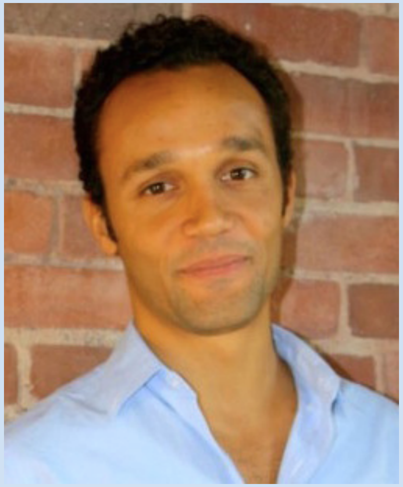
Using video games to reverse engineer human intelligence

Sam Gershman
Department of Psychology and Center for Brain Science
Harvard University



Pedro
Tsividis

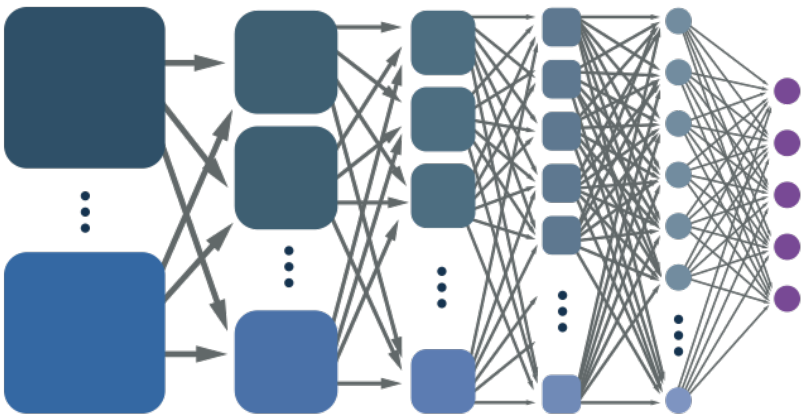
Josh
Tenenbaum

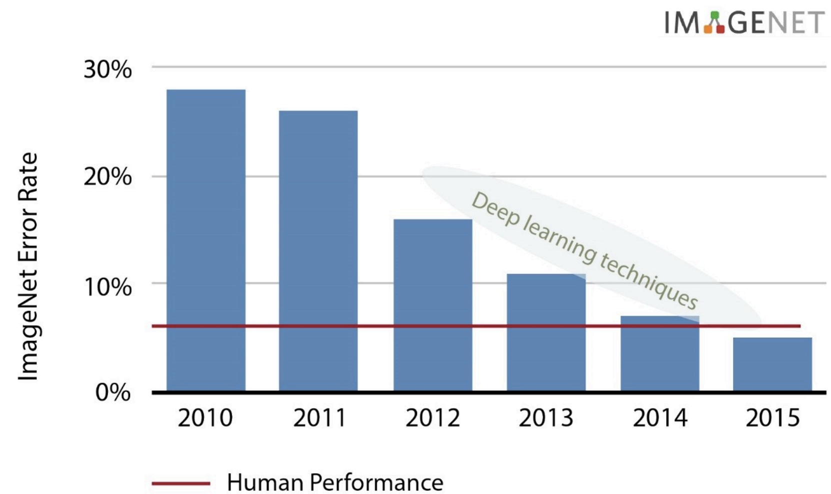
Joao
Loula

Thomas
Pouncy

Funding:

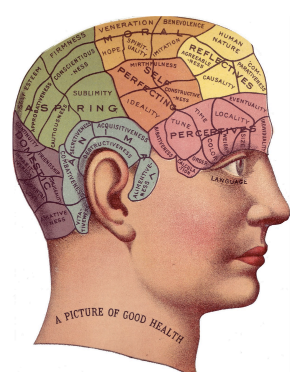
A seductive hypothesis



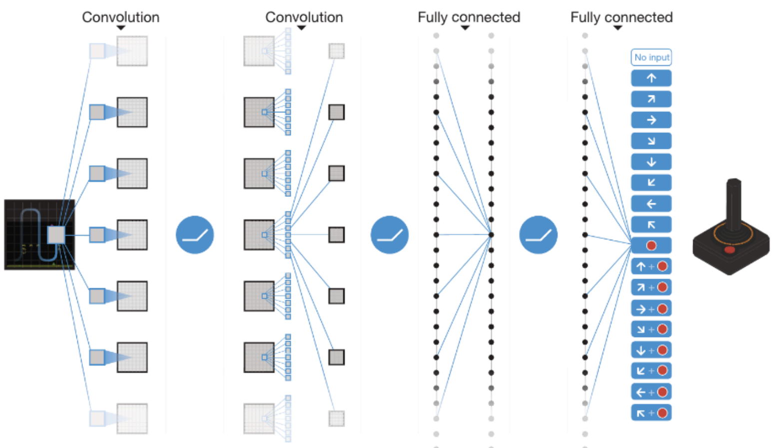


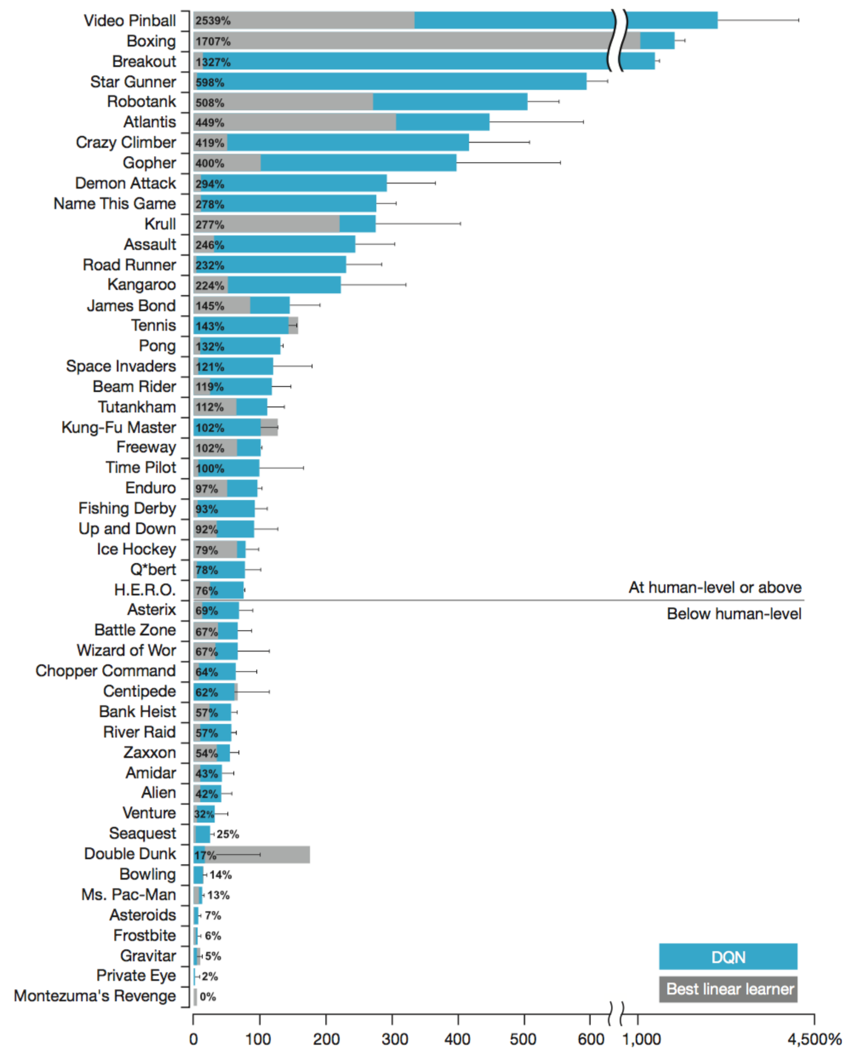
Brain-like computation + Human-level performance

= Human intelligence?



Mastering Atari with deep Q-learning





Mnih et al. (2015)

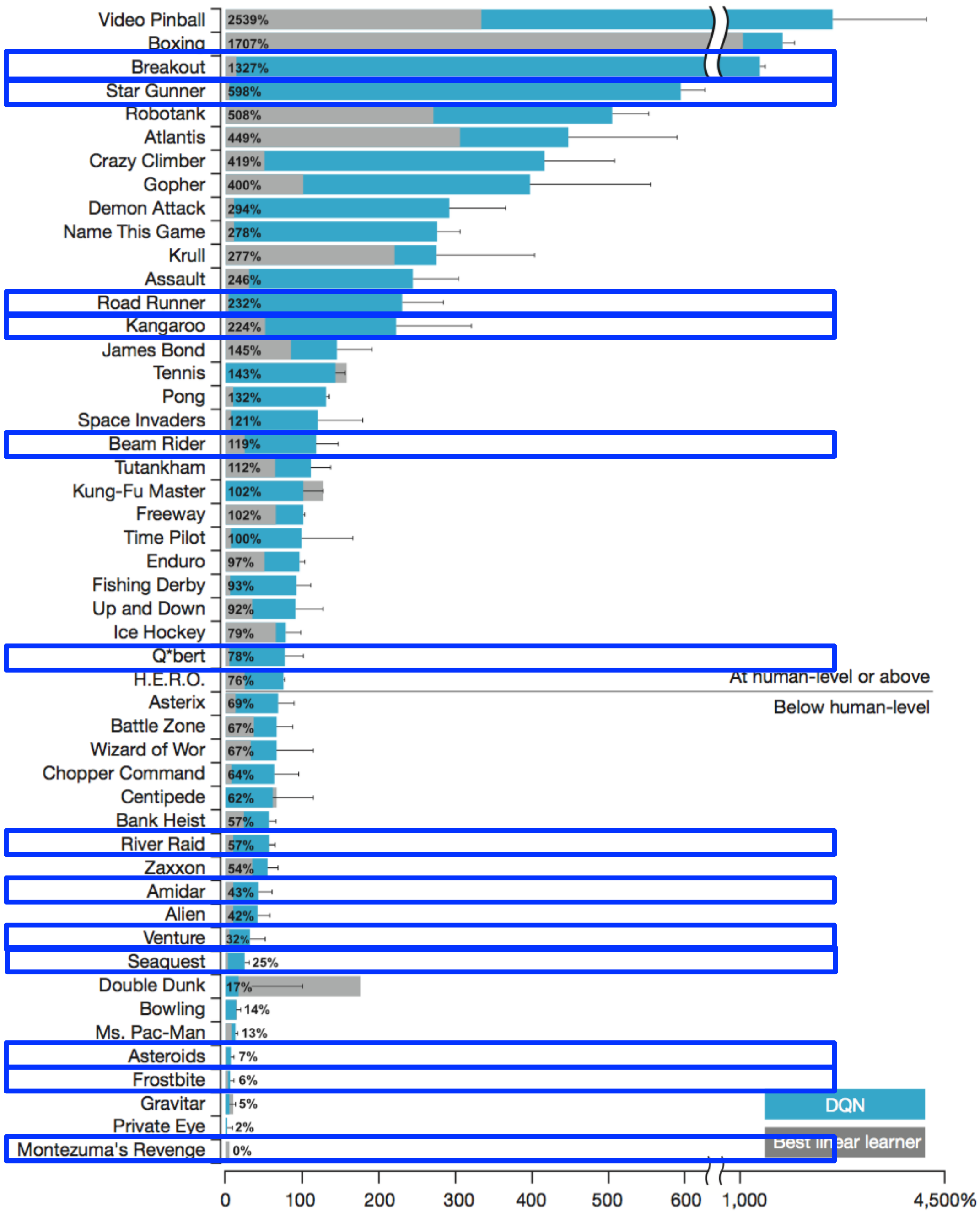
Is this how humans learn?

Is this how humans learn?

Key properties of human intelligence:

1. Rapid learning from few examples.
2. Flexible generalization.

These properties are not yet fully captured by deep learning systems.



Breakout
Star Gunner

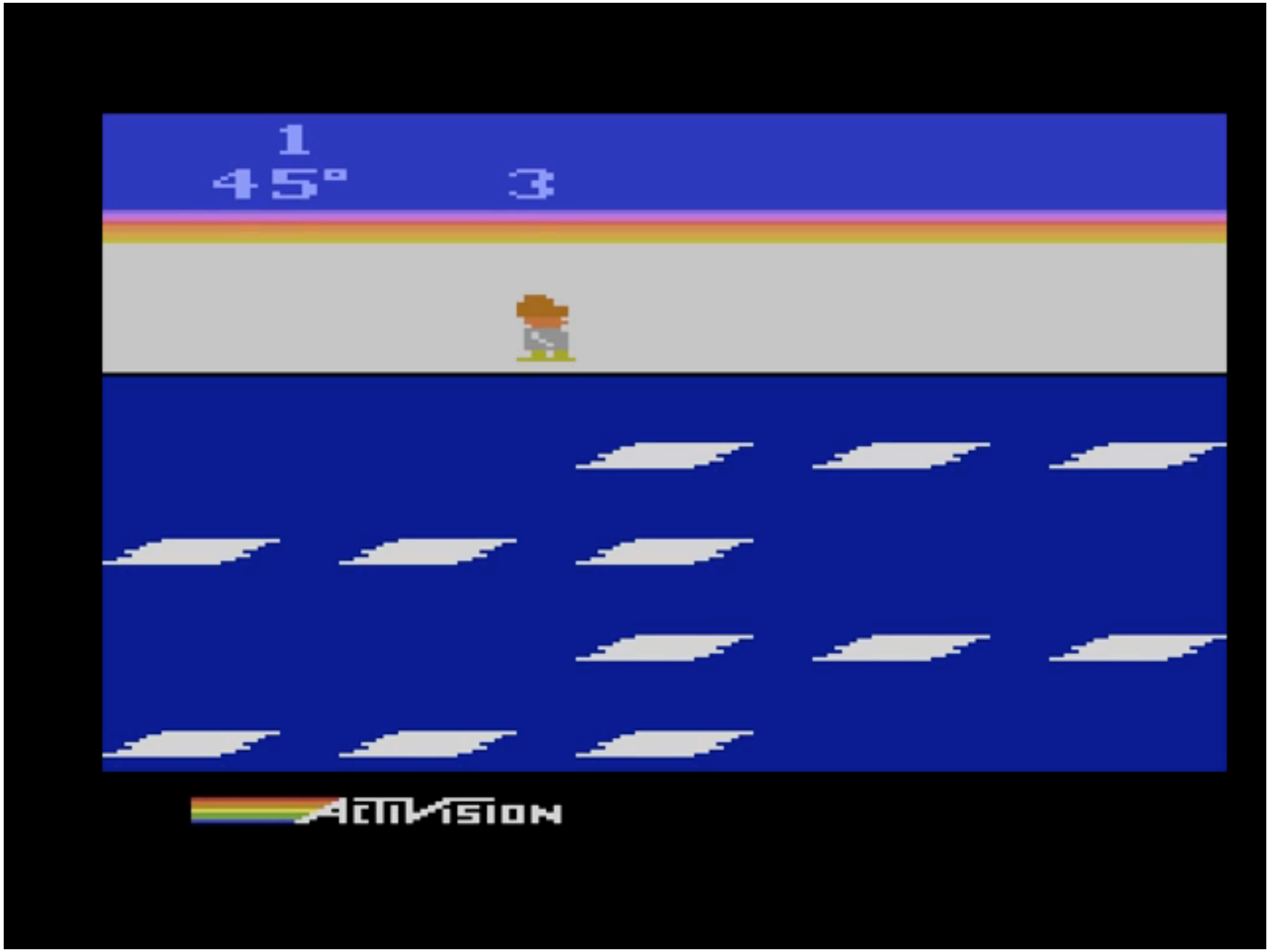
Road Runner
Kangaroo

Beam Rider

Q*bert

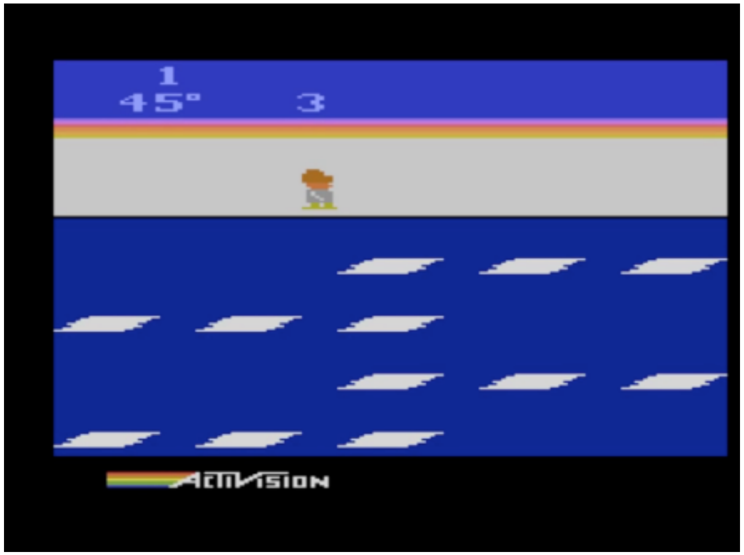
River Raid
Amidar
Venture
Seaquest
Asteroids
Frostbite
Montezuma's Revenge

The “Frostbite challenge”



See Lake, Ullman, Tenenbaum & Gershman (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences.*

The “Frostbite challenge”

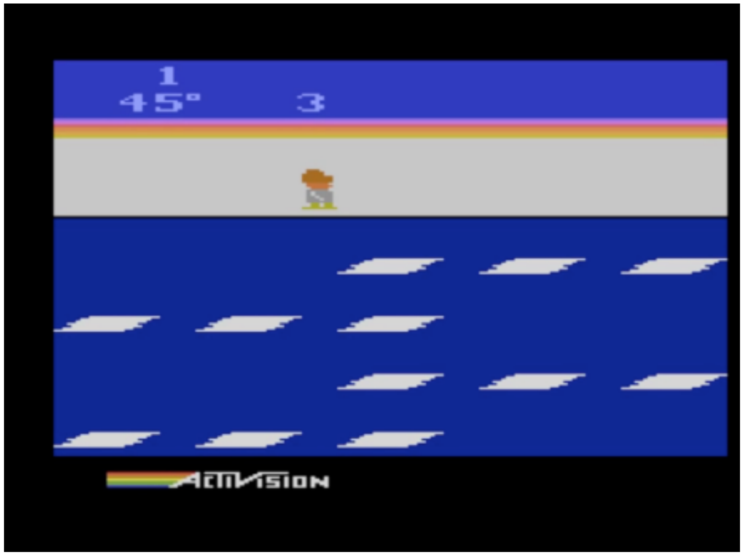


Stage 1:
Reaching basic human-level performance.

Stage 2:
Can we reach human-level performance as *quickly* as people do?

Stage 3:
Can we perform new tasks or goals with little or no retraining?

