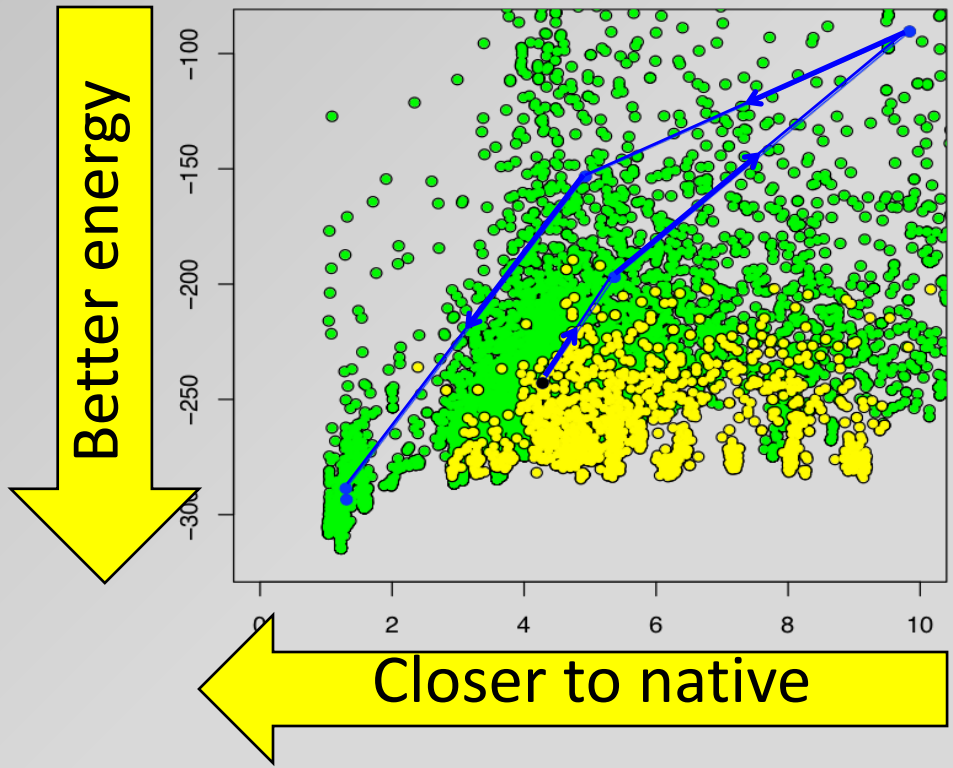


Training

refinement



Comparison to Algorithm



Player solutions

Top player path

Algorithm solutions

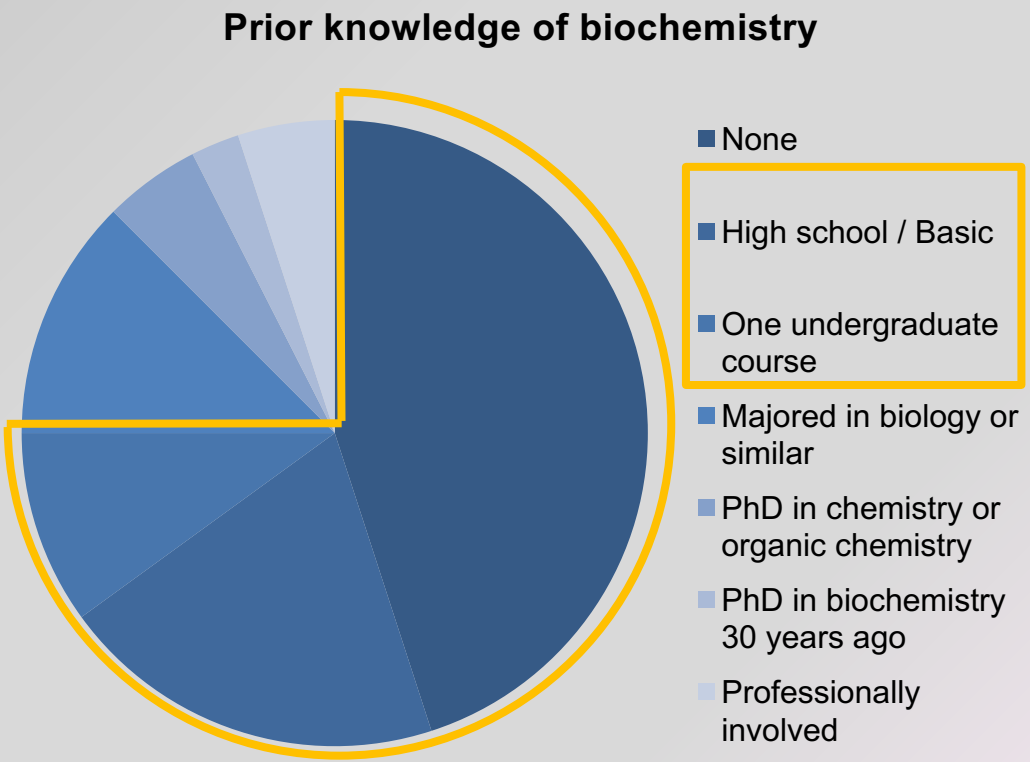
nature
CHEMISTRY IN MOTION
Attosecond snapshots reveal dynamics of valence electrons
HEART DISEASE
Scanning for serum lipid genes
CITIZEN SCIENCE
Crowd-based supercomputing
OCEAN MONITORING
Tap into submarine cables
NATUREJOBS
Marine biology

