Reinforcement-Learning

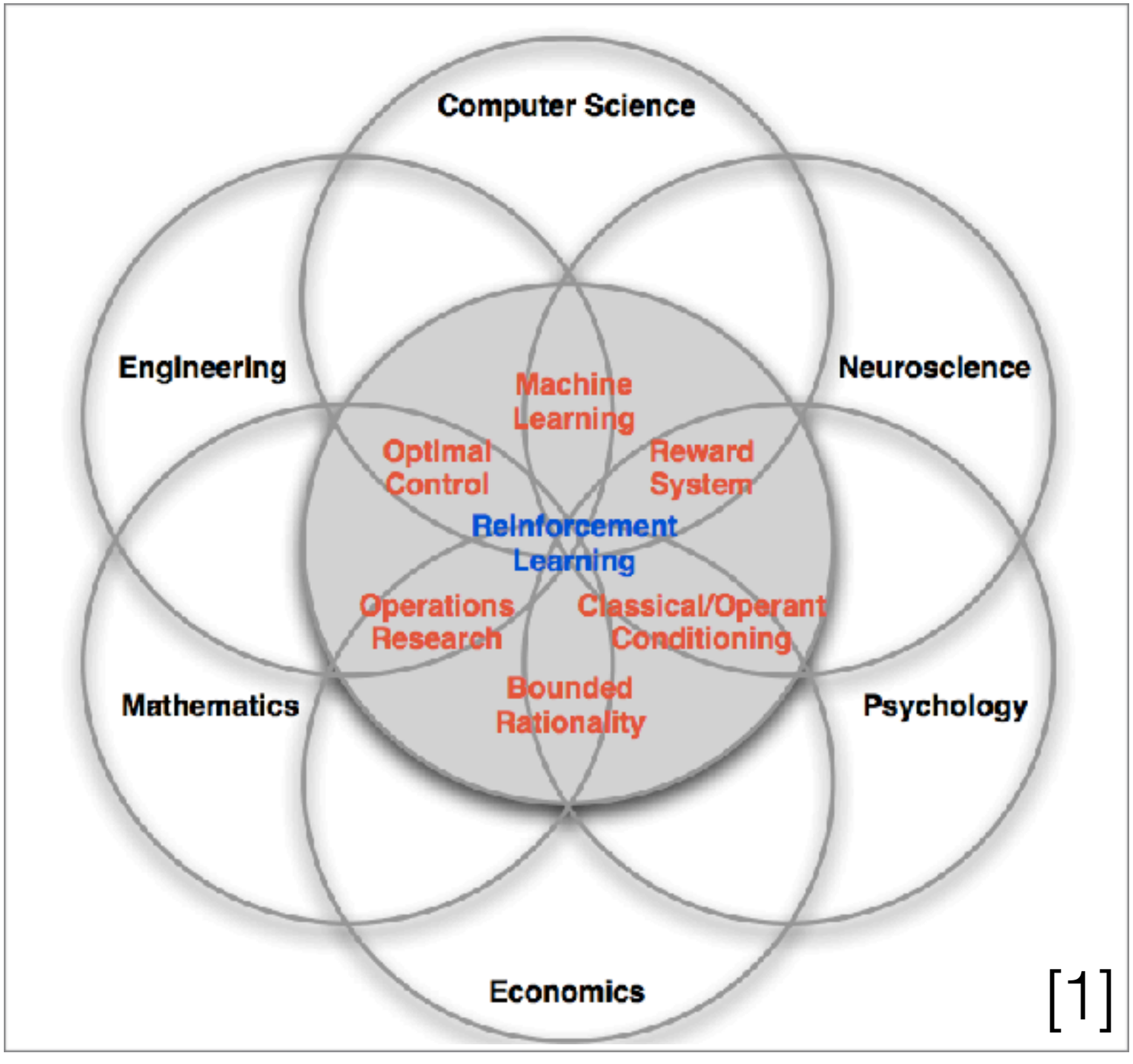
Vortrag von: Fabien Lapok

Betreuer: Prof. Dr. Meisel

Agenda

- Motivation
- Überblick und Probleme von RL
- Aktuelle Forschung
- Mein Vorgehen
- Konferenzen und Quellen