The “Frostbite challenge”

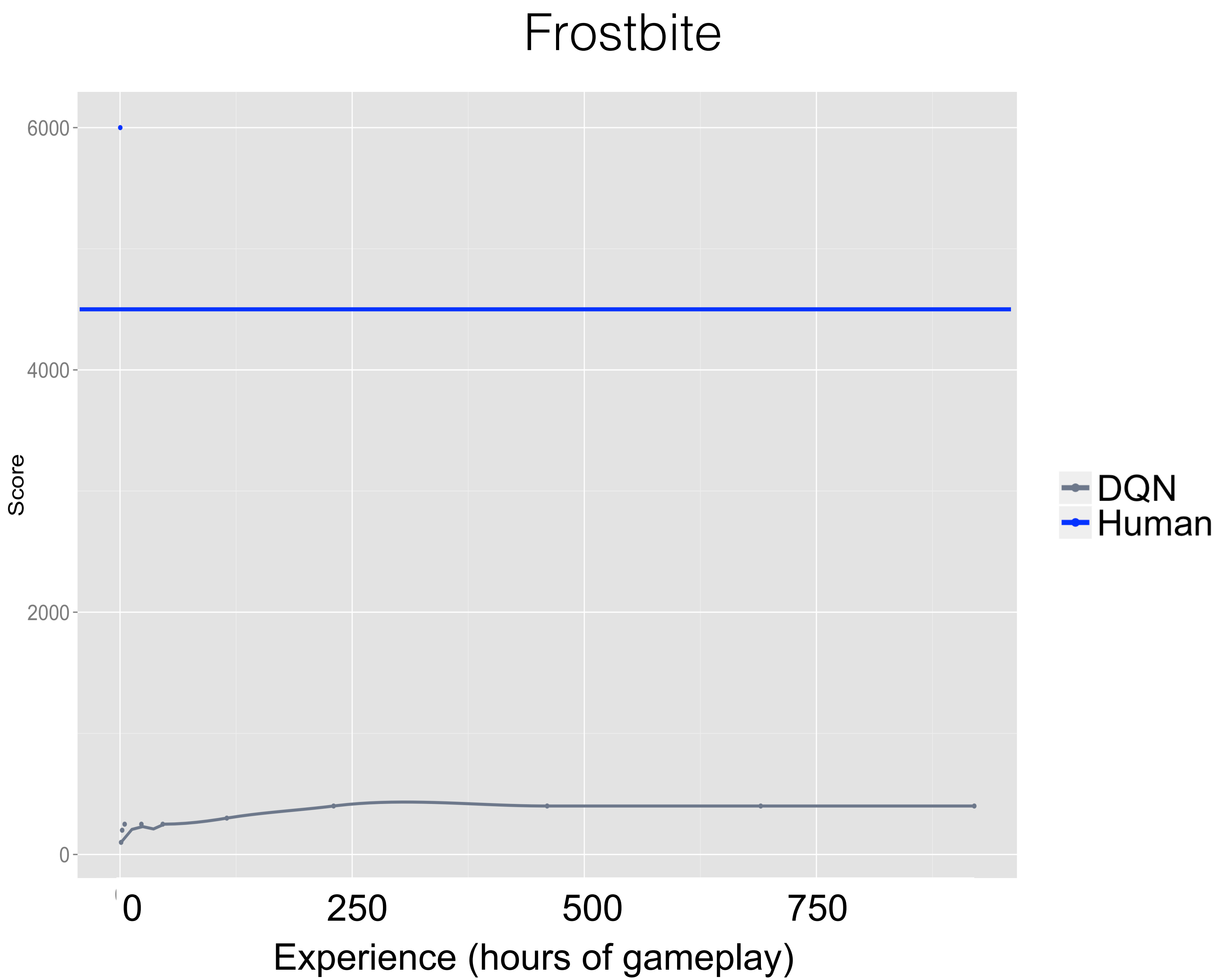


Stage 1:
Reaching basic human-level performance.

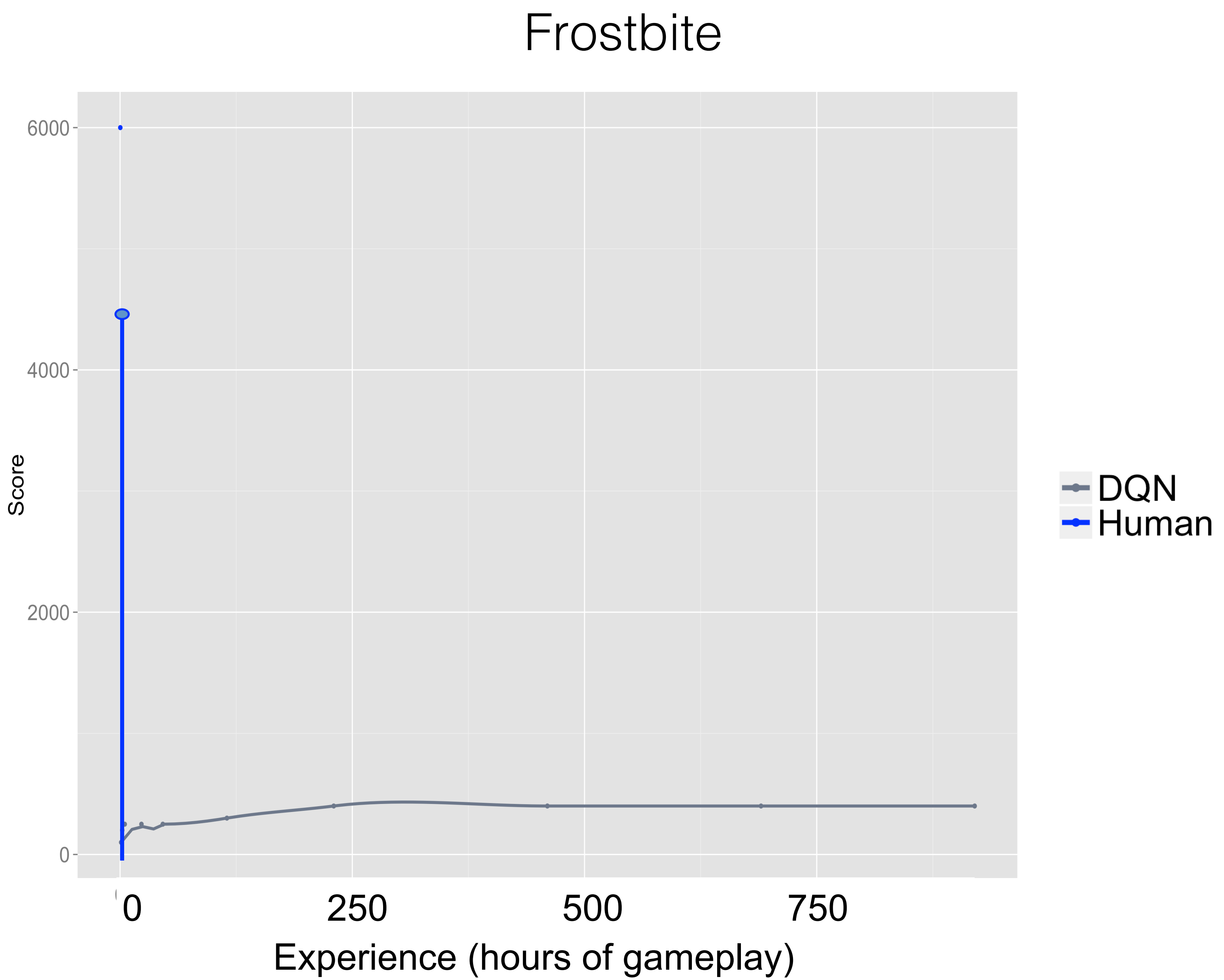
Stage 2:
Can we reach human-level performance as *quickly* as people do?

Stage 3:
Can we perform new tasks or goals with little or no retraining?

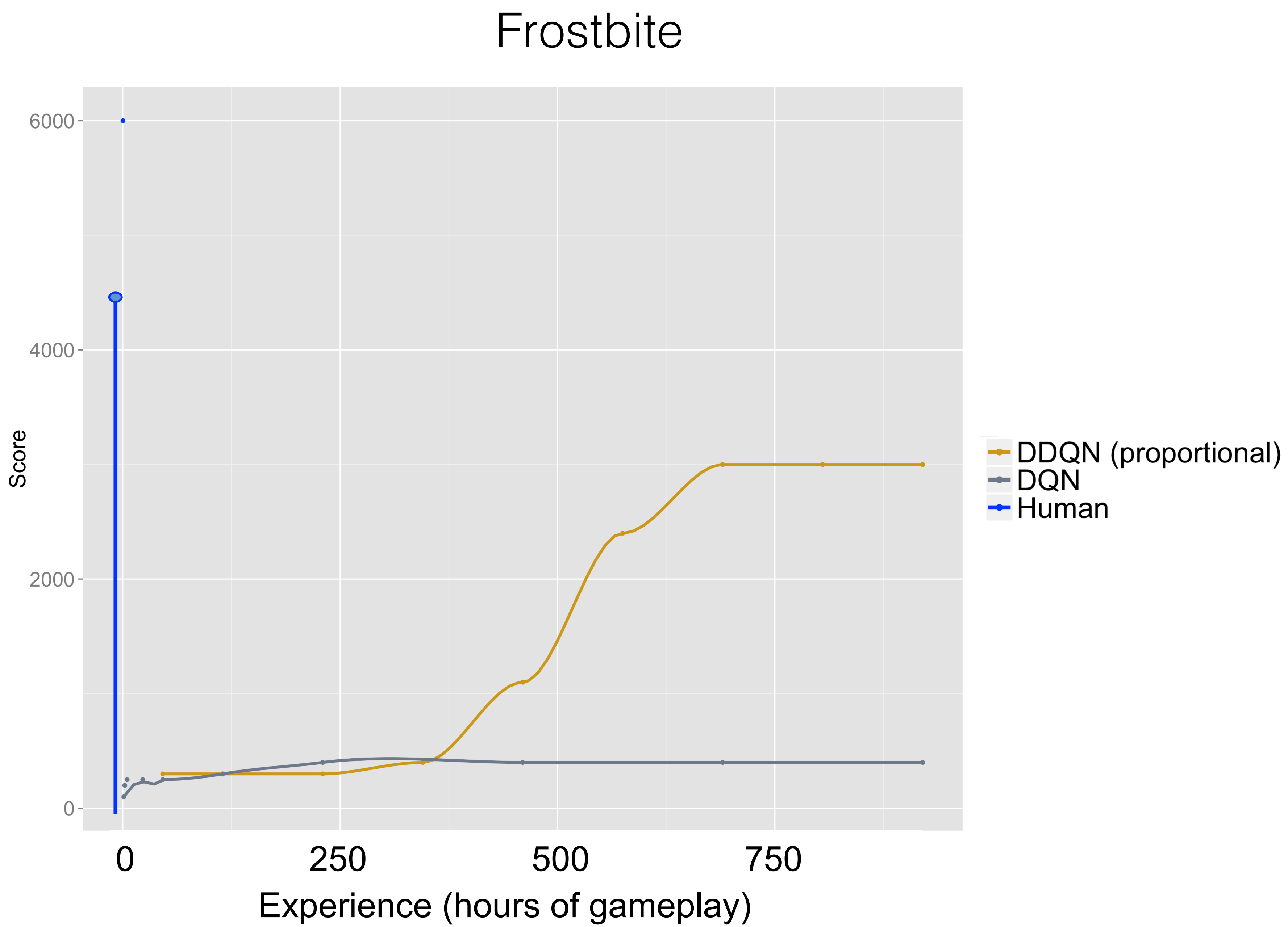
INNOVATIONS IN
The microbiome
nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
LEARNING CURVE
Self-taught AI software attains human-level performance in video games
PAGES 486 & 529
EPIDEMIOLOGY
SHARE DATA IN OUTBREAKS
Forge open access to sequences and more
PAGE 477
COSMOLOGY
A GIANT IN THE EARLY UNIVERSE
A supermassive black hole at a redshift of 6.3
PAGES 490 & 512
QUANTUM PHYSICS
TELEPORTATION FOR TWO
Transferring two properties of a single photon
PAGES 491 & 516
NATURE.COM/NATURE
26 February 2015 £10
Vol. 518, No. 7540

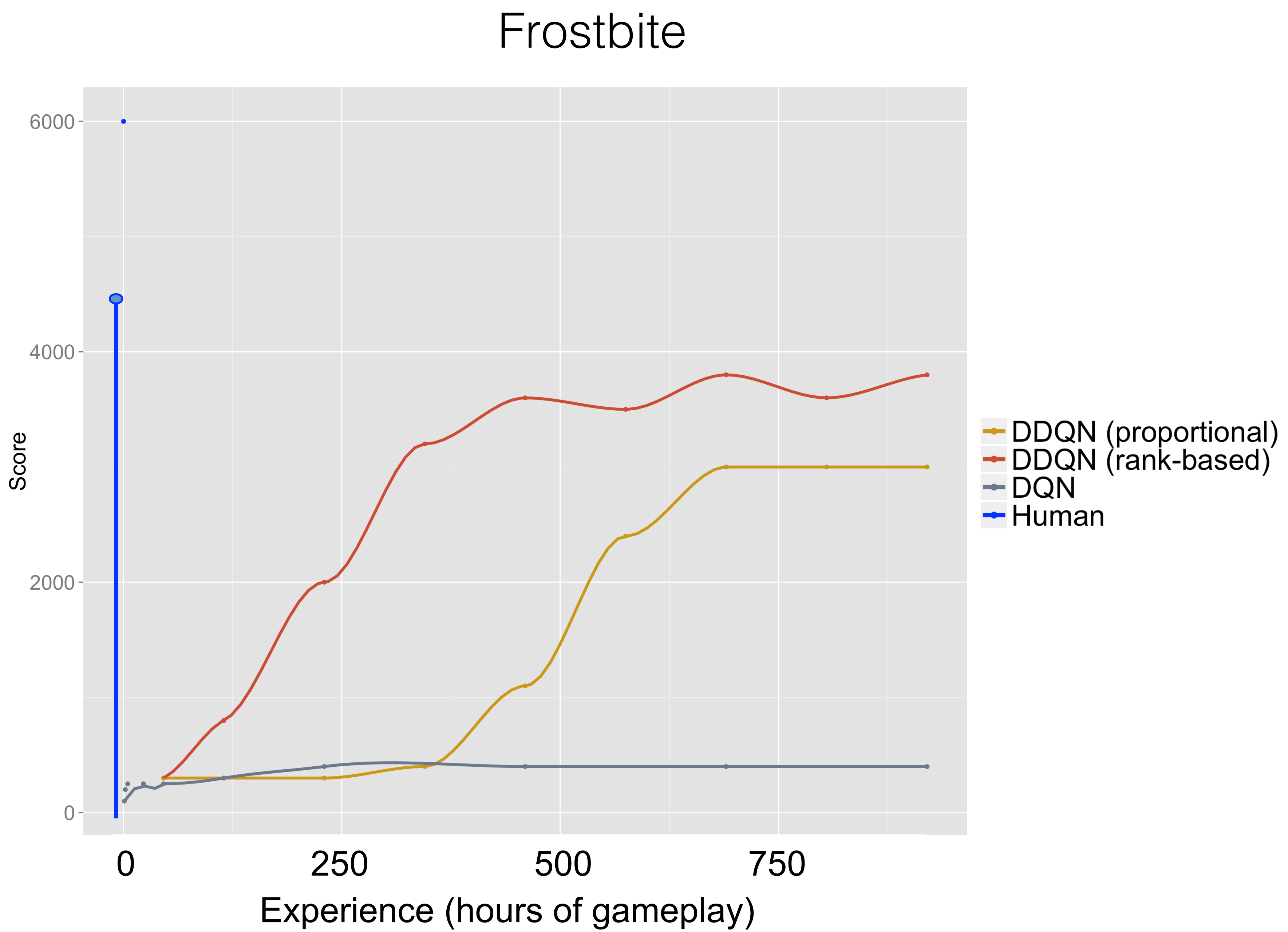


Frostbite
6000
4000
2000
0
Score
0
250
500
750
Experience (hours of gameplay)
DQN
Human

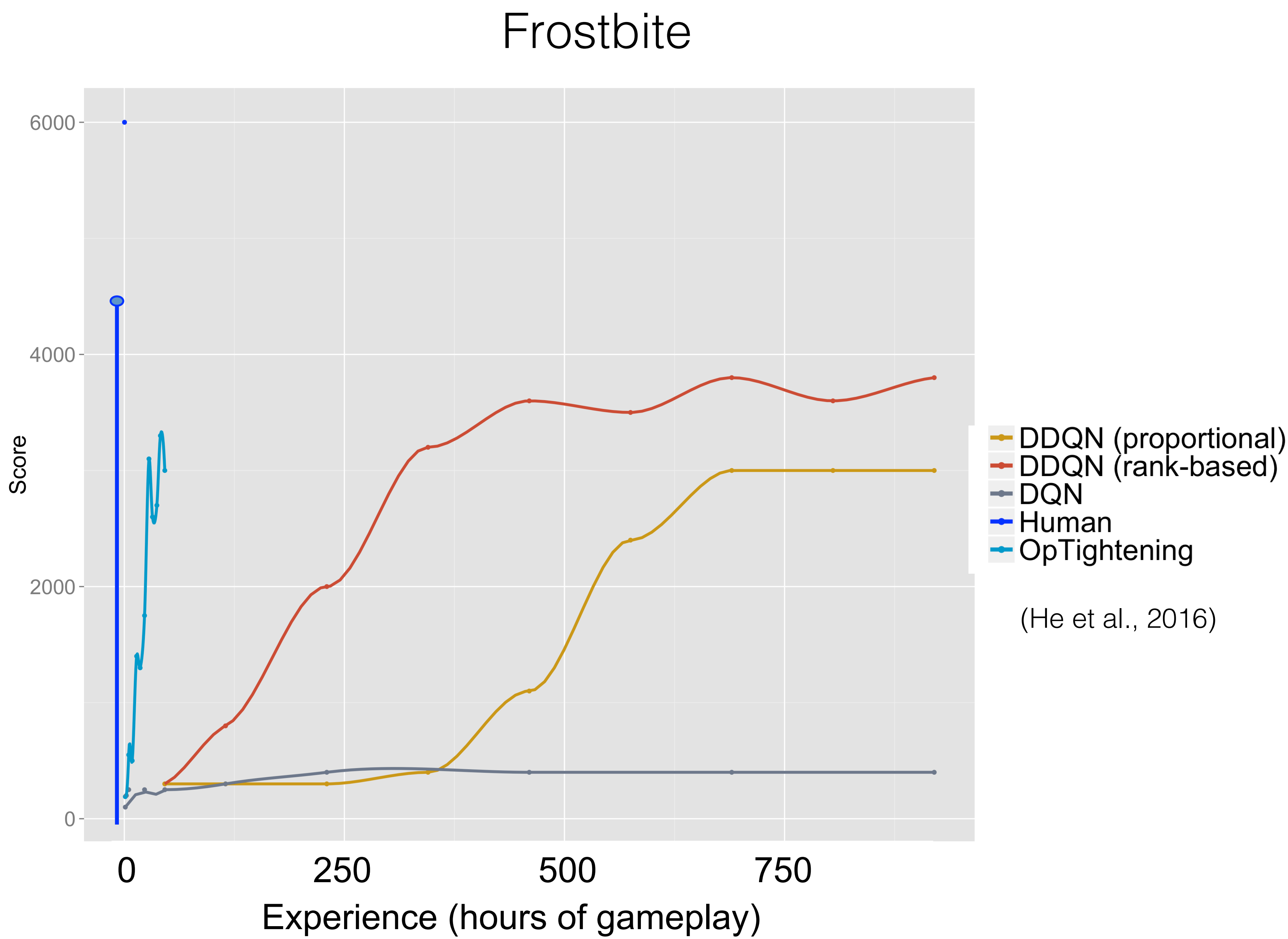


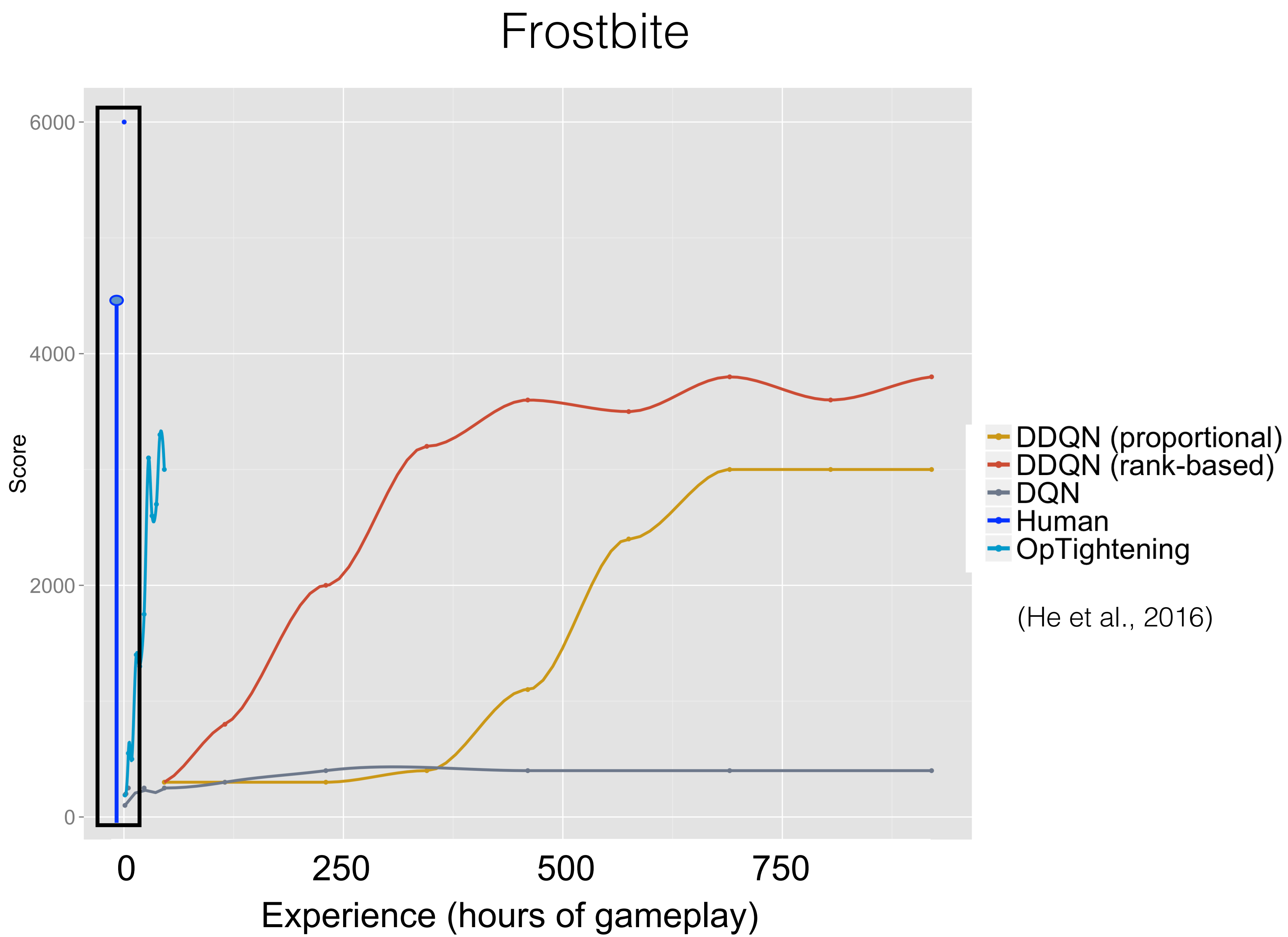
Frostbite
6000
4000
2000
0
Score
0
250
500
750
Experience (hours of gameplay)
DQN
Human