nature biotechnology

nature structural & molecular biology
Competing for RNAs

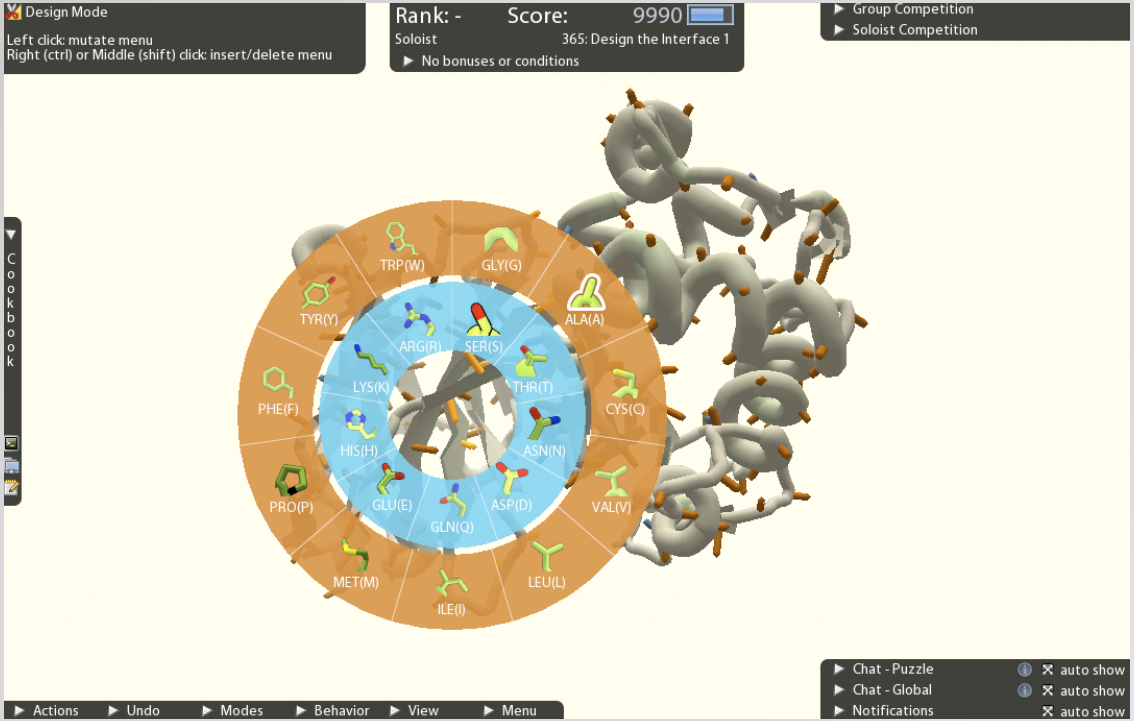
PNAS

Very broad appeal

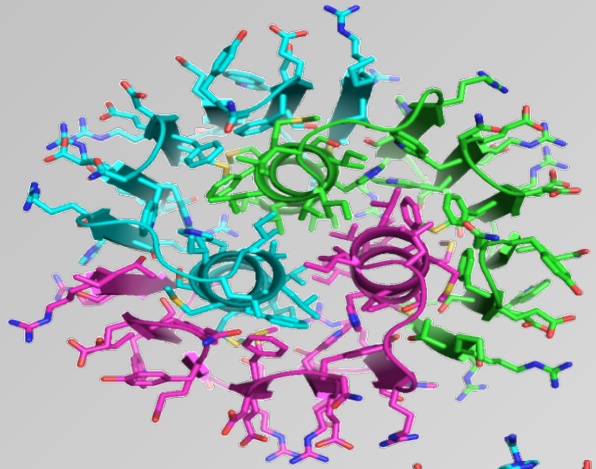


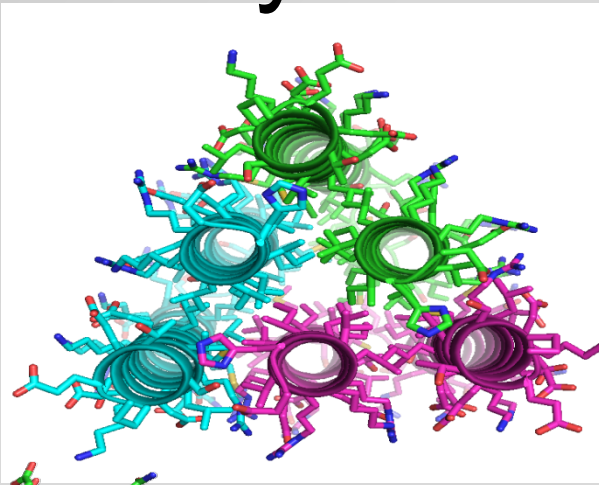
Creativity

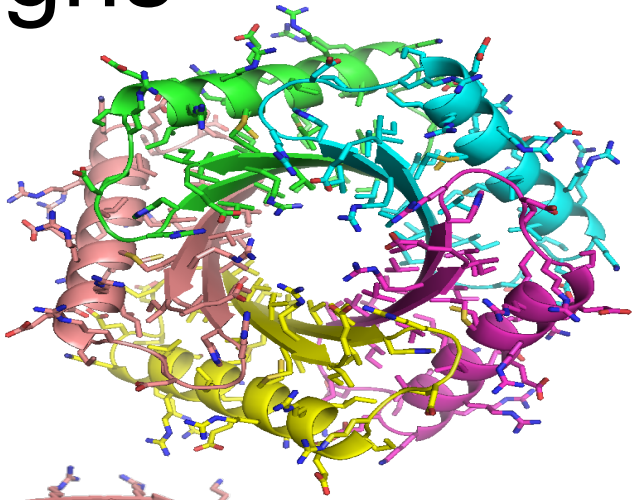
Drug Design

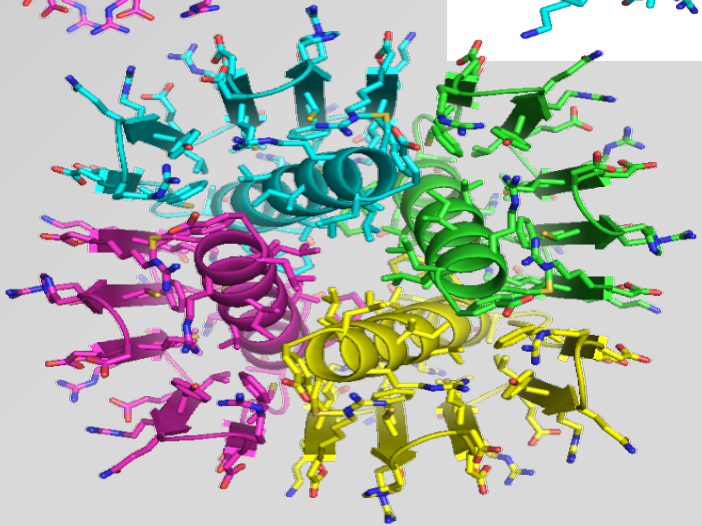

Design Mode
Left click: mutate menu
Right (ctrl) or Middle (shift) click: insert/delete menu
Rank: -
Score: 9990
Soloist
365: Design the Interface 1
No bonuses or conditions
Group Competition
Soloist Competition
Cookbook
TRP(W)
GLY(G)
TYR(Y)
ALA(A)
ARG(R)
SER(S)
LYS(K)
THR(T)
PHE(F)
CYS(C)
HIS(H)
ASN(N)
PRO(P)
GLU(E)
ASP(D)
VAL(V)
GLN(Q)
MET(M)
LEU(L)
ILE(I)
Chat - Puzzle
auto show
Chat - Global
auto show
Notifications
auto show
Actions
Undo
Modes
Behavior
View
Menu

