COMPUTATIONAL APPLICATIONS TO POLICY AND STRATEGY (CAPS)

Session 3 – Building a Learning-Based StarCraft II Bot

Leo Klenner

Outline

1. Recap
2. Machine Learning in StarCraft II
3. Two Types of Machine Learning
4. Testing PySC2
5. Looking Ahead for StarCraft II
6. Further Applications of AI in IR

Recap – Rule-Based StarCraft Bots

- Chains of if-then-else logic based on expert domain knowledge
- CapsBot
  - defeats SC2 built-in AI on `hard`
- TerranBioRush
  - defeats SC2 built-in AI for Zerg and Protoss on `very hard`
  - correction: higher levels of difficulty de facto do not require a non-linear change in strategy
  - defeating Terran on `very hard` requires smart sequencing as that AI plays similar bio rush
- Performance of rule-based agents can be high but robustness remains low

From Rule-Based to Learning Agents

I definitely think the approach [reinforcement learning] itself is very scalable. If you build the bot in the 2010 way that we did at Berkeley [expert system], the bot will do one build order, or two or three, but it doesn't scale very much. At the end of the day, someone can study how it plays, and expose weaknesses. What I like about our approach is that, if it all works out, the agent has learned a very large variety of tactics and counters that couldn't possibly be programmed, in the same way you couldn't program a very good Go player.
- Oriol Vinyals, Research Scientist, DeepMind, Blizzard Interview (3/4/2018)

Enter SC2LE

Screenshot from DeepMind's pysc2 API

Core Research on Learning in StarCraft II

- O. Vinyals, et al. 2017. StarCraft II: A New Challenge for Reinforcement Learning
  - introduces SC2LE, an environment for training reinforcement learning (RL) agents
  - SC2LE provides access to the full game environment and mini games
  - provides baseline results of RL agents that achieve novice human play on mini games

- V. Zambaldi. 2018. Relational Reinforcement Learning
  - augments earlier RL agents of DeepMindwith relational reasoning between entities
  - improves the agents‘ navigation and planning, yield new mini games baselines

Interim Conclusions

“I think the main thing we can confirm is that the learning approach is indeed hard.”

- Timo Ewalds, Software Engineer, DeepMind, Discord (9/20/2018)

StarCraft II – Environment

Properties of the SCII environment	
Type	Real-time strategy (RTS) game
Gameplay	Fast paced micro-actions and need for high-level planning
Action space*	10^8, need for hierarchical actions
Environment states**	$10^{1,685}$
Rule of transition between states	Continuous, based on the agents' actions
Rewards assigned to each state	Unknown
Reward horizon	Long pay-off = strats more important than micro
Mode of information	Fog of war = imperfect information
Mode of action	Simultaneous

* Vinyals, O., et al. 2017. *StarCraft II: A New Challenge for Reinforcement Learning*. https://arxiv.org/abs/1708.04782
** Estimated for StarCraft Brood Wars. Usunier, N., et al. 2016. *Episodic Exploration for Deep Deterministic Policies: An Application to StarCraft Micromanagement Tasks.* https://arxiv.org/abs/1609.02993

Learning in StarCraft II

- Different degrees of learning:
  - Hybrid – partially rule-based, partially learning-based
  - Full learning but structured specifically on StarCraft II
  - Full learning without specific StarCraft II structures

TenCent's Hybrid Approach to the Full Game

- Peng Sun et al. 2018. TStarBots: Defeating the Cheating Level Builtin AI in StarCraft II in the Full Game

How?
- Rule-based aspect 1: hard-coded dependency rules of SCII (i.e. unit **z** requires building **y**, building **y** requires building **x**) are encoded in the learning algorithm
- Rule-based aspect 2: high number of decisions in SCII are trivial (i.e. which worker of workers **w** builds **x**) and unnecessarily reduce learning speed, hence hard-coded embedding

What?
- Learning-based aspects: strategic dimensions such as what to build, when to attack etc.

Two Types of Machine Learning

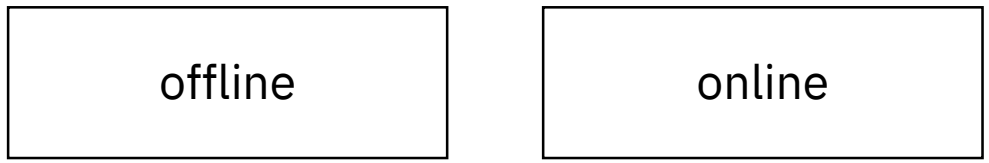
- Supervised Learning (v. unsupervised learning)
  - Strong prior knowledge about output of model (v. no prior knoweldge)

- Reinforcement Learning
  - Sparse prior knowledge about output of model
  - Reinforcement learning is conceptualized between supervised and unsupervised learning

Example of a Machine Learning Problem

- Given data, make statistical model of how player win-rate in StarCraft depends on APMs, number of workers, number of army units, timing of attacks,...
  - Linear regression (simple ML)
  - Modeling a target value based on independent predictors

