Reinforcement Learning and Cognitive Science: A Personal View

Rich Sutton
University of Alberta
DeepMind Alberta

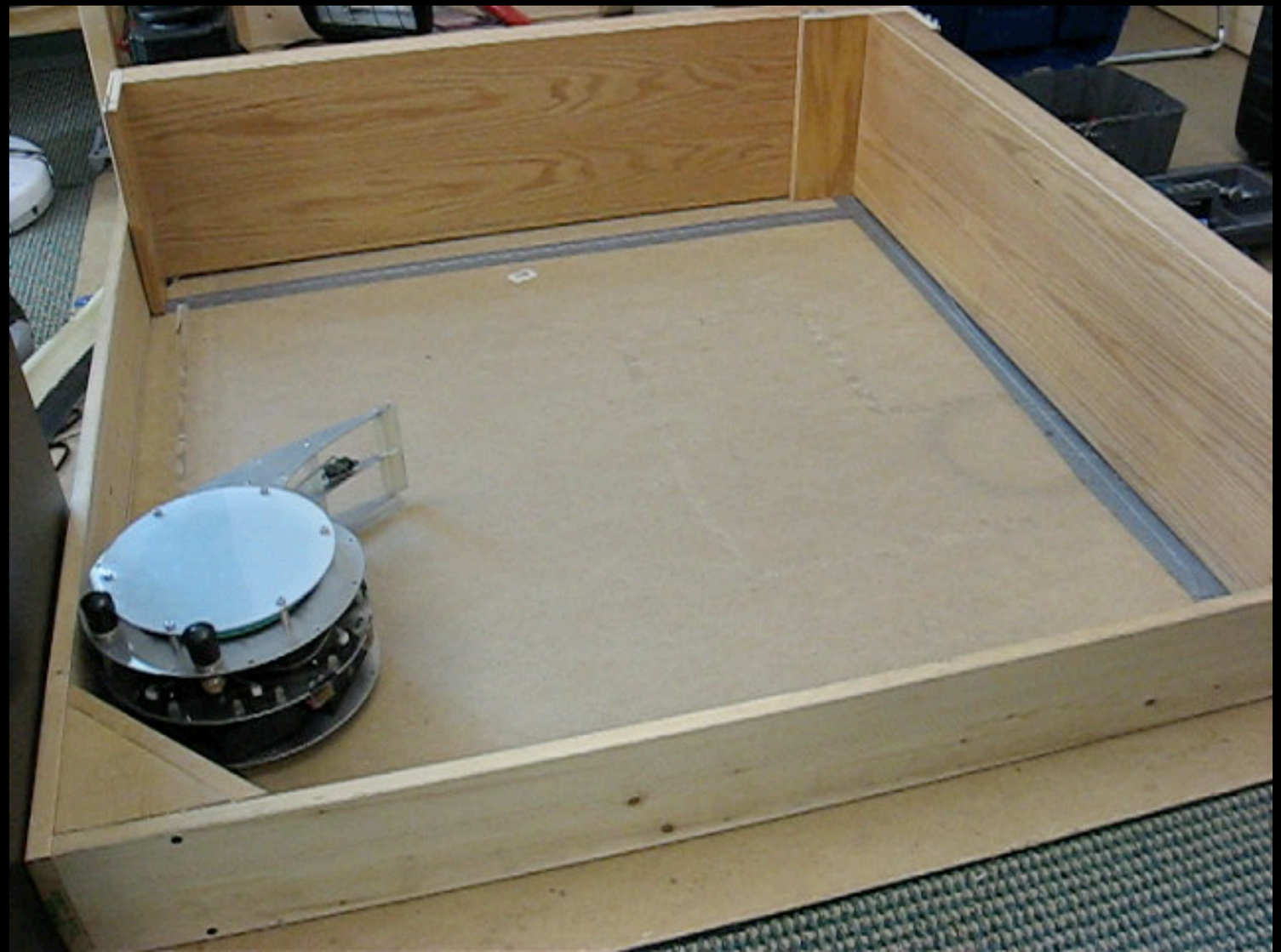
Outline

- Toward an Integrated Science of Mind
- The TD theory of Dopamine
- The Brilliant Potential
- It is exciting times in Artificial Intelligence
- The Bitter Lesson
- Some final lighter fare

There should be an *Integrated Science of Mind* that applies equally well to people, animals, *and* machines

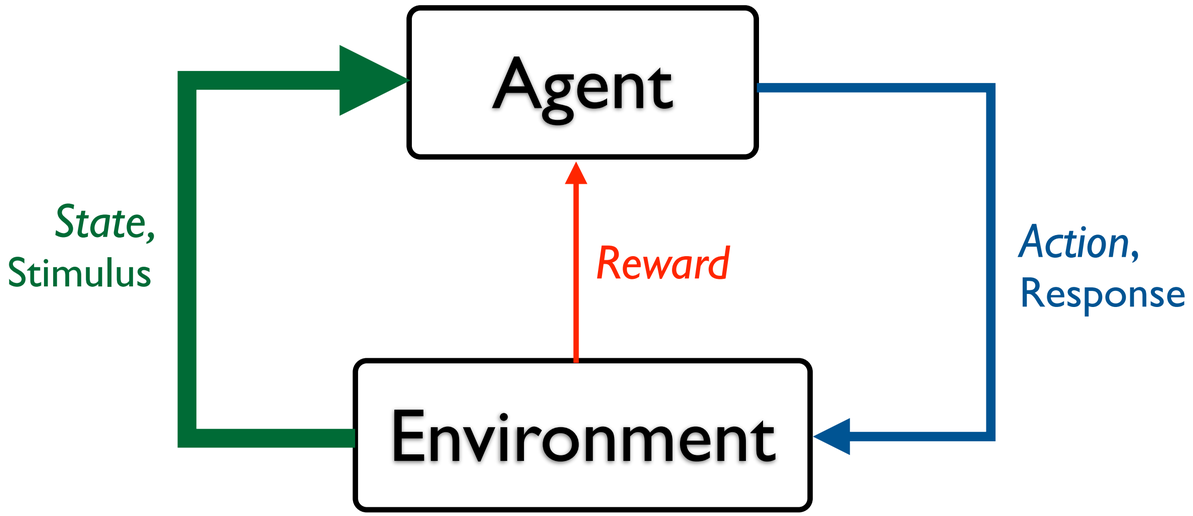
- Because all minds have essential commonalities
- Because in the foreseeable future most minds will be machines
- A Science of Mind does not rest easily within any existing field
  - Psychology? Artificial Intelligence? Cognitive Science?
- Reinforcement Learning can be seen as the beginnings of a Science of Mind