Reinforcement Learning (RL) - die Wissenschaft der Entscheidungsfindung

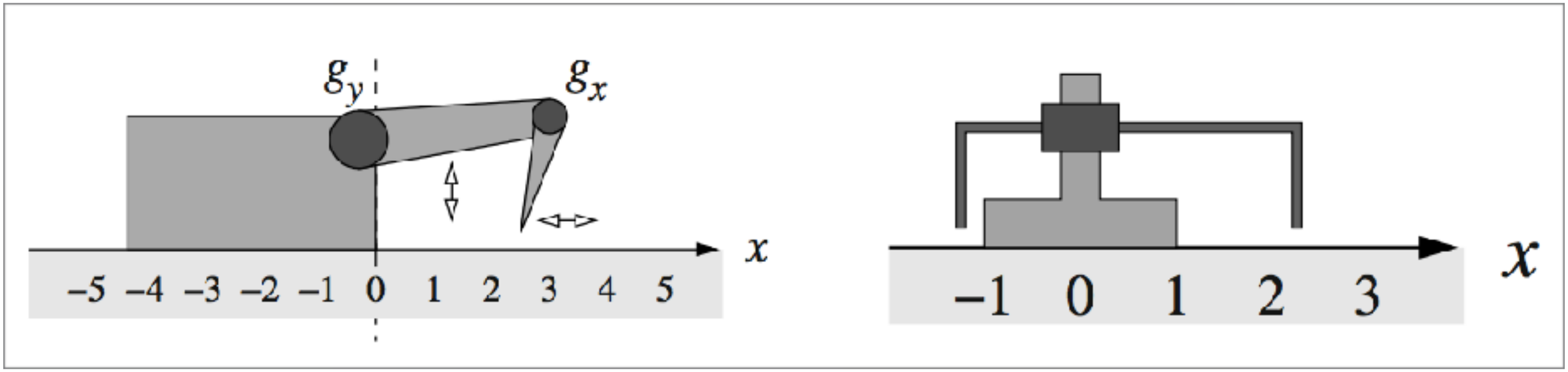


RL zum Lösen von Problemen mit komplexen Zustandsräumen

- Spiele spielen (Atari, Poker, Go ..)
- Mit Nutzern im Internet interagieren
- Steuerung physikalischer Systeme (z.B. Roboter, die Laufen lernen; Drohnen, die Fliegen lernen …)

[Ertel, 2008]

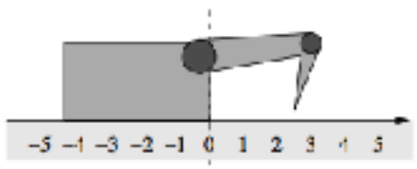
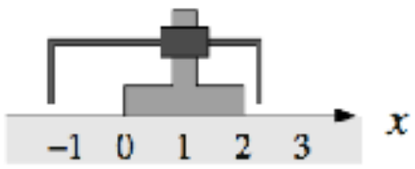
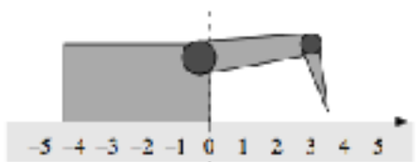
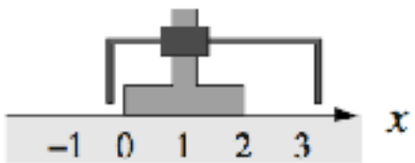
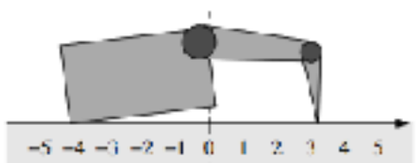
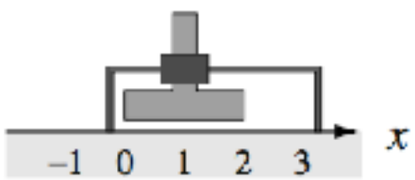
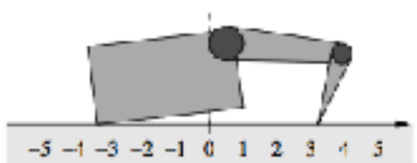
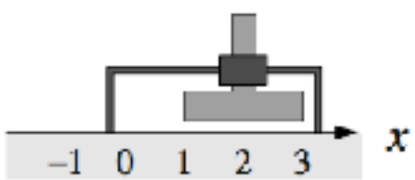
RL am Beispiel eines Krabbelroboters (I)



[Ertel, 2008]

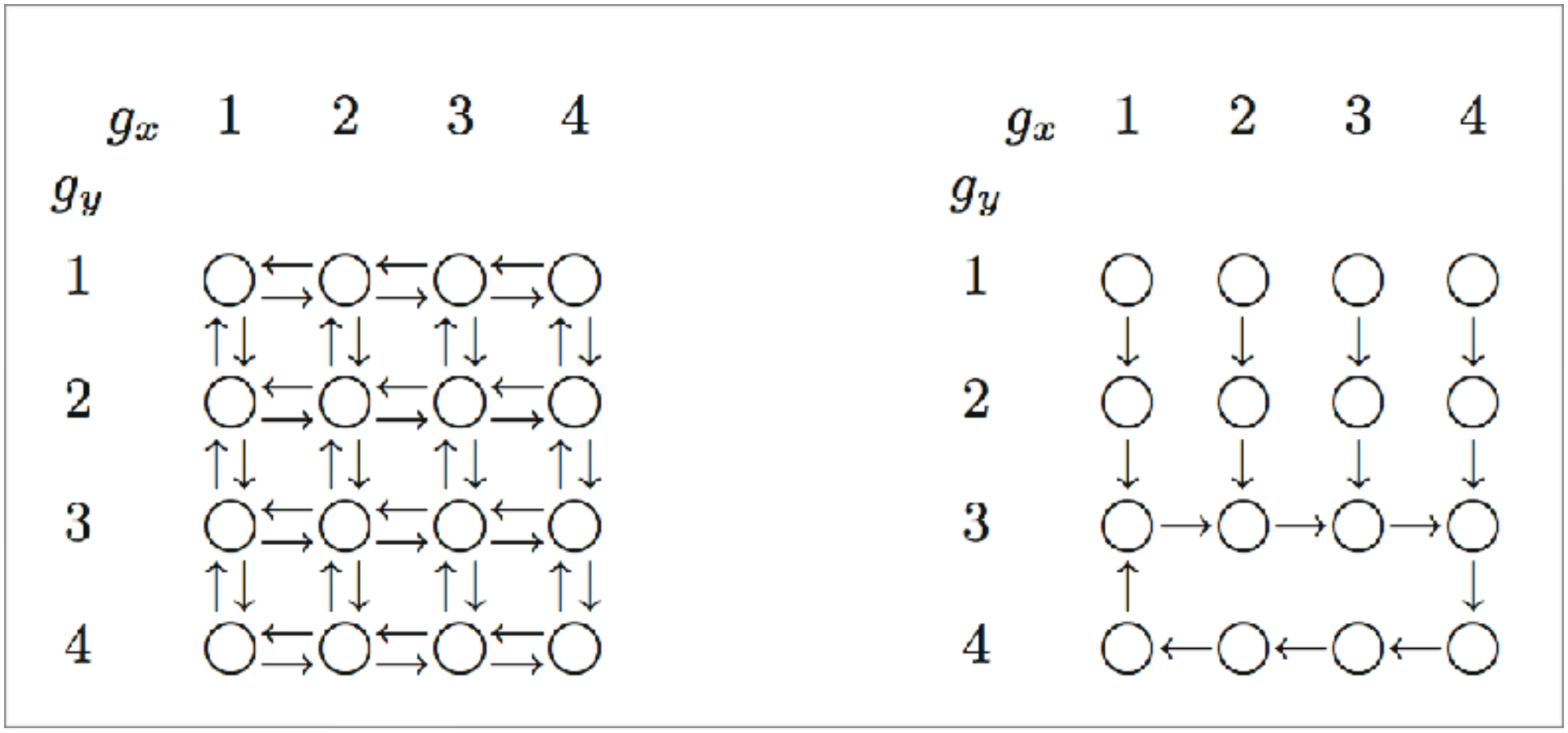
RL am Beispiel eines Krabbelroboters (II)