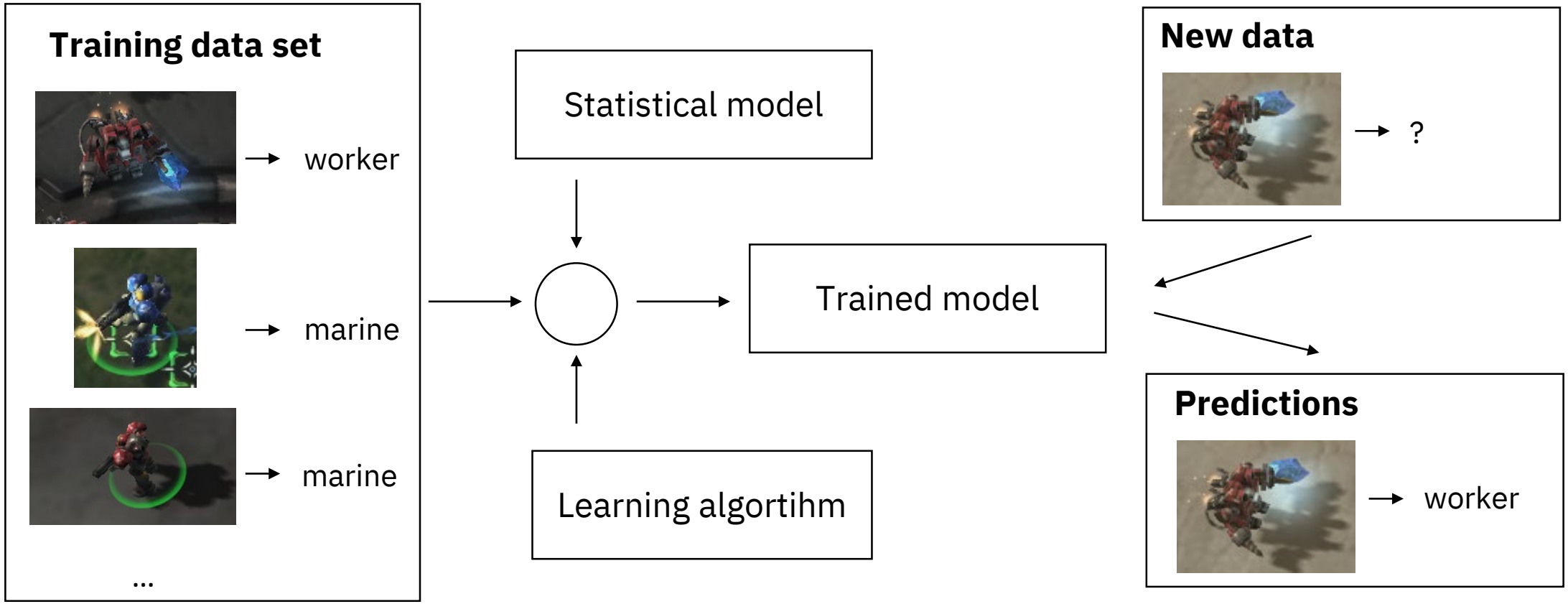
Learn from data to make predictions that can generalize to other data



Supervised Learning

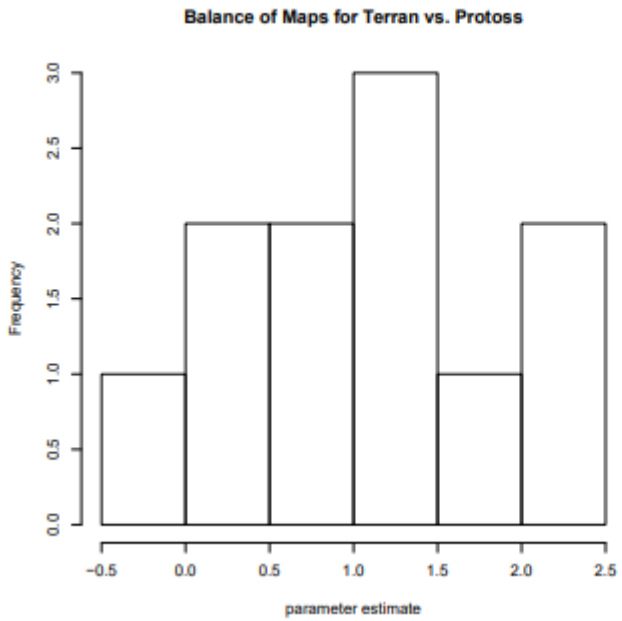
> Learning a mapping from inputs -> outputs