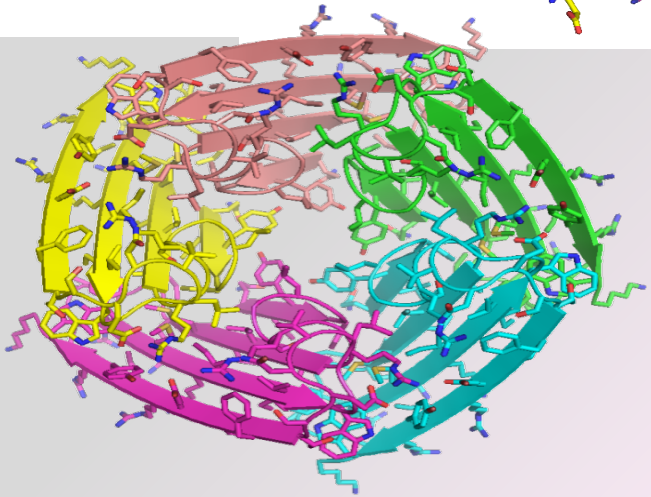
Foldit PlayerDesigns









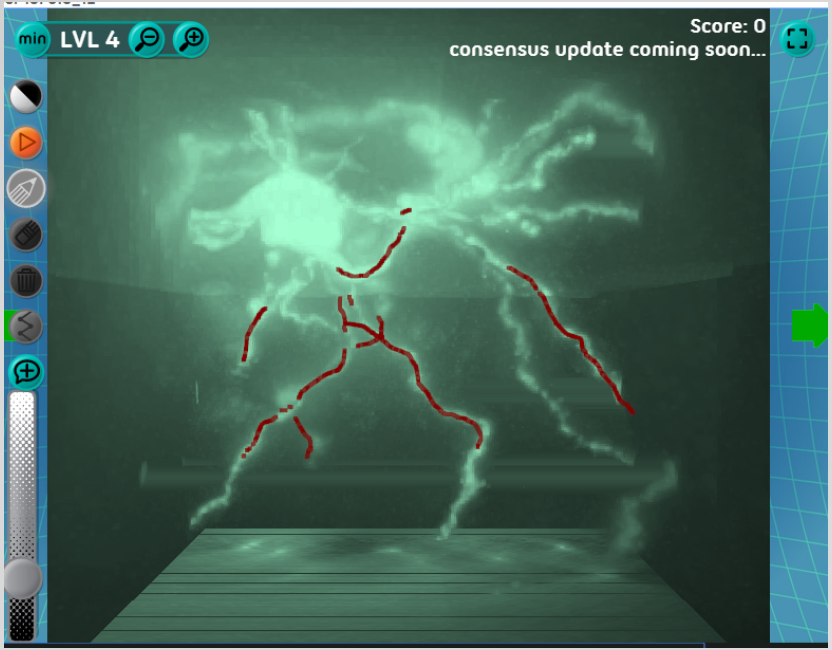


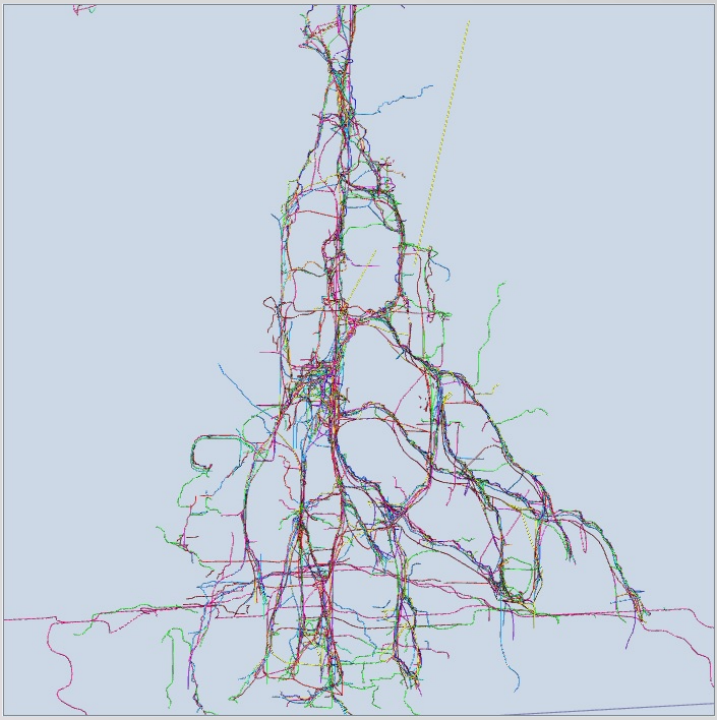
Accelerating Neuroscience

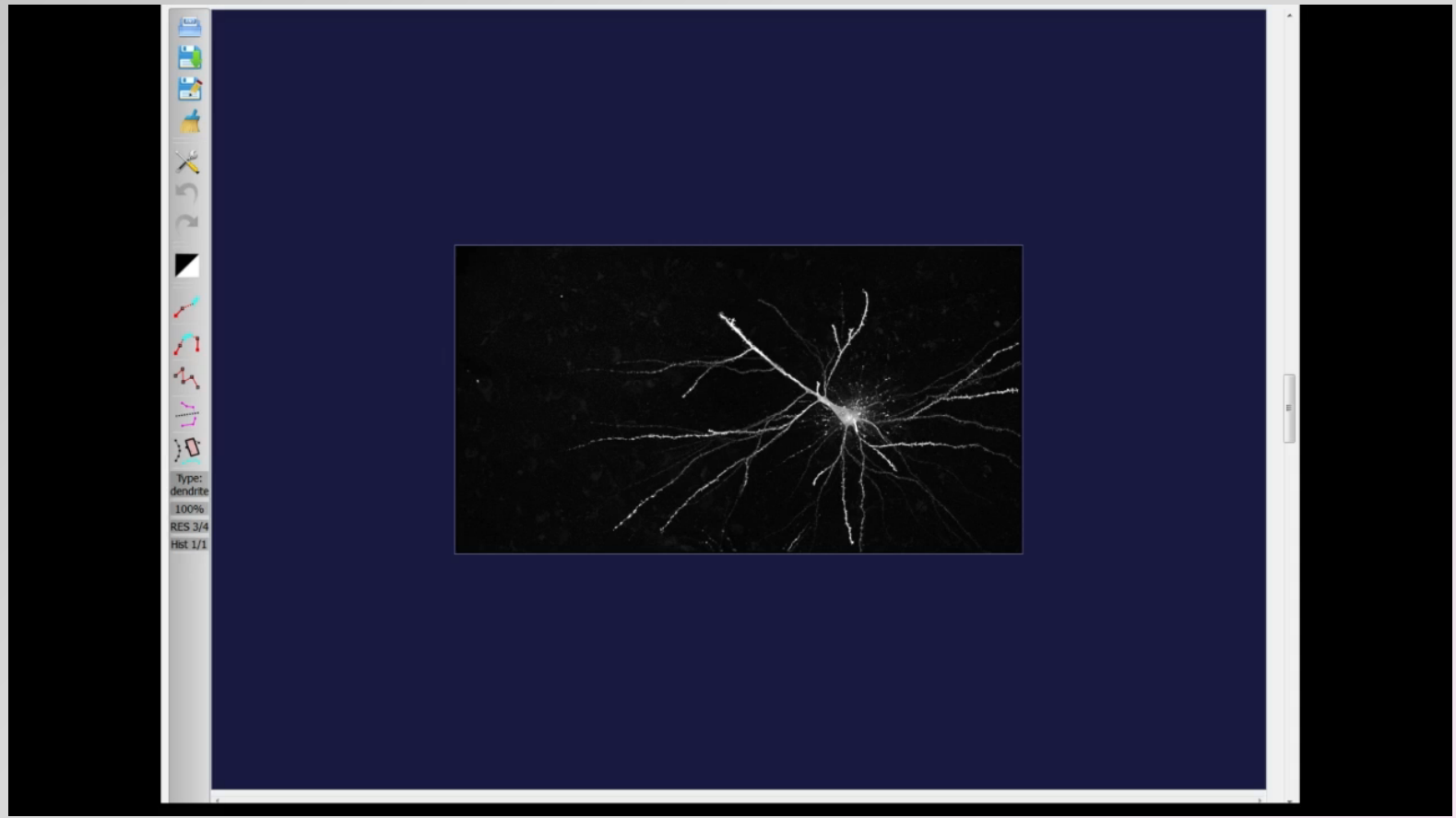
Neuronal Reconstruction

Goal: produce high quality reconstructions at scale by developing novices
and speed up the scientific process **by 100x**