Ein Zyklus einer periodischen Bewegungsfolge mit systematischer Vorwärtsbewegung.

Krabbelroboter	Laufroboter	Zeit t	Zustand g_y	Zustand g_x	Belohnung x	Aktion a_t
		0	oben	links	0	rechts
		1	oben	rechts	0	tief
		2	unten	rechts	0	links
		3	unten	links	1	hoch

[Ertel, 2008]

RL am Beispiel eines Krabbelroboters (III)

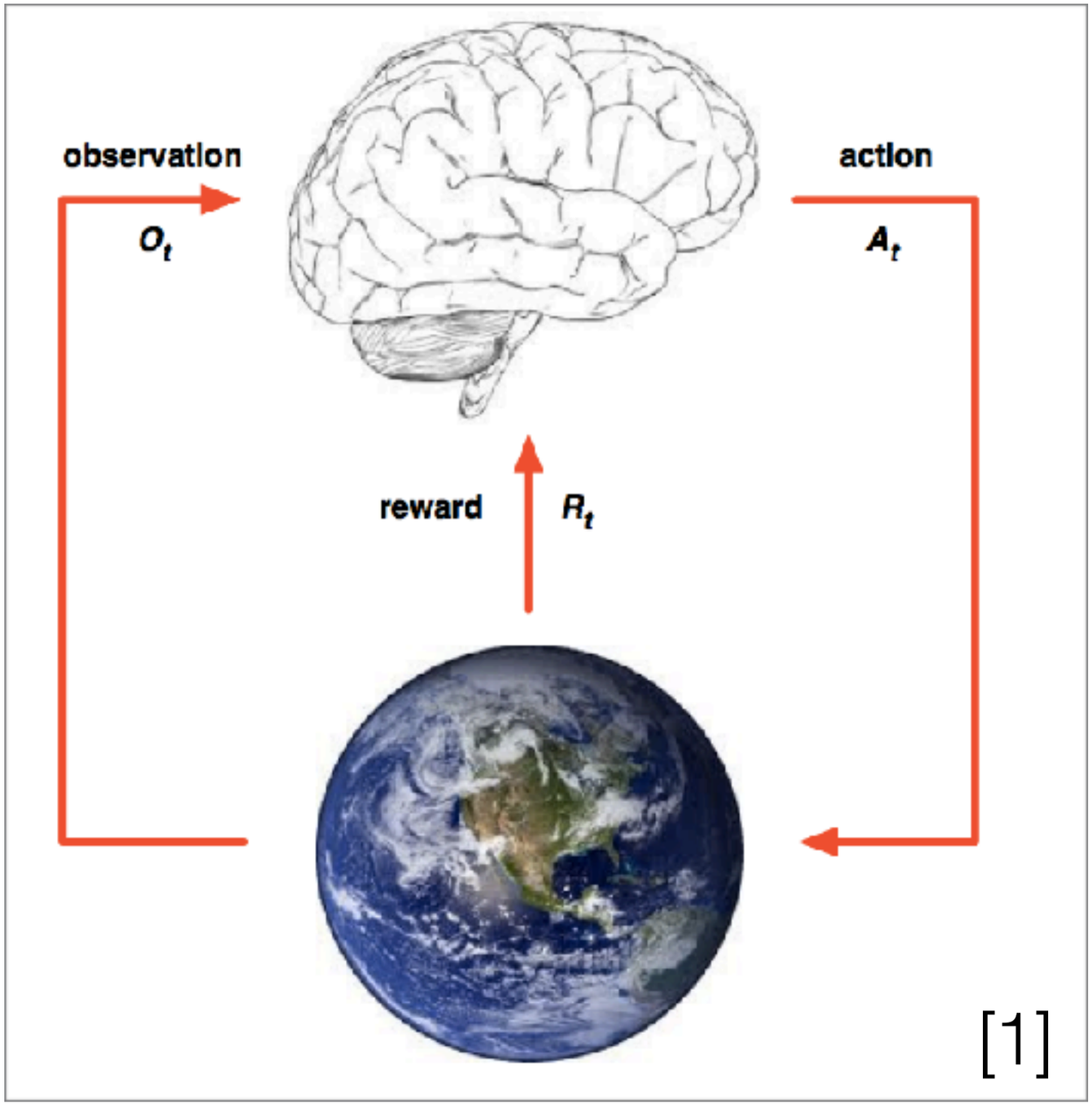


[Ertel, 2008]

RL - Episodisches Lernen durch Versuch und Irrtum

- Anders als andere Lernverfahren, kein überwachtes Lernen

- Durch Versuch und Irrtum lernt das System welche Aktionen in einer Situation gut bzw. schlecht sind



Grenzen von Reinforcement Learning

- Bei zu großen Zustandsräumen kommt das Verfahren zu keinem Ergebnis (”curse of dimensionality”)