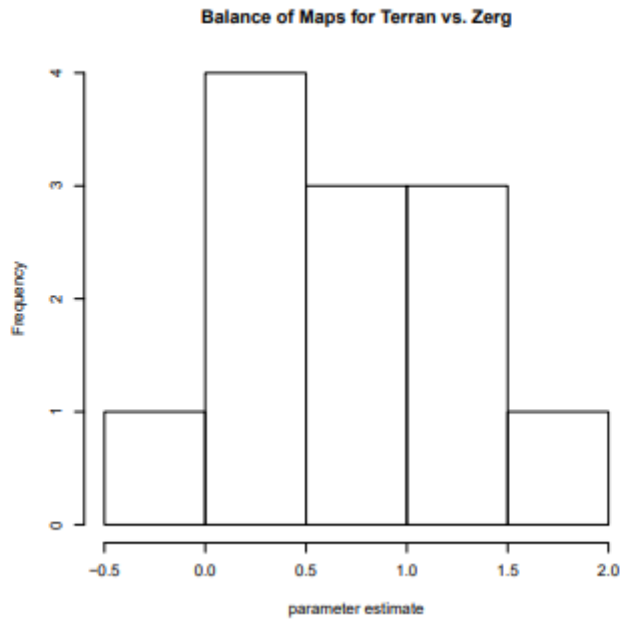
> Example – classification, based on labeled training data set

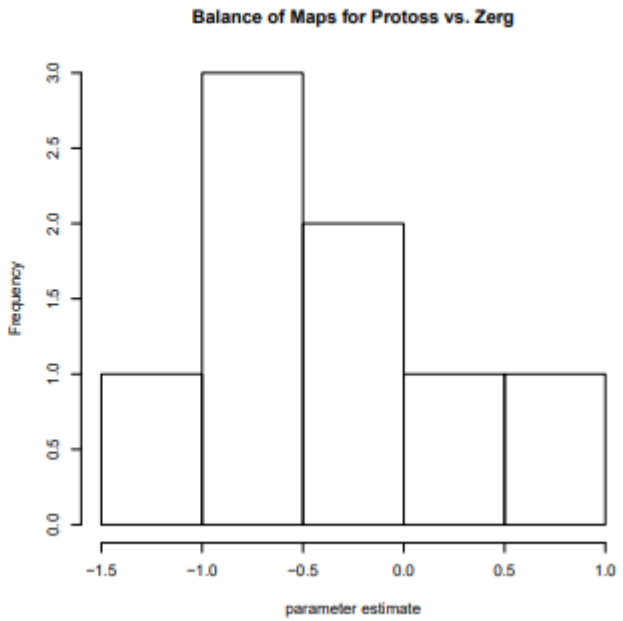


Applications of Supervised Learning to StarCraft

- H. Yun. 2018. Using Logistic Regression to Analyze the Balance of a Game – The Case of StarCraft II.
  - Different maps are balanced towards different races (T, P, Z)