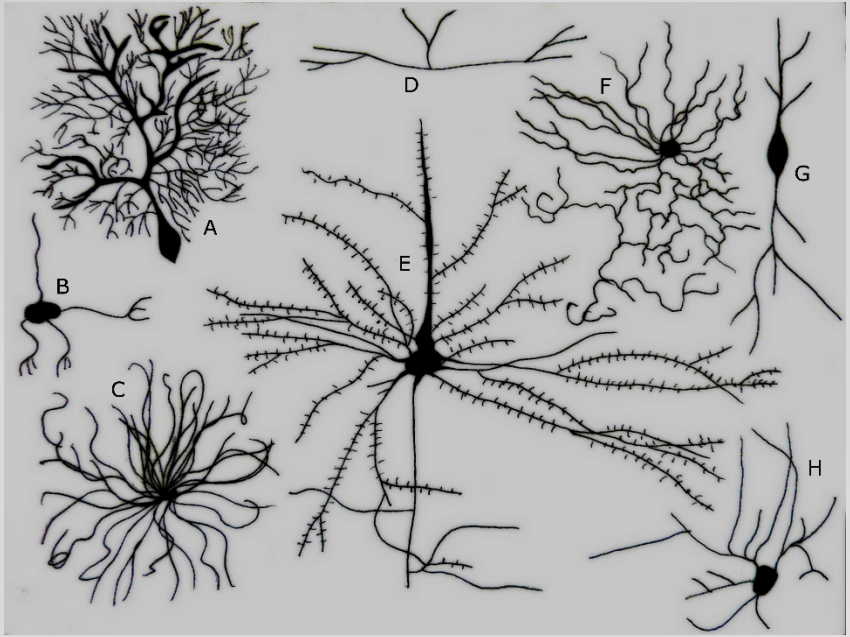
http://mozak.science

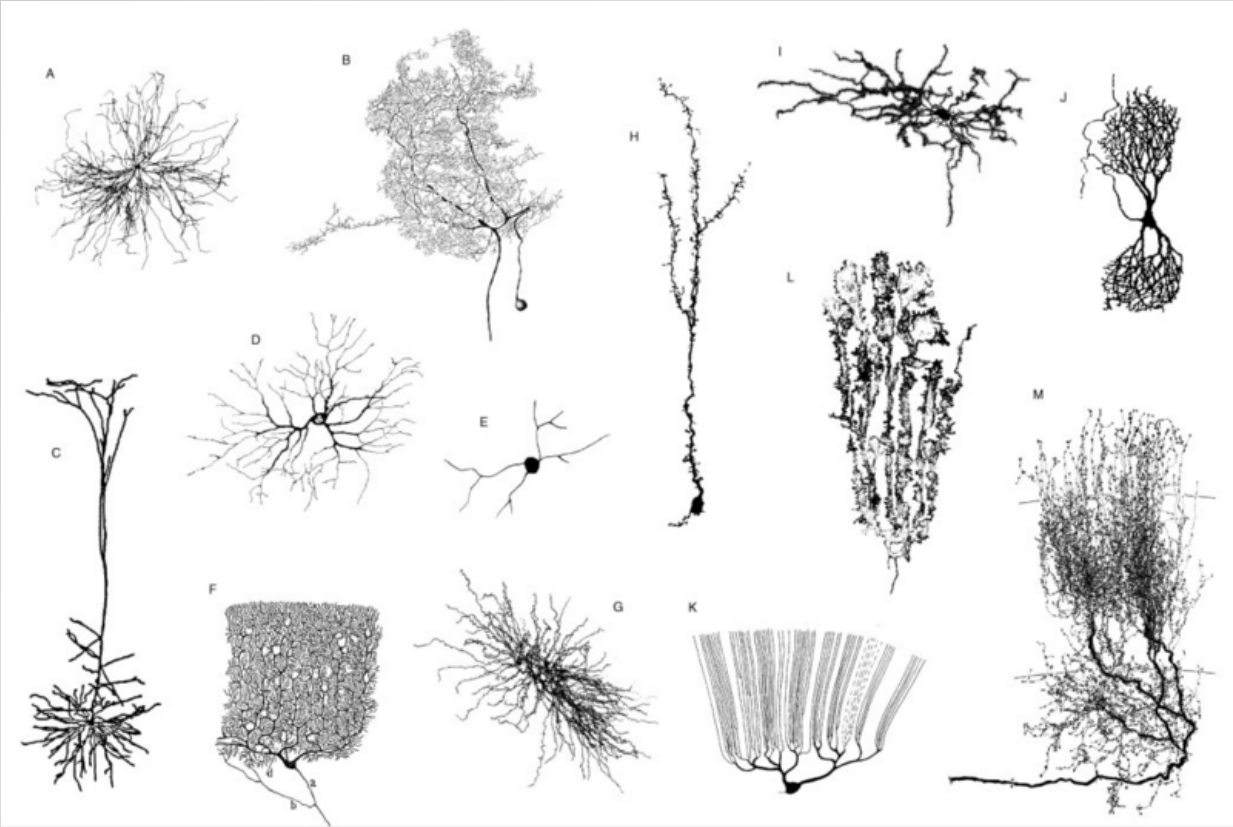





Type:
dendrite
100%
RES 3/4
Hst 1/1

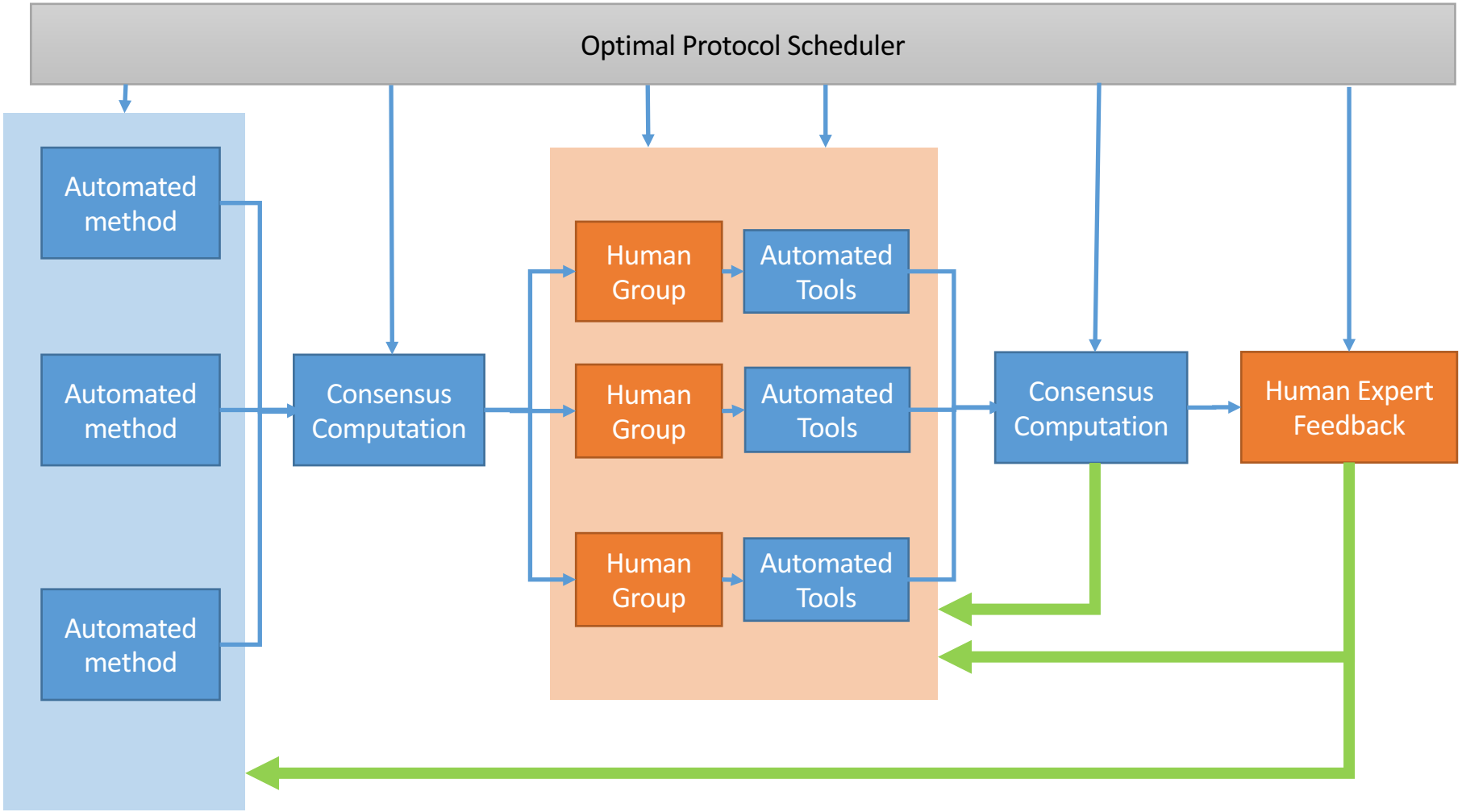
Ontology Classification



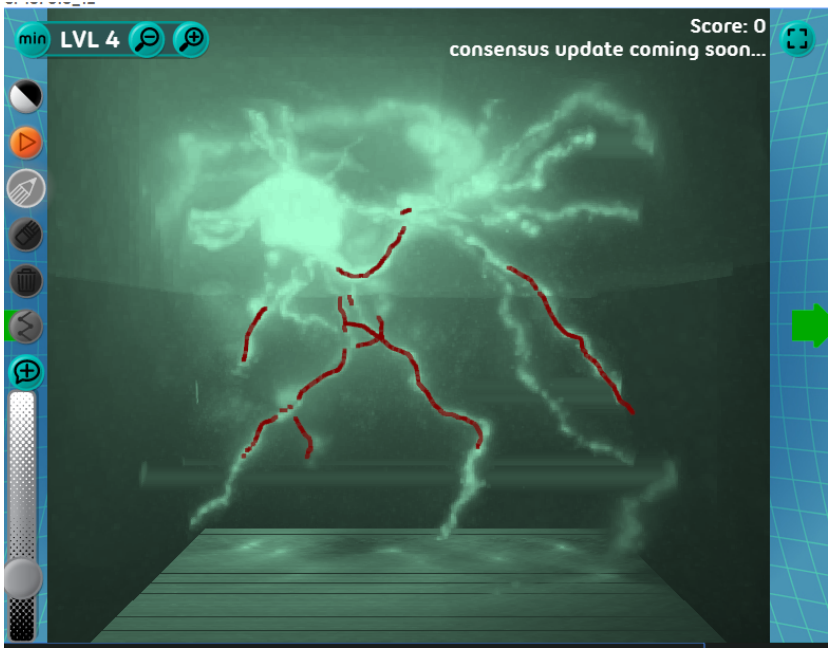

A
B
H
I
J
L
D
E
M
C
F
G
K

Temporal Neuron function

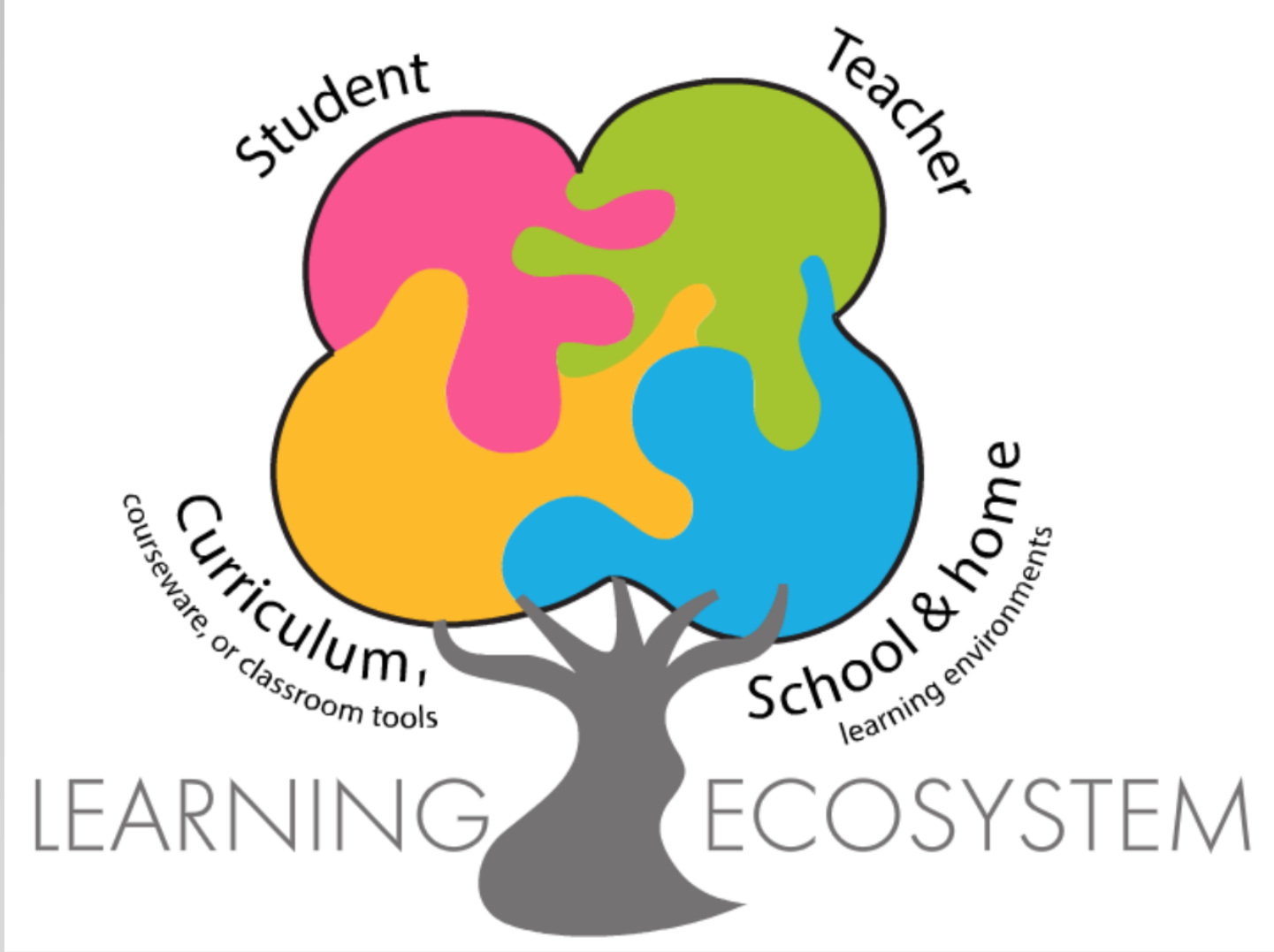
Modeling conectome network


Optimal Protocol Scheduler
Automated method
Automated method
Automated method
Consensus Computation
Human Group
Automated Tools
Human Group
Automated Tools
Human Group
Automated Tools
Consensus Computation
Human Expert Feedback

http://mozak.science



Rapid Mastery


Student
Teacher
Curriculum,
courseware, or classroom tools
School & home
learning environments
LEARNING ECOSYSTEM

Why Current ML Methods Fall Short

1. Ecosystem models do not exist
2. State space orders of magnitude larger than what RL methods can handle
3. Undesirable exploration sampling

Research Frontiers

1. Parametric synthesis of content curriculum
2. Human-in-the-loop ML for model refinement
3. Policy exploration in high-stakes environment

Ideal Action Space: Infinite Curriculum

Content Amplification

- Encodes desired thought processes which automatically generate content
- Produces deep learning maps with knowledge connections

NS3-9: Comparing and Ordering Numbers

1. Write the **value** of each digit. Then complete the sentence.

a) 3 7 → ☐ → ☐

4 7 → ☐ → ☐

_____ is greater than _____

b) 6 4 7 → ☐ → ☐ → ☐

5 4 7 → ☐ → ☐ → ☐

_____ is greater than _____

2. Circle the pair of digits that are different in each pair of numbers.
Then write the greater number in the box.

a) 475
465
[475]

b) 360
260
[ ]

c) 852
858
[ ]

d) 136
126
[ ]

3. Read the numbers from left to right.
Circle the first pair of digits you find that are different. Then write the greater number in the box.

a) 583
597
[597]

b) 629
654
[ ]

c) 576
603
[ ]

d) 432
431
[ ]

e) 384
597
[ ]

f) 906
904
[ ]

g) 875
869
[ ]

h) 238
221
[ ]

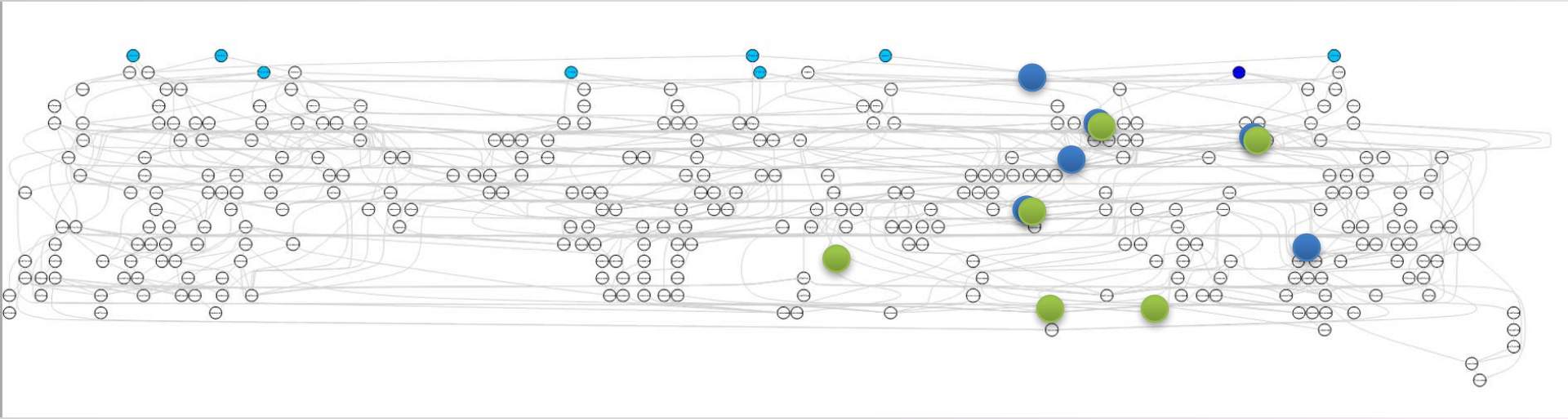
4. Circle the greater number.