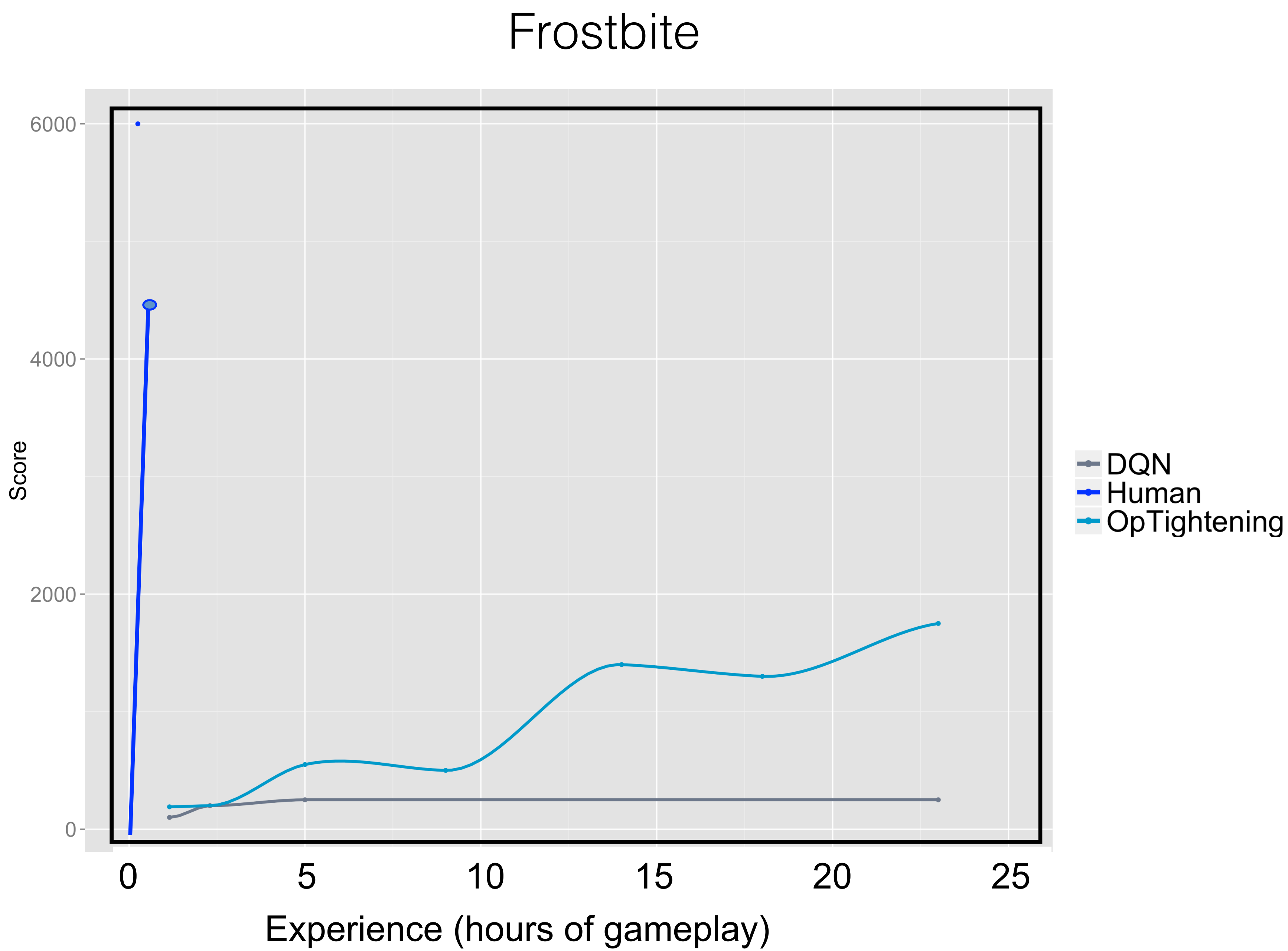

Frostbite
Score
6000
4000
2000
0
0
250
500
750
Experience (hours of gameplay)
DDQN (proportional)
DQN
Human



Frostbite
Score
6000
4000
2000
0
0
250
500
750
Experience (hours of gameplay)
DDQN (proportional)
DDQN (rank-based)
DQN
Human


Frostbite
Score
6000
4000
2000
0
0
250
500
750
Experience (hours of gameplay)
DDQN (proportional)
DDQN (rank-based)
DQN
Human
OpTightening
(He et al., 2016)

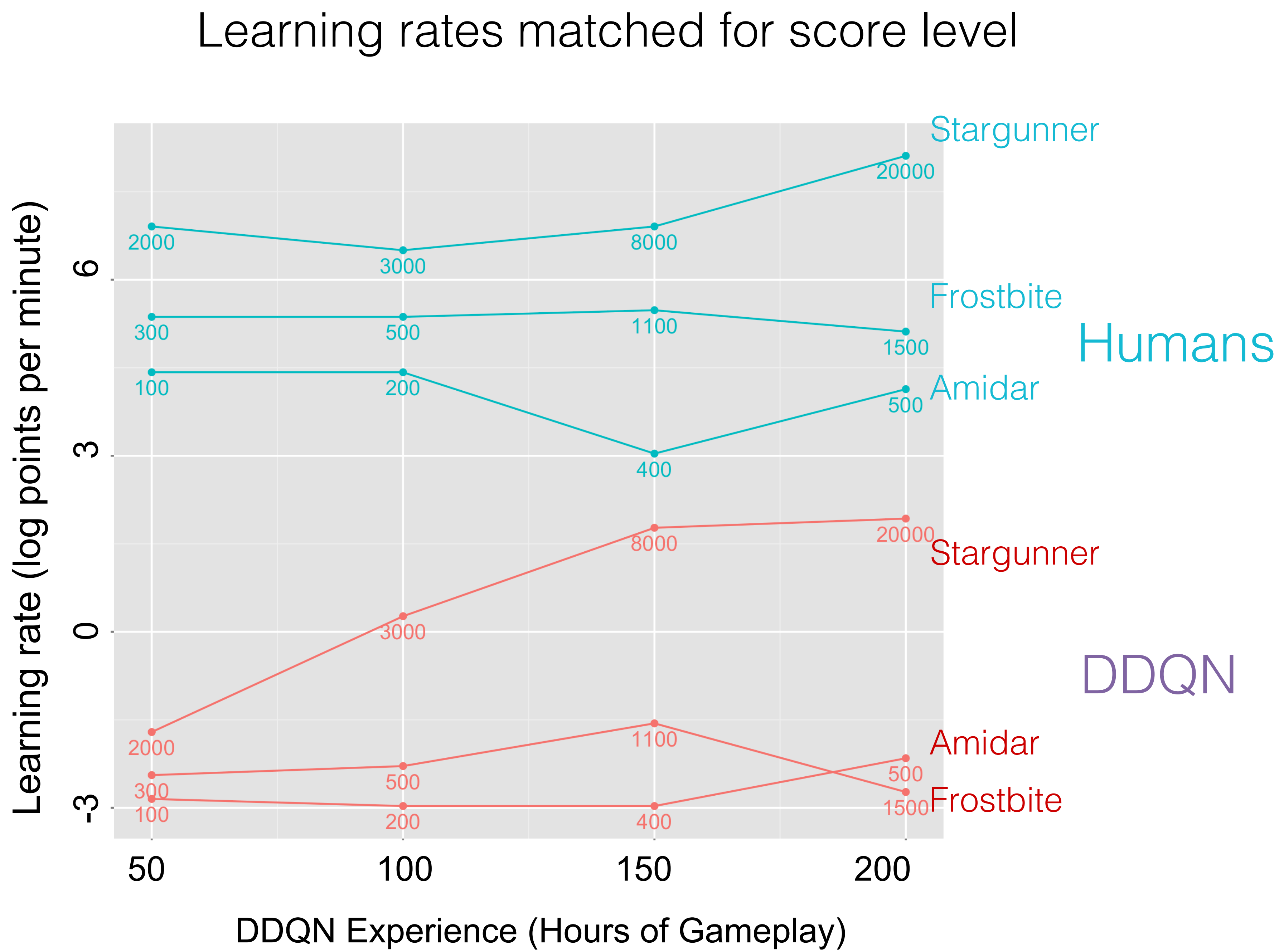

Frostbite
Score
6000
4000
2000
0
0
250
500
750
Experience (hours of gameplay)
DDQN (proportional)
DDQN (rank-based)
DQN
Human
OpTightening
(He et al., 2016)



Frostbite
6000
4000
2000
0
Score
0
5
10
15
20
25
Experience (hours of gameplay)
DQN
Human
OpTightening

Unfair comparison

- Deep neural networks (at least in the way they’re typically trained) must learn their entire visual system from scratch.
- Humans have their entire childhoods plus hundreds of thousands of years of evolution.
- Maybe deep neural networks learn like humans, but their learning curve is just shifted.

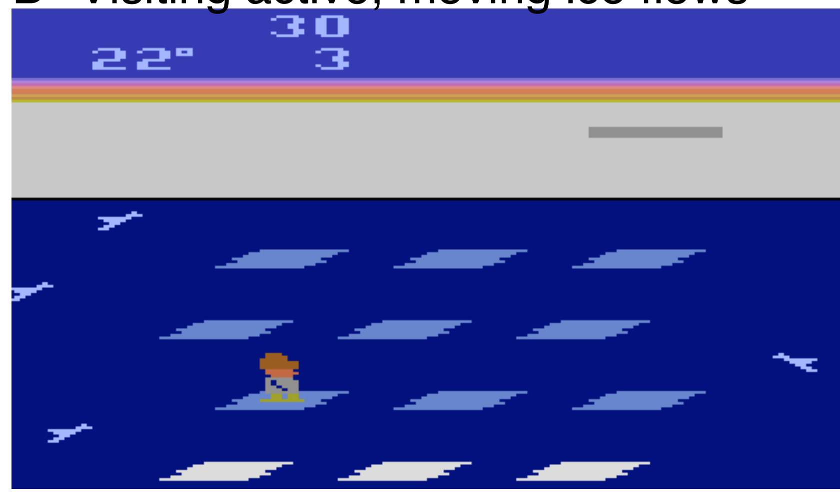


Learning rates matched for score level
Stargunner
Frostbite
Humans
Amidar
Stargunner
DDQN
Amidar
Frostbite
2000
3000
8000
20000
300
500
1100
1500
100
200
400
500
8000
20000
3000
2000
1100
500
300
500
1500
100
200
400
6
3
0
-3
50
100
150
200
Learning rate (log points per minute)
DDQN Experience (Hours of Gameplay)

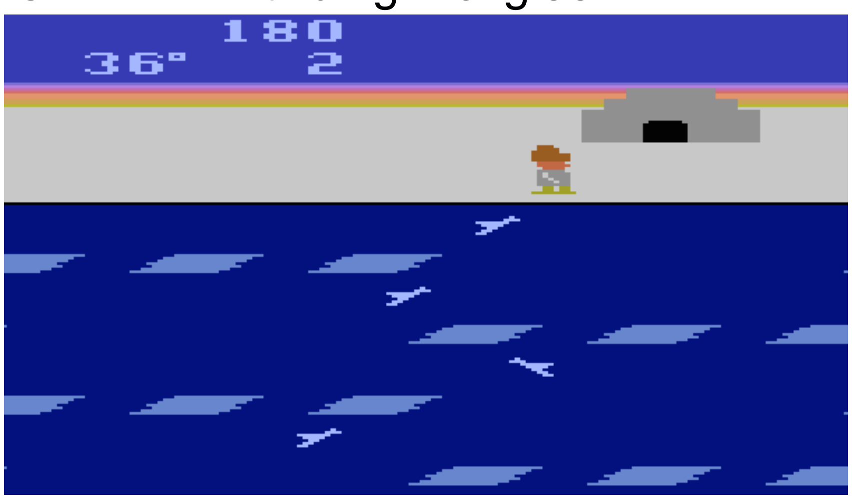
From the very beginning of play, people see objects, agents, physics. Actively explore possible object-relational goals, and soon come to multistep plans that exploit what they have learned.

A How to play Frostbite: Initial setup

B Visiting active, moving ice flows



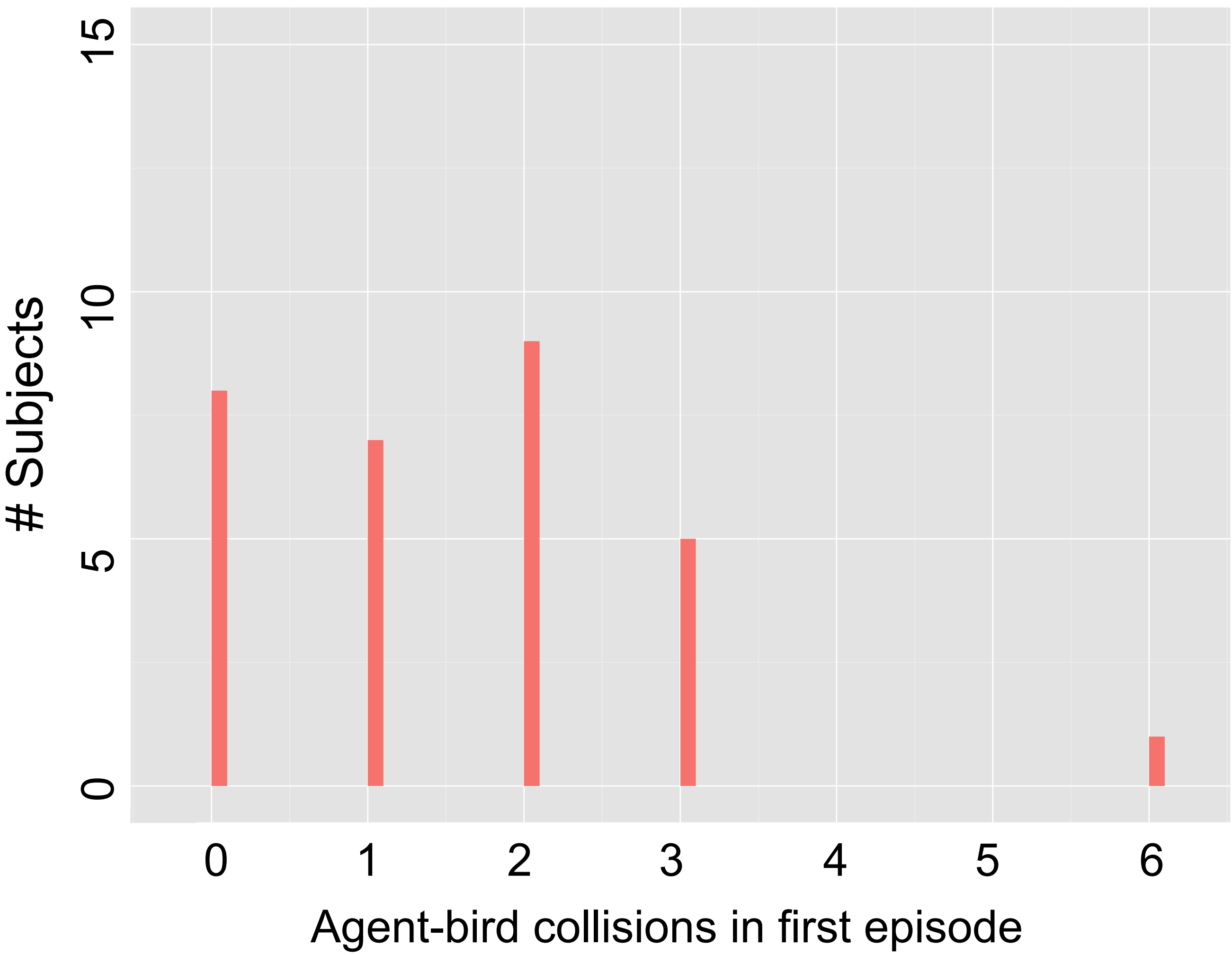
C Building the igloo



D Obstacles on later levels

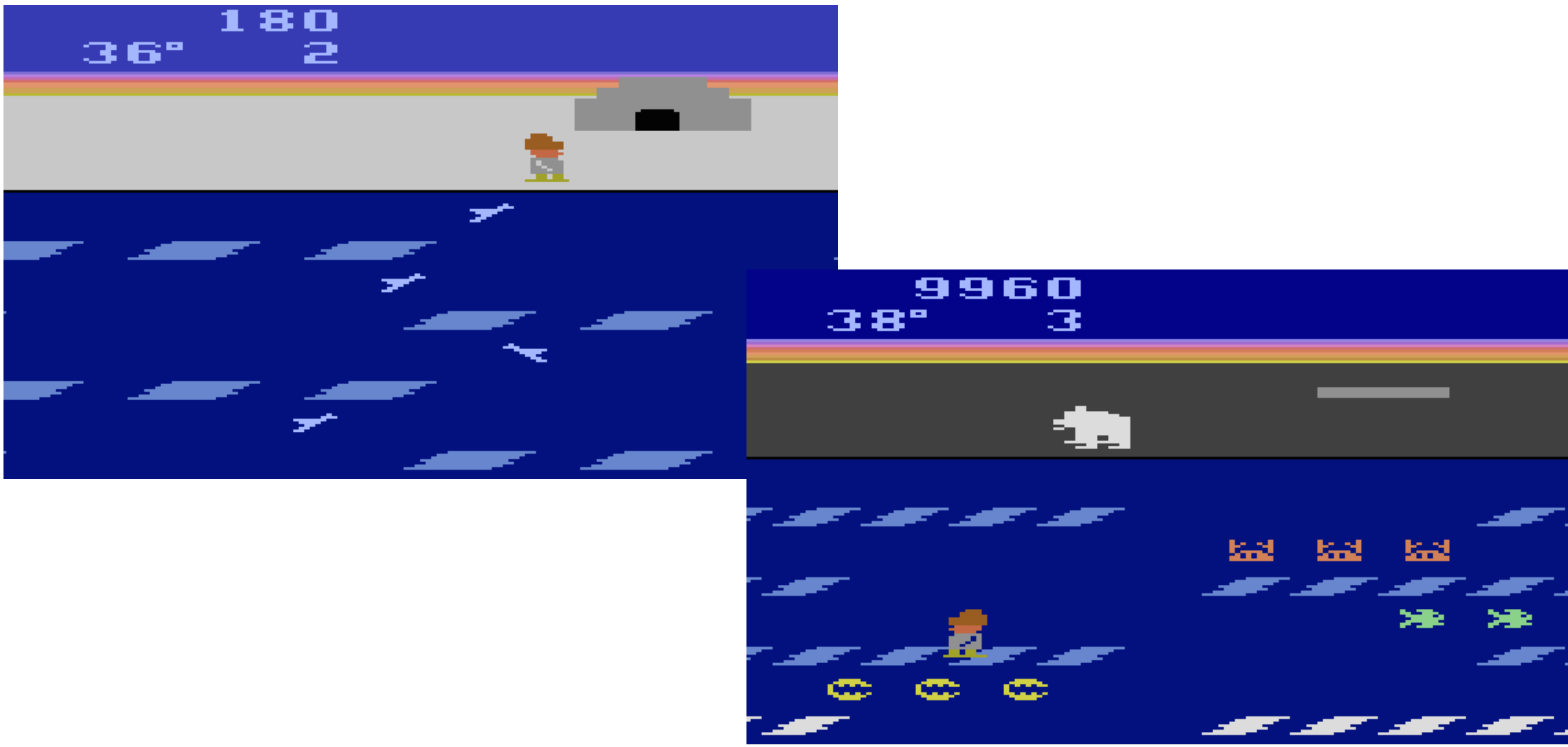
What drives such rapid learning?