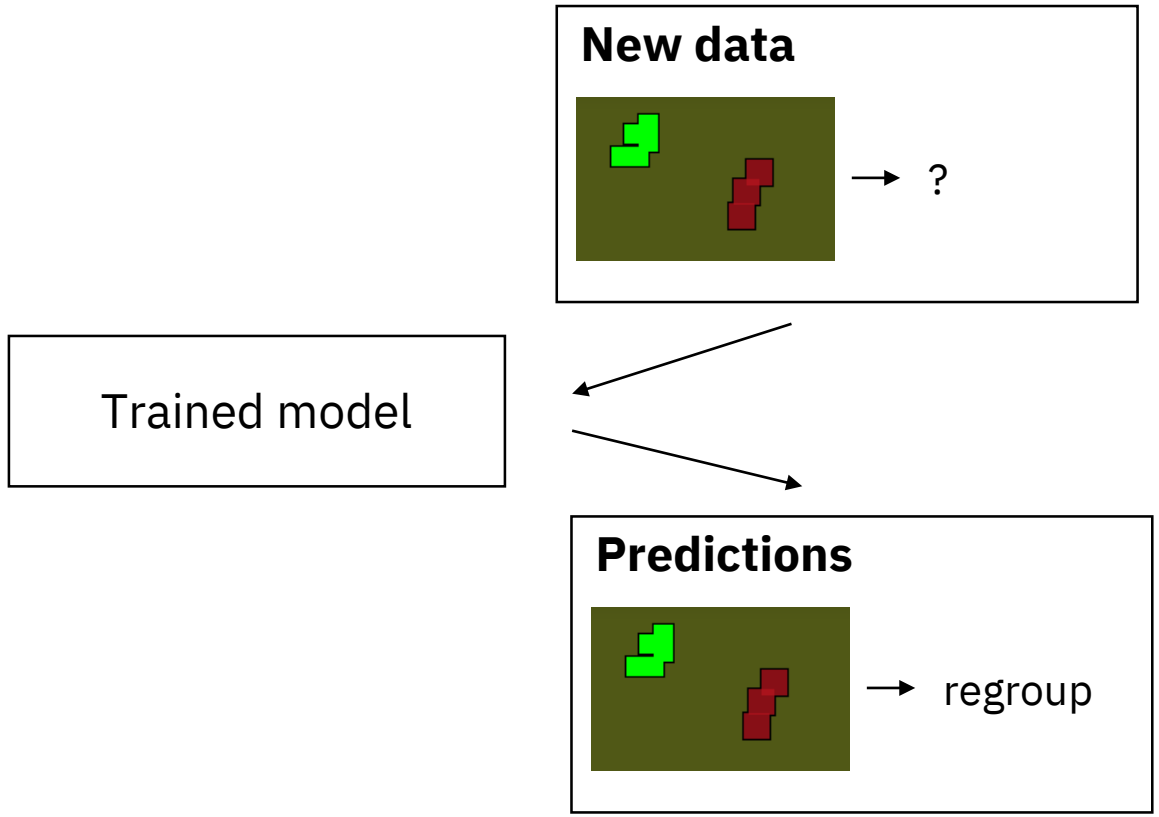






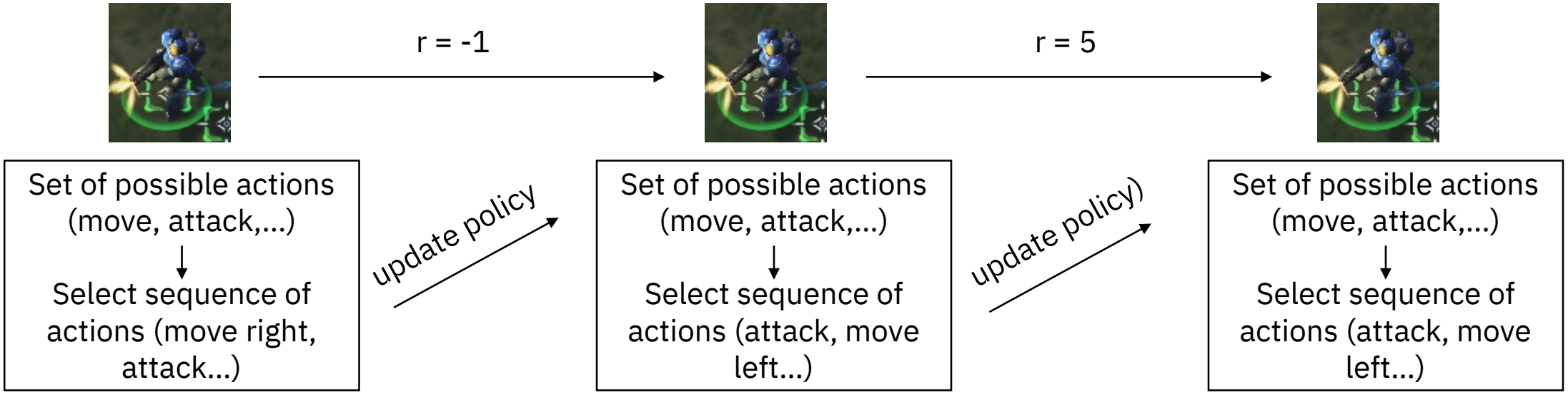
Supervised Learning for Agent‘s Decisionmaking

- Assume data set of StarCraft replays
- Map input (state of the game) to output labels (optimal action in that state)
- Is this feasible? What are trade-offs?



Reinforcement Learning

- Reiforcement learning works through trial-and-error based on specified reward function
  - One label for each training example => supervised learning
  - Sparse and delayed labels (rewards) => reinforcement learning



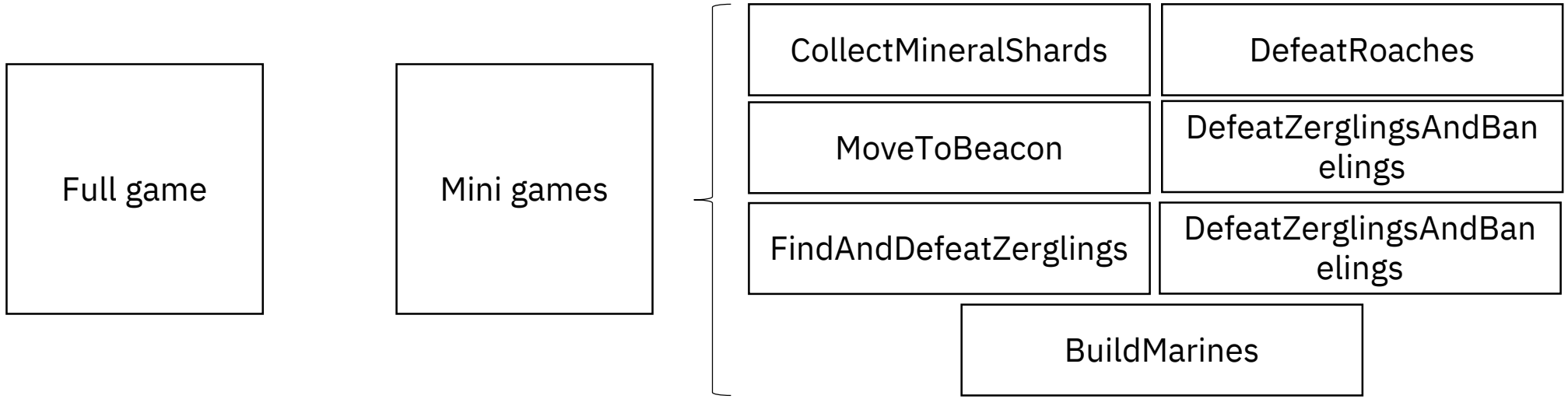
Challenges of Reinforcement Learning

- Credit assignment problem
  - which of the preceeding actions was responsible for getting the reward and to what extent?
  - especially difficult when reward is delayed on long horizon and binary (win game, lose game)
- Explore-exploit dilemma
  - should you exploit the known working strategy or explore other, possibly better strategies?
- Sparse reward signal problem
  - if no reward signal can be picked up, how can you learn?