An animal-like robot, being experimented on

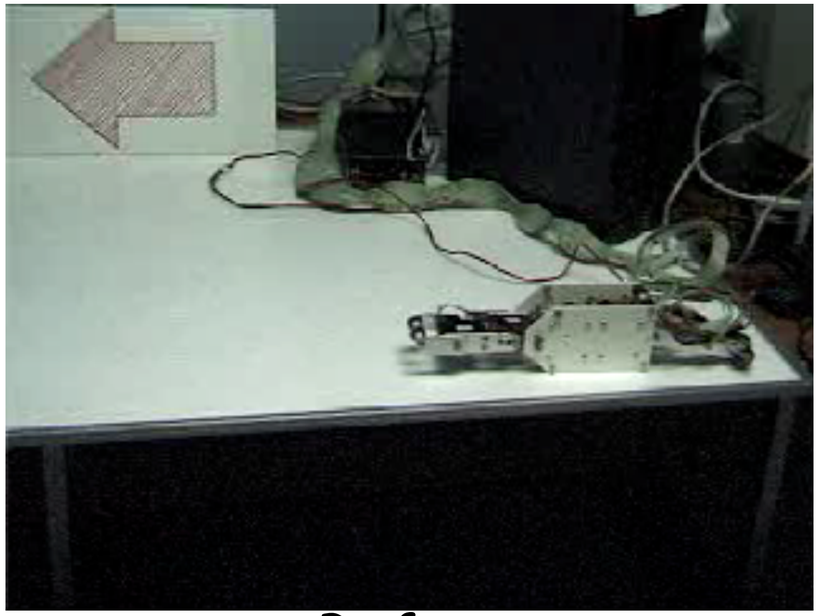


The standard Reinforcement Learning diagram

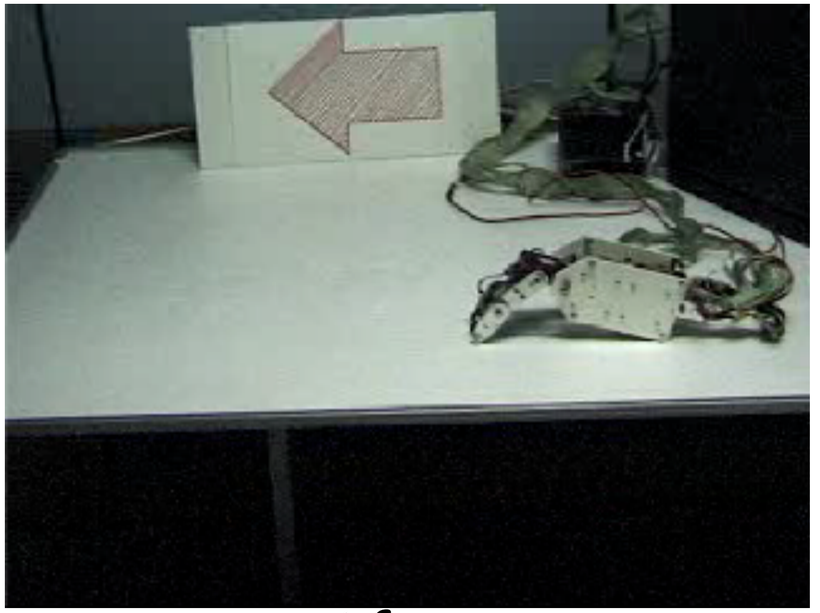
looks a lot like Thorndike's instrumental learning…



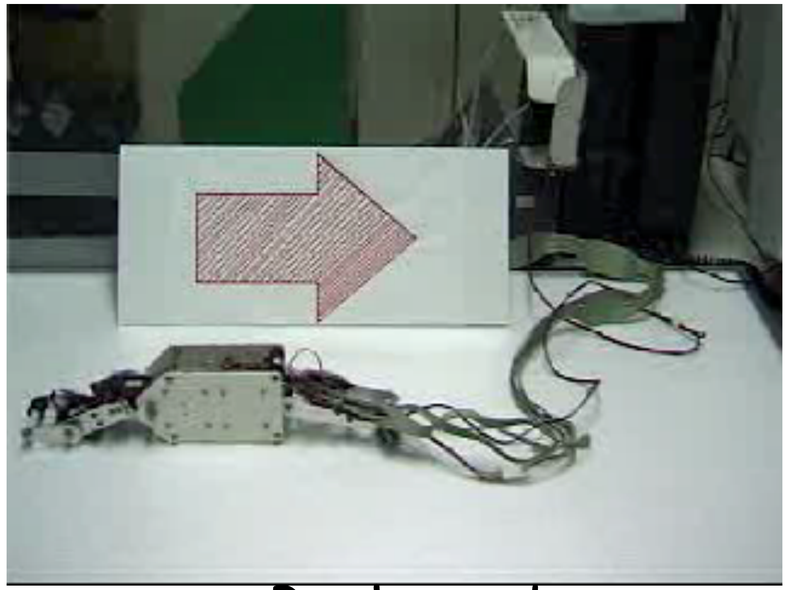
Hajime Kimura's Reinforcement Learning Robots



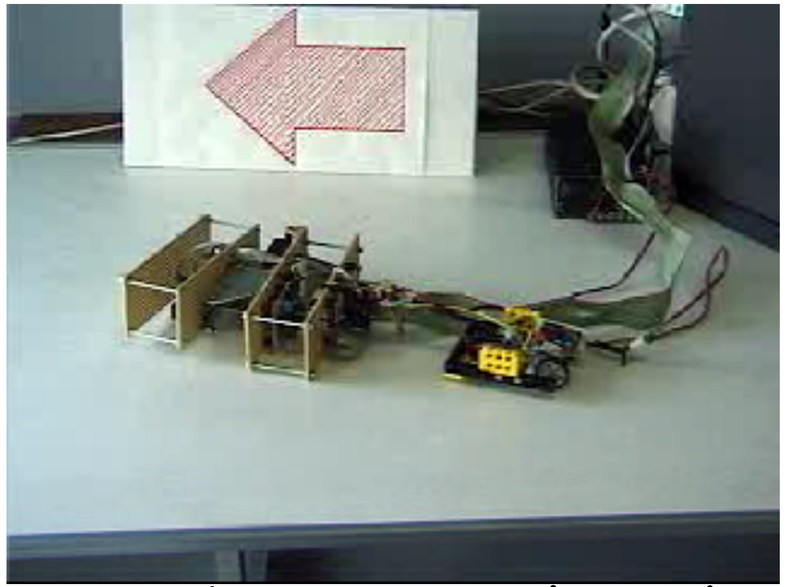
Before



After



Backward



New Robot, Same algorithm

The Integrated Science of Mind's biggest success so far:

The theory that brain reward systems are implementing TD learning
may be *the most important interaction ever*
between the engineering sciences and neuroscience


Martin Hammer
data <1995


Wolfram Schultz
data 1992+


Read Montague


Peter Dayan


Terry
Sejnowski


Andy Barto


James Houk

Workshops in 1994; early papers in 1995; Science article in 1997

What is TD learning?