- Systeme mit spärlichem und spätem Feedback sind eine Herausforderung für RL Algorithmen

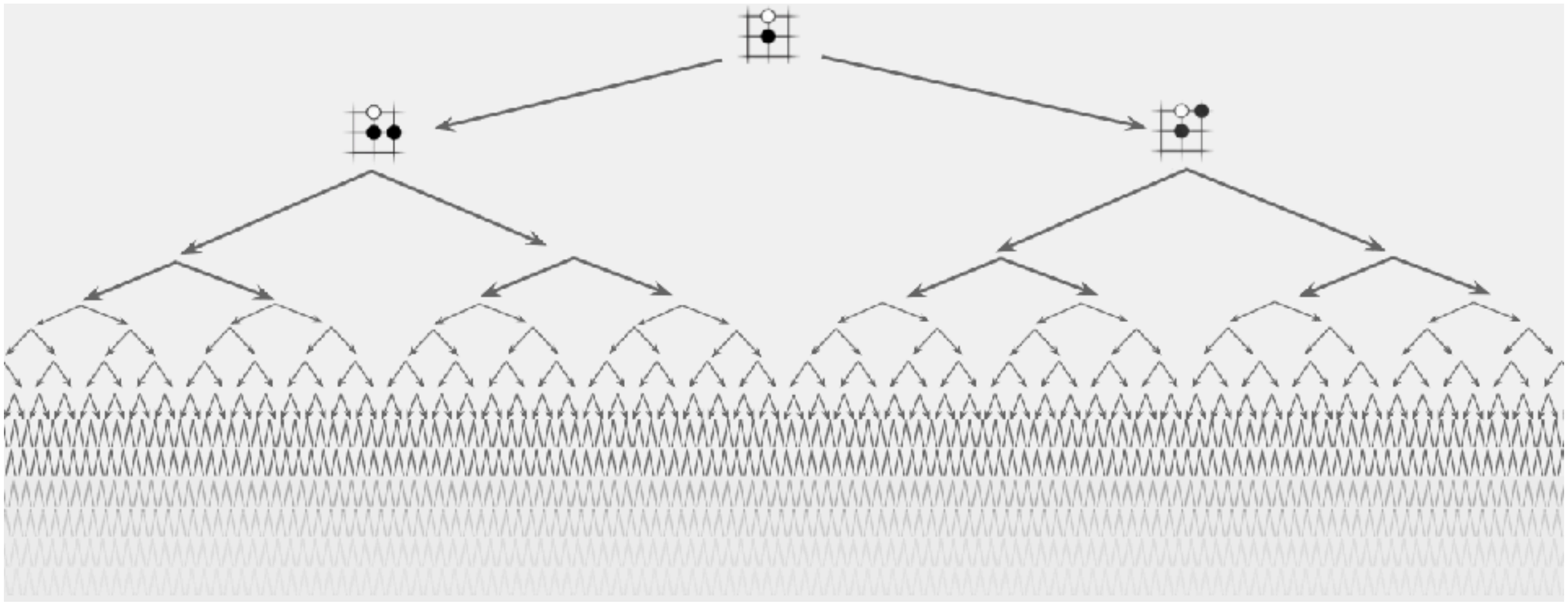
Lösungsansätze in der aktuelle Forschung

- Deep Reinforcement Learning als Ansatz zum Bewältigen von großen Zustandsräumen
  - Paper: Mastering the Game of Go with Deep Neural Networks and Tree Search [Silver, Huang et al, 2016]
- Hierarchisches Reinforcement Learning für den Umgang mit spätem Feedback
  - Paper: Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation [Kulkarni, Narasimhan et al, 2016]

Mastering the Game of Go with Deep Neural Networks and Tree Search [Silver, Huang et al, 2016]

Mastering the Game of Go with Deep Neural Networks and Tree Search (I)