Applications of Reinforcement Learning to StarCraft

- RL powers most of the learning-based StarCraft bots based on SC2LE
  - wide range of RL algortihms (DQN, AC2, ...) that differ broadly in how they evaluate optimal state-action pairs

- SC2LE provides different environments to train agents



Evaluation of Minimap Agents

> Compare scripted, random and DQN/AC2 agents on three mini games

> Implementations for DQN/AC2 from https://github.com/rayheberer/SC2Agents

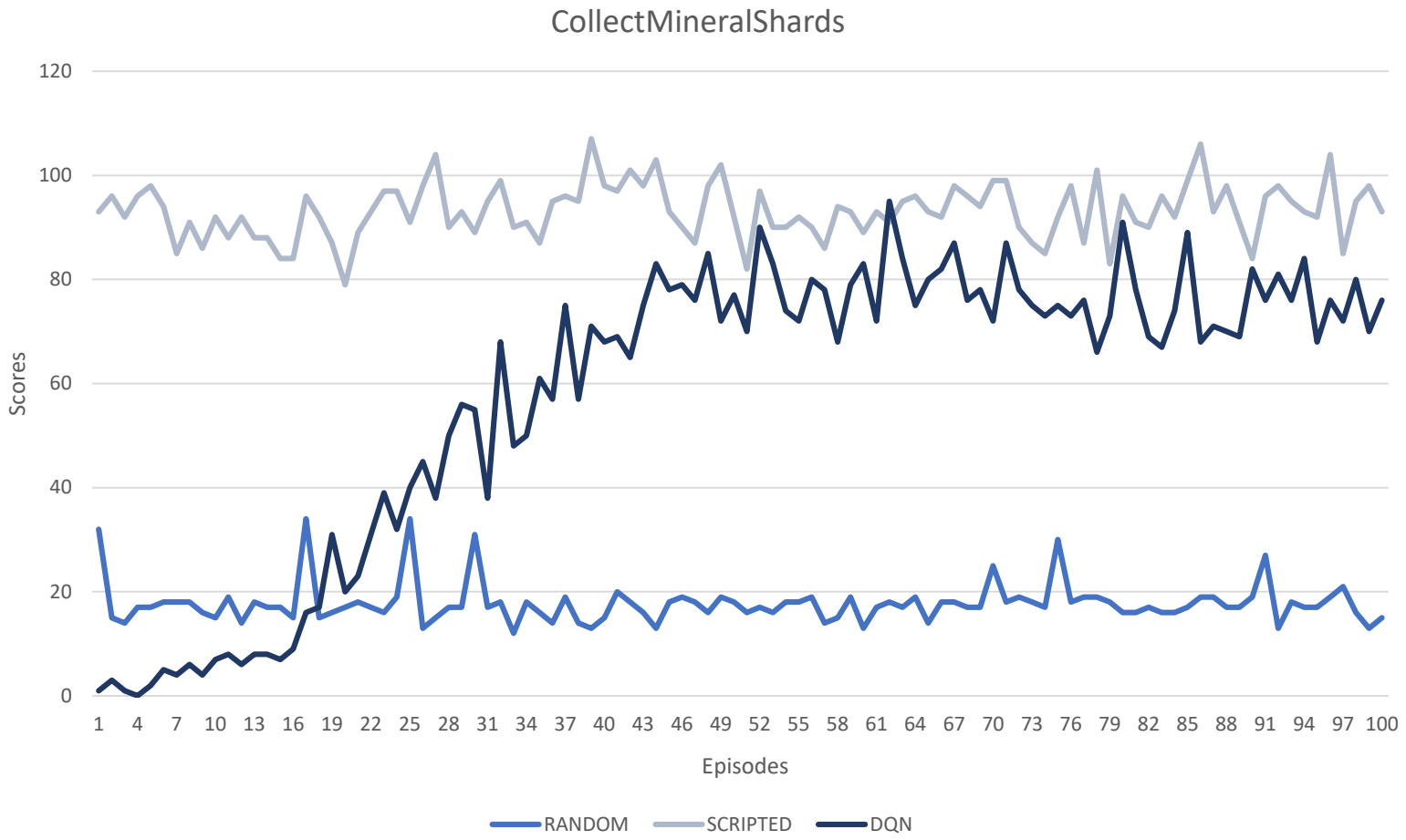