One-shot (or few-shot) learning about harmful actions and outcomes:



What drives such rapid learning?

To what extent is rapid learning dependent on prior knowledge about real-world objects, actions, and consequences?

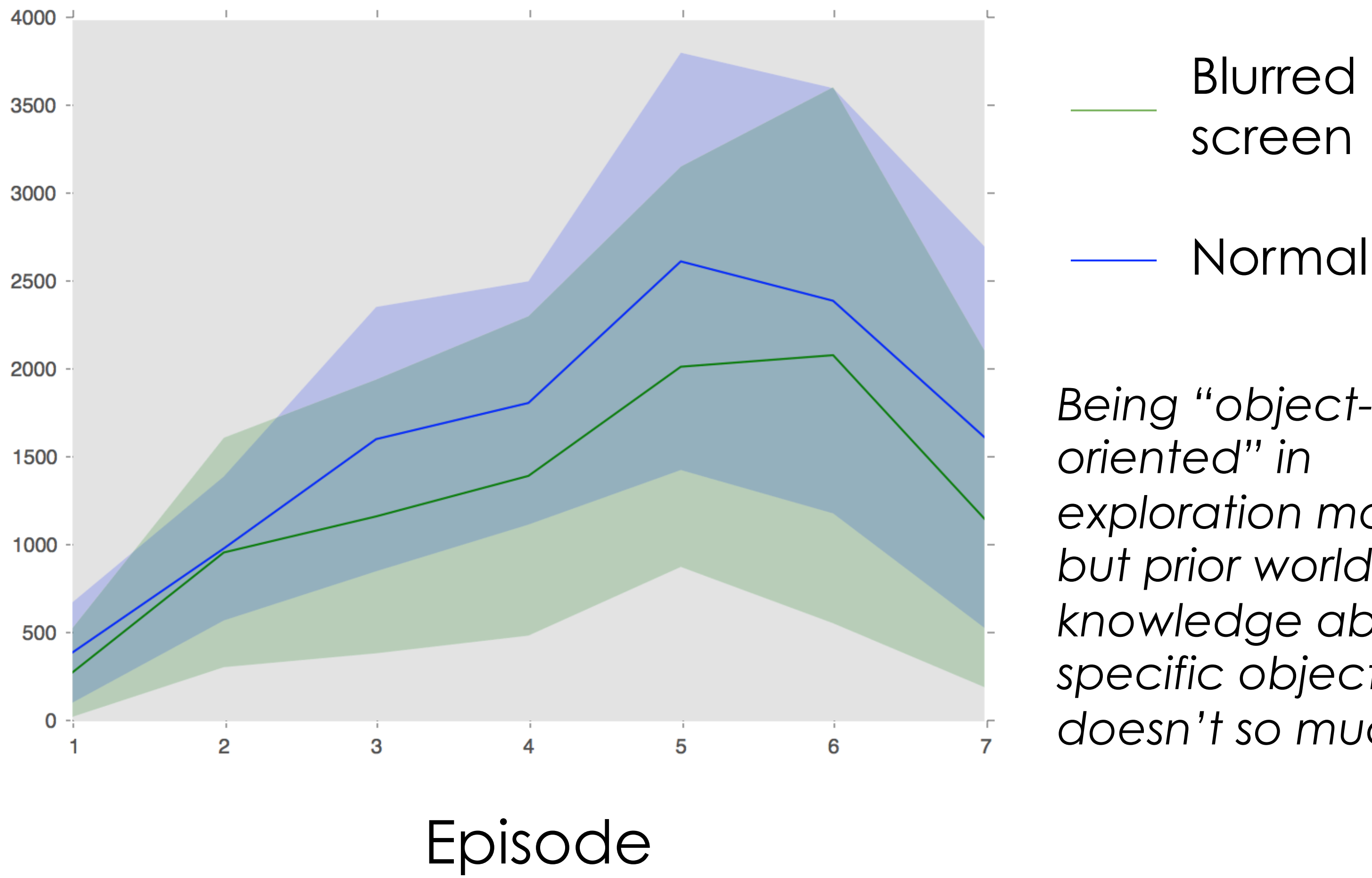


What drives such rapid learning?

To what extent is rapid learning dependent on prior knowledge about real-world objects, actions, and consequences?

What drives such rapid learning?

To what extent is rapid learning dependent on prior knowledge about real-world objects, actions, and consequences?



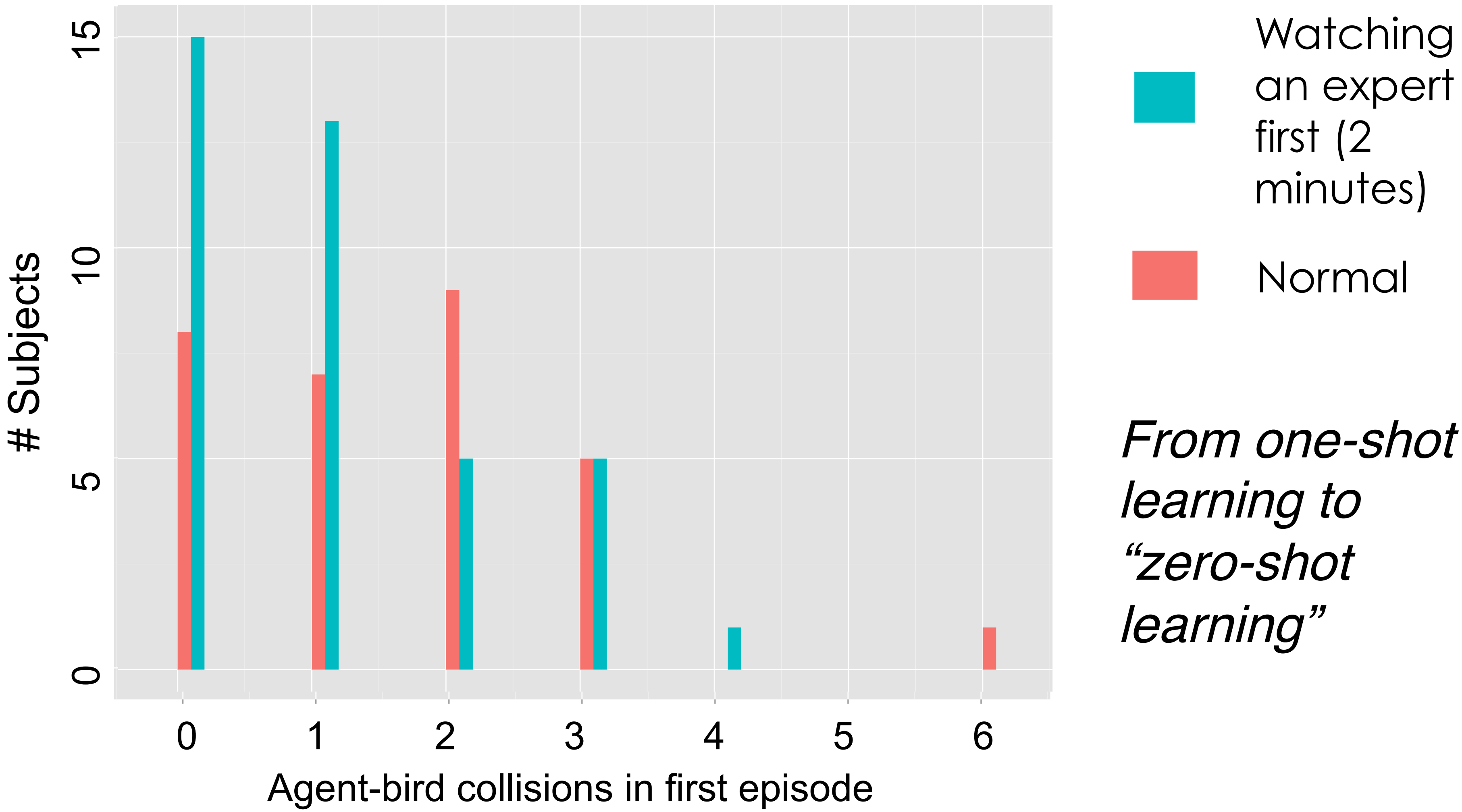
Being "object-oriented" in exploration matters, but prior world knowledge about specific object types doesn't so much!

What drives such rapid learning?

People can learn even faster if they combine their own experience with just a little help from others:

What drives such rapid learning?

People can learn even faster if they combine their own experience with just a little help from others:



From one-shot learning to "zero-shot learning"

FROSTBITE BASICS

The object of the game is to help Frostbite Bailey build igloos by jumping on floating blocks of ice. Be careful to avoid these deadly hazards: killer clams, snow geese, Alaskan king crab, grizzly polar bears and the rapidly dropping temperature.

To move Frostbite Bailey up, down, left or right, use the arrow keys. To reverse the direction of the ice floe you are standing on, press the spacebar. But remember, each time you do, your igloo will lose a block, unless it is completely built.

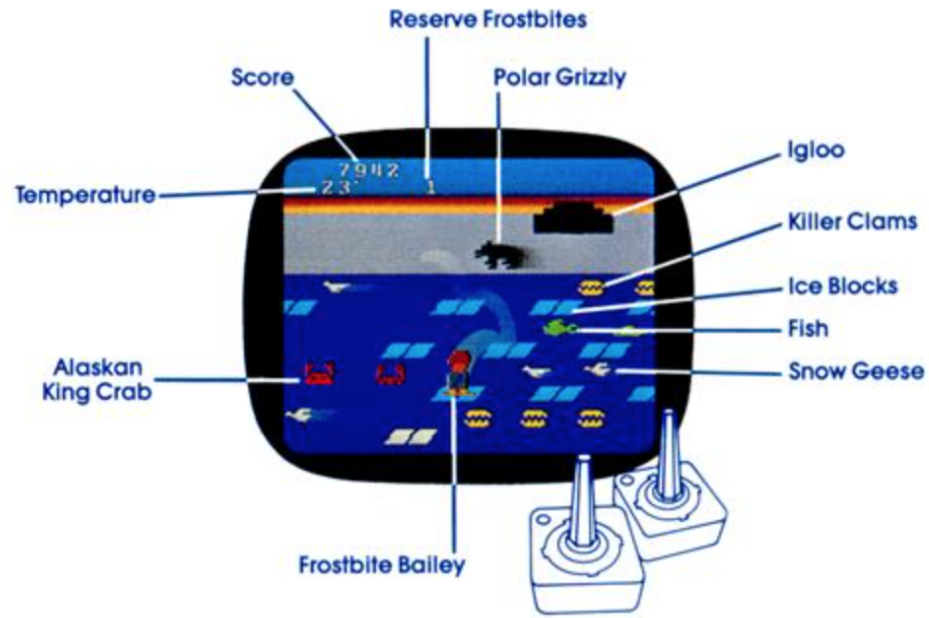
You begin the game with one active Frostbite Bailey and three on reserve. With each increase of 5,000 points, a bonus Frostbite is added to your reserves (up to a maximum of nine).

Frostbite gets lost each time he falls into the Arctic Sea, gets chased away by a Polar Grizzly or gets caught outside when the temperature drops to zero.

The game ends when your reserves have been exhausted and Frostbite is 'retired' from the construction business.

IGLOO CONSTRUCTION

Building codes. Each time Frostbite Bailey jumps onto a white ice floe, a "block" is added to the igloo. Once jumped upon, the white ice turns blue. It can still be jumped on, but won't add points to your score or blocks to your igloo. When all four rows are blue, they will turn white again. The igloo is complete when a door appears. Frostbite may then jump into it.

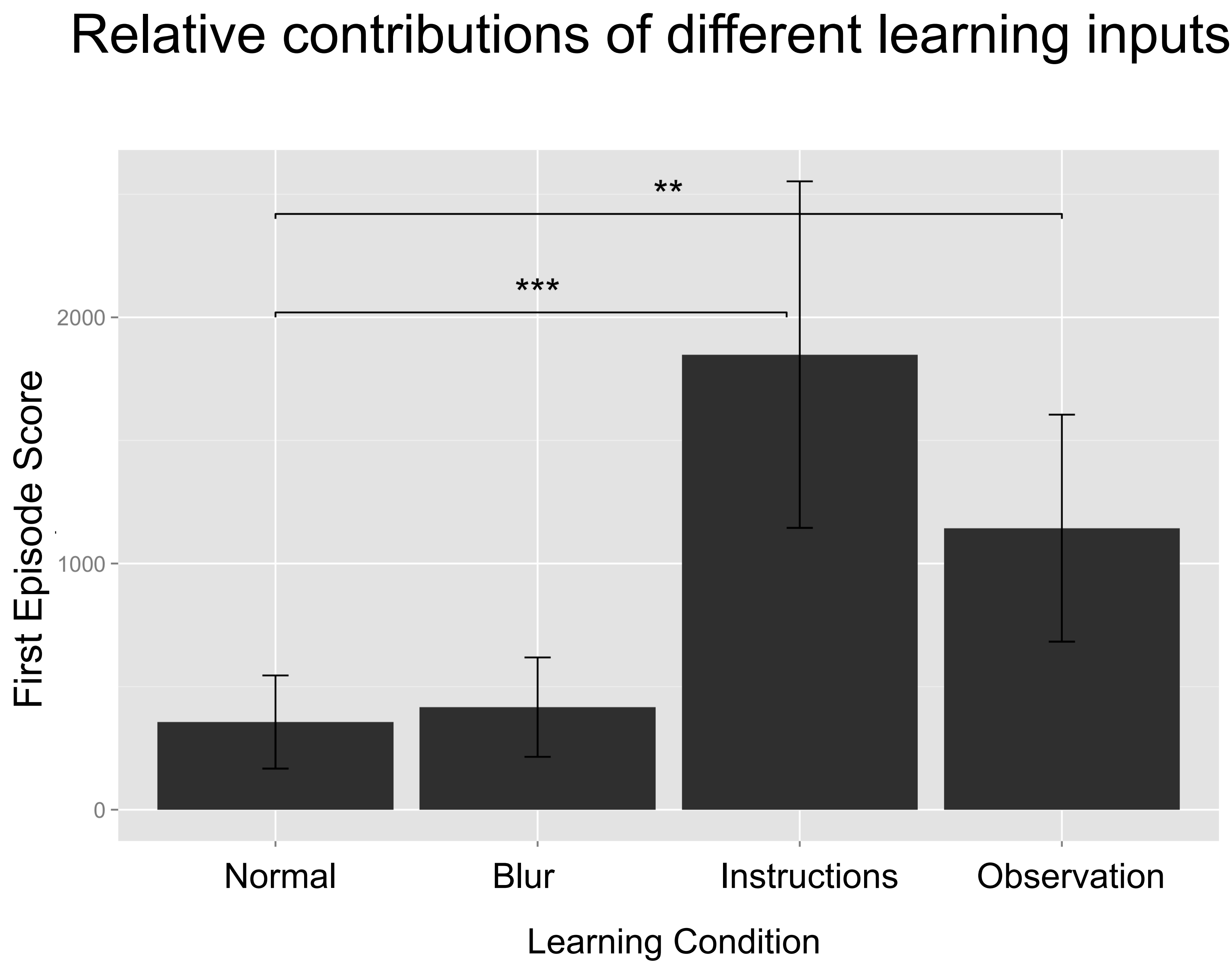


Work hazards. Avoid contact with Alaskan King Crabs, snow geese, and killer clams, as they will push Frostbite Bailey into the fatal Arctic Sea. The Polar Grizzlies come out of hibernation at level 4 and, upon contact, will chase Frostbite right off-screen.

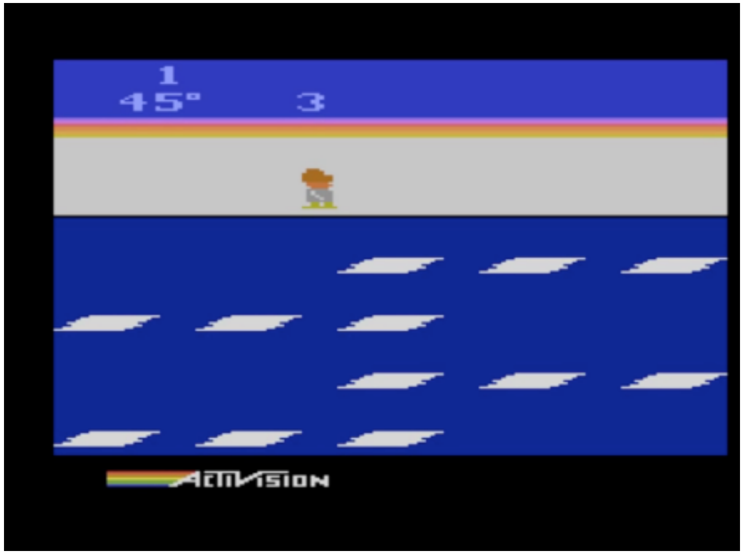
No Overtime Allowed. Frostbite always starts working when it's 45 degrees outside. You'll notice this steadily falling temperature at the upper left corner of the screen. Frostbite must build and enter the igloo before the temperature drops to 0 degrees, or else he'll turn into blue ice!

SPECIAL FEATURES OF FROSTBITE

Fresh Fish swim by regularly. They are Frostbite Bailey's only food and, as such, are also additives to your score. Catch' em if you can.


Relative contributions of different learning inputs
**
***
2000
1000
0
First Episode Score
Normal
Blur
Instructions
Observation
Learning Condition

The “Frostbite challenge”



Stage 1:
Reaching basic human-level performance.

Stage 2:
Can we reach human-level performance as *quickly* as people do?

Stage 3:
Can we perform new tasks or goals with little or no retraining?

The flexibility of human goals

In Frostbite, or any video game, people can easily reuse their learned models for an endless range of new tasks and goals:

The flexibility of human goals

In Frostbite, or any video game, people can easily reuse their learned models for an endless range of new tasks and goals:

- Get the lowest possible score.

The flexibility of human goals

In Frostbite, or any video game, people can easily reuse their learned models for an endless range of new tasks and goals:

- Get the lowest possible score.
- Get closest to 100, or 300, or 1000, or 3000, or any level, without going over.