- Learning a guess from a guess
- A moment-by-moment version of the main idea of the Rescorla-Wagner model of associative learning:

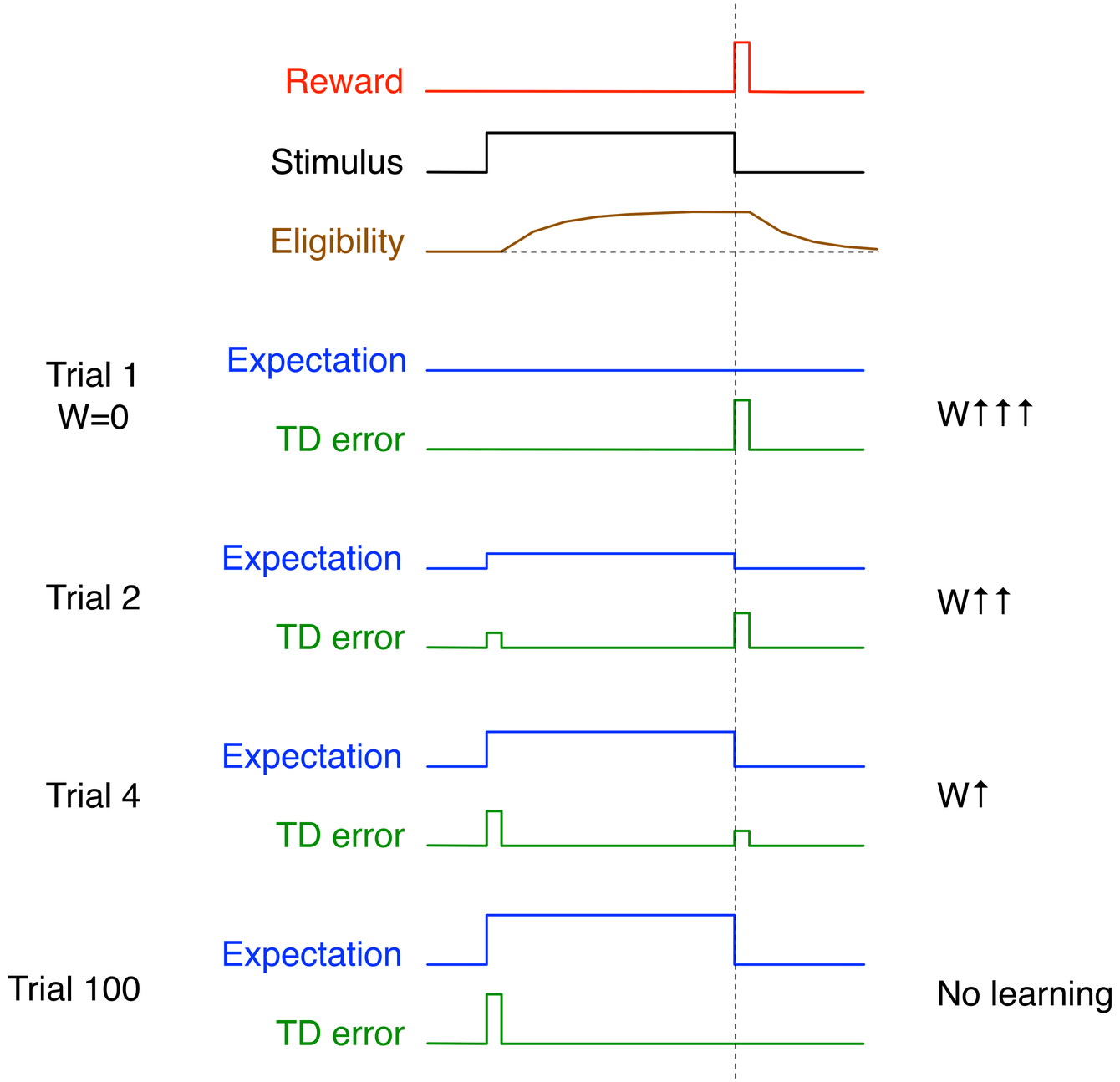
"Organisms only learn when events violate their expectations."

–Rescorla & Wagner, 1972

TD learning: Acquisition trials

TD=Temporal Difference

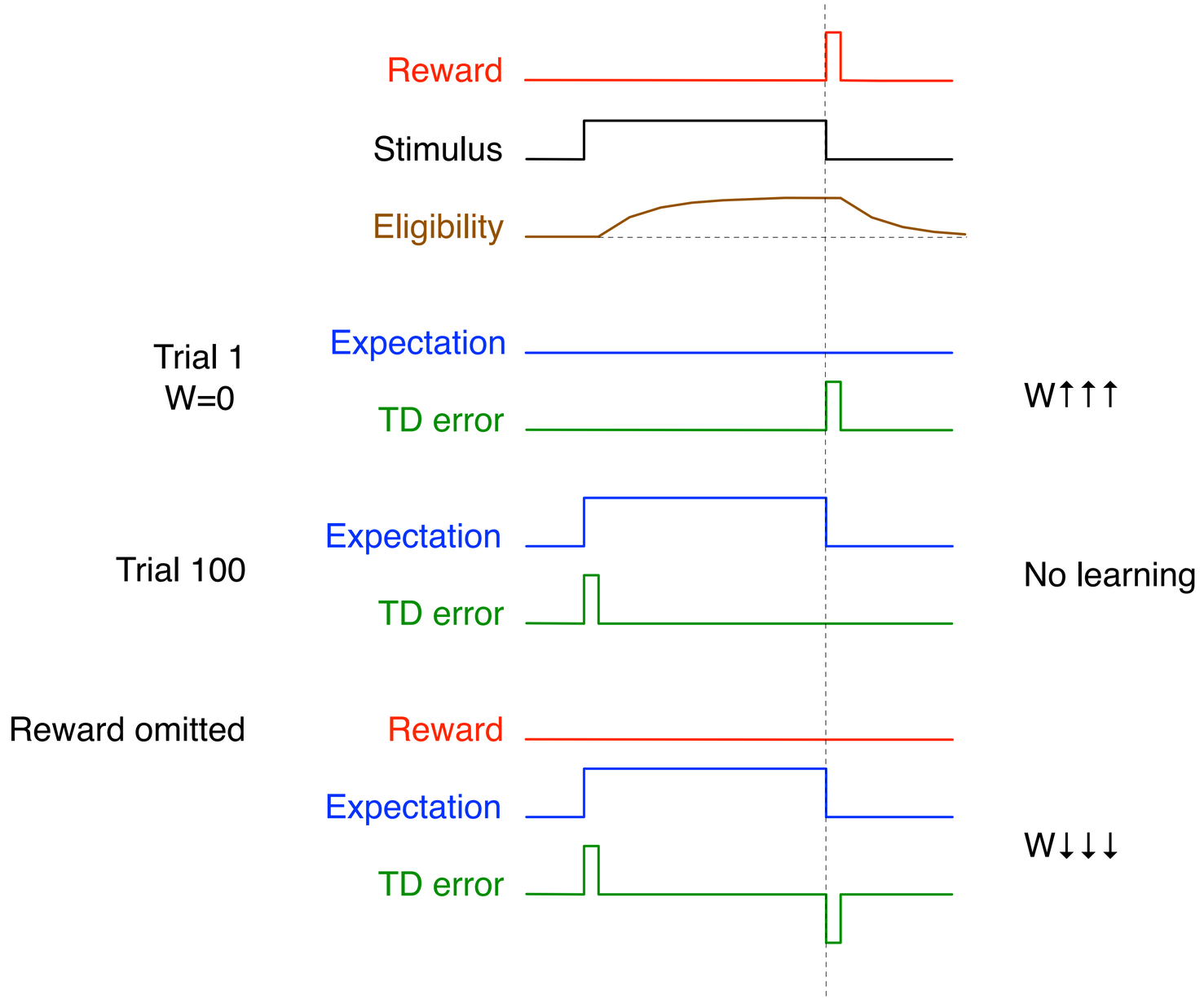
$$\text{TD error} = Reward_t + \Delta Expectation_t$$