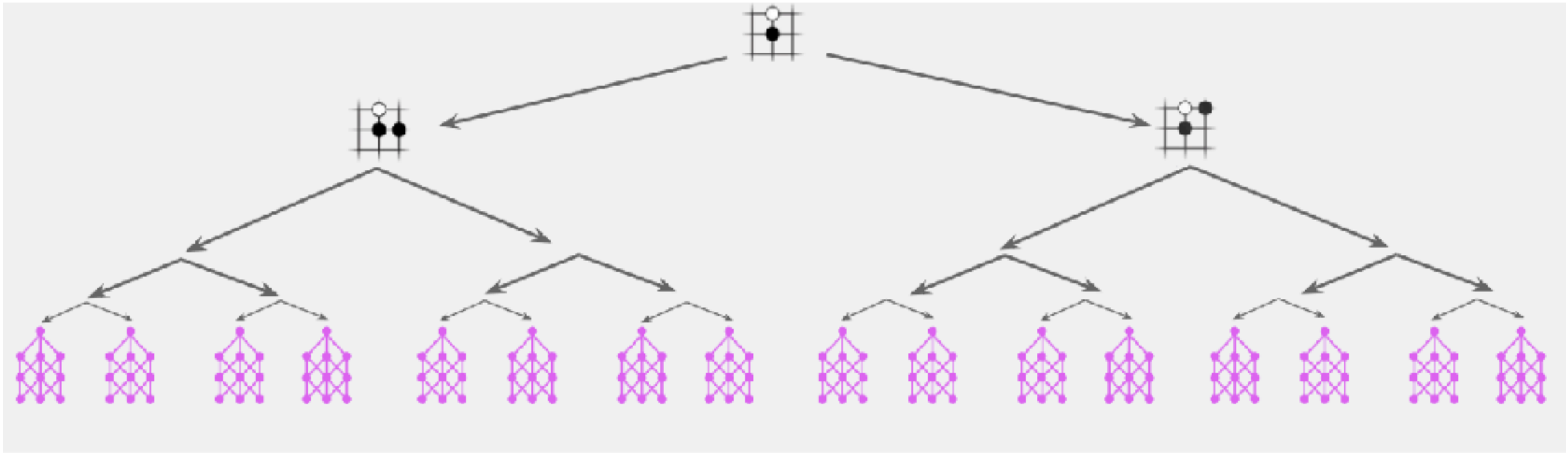
Beispielhafter Suchraum



[2]

Mastering the Game of Go with Deep Neural Networks and Tree Search (II)

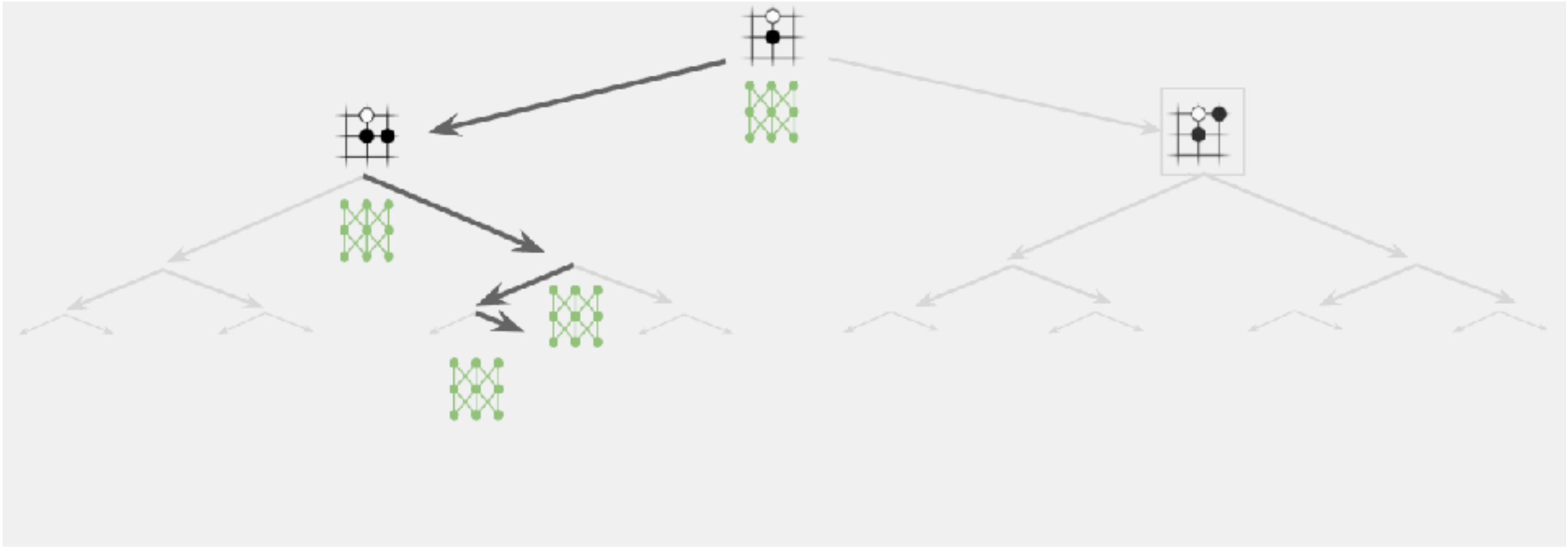
Minimieren der Tiefe des Suchraums mithilfe einer Approximationsfunktion