The flexibility of human goals

In Frostbite, or any video game, people can easily reuse their learned models for an endless range of new tasks and goals:

- Get the lowest possible score.
- Get closest to 100, or 300, or 1000, or 3000, or any level, without going over.
- Beat your friend, who's playing next to you, but just barely, not by too much, so as not to embarrass them.

The flexibility of human goals

In Frostbite, or any video game, people can easily reuse their learned models for an endless range of new tasks and goals:

- Get the lowest possible score.
- Get closest to 100, or 300, or 1000, or 3000, or any level, without going over.
- Beat your friend, who’s playing next to you, but just barely, not by too much, so as not to embarrass them.
- Go as long as you can without dying.

The flexibility of human goals

In Frostbite, or any video game, people can easily reuse their learned models for an endless range of new tasks and goals:

- Get the lowest possible score.
- Get closest to 100, or 300, or 1000, or 3000, or any level, without going over.
- Beat your friend, who's playing next to you, but just barely, not by too much, so as not to embarrass them.
- Go as long as you can without dying.
- Die as quickly as you can.

The flexibility of human goals

In Frostbite, or any video game, people can easily reuse their learned models for an endless range of new tasks and goals:

- Get the lowest possible score.
- Get closest to 100, or 300, or 1000, or 3000, or any level, without going over.
- Beat your friend, who's playing next to you, but just barely, not by too much, so as not to embarrass them.
- Go as long as you can without dying.
- Die as quickly as you can.
- Pass each level at the last possible minute, right before the temperature timer hits zero and you die (i.e., come as close as you can to dying from frostbite without actually dying).

The flexibility of human goals

In Frostbite, or any video game, people can easily reuse their learned models for an endless range of new tasks and goals:

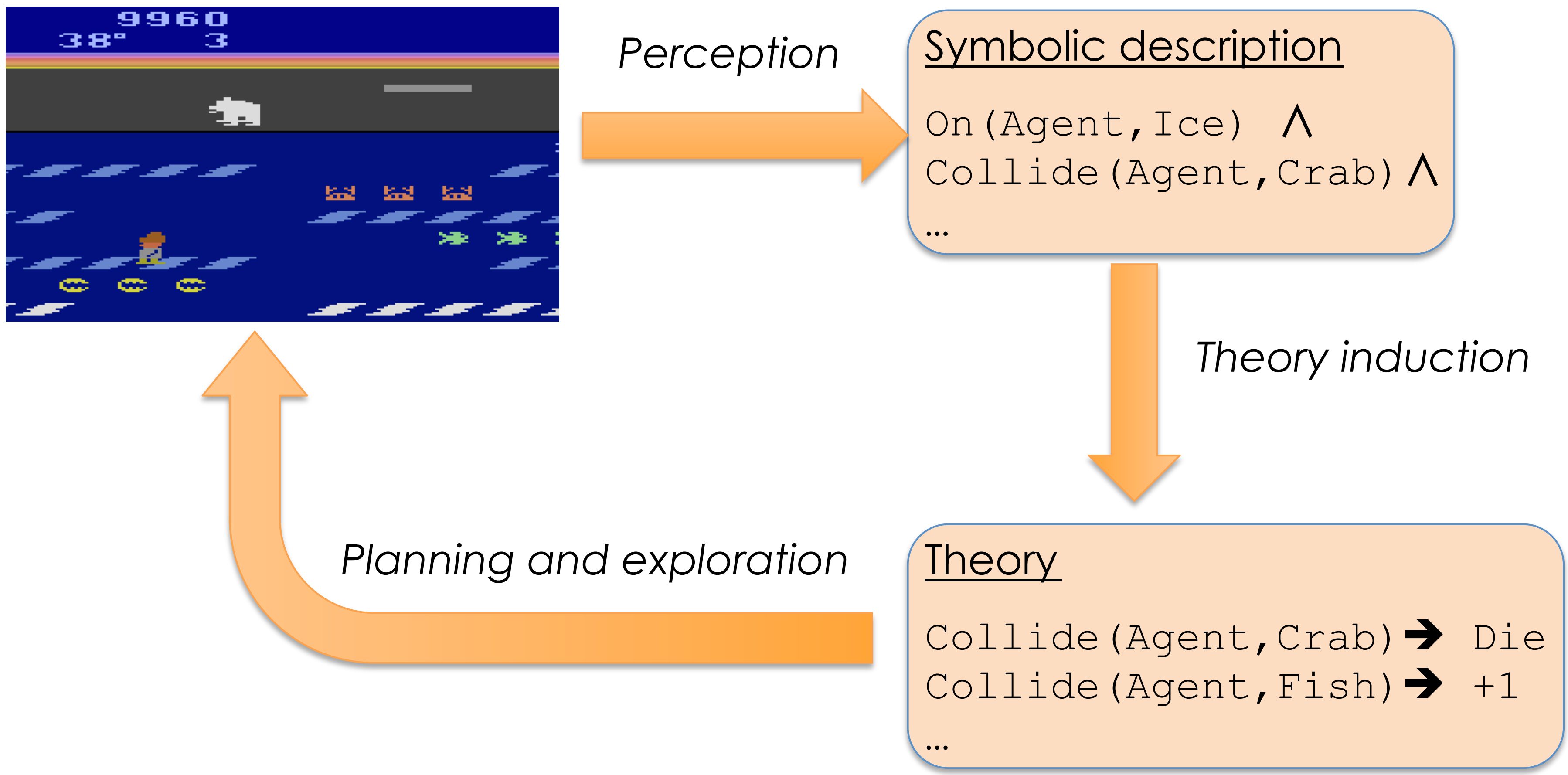
- Get the lowest possible score.
- Get closest to 100, or 300, or 1000, or 3000, or any level, without going over.
- Beat your friend, who's playing next to you, but just barely, not by too much, so as not to embarrass them.
- Go as long as you can without dying.
- Die as quickly as you can.
- Pass each level at the last possible minute, right before the temperature timer hits zero and you die (i.e., come as close as you can to dying from frostbite without actually dying).
- Get to the furthest unexplored level without regard for your score.
- See if you can discover secret Easter eggs.
- Get as many fish as you can.
- Touch all the individual ice floes on screen once and only once.
- Teach your friend how to play as efficiently as possible.

Towards more human-like RL agents

Humans don’t just do pattern recognition and function approximation; they learn “theories” of the game.

These theories support rapid learning, efficient planning, and flexible generalization.

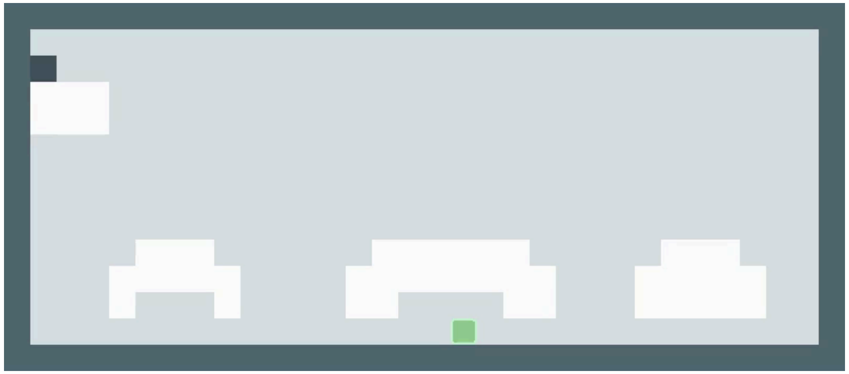
Explore, Model, Plan agent (EMPA)



Video game description language (VGDL)

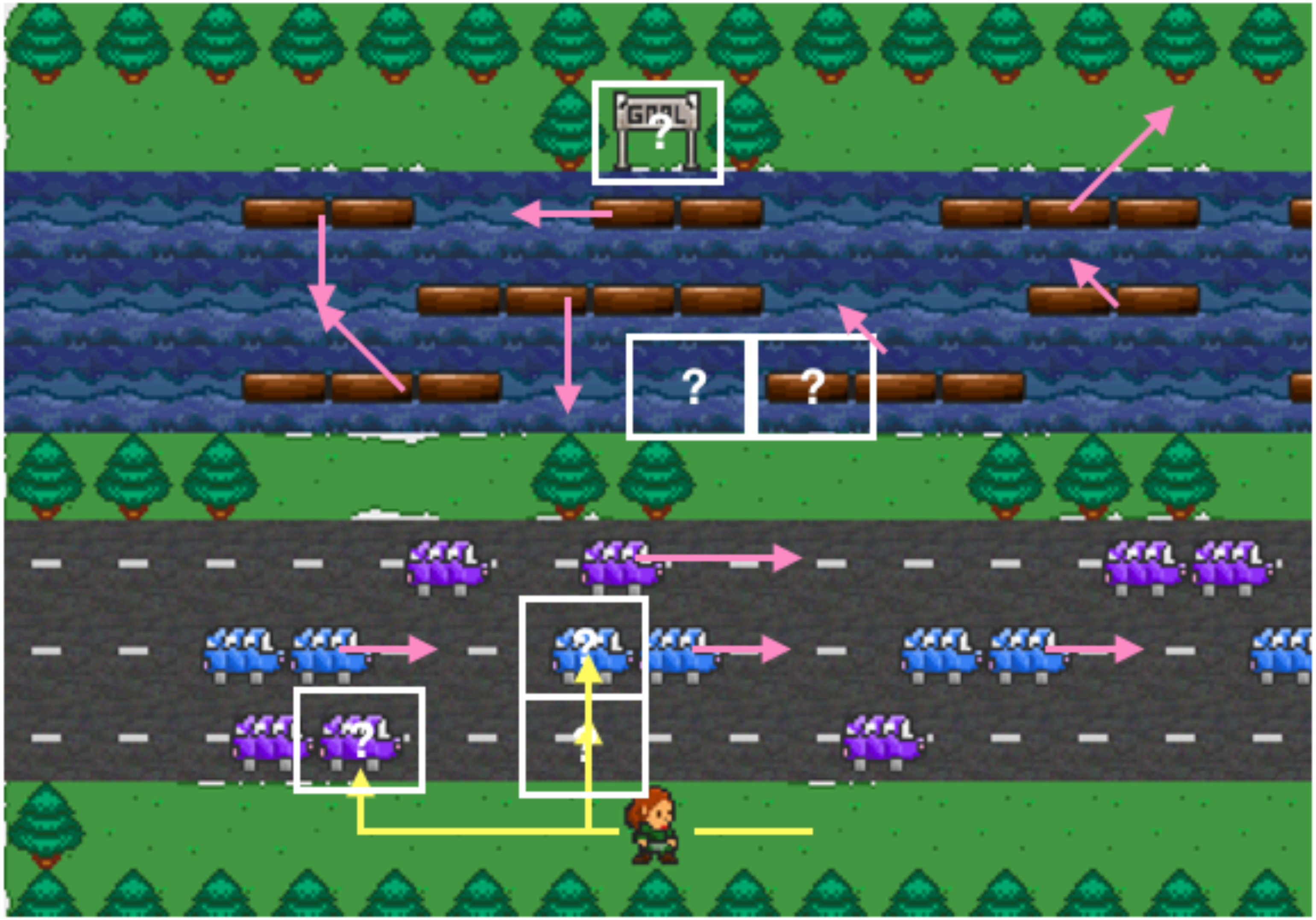
```
BasicGame
    SpriteSet
        base    > Immovable    color=WHITE
        avatar  > FlakAvatar stype=sam
        missile > Missile
            sam  > orientation=UP    color=BLUE singleton=True
            bomb > orientation=DOWN  color=RED  speed=0.5
        alien   > Bomber       stype=bomb   prob=0  cooldown=3 speed=0.75
        portal  > SpawnPoint   stype=alien  cooldown=10   total=3

    LevelMapping
        0 > base
        1 > portal

    InteractionSet
        avatar  EOS  > stepBack
        alien   EOS  > turnAround
        missile EOS  > killSprite
        missile base > killSprite
        base missile > killSprite
        base   alien > killSprite
        avatar alien > killSprite
        avatar bomb  > killSprite
        alien  sam   > killSprite

    TerminationSet
        SpriteCounter      stype=avatar               limit=0 win=False
        MultiSpriteCounter stype1=portal stype2=alien limit=0 win=True
```

```
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
w                              w
w1                             w
w000                           w
w000                           w
w                              w
w                              w
w                              w
w                              w
w    000      000000     000   w
w   00000    00000000   00000  w
w   0   0    00    00   00000  w
w                A             w
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
```



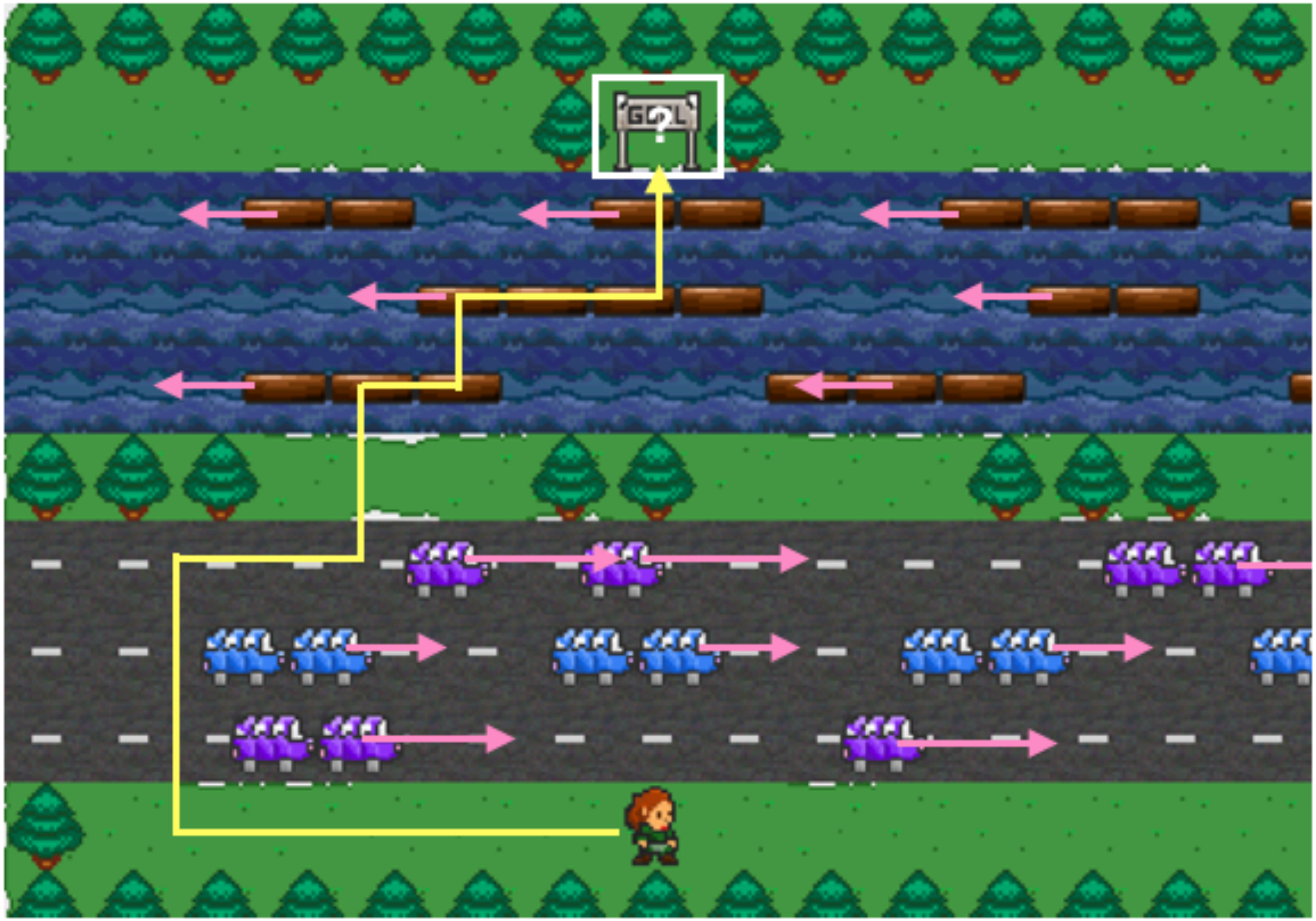
Schaul (2013)