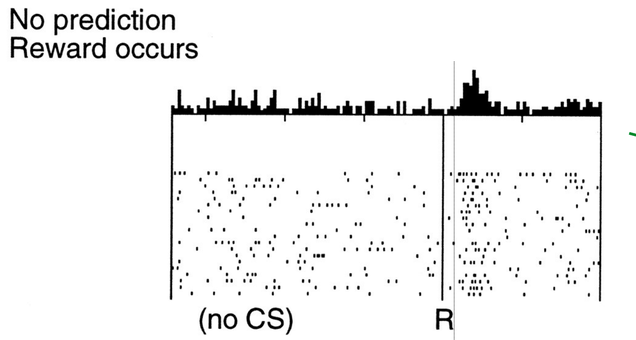
TD learning: Acquisition then extinction

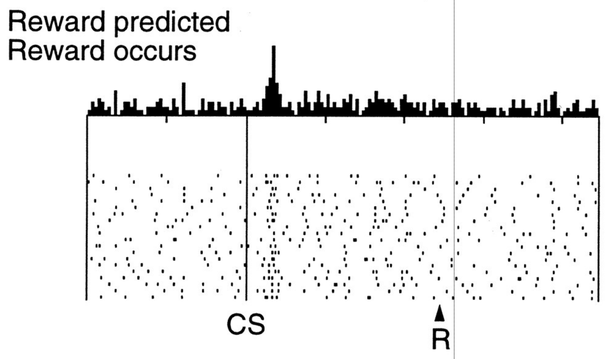
TD=Temporal Difference

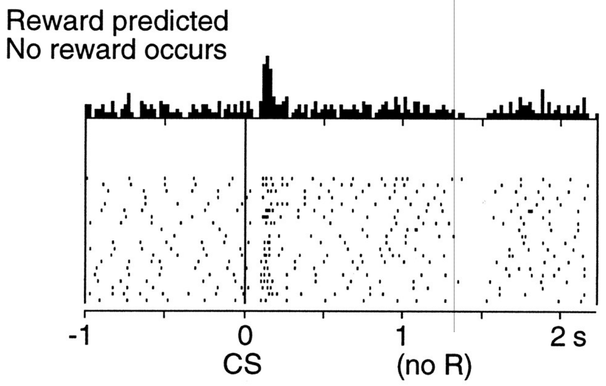
$$\text{TD error} = Reward_t + \Delta Expectation_t$$



cf. Recordings from monkey Dopamine neurons



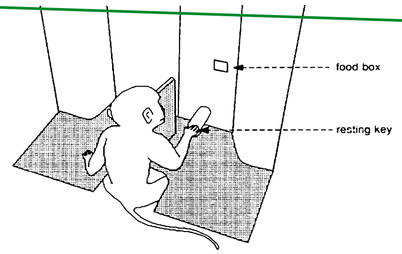




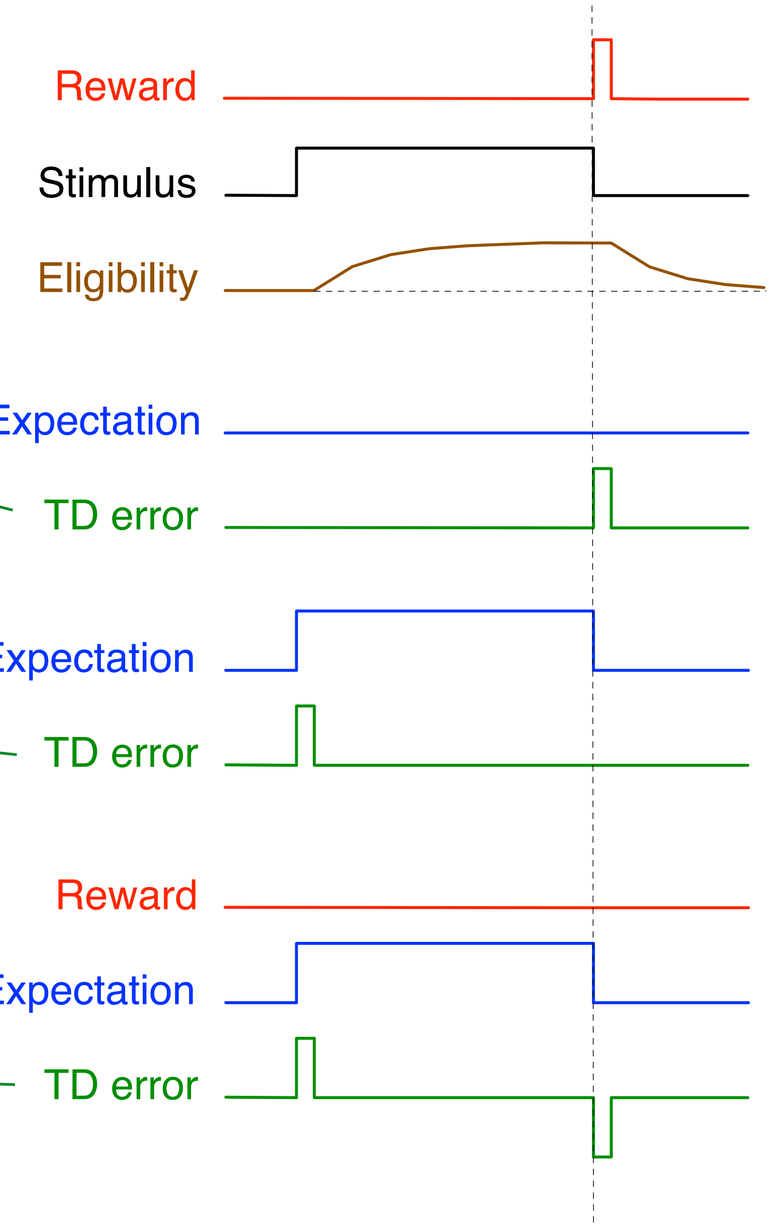
Trial 1
W=0

Trial 100

Reward omitted



Wolfram Schultz, et al.



Dopamine encodes TD error

The Integrated Science of Mind's biggest success so far:

The theory that brain reward systems are implementing TD learning
may be *the most important interaction ever*
between the engineering sciences and neuroscience