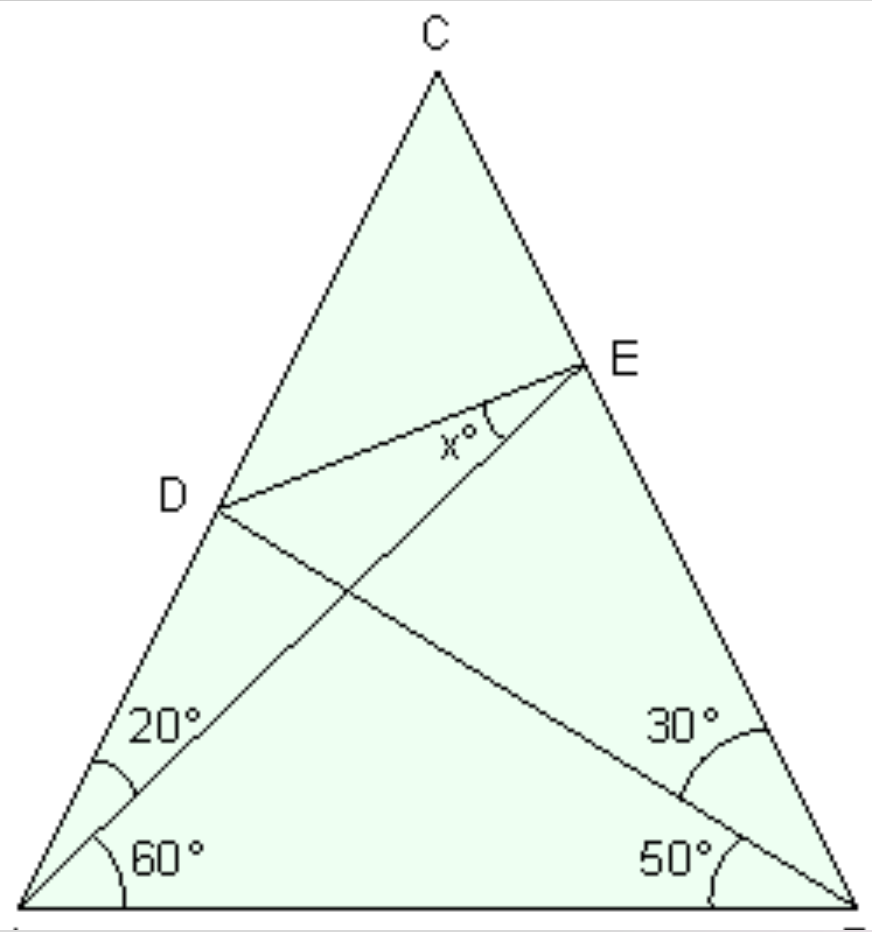
a) 111 or 311
b) 625 or 525
c) 321 or 721
d) 843 or 867
e) 480 or 412
f) 219 or 220
g) 125 or 122
h) 854 or 859
i) 336 or 330

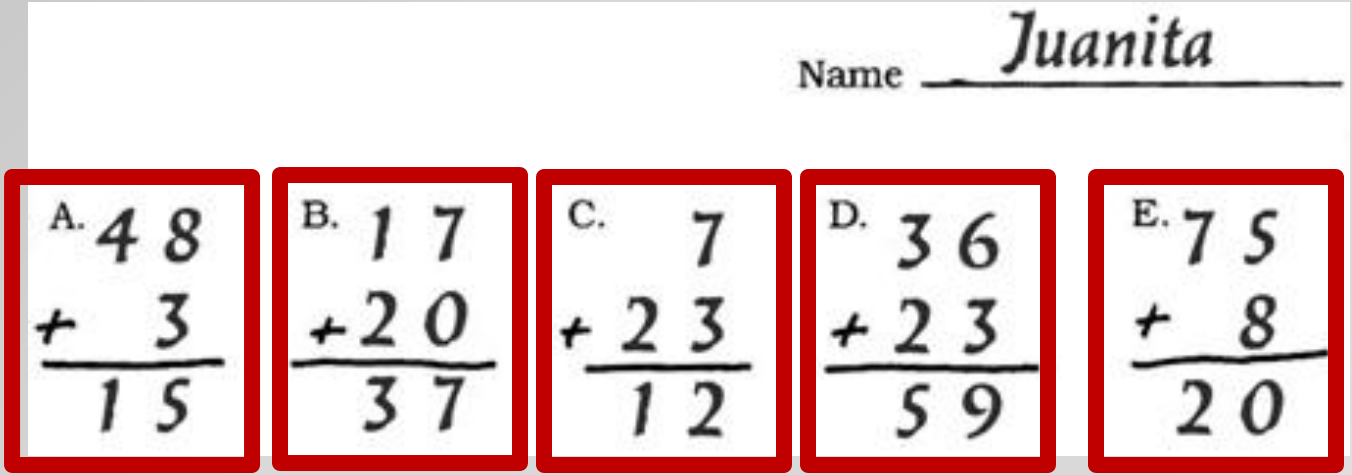
JumpMath and Singapore Math cover a tiny part of the full space

Green: Singapore Math

Blue: JumpMath




C
E
D
x°
20°
30°
60°
50°



Question: can we diagnose such errors automatically?

A.	B.	C.	D.	E.
48	17	7	36	75
+ 3	+ 20	+ 23	+ 23	+ 8
15	37	12	59	20

if there are only three digits:

add all digits

else:

for each column:

add column

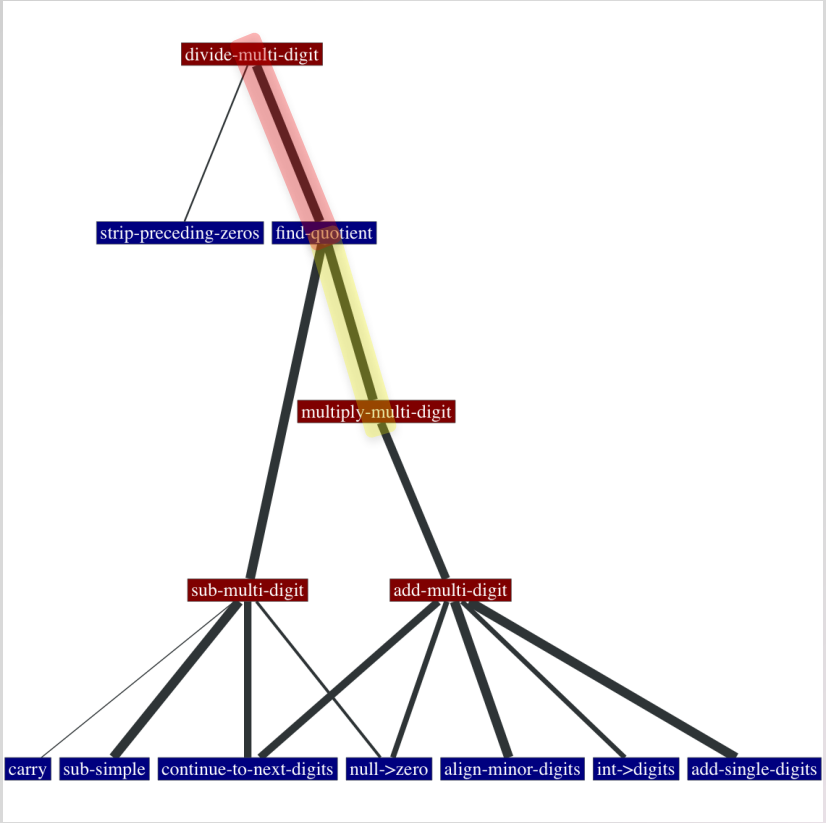
Tractability via Model Design

Three-Level Specialization

1. Concept/Skill Level
2. Practice Problem Progression
3. In-problem through process scaffolding

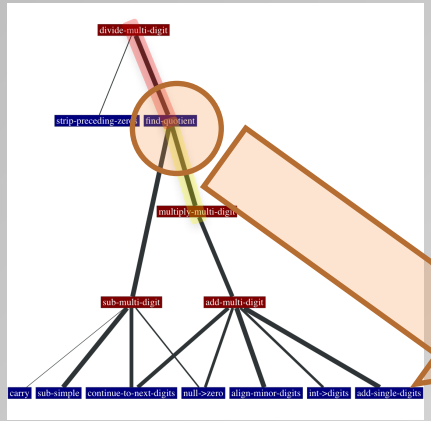
Level 1

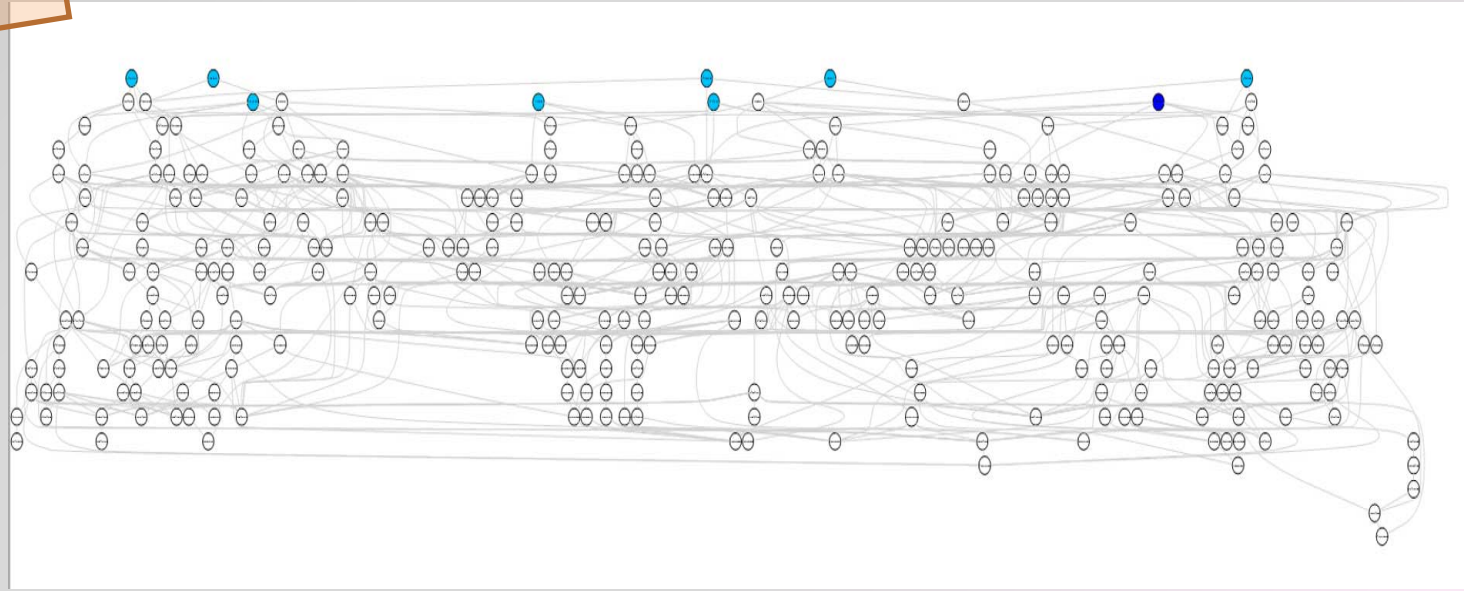
Personalized progression through skill map



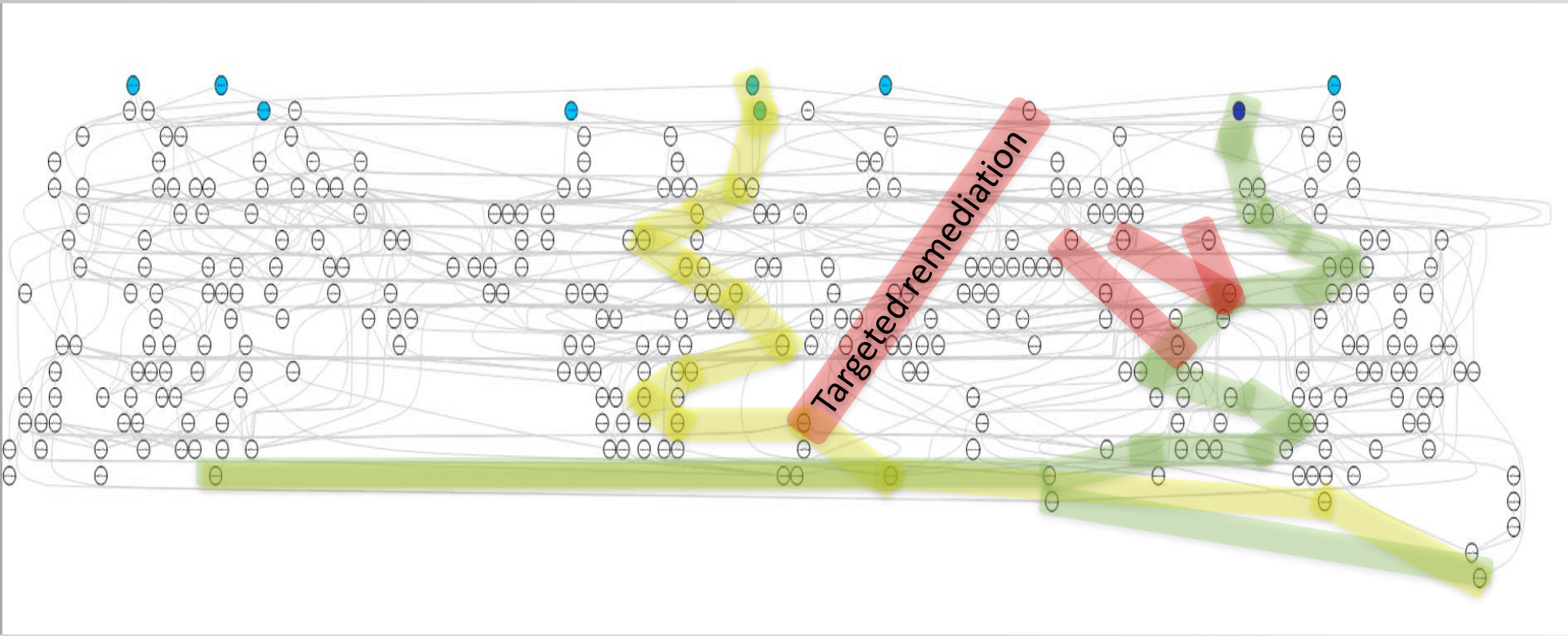
Level 2

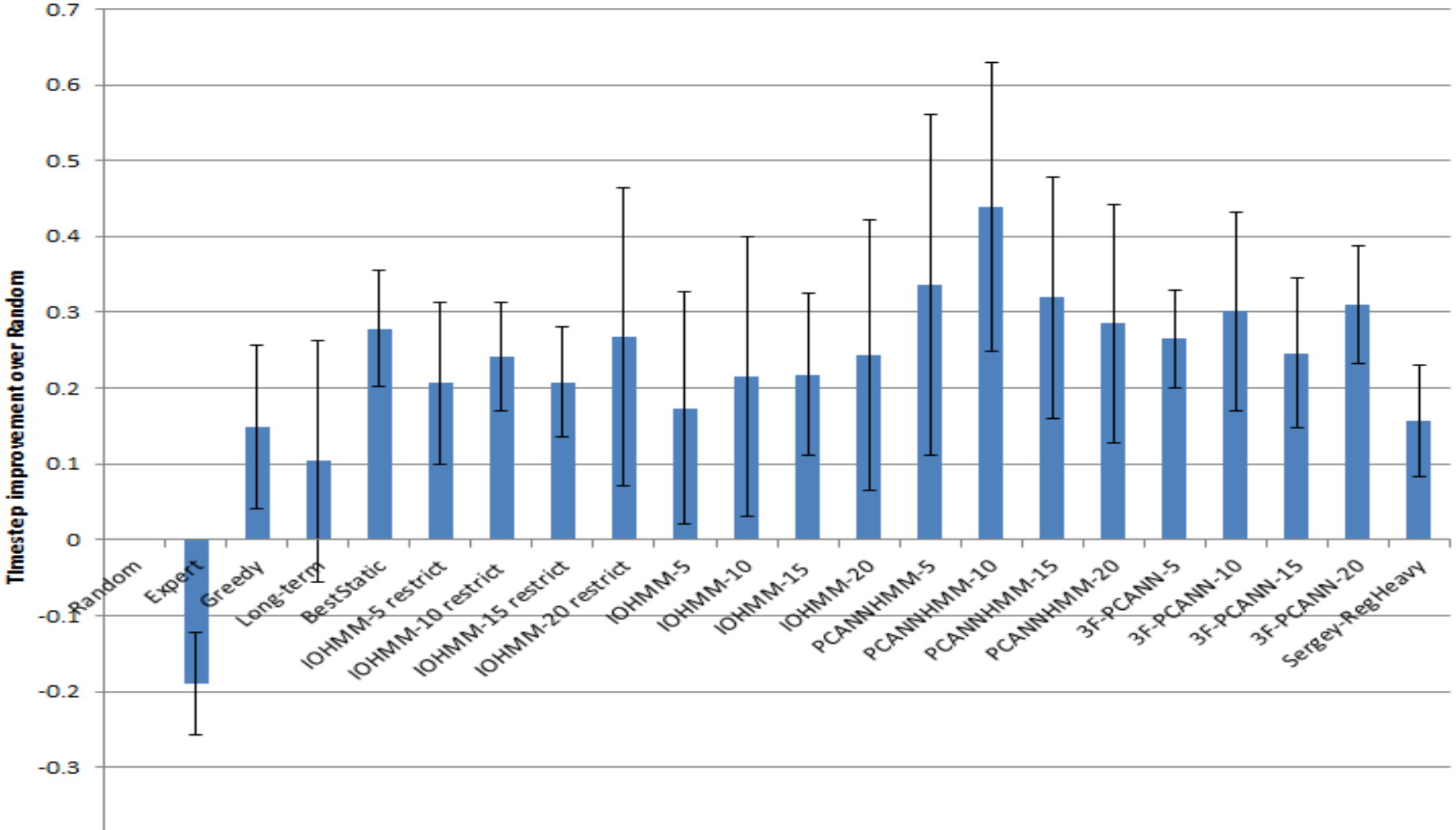
Each skill has a detailed thinking process map





Personalized progressions for two students




Timestep improvement over Random
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
Random
Expert
Greedy
Long-term
BestStatic
IOHMM-5 restrict
IOHMM-10 restrict
IOHMM-15 restrict
IOHMM-20 restrict
IOHMM-5
IOHMM-10
IOHMM-15
IOHMM-20
PCANNHMM-5
PCANNHMM-10
PCANNHMM-15
PCANNHMM-20
3F-PCANN-5
3F-PCANN-10
3F-PCANN-15
3F-PCANN-20
Sergey-RegHeavy

Level 3

Real-time Actions

Find the greatest common factor
of these two numbers.