CollectMineralShards

MoveToBeacon

DefeatRoaches

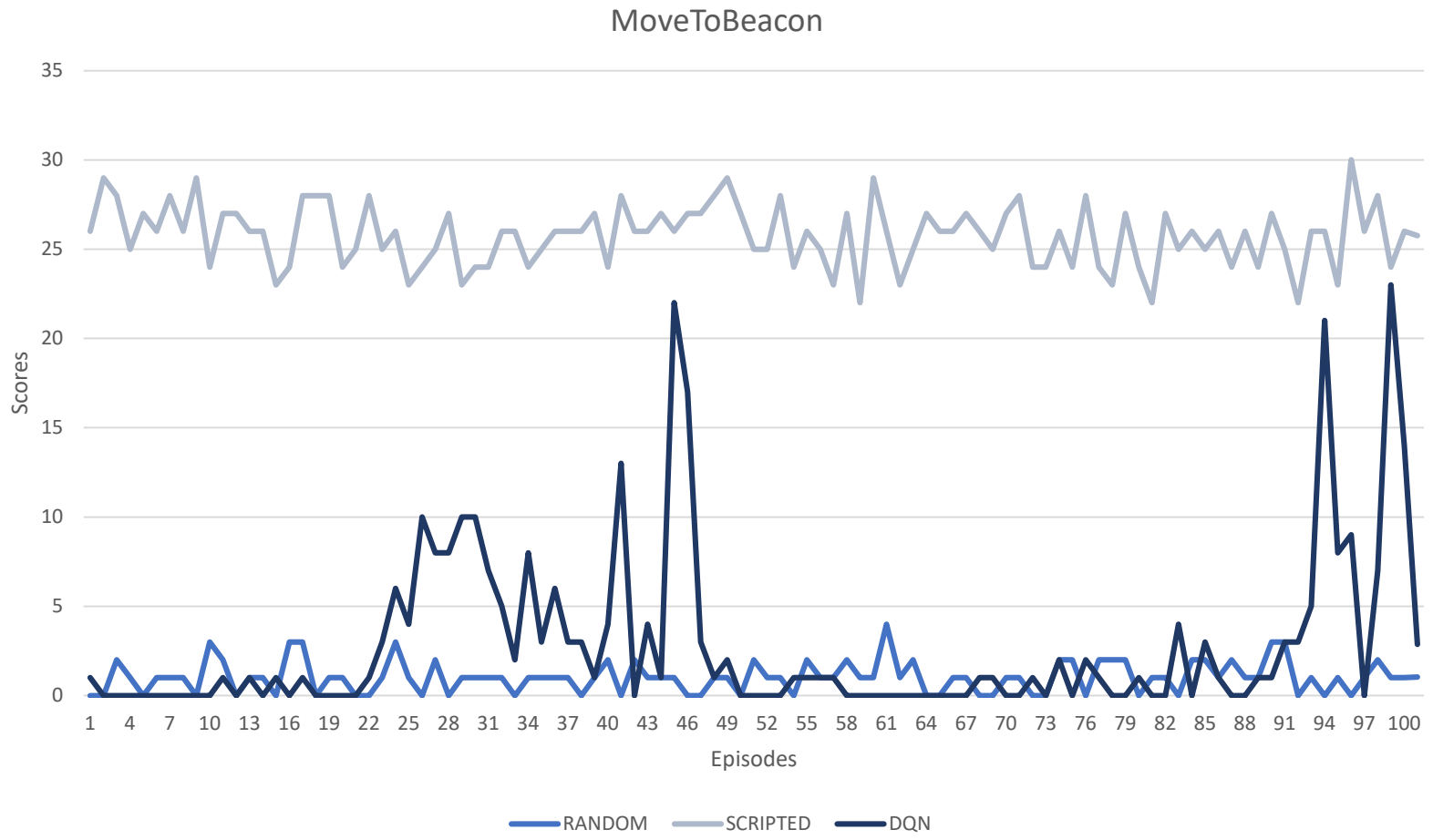
CollectMineralShards

> 2 marines, 20 mineral shards (endless supply), r = +1 for each shard collected



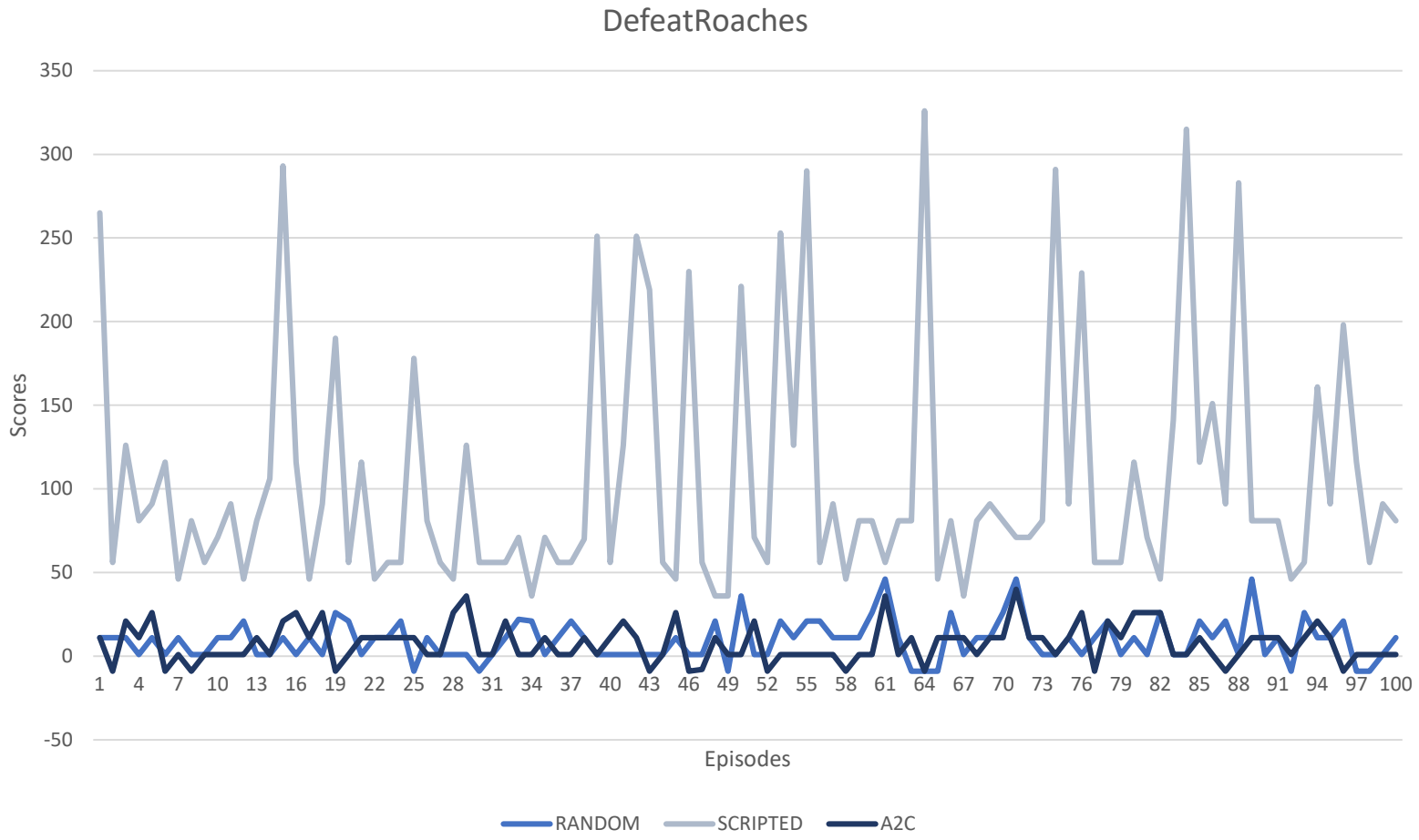
MoveToBeacon

> 1 marines, 1 beacon , r = +1 for reaching the beacon



DefeatRoaches

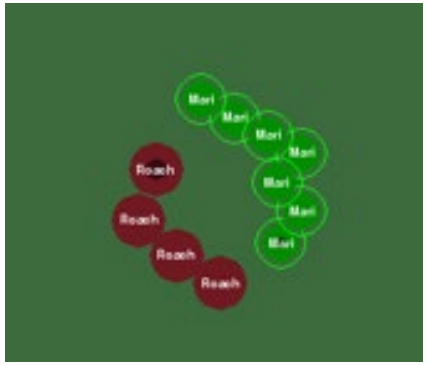
> 9 marines, 4 roaches , r = +10 for defeating roach, r = -1 for losing marine

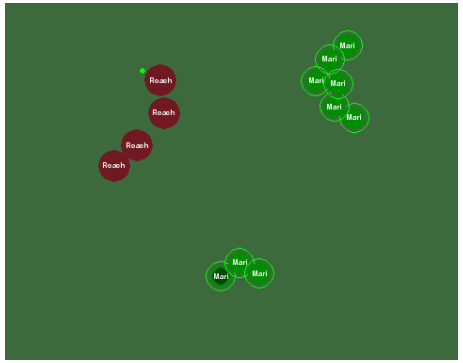


Unit Formation for A2C on DefeatRoaches

> Grouping of units is important for combat effectiveness

> Marines should not be split up but aligned to focus fire (scripted agent)





Evaluating our Agents‘ Performance

Mean

	CollectMineralShards	**MoveToBeacon**	**DefeatRoaches**
Random	17.76	1.03	9.31
Scripted	93.13	25.76	105.6
DQN	57.65	2.86	na
AC2	na	na	7.45

Max

	CollectMineralShards	**MoveToBeacon**	**DefeatRoaches**
Random	34	4	46
Scripted	107	30	326
DQN	95	23	na
AC2	na	na	40