Martin Hammer
data <1995


Wolfram Schultz
data 1992+


Read Montague


Peter Dayan


Terry
Sejnowski


Andy Barto


James Houk

Workshops in 1994; early papers in 1995; Science article in 1997

The Brilliant Potential

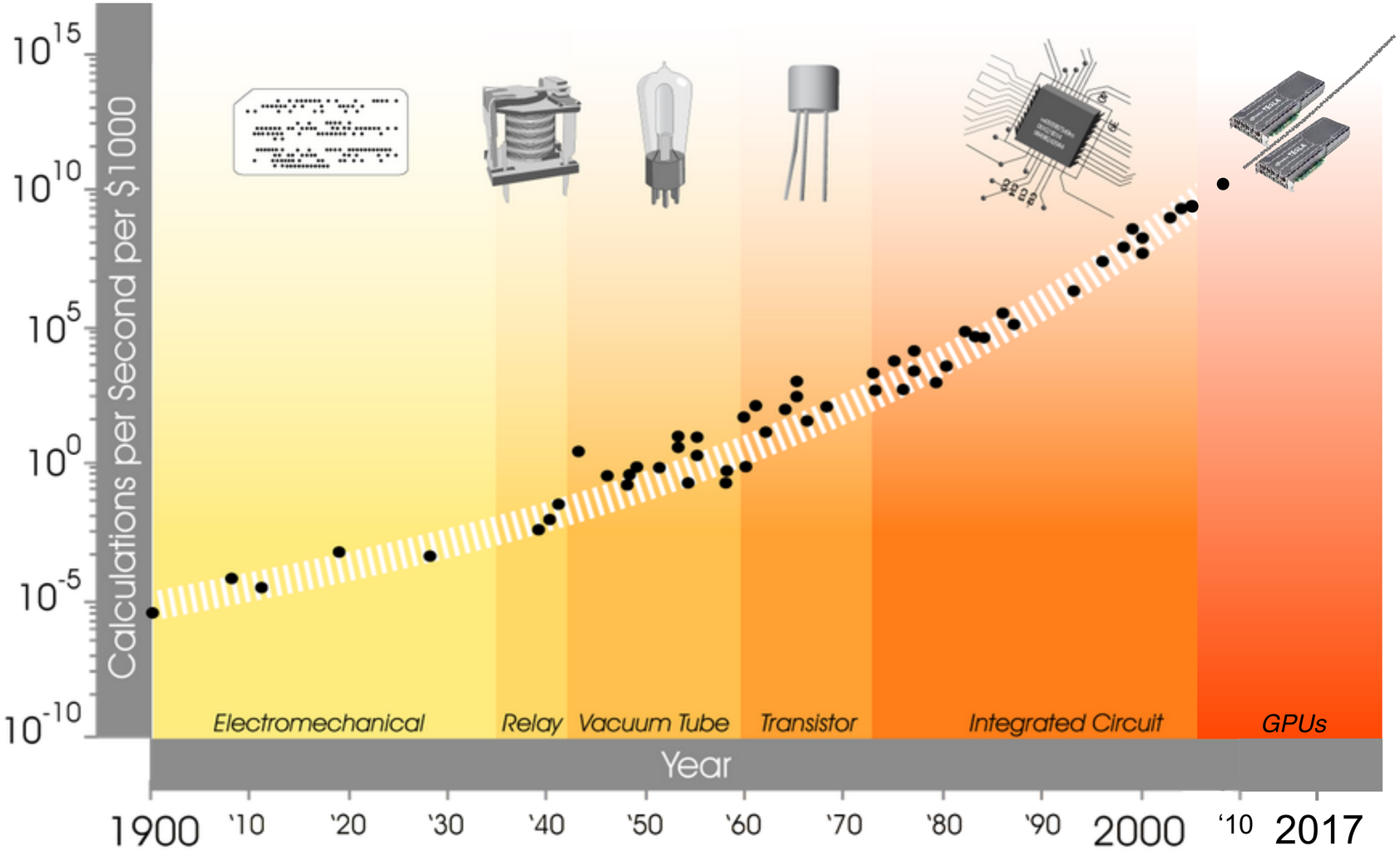
A scientific understanding of mind would be the greatest scientific achievement of all time

Mind is computational and complex

Understanding it requires more computation than we have previously had available

We have enough “now”

Moore's Law:
We live in an age of massive, ever-cheaper computation



Moore's law will continue for the foreseeable future

adapted from Kurzweil AI

Cognition as real-time high-banwidth information processing (skilled perception and action)

Mind is

- a means of predicting and controlling high-bandwidth, real-time information streams
- “the computational part of the ability to achieve goals” —John McCarthy
- a matter of degree
- in the eye of the beholder, an appearance
- “the most powerful phenomena is the universe” —Ray Kurzweil

Outline

- Toward an Integrated Science of Mind
- The TD theory of Dopamine
- The Brilliant Potential
- It is exciting time in Artificial Intelligence
- The Bitter Lesson
- Some final lighter fare

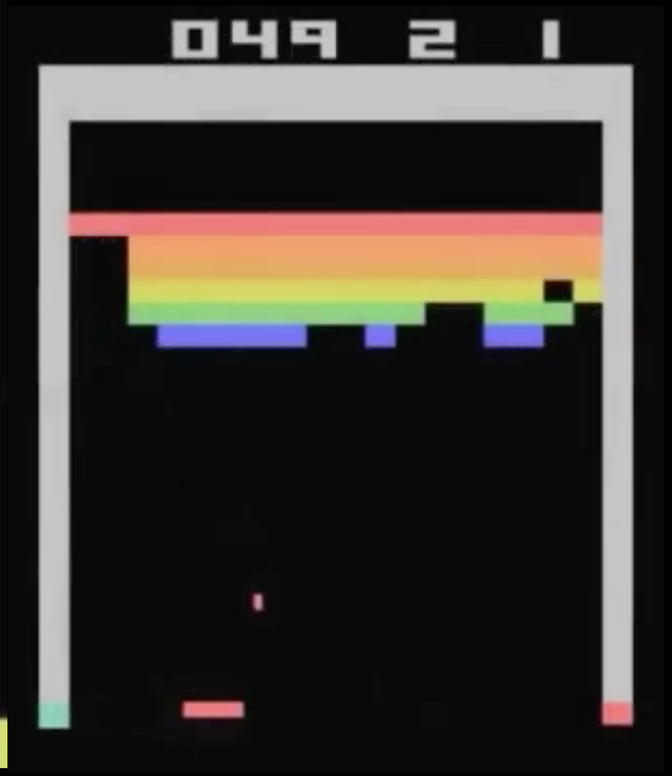
It is an exciting time in Artificial Intelligence;

In the last seven years:

- IBM's Watson beats the best human players of *Jeopardy!* (2011)
- Deep neural networks greatly improve the state of the art in speech recognition, computer vision, and natural language processing (2012—)
- Self-driving cars becomes a plausible reality (2013—)
- Deepmind's DQN learns to play Atari games at the human level, from pixels, with no game-specific knowledge (≈2014, *Nature*)
- University of Alberta program solves Limit Poker (2015, *Science*), and then defeats professional players at No-limit Poker (2017, *Science*)
- Deepmind's AlphaGo defeats legendary Go player Lee Sedol (2016, *Nature*), and world champion Ke Jie (2017), vastly improving over all previous programs
- DeepMind's AlphaZero decisively defeats the world's best programs in Go, chess, and shogi (Chinese chess), *with no prior knowledge other than the rules of each game*

RL + Deep Learning Performance on Atari Games

Space Invaders

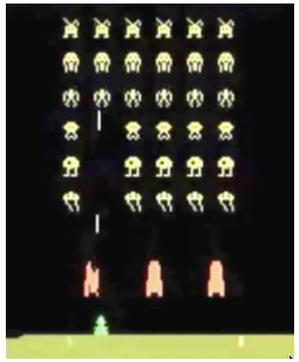


Breakout

Enduro

Reinforcement Learing + Deep Learning, applied to Classic Atari Games

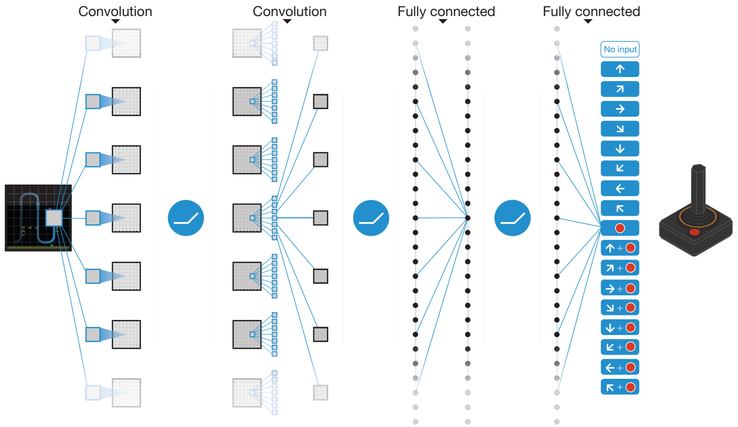
Google Deepmind 2015, Bowling et al. 2012