[2]

Mastering the Game of Go with Deep Neural Networks and Tree Search (III)

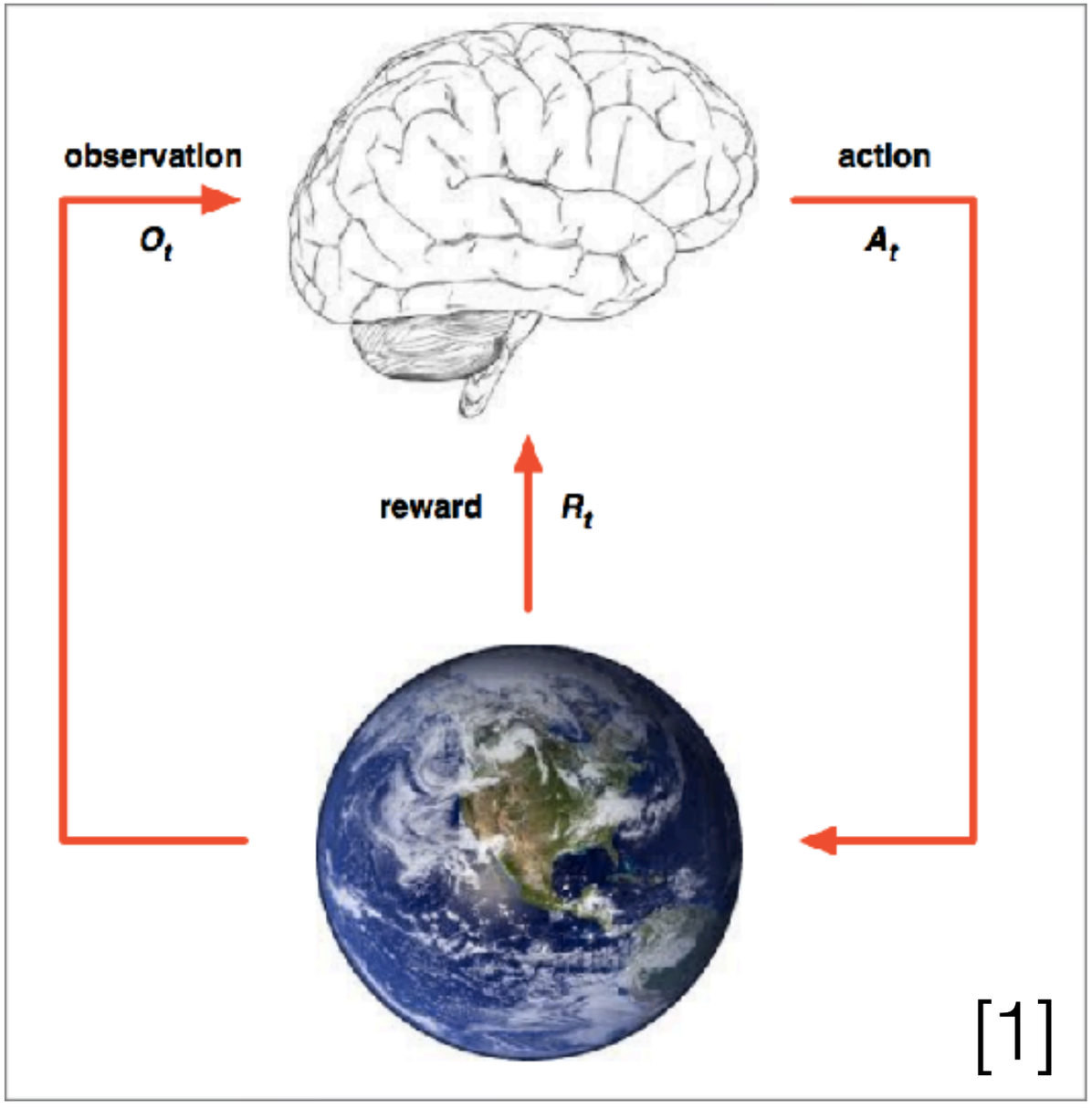
Minimieren der Entscheidungsmöglichkeiten mithilfe einer Policy-Funktion



[2]

Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation [Kulkarni, Narasimhan et al, 2016]

Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation (I)