DeepMind's Performance

	Mini-game						
Agent	(1)	(2)	(3)	(4)	(5)	(6)	(7)
DeepMind Human Player [15]	26	133	46	41	729	6880	138
StarCraft Grandmaster [15]	28	177	61	215	727	7566	133
Random Policy [15]	1	17	4	1	23	12	< 1
FullyConv LSTM [15]	26	104	44	98	96	3351	6
PBT-A3C [33]	–	101	50	132	125	3345	0
Relational agent	**27**	**196** ↑	**62** ↑	**303** ↑	**736** ↑	4906	**123**
Control agent	**27**	187 ↑	61	295 ↑	602	**5055**	120

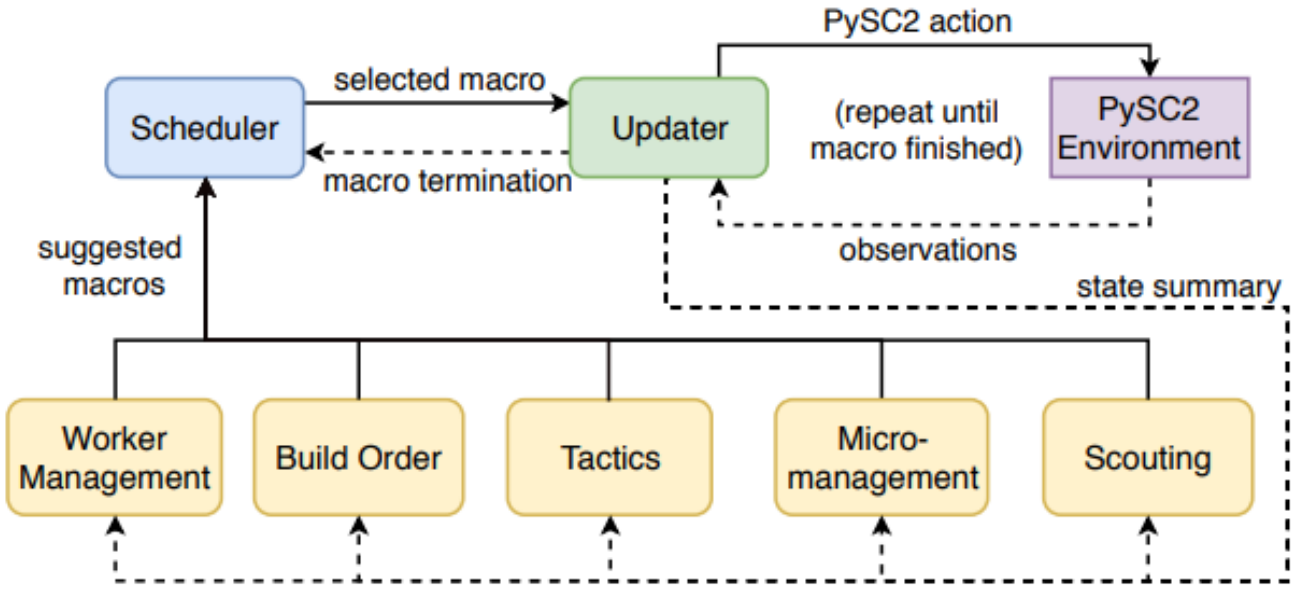
Table 1: Mean scores achieved in the StarCraft II mini-games using full action set. ↑ denotes a score that is higher than a StarCraft Grandmaster. Mini-games: (1) Move To Beacon, (2) Collect Mineral Shards, (3) Find And Defeat Zerglings, (4) Defeat Roaches, (5) Defeat Zerglings And Banelings, (6) Collect Minerals And Gas, (7) Build Marines.

Looking Ahead for StarCraft II

> Hybrid approaches open up the full game

> Gradual transition from hybrid to pure RL

> D. Lee, et al. 2018. Modular Architecture for StarCraft II with Deep Reinforcement Learning



Further Applications of RL in IR

- S. van der Hoog. 2017. Deep Learning in (and of) Agent-Based Models: A Prospectus
  - Reinforcement learning as policy design
  - "A government or central bank agent may be given certain goals (such as a stable price level, low unemployment rates, or macrofinancial stability), rather than hand-crafted rules. Using reinforcement learning techniques, an agent starts with little knowledge of the world, but given a reward function that models those goals, the agent learns to perform better over time."
  - "This may lead to more flexible policies and more adaptive behavior on the part of the policy agent, as it allows for more flexible, discretionary policy setting behavior, rather than using a fixed rule-based policy. As the policy agent learns how to set policies optimally, it must adapt to the behavioral changes of the other agents, who might change their behavior in response to the policy."

- In the context of RL techniques, what does it mean to have a rules-based international system?

Summary

- We reviewed supervised learning and reinforcement learning as examples of machine learning techniques
- Reinforcement learning is a powerful tool to build agents capable of decisionmaking but faces intrinsic challenges
- These challenges (credit assignment, explore-exploit, sparse reward signal) are all present in StarCraft II and are amplified by the game‘s complexity
- Different mini games allow us to test and evaluate agents in simplified environments, which cannot be scaled to full game complexity
- RL-based agents performance is heavily dependent on the type of environment, especially given the short training of 100 iterations
- Learned decisionmaking is fundamentally different from rule-based decisionmaking and brings new applications to the international arena