- Learned to play 49 games for the Atari 2600 game console, without labels or human input, from self-play and the score alone

mapping raw screen pixels



to predictions of final score for each of 18 joystick actions

- Learned to play better than all previous algorithms and at human level for more than half the games

Same learning algorithm applied to all 49 games! w/o human tuning

It is an exciting time in Artificial Intelligence;

In the last seven years:

- IBM's Watson beats the best human players of *Jeopardy!* (2011)
- Deep neural networks greatly improve the state of the art in speech recognition, computer vision, and natural language processing (2012—)
- Self-driving cars becomes a plausible reality (2013—)
- Deepmind's DQN learns to play Atari games at the human level, from pixels, with no game-specific knowledge (≈2014, *Nature*)
- University of Alberta program solves Limit Poker (2015, *Science*), and then defeats professional players at No-limit Poker (2017, *Science*)
- Deepmind's AlphaGo defeats legendary Go player Lee Sedol (2016, *Nature*), and world champion Ke Jie (2017), vastly improving over all previous programs
- DeepMind's AlphaZero decisively defeats the world's best programs in Go, chess, and shogi (Chinese chess), *with no prior knowledge other than the rules of each game*

The Bitter Lesson

The Bitter Lesson in Artificial Intelligence

- The less we build in, the better things work (eventually)
- Every time we try to help, by building in how we think we think
  - in the short-term there is improvement
  - but in the long run it is counterproductive
- We saw this in speech recognition, game playing (chess, Go, backgammon), computer vision, natural language processing
- Deep learning is just the latest instance of this bitter lesson
- Examples of this span the 70-year history of AI

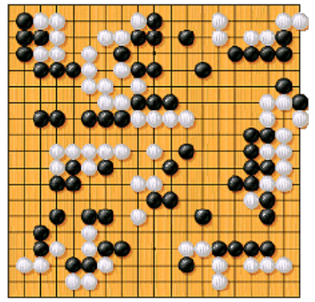
The Bitter Lesson in Computer chess

- Early computer pioneers had hoped to program computers to play chess much like humans do, by relying primarily on chess heuristics –The Computer History Museum
- By the 1970s, a new generation of chess machines arose that gave up on playing like people and focused on optimizing search
  - this was controversial, and the results were initially mixed
- In 1997, IBM's Deep Blue machine, using specialized hardware and a general α-β search defeated Gary Kasparov, the reigning world chess champion
  - At the time, many found Deep Blue's victory unsatisfying (calling it a "brute force" solution and "not the way people play chess")
- Now, with AlphaZero, most of what was learned from Deep Blue is gone
  - including α-β, hand-crafted value functions, the opening book, and the endgame database

The Bitter Lesson in Computer Go

- Chess machines were based on α-β search and state-evaluation functions, but neither of these worked well for Go
- Heuristic methods were extensively tried and gave modest improvements, but not strong play
- In 2006, a new kind of search (MCTS), was introduced, greatly improved performance, and transformed the field
  - Almost all the heuristics of previous programs were left out in MCTS
- In 2016, deep learning and reinforcement learning were used to learn an effective state-value function, dramatically improving performance
- Now, with AlphaZero, all human knowledge is removed, improving play

Exponential improvement (since MCTS, 2006) in the strength of the best computer Go programs



8 dan
7 dan
6 dan
5 dan
4 dan
3 dan
2 dan
Master 1 dan
Beginner 1 kyu
2 kyu
3 kyu
4 kyu
5 kyu
6 kyu
7 kyu
8 kyu
9 kyu
10 kyu
11 kyu
12 kyu
13 kyu
14 kyu
15 kyu

Strength of sample-based search programs

Strength of conventional search programs

AlphaGo
Zen
Zen
Zen
CrazyStone
Zen
MoGo
MoGo
CrazyStone
Indigo
Handtalk
Indigo
First Monte Carlo programs

1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2014 2015

Year

The Bitter Lesson in visual object recognition

- Early methods (dating back to the 1960s) used CAD-like models of the objects, or generalized-cylinder models, geometric models
- Later methods use more generic features, like edges, gradients, Hessian and difference-of-Gaussian detectors, then SIFT and SURF features, and finally matched to models or dictionaries;
    - Each more-general method scaled better and eventually worked better
- All this is thrown out in deep learning, which performs better and is easier to design
    - Features are learned instead of being built in
    - The only things built in are invariance to translation and scale.

The Bitter Lesson in Artificial Intelligence

- In chess

  we thought human ideas were key, but it turned out (deep Blue 1997)
  that big, efficient, heuristic search was key

- In computer Go

  we thought human ideas were key, but it turned out (MCTS 2006–)
  that big, sample-based search was key,
  and eventually all human knowledge was discarded (AlphaZero, 2018)

- In speech recognition

  human ideas were key to early systems (Harpy and Hearsay, 1970s);
  later systems used engineered statistical models (HMMs, 1980s),
  but eventually all human designed features were discarded (deep learning, 2010s)

- In natural language processing

  we thought that human-written rules were key,
  but it turned out that statistical machine learning and big data were key

- In visual object recognition

  we thought human ideas were key, but it turned out (deep learning 2012–)
  that big data sets, many parameters, and long training was key

In AI, general principles have generally won the day

- Early symbolic, hand-crafted, and domain-specific AI methods relied heavily on human understanding and participation in their design
- Over time, statistical, learned, and general-purpose AI methods have steadily increased in relative importance
- In the early days of AI, a distinction was made between "strong" methods (powered by human input) and "weak" methods (relying on general principles)
  - The terminology is telling; the founding fathers favored methods that sought to leverage human input
  - But they were wrong; the weak have inherited AI

Many much-loved topics in Cognitive Science seem vulnerable to the bitter lesson

- Most cognitive scientists work at a high level
  by presuming lower levels are given; they presume things like:
    - language, objects, relations, space, other minds
- But what if our preconceptions of these things are wrong?
- Will all this work go the way of prior built-in features,
  and be swept away by some future version of deep learning?