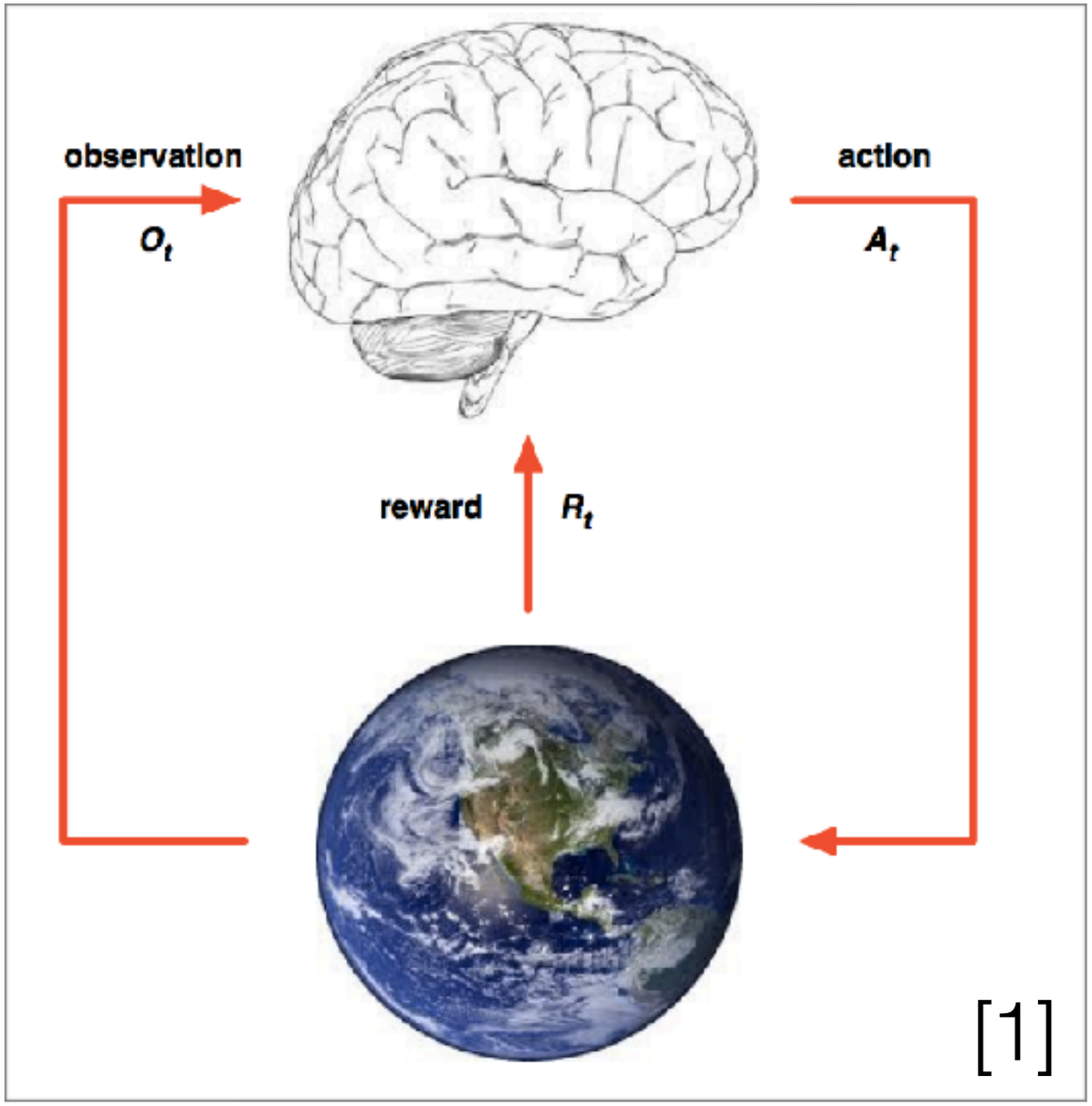


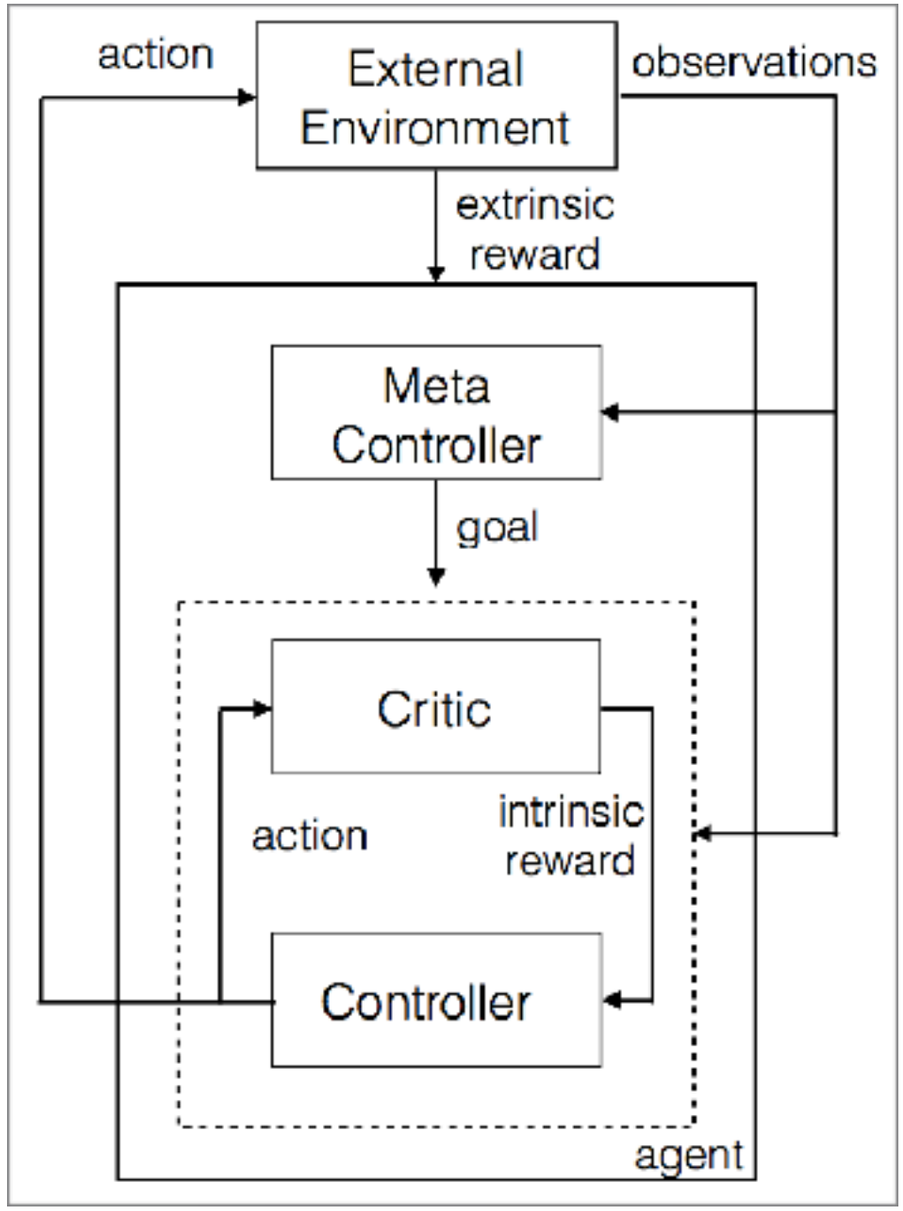
[1]