9 6

Next Problem

Next

Teacher Actions

Chi Weng
High five. Breakthrough in scaling fractions.
LIGHT BULB MOMENT

Daniel Wetzel
Adds denominators. Use visual to explain.
CONCEPT HELP

Wendy Merchant, Katie Richards, Lila Fleck
Large denominators.
GROUP WORK

Main Arrangement Normal Custom

Mike Jinsen								
Steven Gillstein								
Amanda Hintz								
Justin McLander								
Blake Neiderman								
Daniel Wetzel								
Andy Blankenship	off							
Anthony McWitt	off							
Aubrey Johnson	off							

Finish Exercise

Close

Main	playing copilottt3
Amanda Hintz	
Andy Blankenship	Offline
Anthony McWitt	Offline
Aubrey Johnson	
Blake Neiderman	Offline
Chelsey Laster	Offline
Chi Weng	
Daniel Wetzel	Offline
Darren Singer	Offline
Franco Slatery	
Grant Leighy	Offline
Hank Fremont	Offline
Isaac Runti	Offline
Jimena Masi	Offline
Justin McLander	
Kat Stover	Offline
Katie Richards	Offline
Kirk Brown	Offline
Lila Fleck	
Madi Klein	Offline
Mike Jinsen	
Patricia Morany	Offline
Pedro Mendez	Offline
Peter Gerting	Offline
Phil Croster	Offline
Richard Thompson	
Steven Gillstein	
Vince Allister	
Wendy Merchant	
Zeek Risenhouse	Offline

Katie R.
High five. Breakthrough in scaling fractions.

LIGHT BULB MOMENT

Algebra Mastery in Elementary School

3 equations without errors with behaviors that indicate fluency

$$ax + b = \frac{c + d}{e}$$

$$\frac{abx}{b + c} + c + 0 = d + yz$$

$$\frac{b - c}{x} = c + d \ - c$$

CHALLENGE WEEK **June 3 - 7**

Washington State ALGEBRA CHALLENGE 2013

We did it!

EQUATIONS: **390,935**

TOTAL EFFORT: **6 months 28 days 2 hours**

KIDS ARE STILL PLAYING:

EQUATIONS SOLVED:	TOTAL EFFORT:
429,053	7m 20d 07

Winners Announced!

This year's Washington State Algebra Challenge was a fantastic success. Not only did we exceed our original goal of 250,000 equations, but we had participation from a total of 4,192 students representing public and private schools, home school programs, and after school groups. Check out the winners by clicking the button below!

VIEW WINNERS!

NORWAY ALGEBRA CHALLENGE 2014

7 700 000
Likninger løst

DET HENDTE:
13. - 17. Januar 2014

36 110 elever løste likninger sammen
1711 klasser deltok i utfordringen
93% oppnådde "mestring" innen 1½ time

En uforglemmelig matematikktime!

Fra 13. til 17. Januar 2014 ble en tilpasset versjon av DragonBox gjort gratis tilgjengelig for alle skoler i Norge. Les om hvordan det gikk her.

Kontakt oss!

ALGEBRA CHALLENGE 2014

Minnesota Algebra Challenge

We did it!

EQUATIONS: **644,936**

TOTAL EFFORT: **9 months 29 days 22 hours**

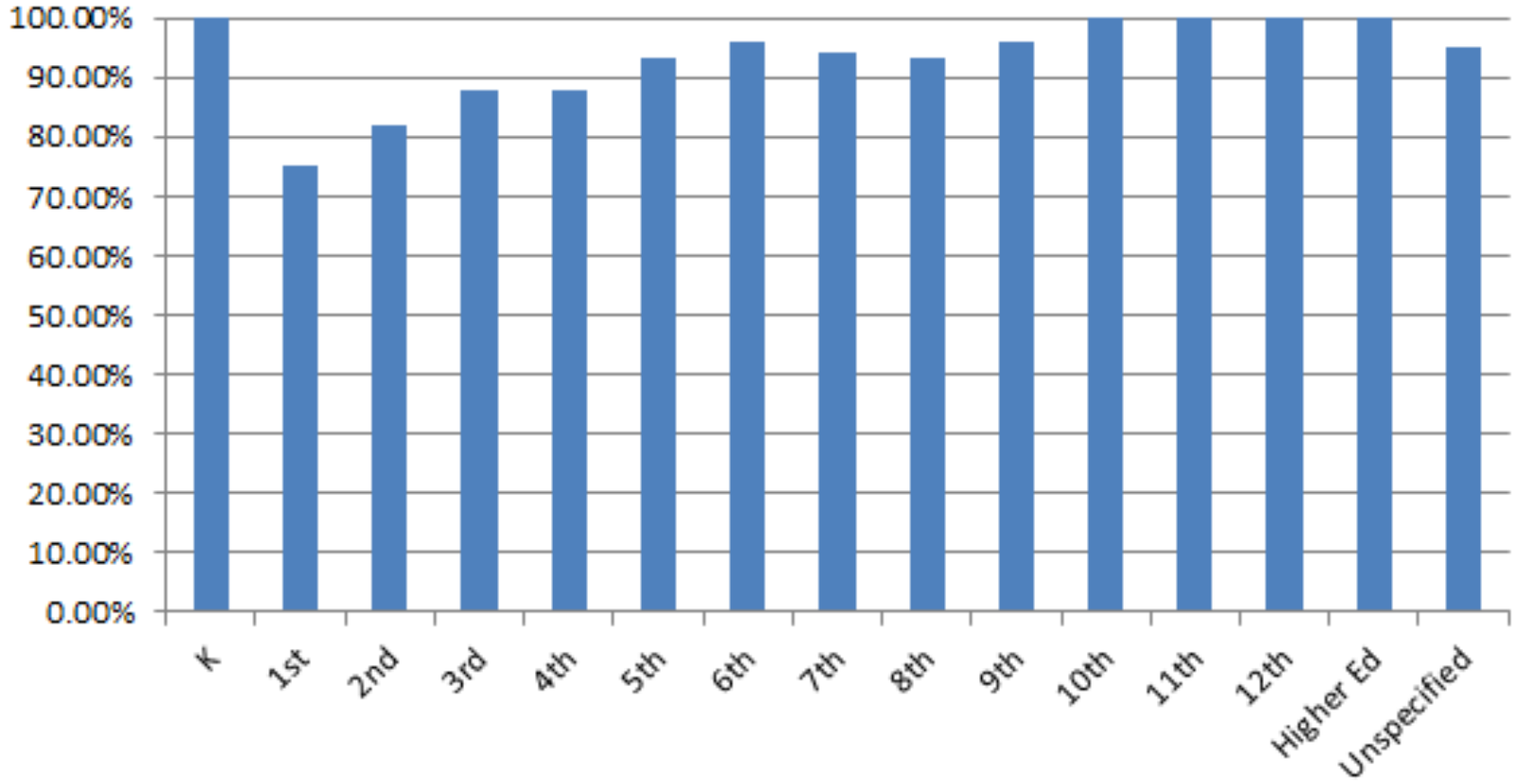
	EQUATIONS SOLVED:	TOTAL EFFORT:
KIDS ARE STILL PLAYING:	697,292	10m 25d 13

Congratulations!

Over 6000 students completed over 600,000 equations – most of those during the first 5 days! Total student playing time reached almost 10 full months (24 hour per day)!! An amazing effort by all. Quite a few classrooms achieved over 95% mastery and all the students were exposed to (and hopefully excited by) the conceptual ideas involved in solving algebraic equations. Thanks to all who participated.