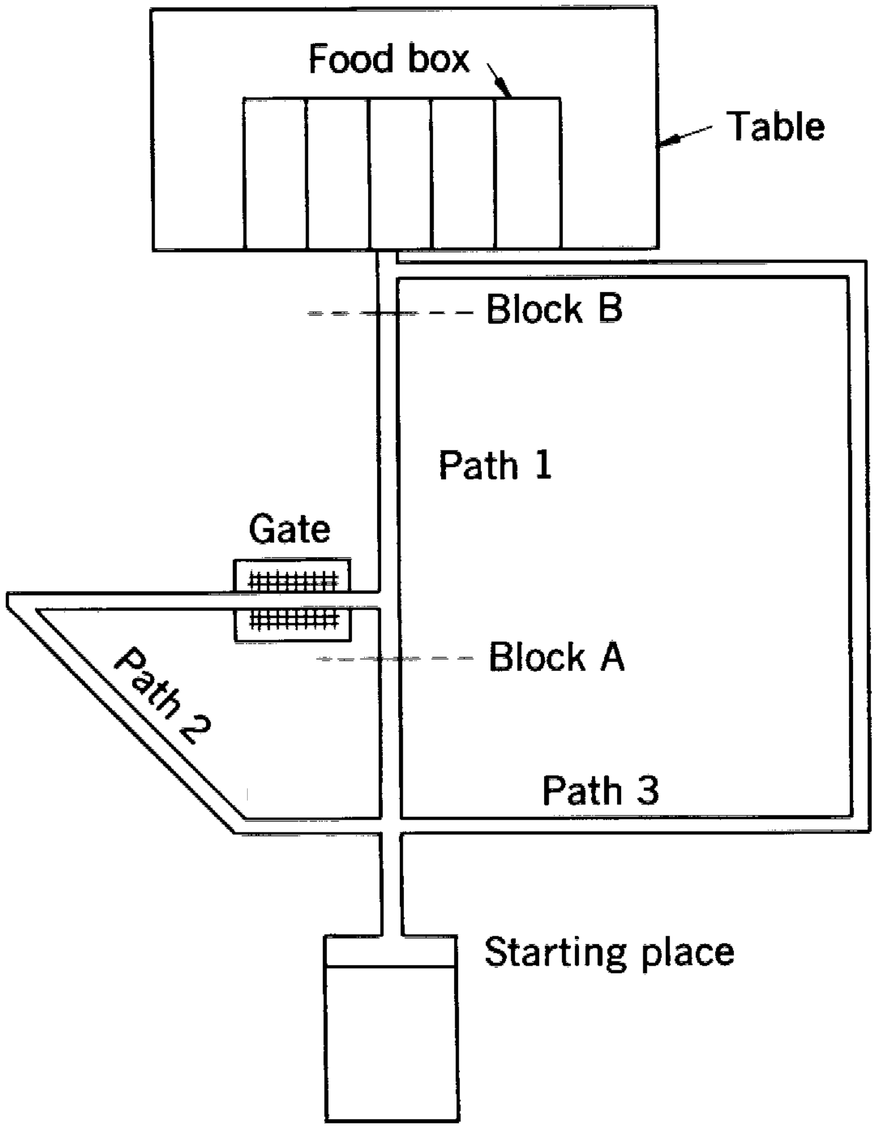
The *contents* of minds are irredeemably complex; we should stop trying to understanding them!

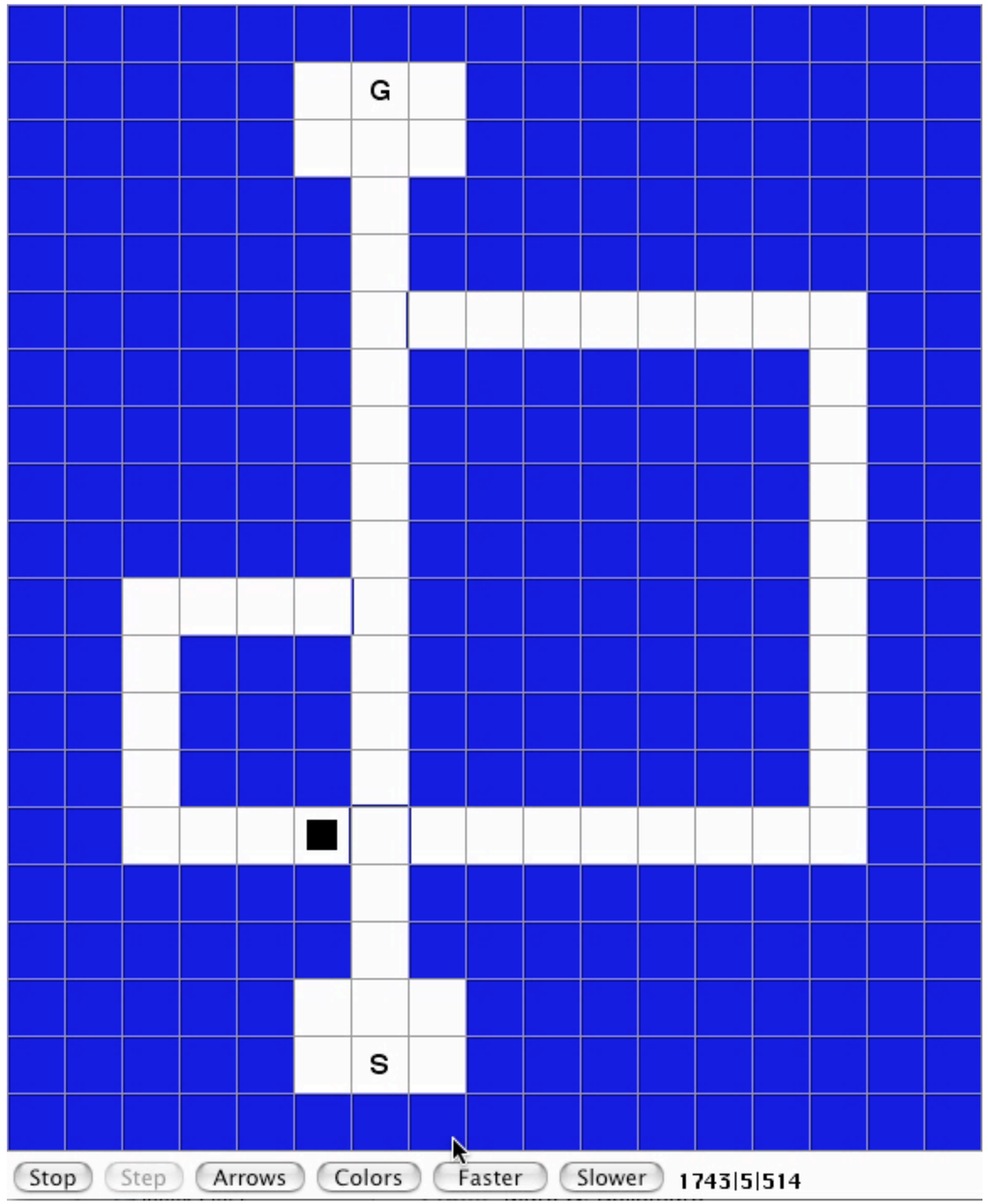
Instead, we should understand the meta-methods for *finding the contents*

Outline

- Toward an Integrated Science of Mind
- The TD theory of Dopamine
- The Brilliant Potential
- It is exciting time in Artificial Intelligence
- The Bitter Lesson
- Some final lighter fare

Tolman & Honzik (1930) "Insight in Rats"