Early-stage

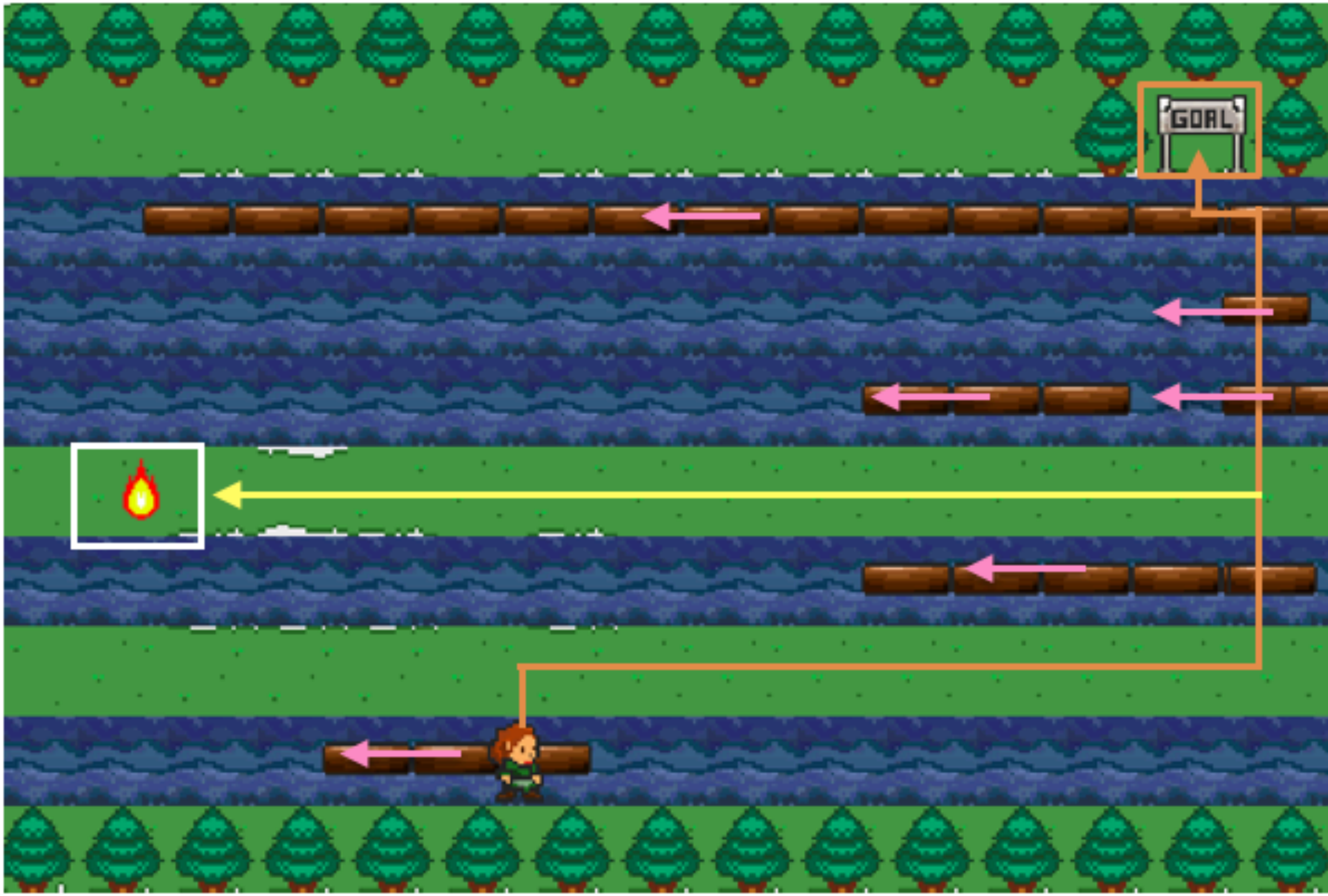


Mid-stage

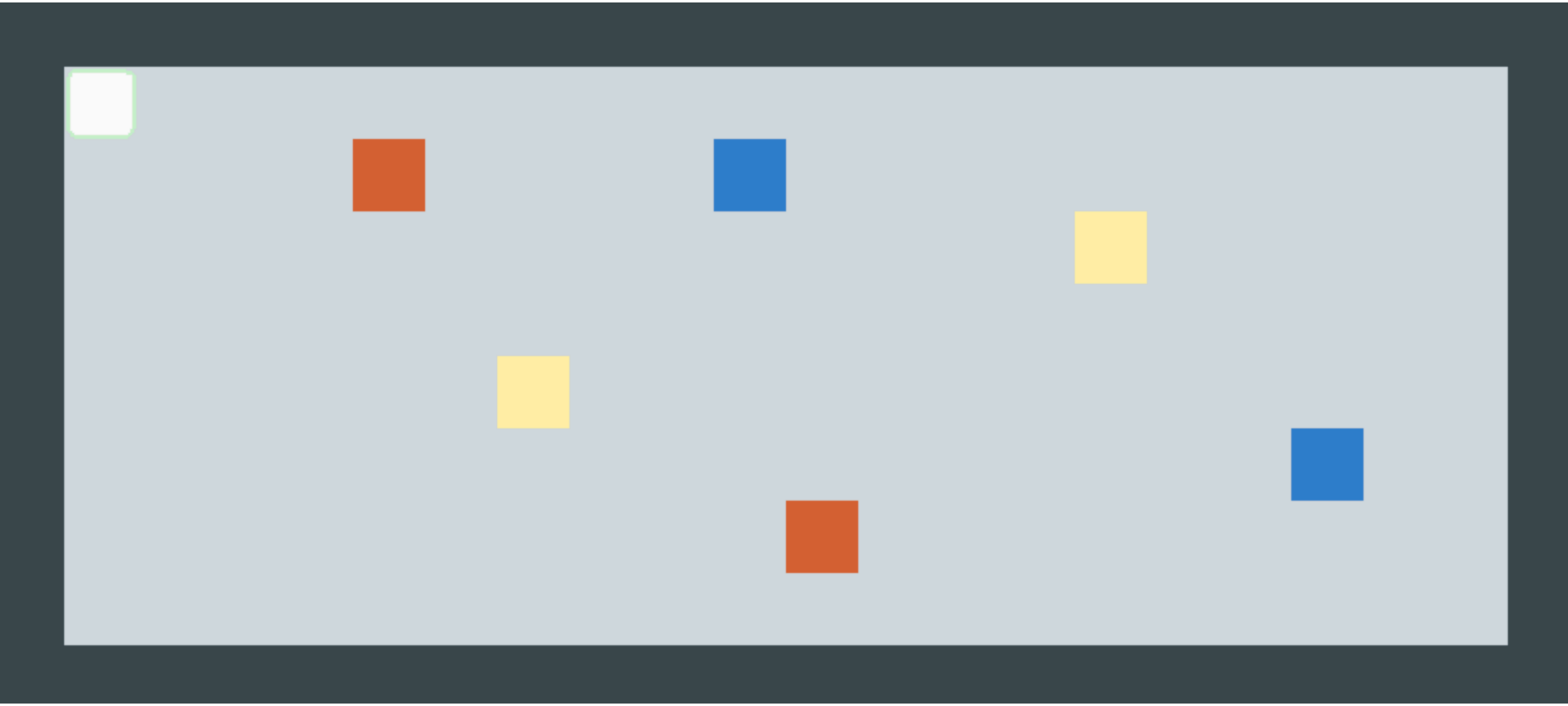
Late-stage



Generalization

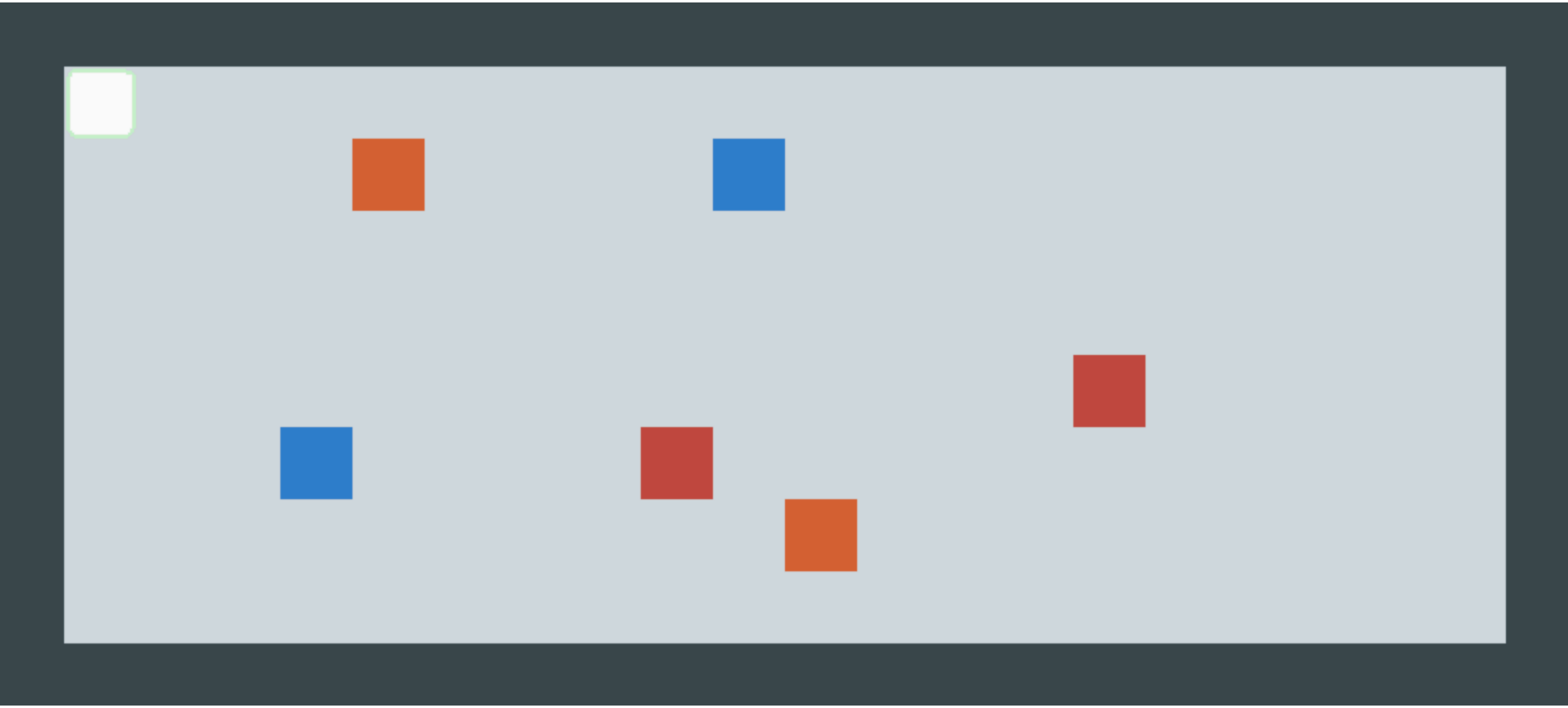


Relational structure: level 1



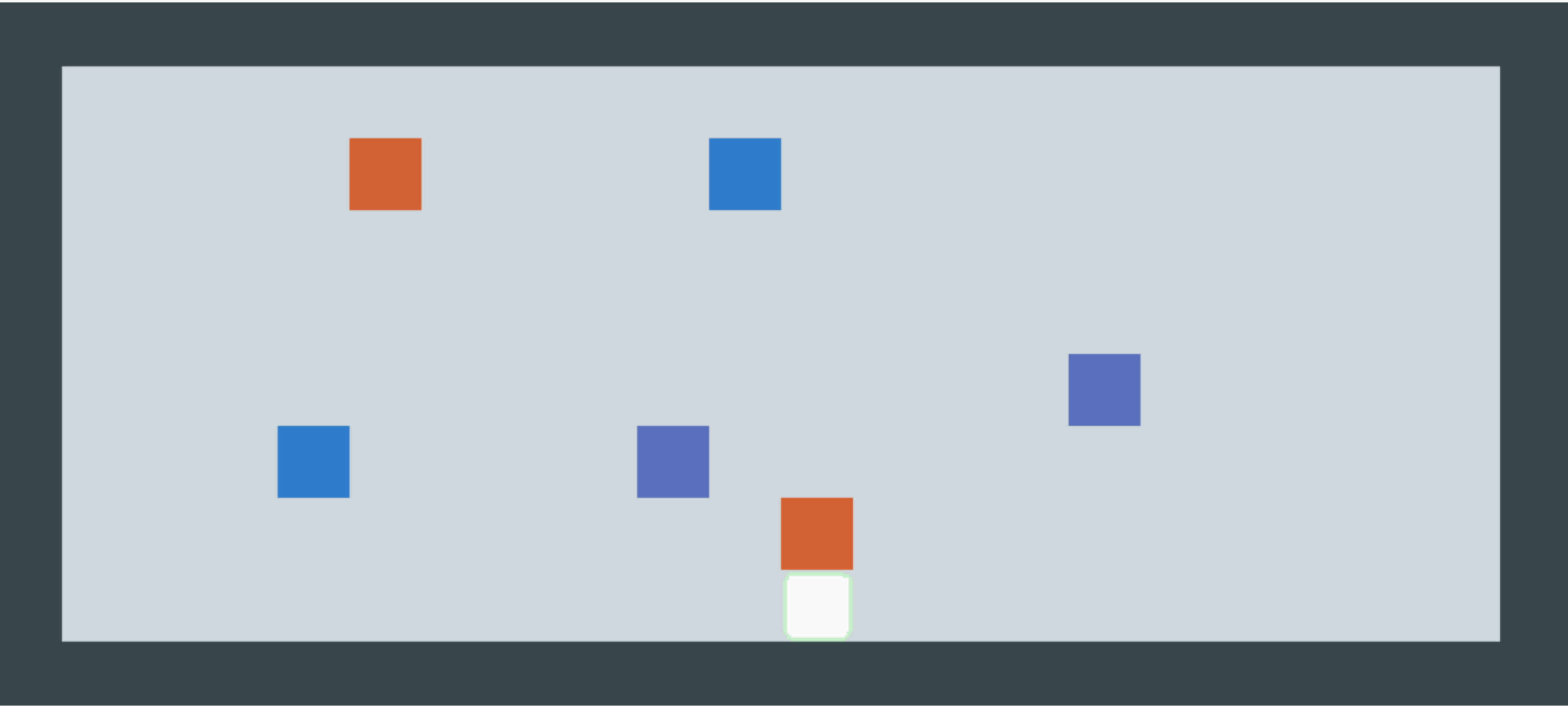
Agent wins by making all blue disappear,
by pushing blue into yellow.

Relational structure: level 2



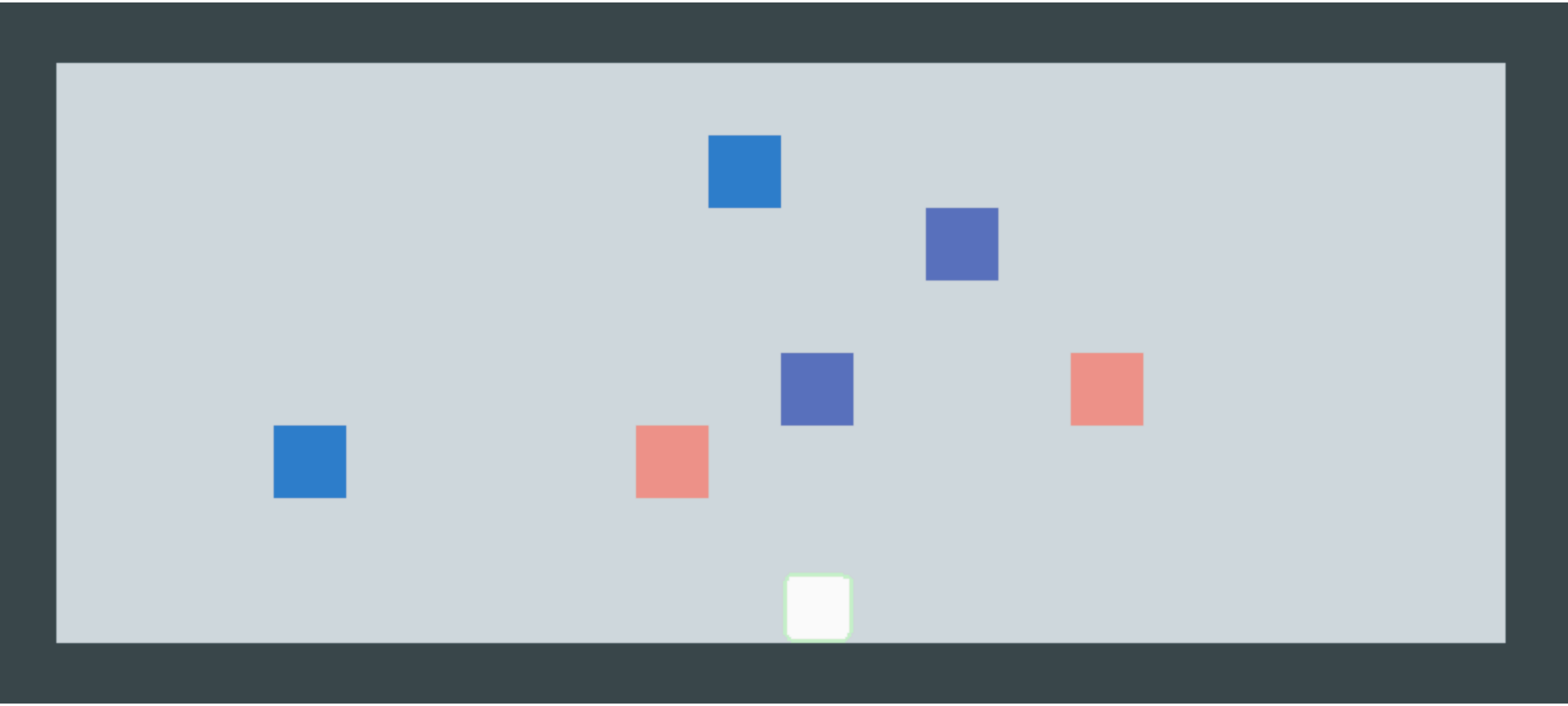
Touching red turns red into yellow.

Relational structure: level 3



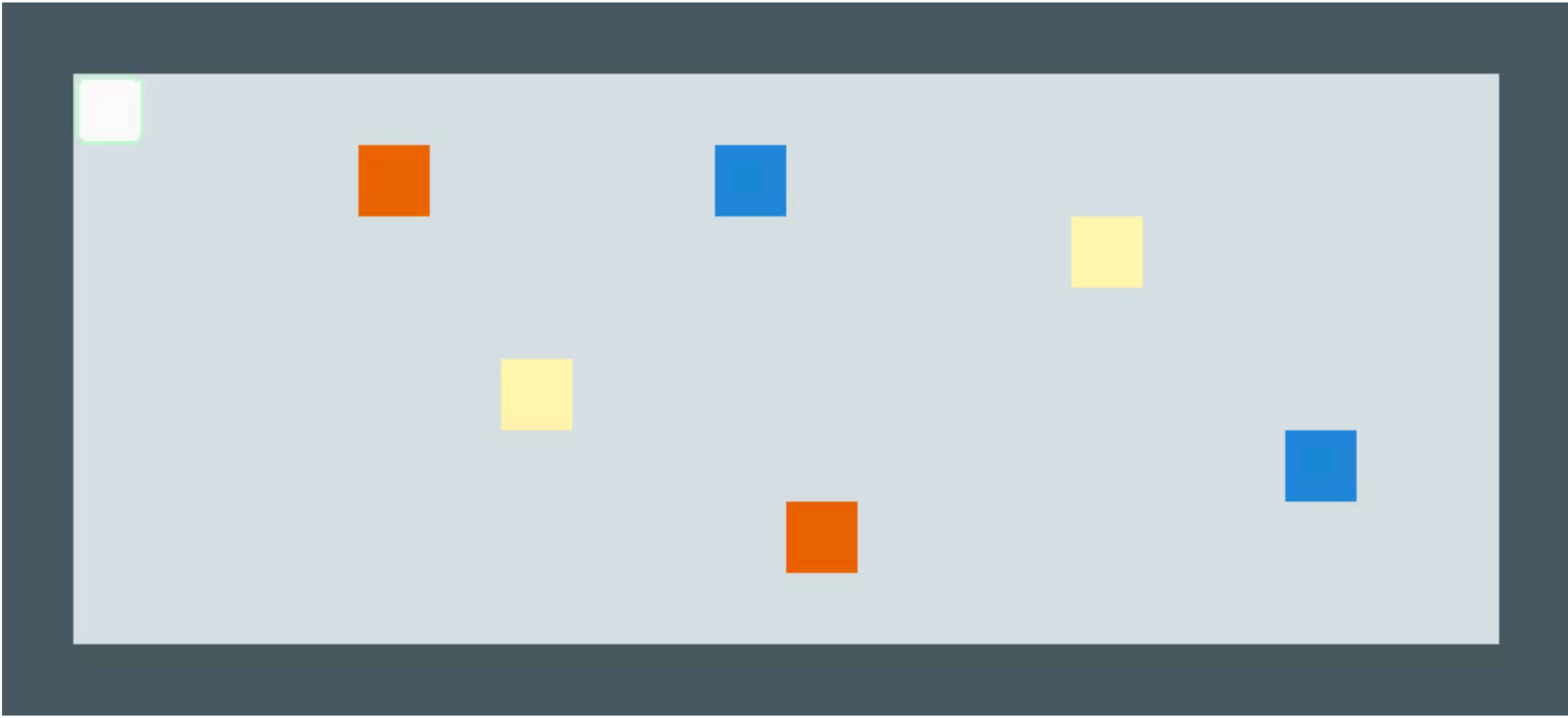
Pushing orange into purple makes orange disappear and turns purple into yellow.

Relational structure: level 4



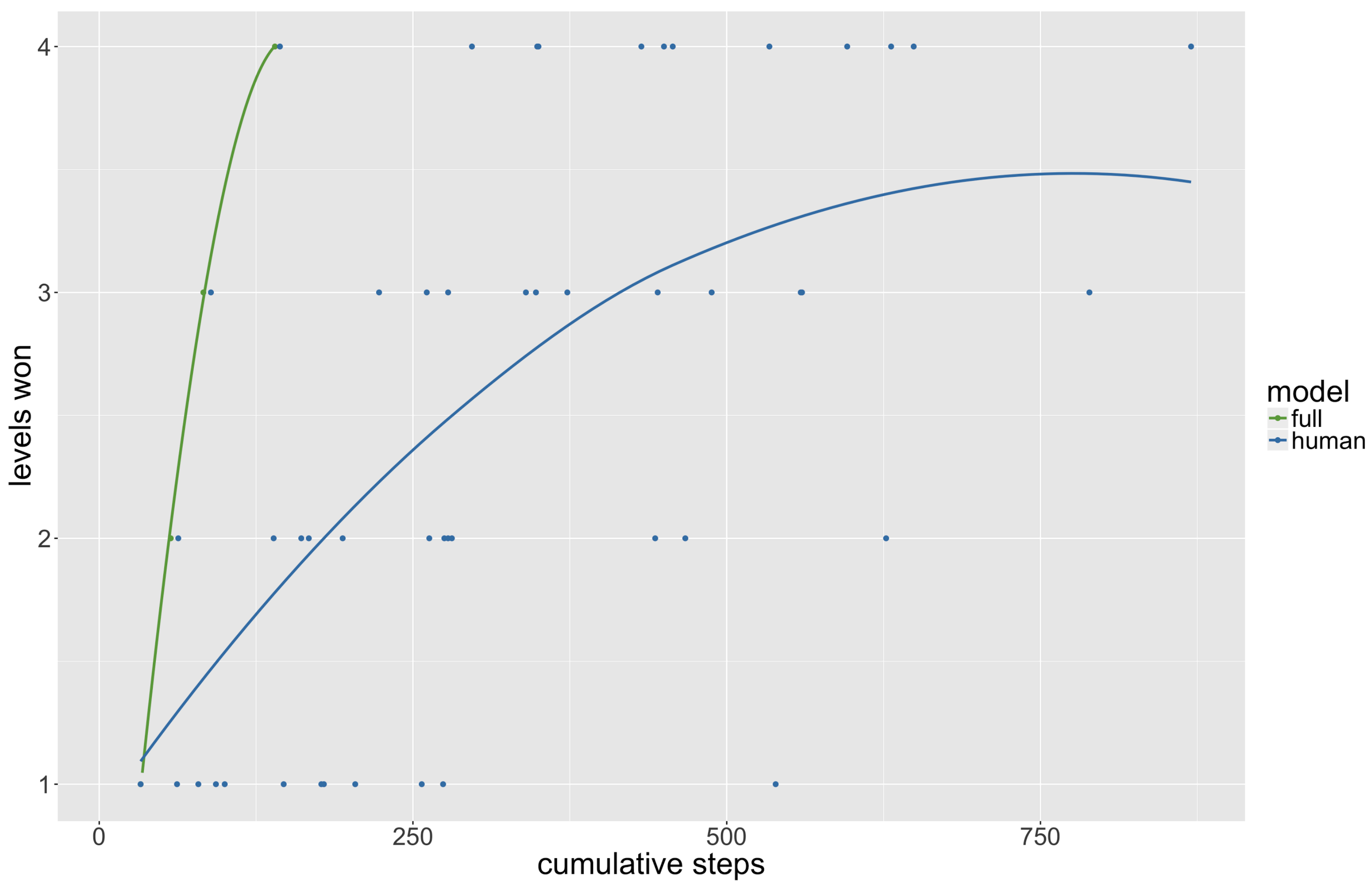
Touching pink turns pink into orange.