ge.org

1.5 Hours of Play / Percent Acheived Mastery



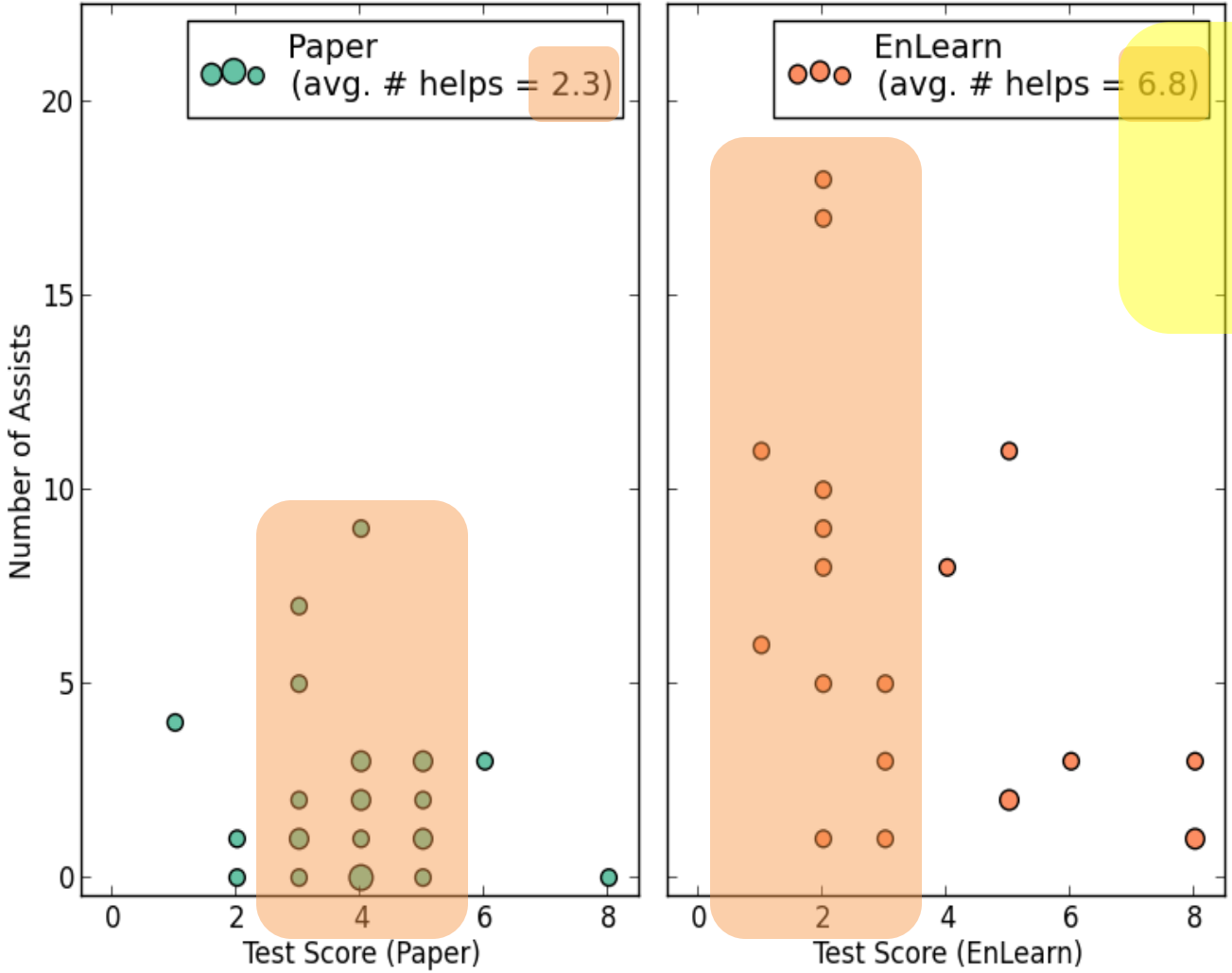
Norway Challenge

- 8 million equations
- 42% of all work is done in after school hours

Application to full curricula

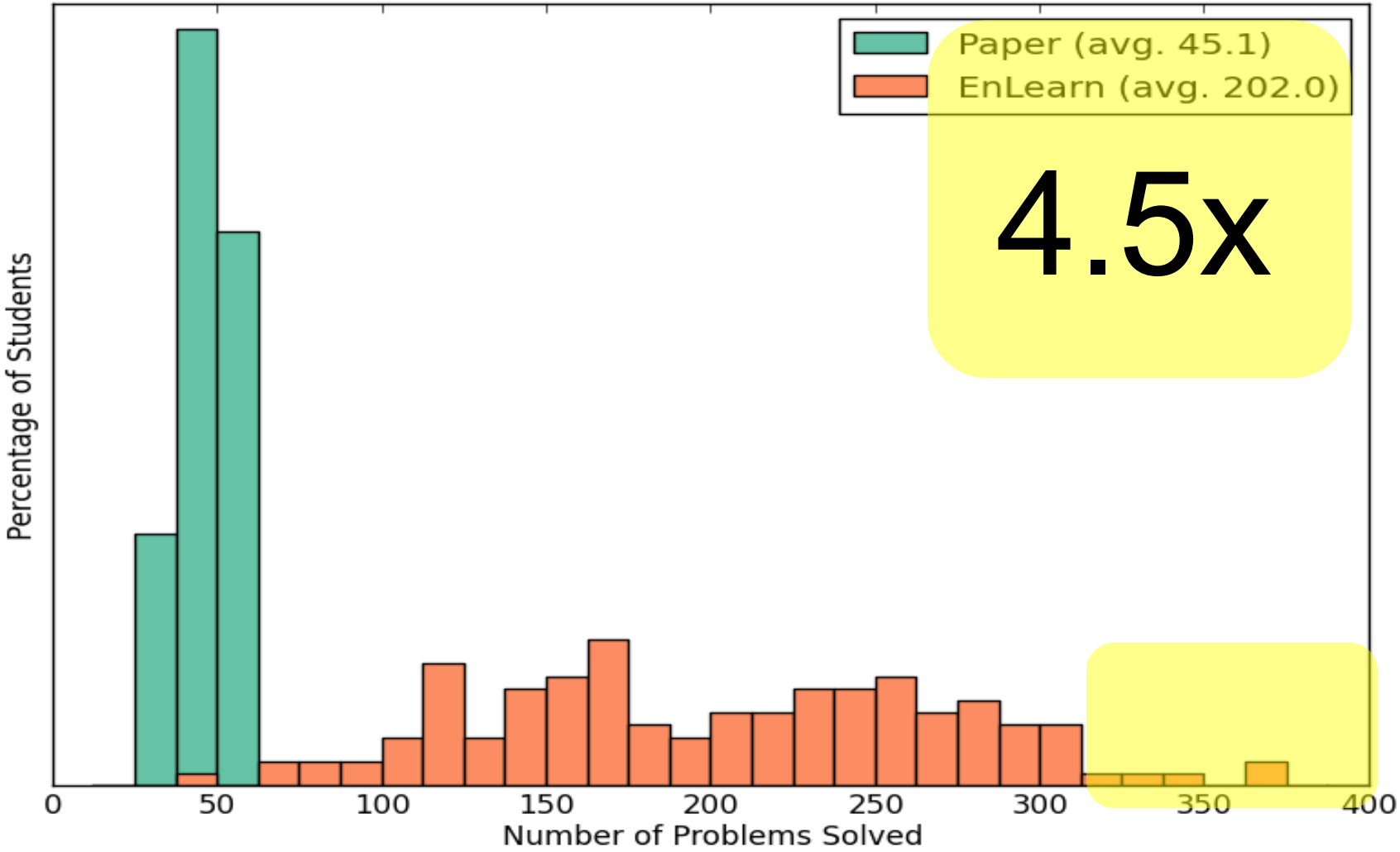
- Same teacher
- Same curriculum
- Paper and Enlearn platform

Teacher Effects


Paper
(avg. # helps = 2.3)
EnLearn
(avg. # helps = 6.8)
Number of Assists
Test Score (Paper)
Test Score (EnLearn)
0
5
10
15
20
0
2
4
6
8

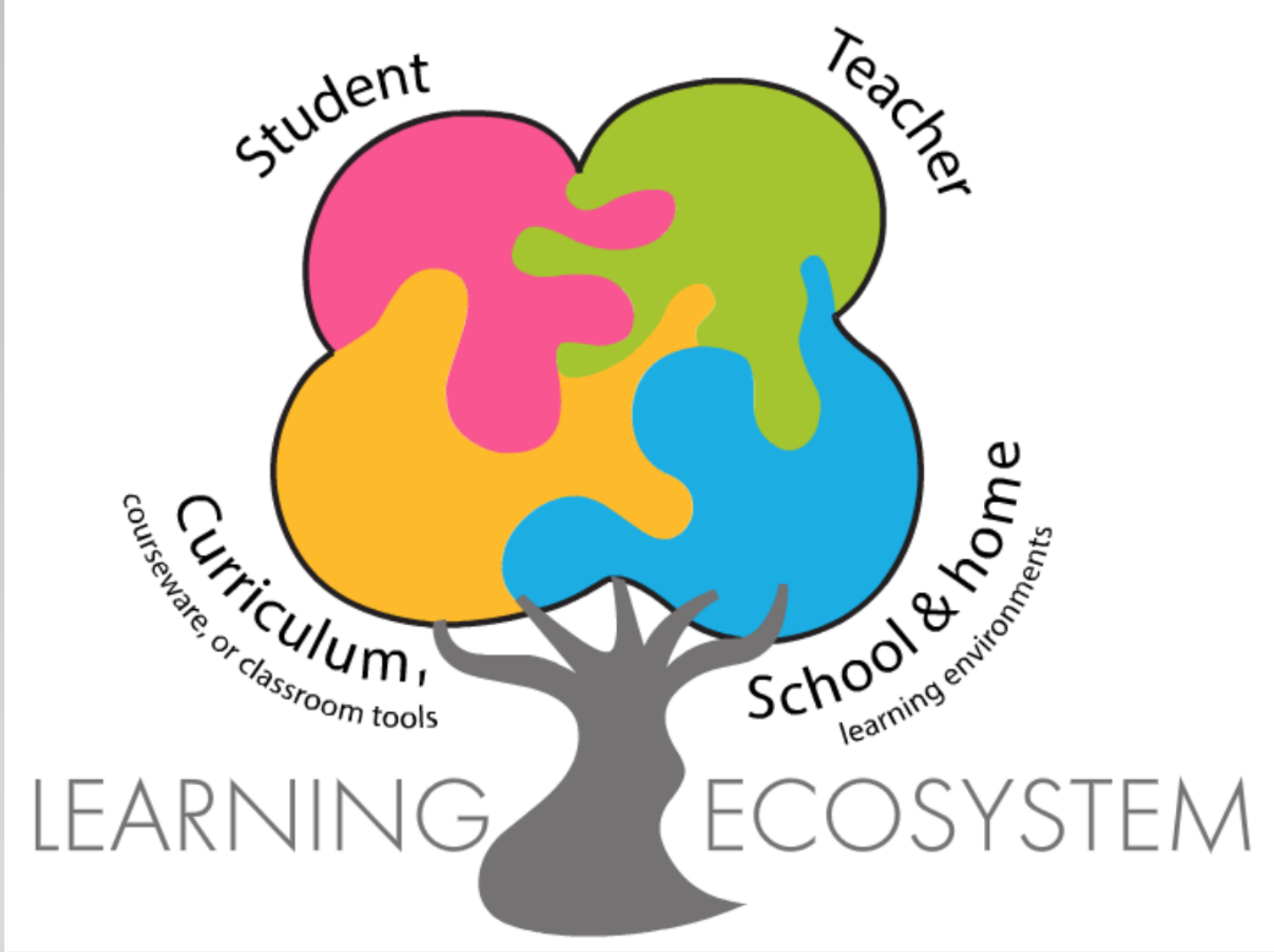
3x

Student Effects


Paper (avg. 45.1)
EnLearn (avg. 202.0)
4.5x
Percentage of Students
0
50
100
150
200
250
300
350
400
Number of Problems Solved

Current Deployment

- Voyager Sopris K-5 ELA and Math
- 100 million first generation learners in India
  - basic numeracy and literacy
  - Deployed by Indian non-profit Ekstep
- One of the largest assessment companies in US
  - Generation of test items from exemplars
- PBS Kids Scientific Inquiry (4 years of content)


Student
Teacher
Curriculum,
courseware, or classroom tools
School & home
learning environments
LEARNING ECOSYSTEM

Key Frontiers

- AI on Human-Computer Symbiosis Architecture
- Focus on model design/refinement
- Optimize the entire ecosystem