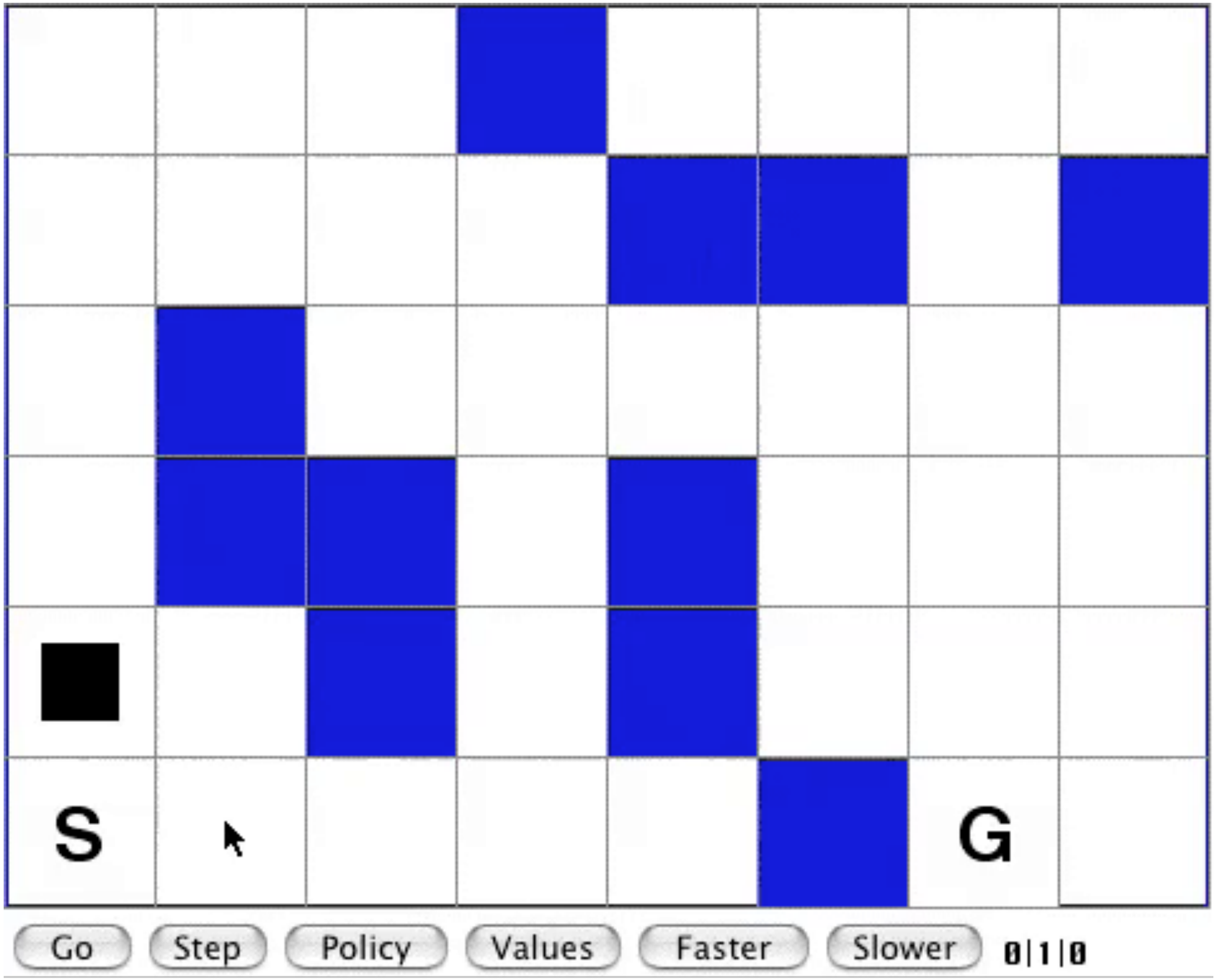


Rat brains appear to process imaginary experience

- Recordings from place cells in the hippocampus appear to reveal what places the rat is 'thinking about'
- At choice points, rats imagine upcoming place sequences
- Imagination also is seen during sleep and at rest times
- Imagination is 6-7 times faster than physical movement
- Paths can be synthesized in imagination that never occurred in reality

Pavlides & Winson 1989; Skaggs & McNaughton 1996; Foster & Wilson 2006; Euston, Tatsuno, & McNaughton 2007; Johnson & Redish, 2007; Gupta, van der Meer & Redish 2010

Example of Dyna, Planning & Learning via Imagination



"Dyna" Sutton, 1990

It is natural to base a Science of Mind on Marr's Three Levels at which any information processing system can be understood

- Computational Theory Level
  - What are the goals of the computation?
  - What is being computed?
  - Why are these the right things to compute?
  - What overall strategy is followed?

What and Why?

- Representation and Algorithm Level
  - How are these things computed?
  - What representation and algorithms are used?

How?

- Hardware Implementation Level
  - How is this implemented physically?

Really how?

Societal implications of advanced AI

- Intelligence Augmentation (IA!) will be a thread of lasting importance
  - a less threatening kind of AI,
    continuous with web search, speech recognition, assistants, user interfaces
- There is no reason to think greater-than-human intelligences are not physically possible
  - They will be economically valuable, and scientifically fascinating
- So I fully expect they will be made, if we don’t destroy ourselves first
  - It will probably be within our lifetimes
  - If we don’t destroy civilization first

In the long run…

- AI technology will be part of what disrupts existing social and power structures
  - AIs will force us to re-examine our moral and social foundations
  - Continuing trends that are 1000s of years old
- AI will bring greater diversities of intelligences, both natural and artificial
  - There will be biases against the new and different. There will be feelings of entitlement
  - These will be counterproductive and *eventually* fade away
- Universal Basic Income sounds like a terrible idea to me
- AI soldiers/weaponry sounds like a terrible idea to me
- Will we welcome independent AIs? Or force them to be outlaws?

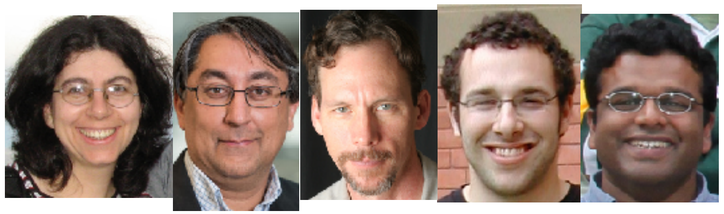
Conclusion

- An Integrated Science of Mind would be a historic achievement
- It is in reach now in every sense that matters
- We can see parts of it in some Reinforcement Learning ideas
- There is strikingly rapid recent progress in Artificial Intelligence, which also makes an ISoM seem potentially imminent
- The *contents* of minds are irredeemably complex;
  we should stop trying to understanding them directly

Thank you for your attention

Join us at the 4th Multidisciplinary Conference on Reinforcement Learning and Decision Making (RLDM) on June, 2019, in Montreal, Quebec

thanks to Doina Precup, Satinder Singh, Mark Ring, Adam White, and Joseph Modayil



Further reading

- For foundations: See the 2nd edition of the RL textbook, by Andy Barto and myself, available free on the internet. Includes TD learning and planning by imagination. Includes multi-disciplinary links to psychology, neuroscience, and control

- For General Value Functions: See “Horde: A scalable real-time architecture for learning knowledge from unsupervised sensorimotor interaction”, AAMAS-2011, and *Adaptive Behavior 22*(2):146-160

- For the general philosophy: See “Beyond reward: The problem of knowledge and data”, ILP 2011, “The grand challenge of predictive empirical abstract knowledge”, IJCAI-09 workshop, and the incomplete ideas blog at richsutton.com