Relational structure: model

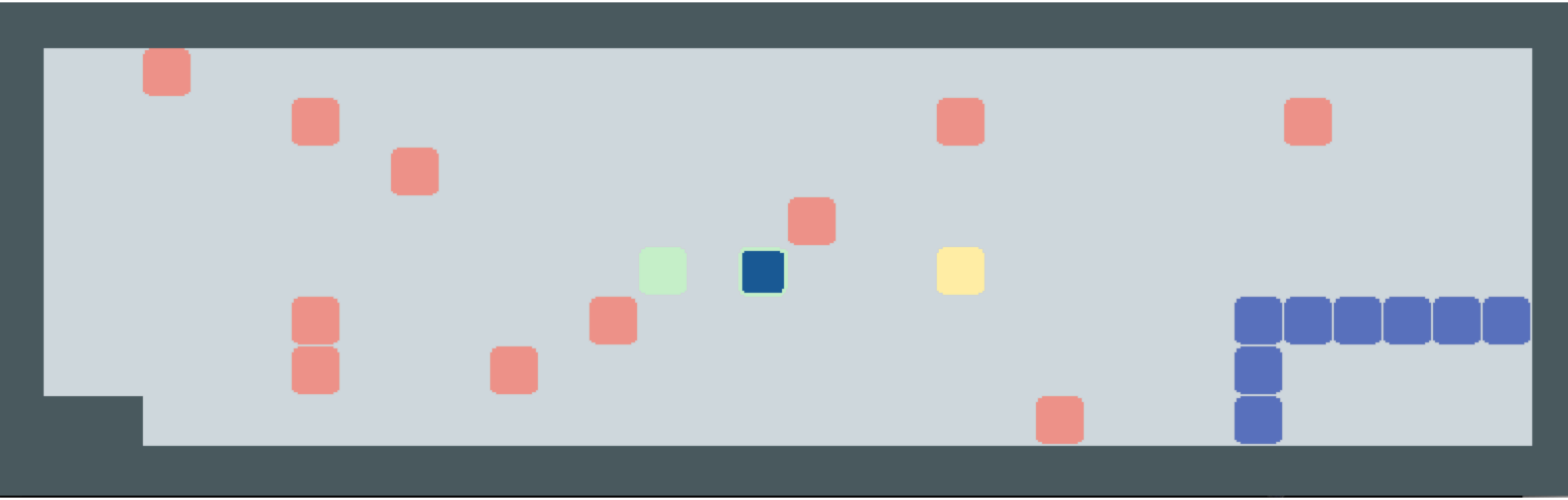


Relational structure: humans

Relational structure

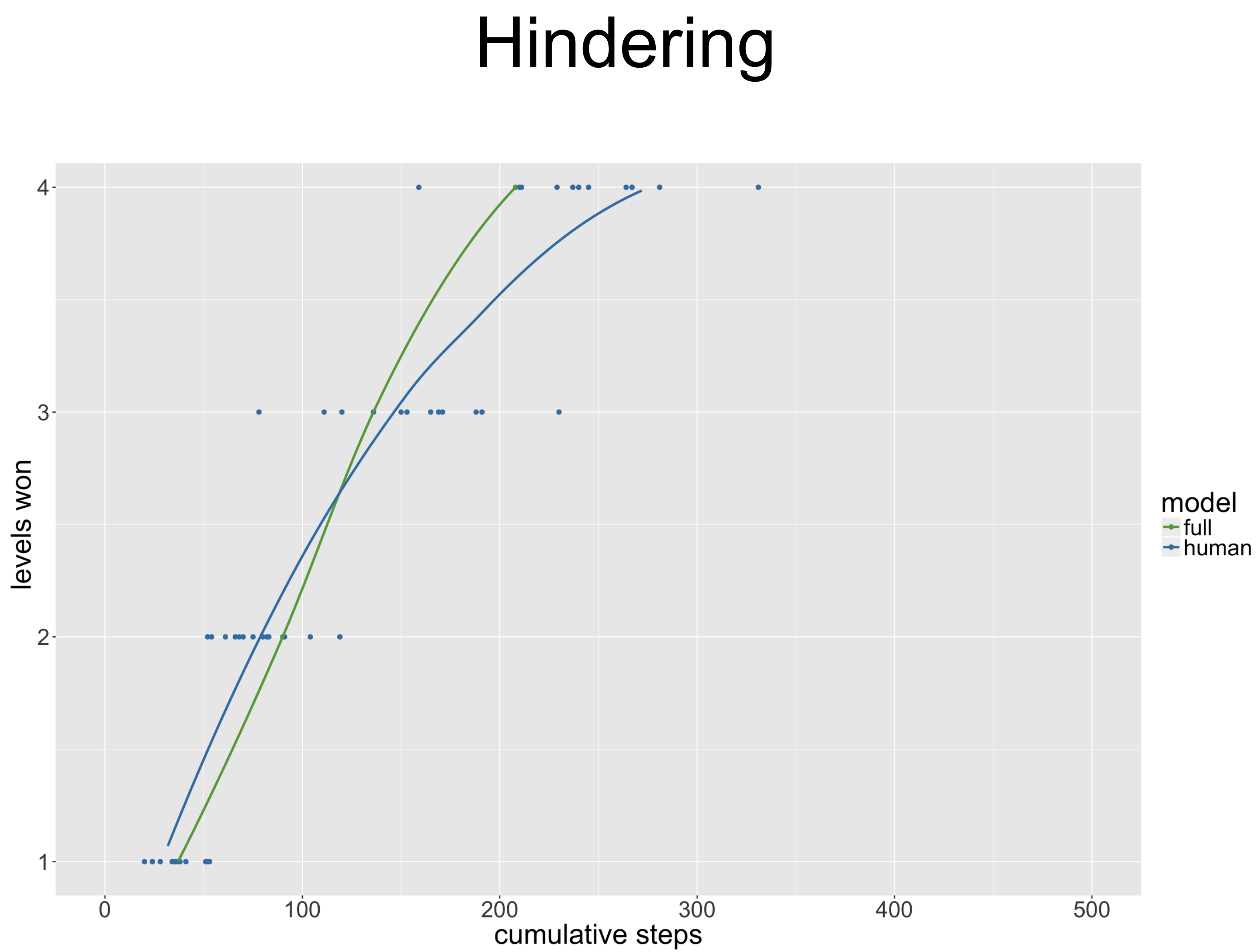


Hindering

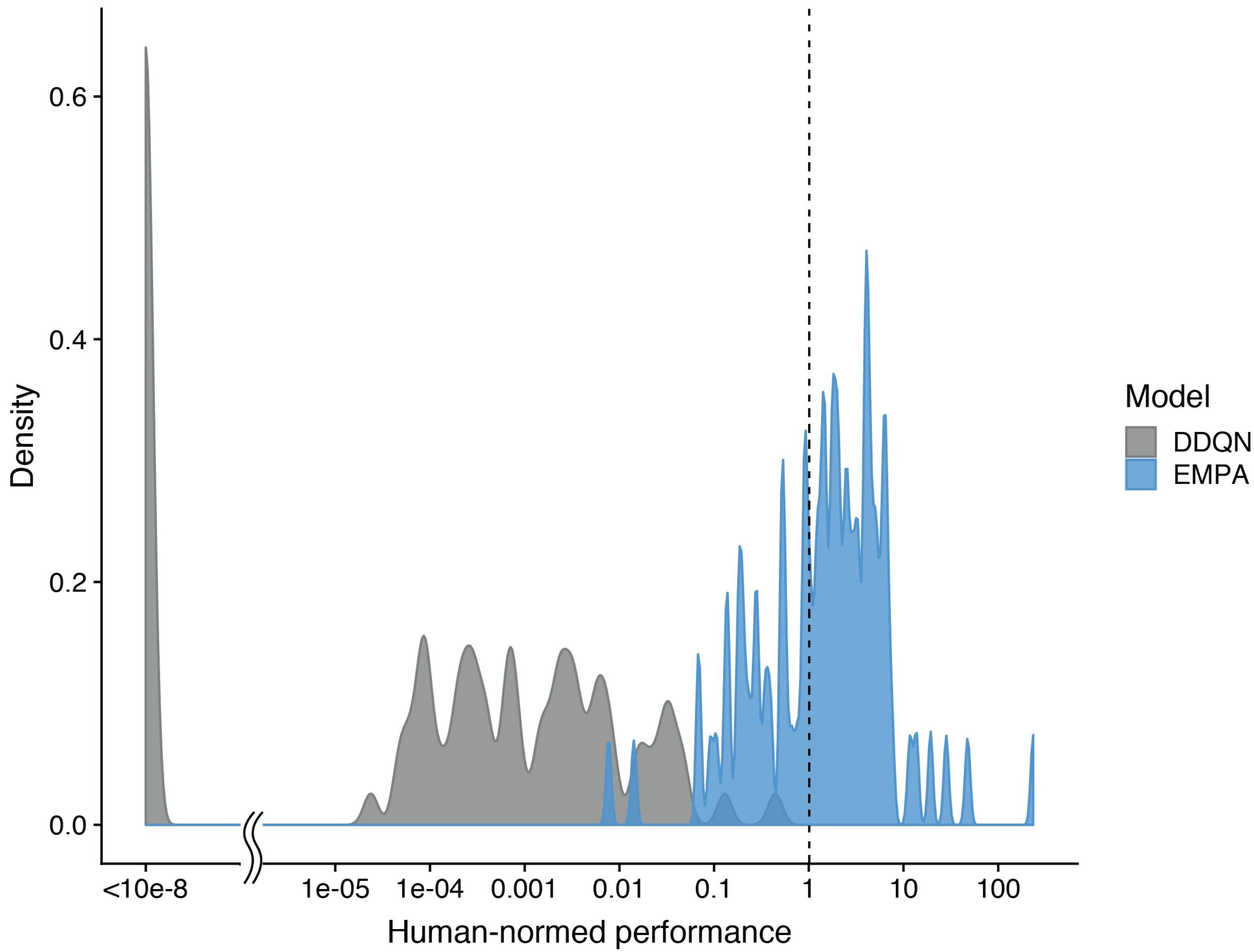


Agent has to pick up all the pink boxes to win. The Chaser (green) tries to pick up the yellow box (a termination condition). Purple fence that agent can go through and the yellow can be pushed through, that the Chaser can't pass.

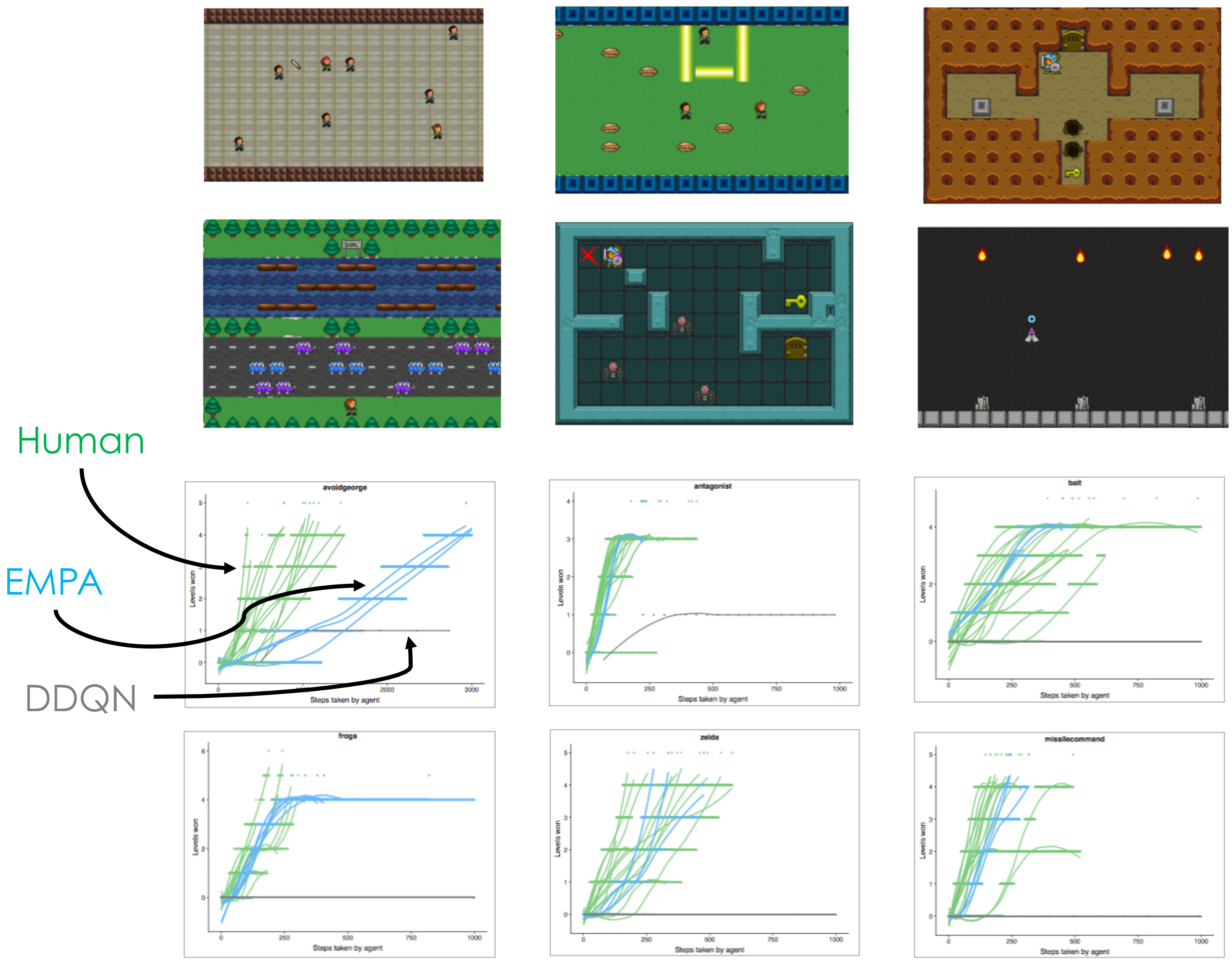
Hindering

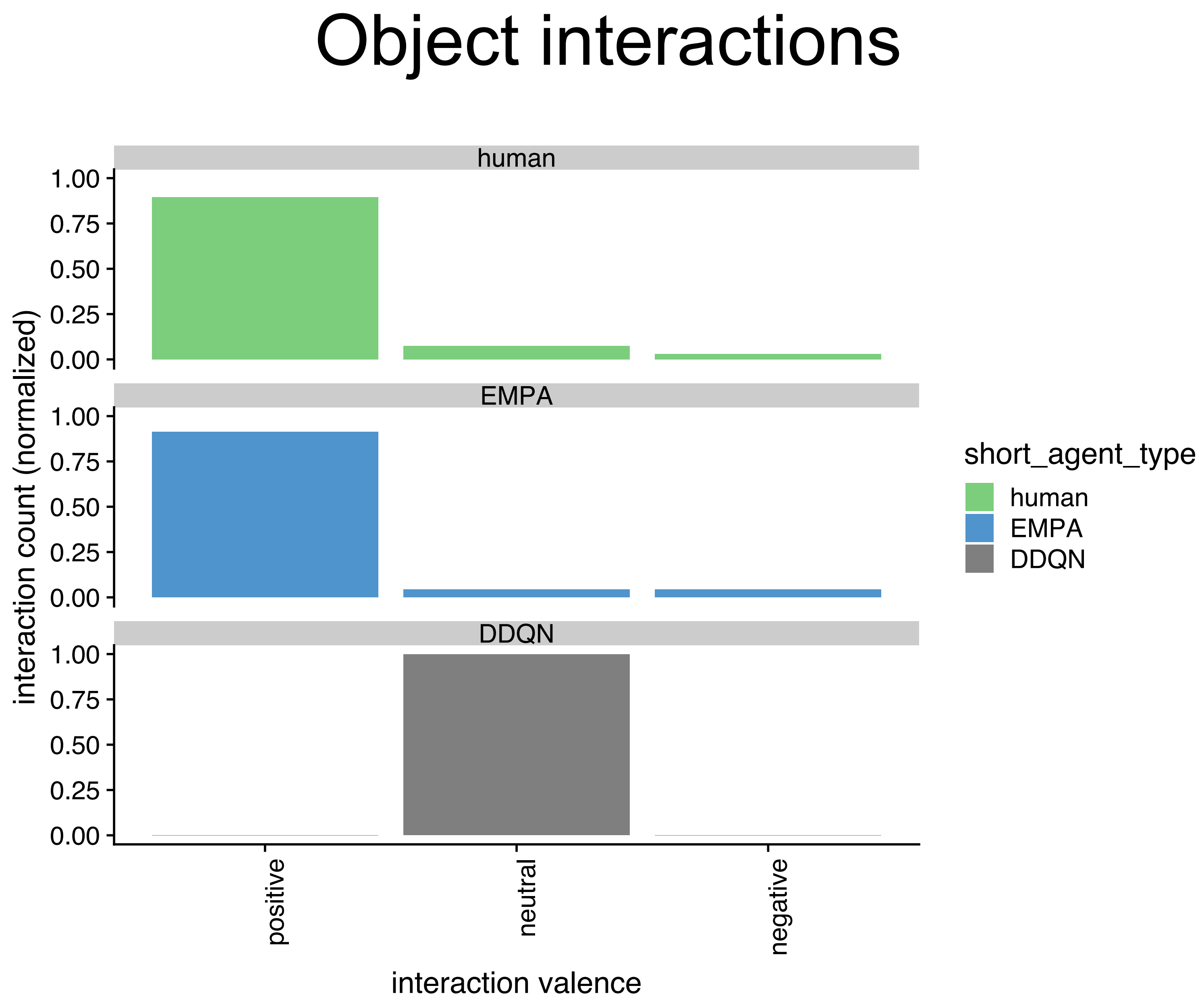

Hindering
levels won
4
3
2
1
0
100
200
300
400
500
cumulative steps
model
full
human

Comparison with humans on 90 challenging games



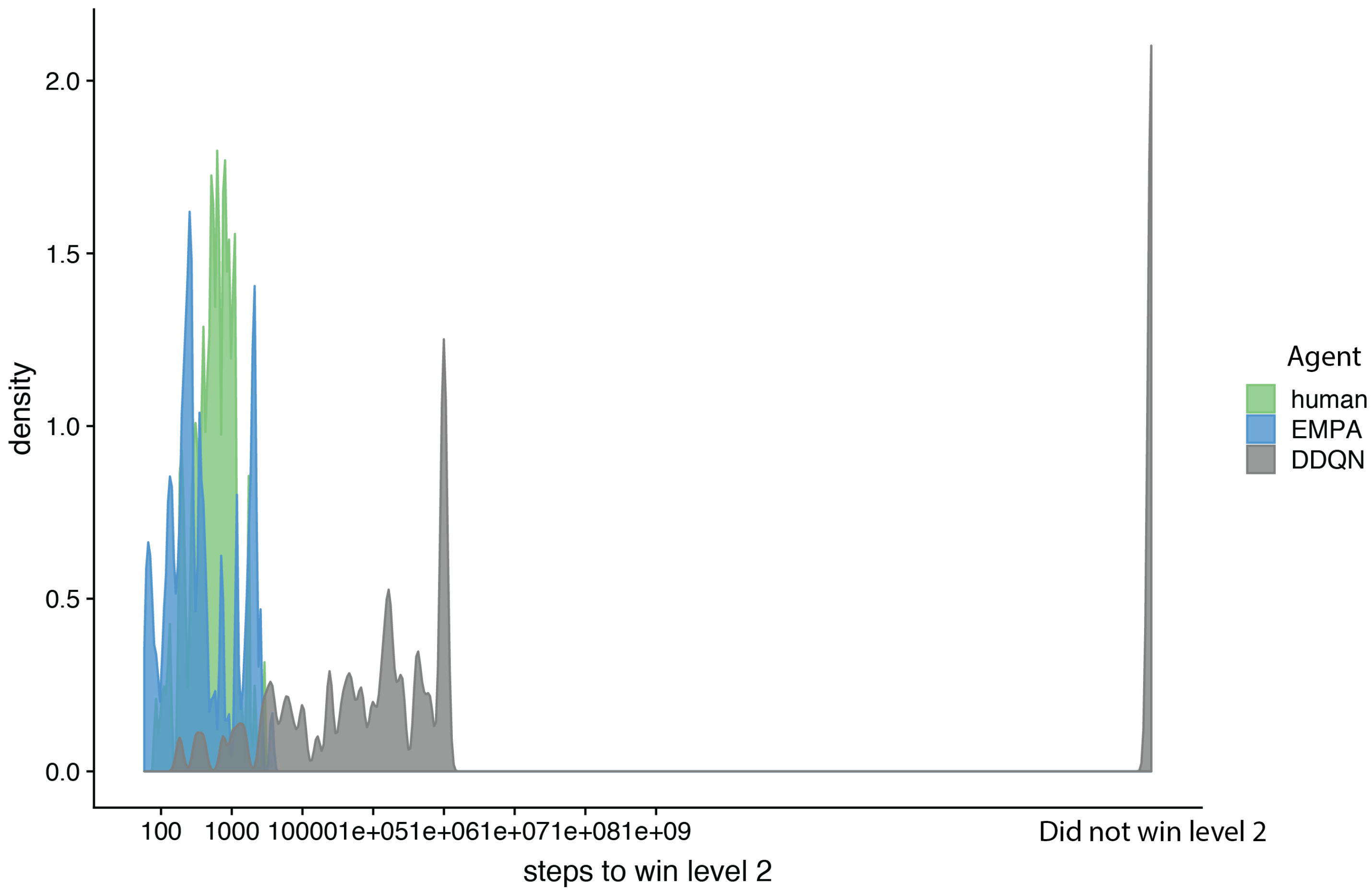
Learning curves



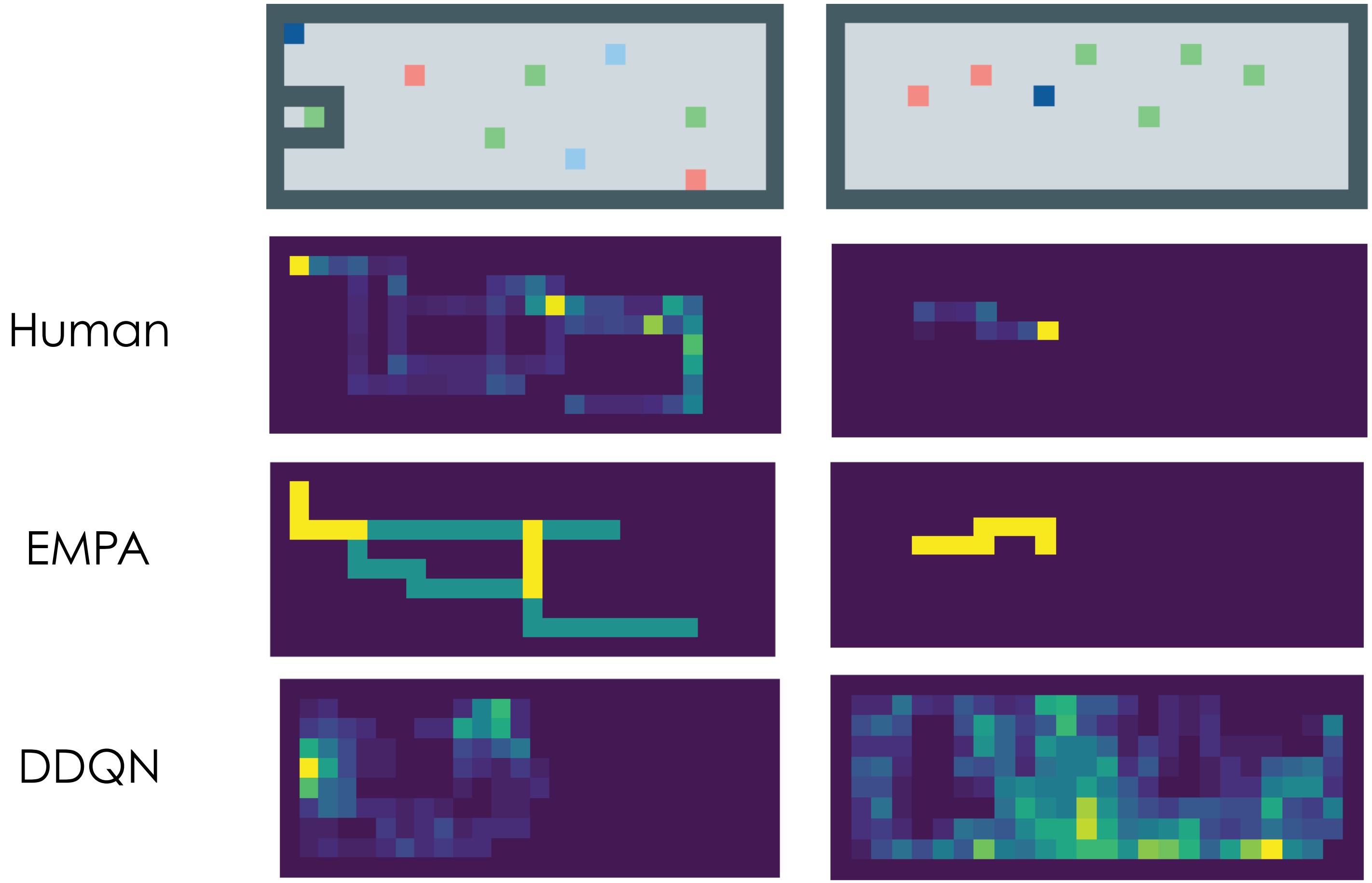


Object interactions
human
EMPA
DDQN
1.00
0.75
0.50
0.25
0.00
interaction count (normalized)
positive
neutral
negative
interaction valence
short_agent_type
human
EMPA
DDQN

Learning to learn



Exploration



Summary of experiments and modeling

A theory-based RL agent can play games in human-like ways (not just asymptotically the same performance).

Object-oriented, relational representation is key, combined with a theory induction algorithm for sample-efficient learning.

What is the “model” in model-based RL?

Most research in psychology and neuroscience has made fairly simplistic (e.g., tabular) assumptions about model-based RL in the brain.

We argue that human model-based RL uses structured, object-oriented programs that are learned from experience. Structure reduces both sample complexity and planning complexity.

What about deep learning?

Perception, theory induction and planning are all computationally expensive. Deep learning can make these more efficient.

- Learning fast pixel-to-symbol mappings.
- Finding good theories quickly using neural program search.
- Using neural value approximations as heuristics to guide planning.

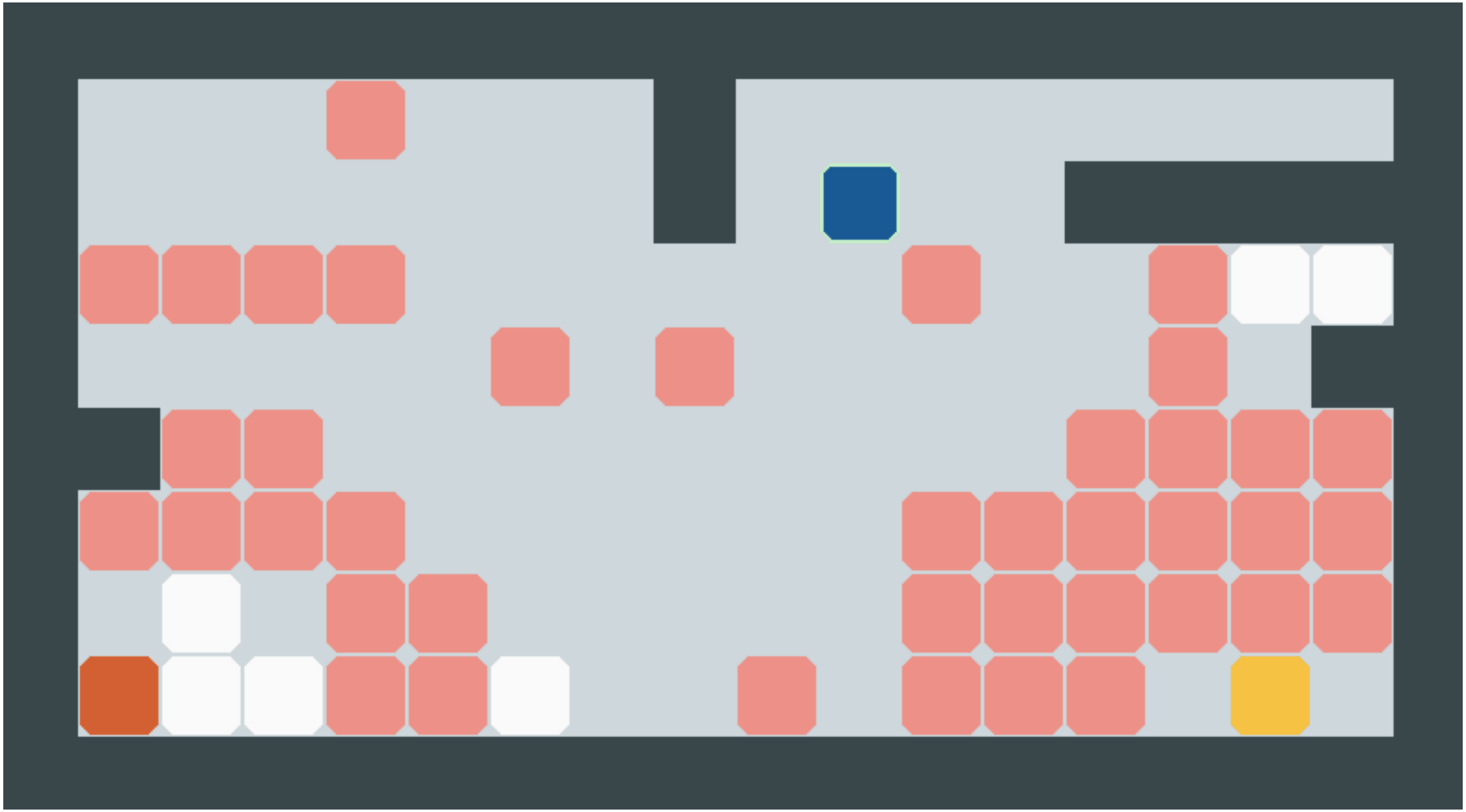
Conclusions

Human video game learning as theory-building, not pattern recognition.

Using theory induction methods in a simple but rich description language for games can do better, but there is still much work to be done.

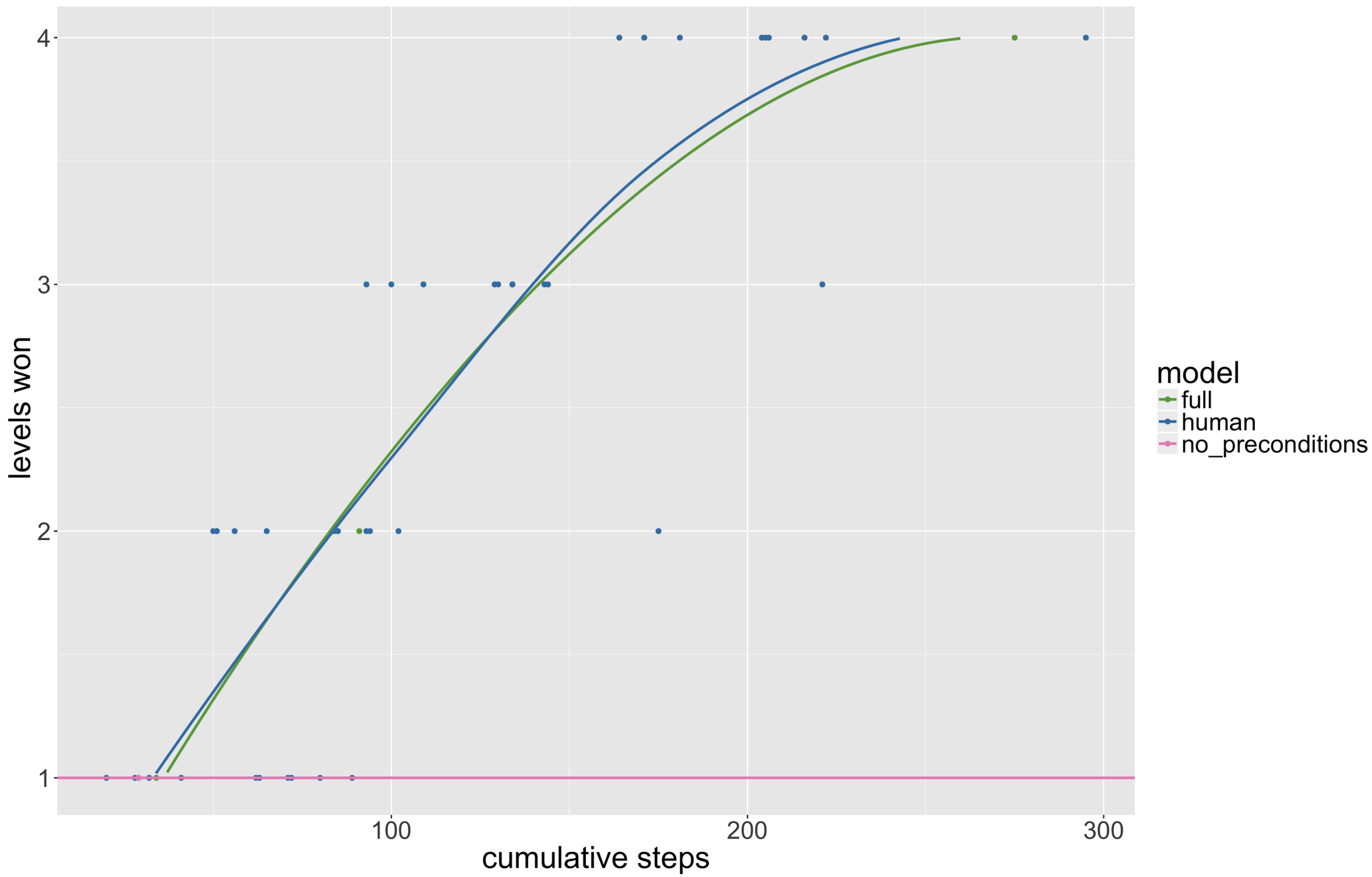
Extra slides

Preconditions



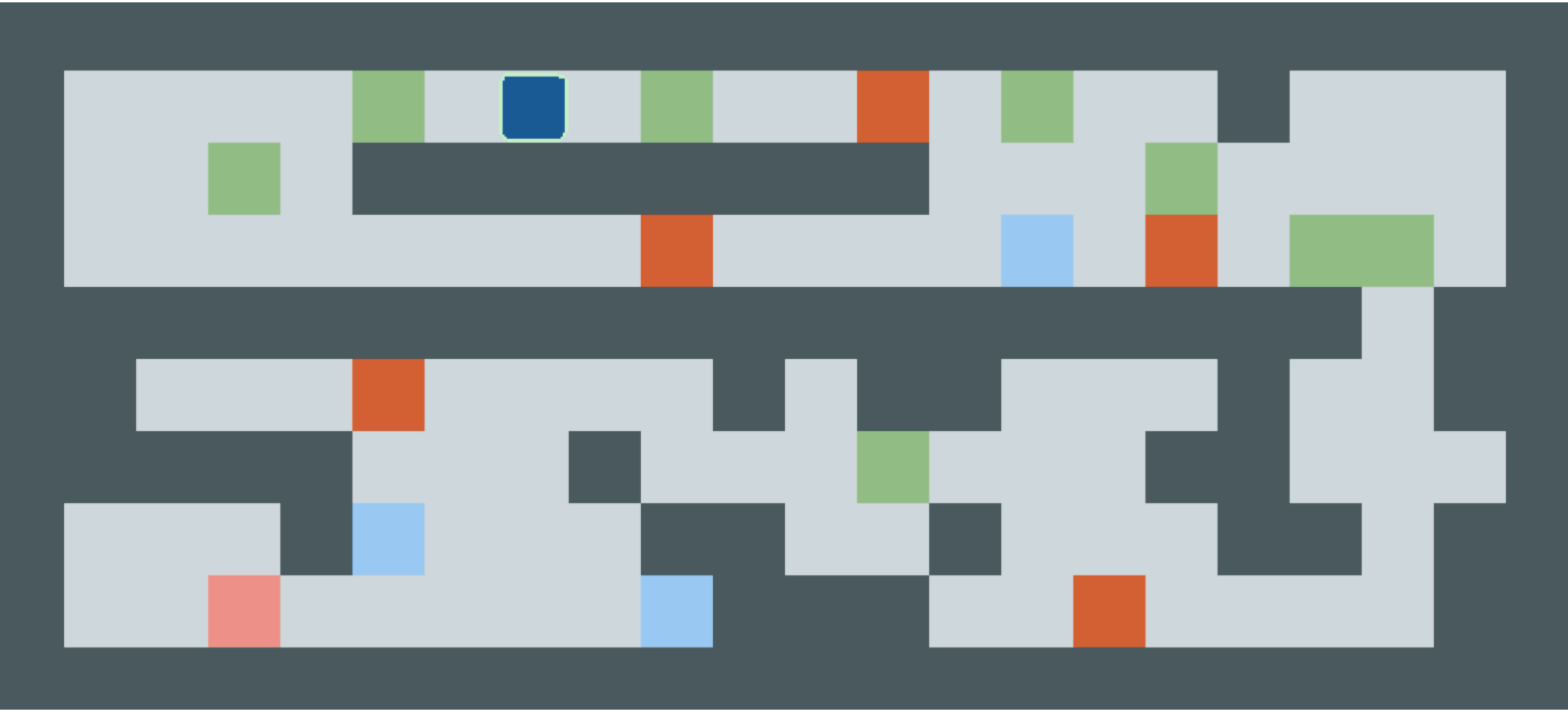
Win by getting the yellow item. Touching pink kills the agent, unless the agent has picked up a white box. Touching pink while holding a white box eliminates that white box from the agent's possession.

Preconditions



No_preconditions lesion: Can't learn that there can be modifications to the rule, "pink kills agent".

"Pushing boulders"



Agent (dark blue) wins when it touches the pink square. Green can be pushed. Orange kills the agent, but green can be pushed into orange to destroy it.

“Pushing boulders”