- Hinzufügen einer weiteren Abstraktionsschicht im Agenten

- „Intrinsische“ Motivation zum erkennen von Subgoals

Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation (II)

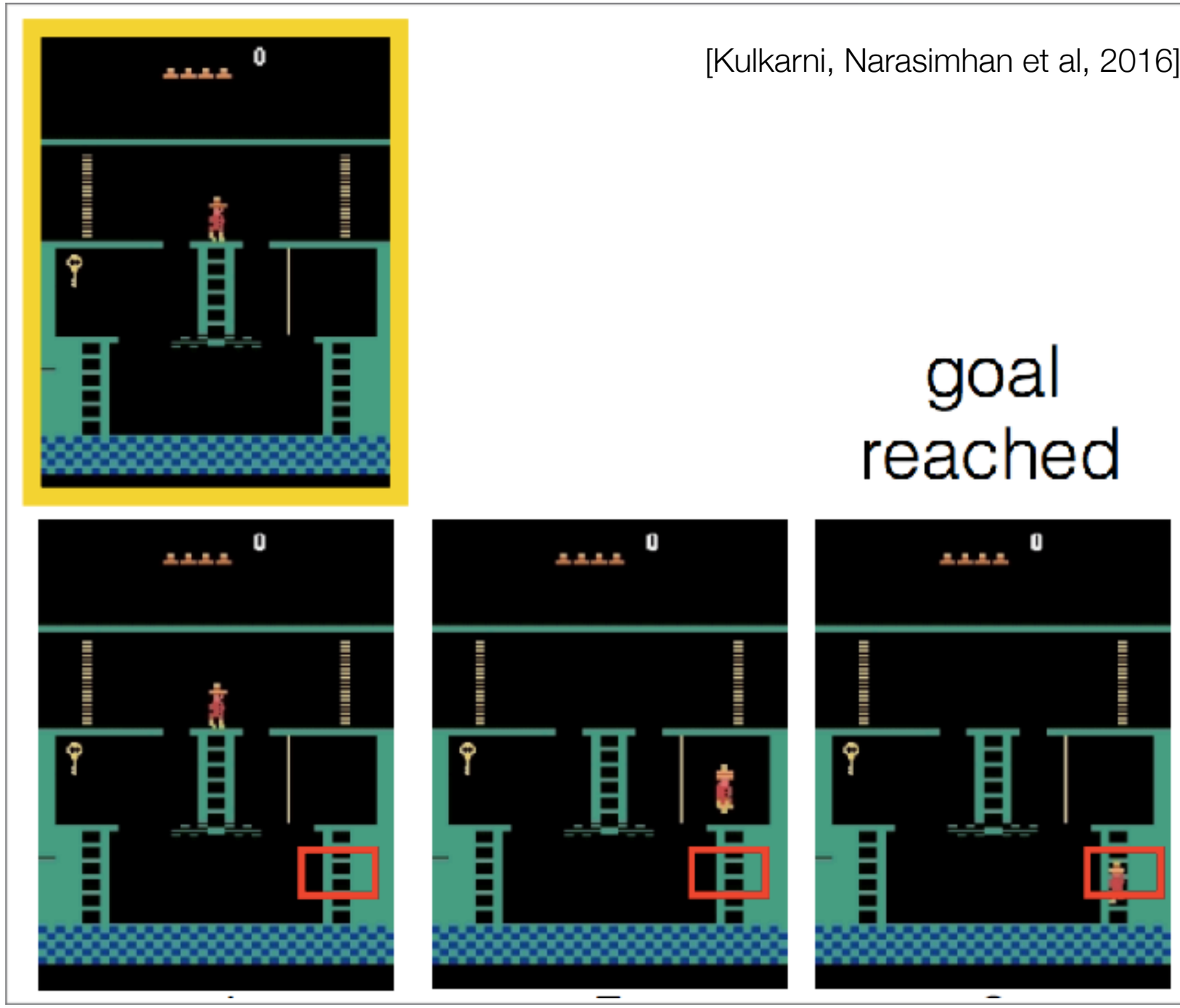




[Kulkarni, Narasimhan et al, 2016]

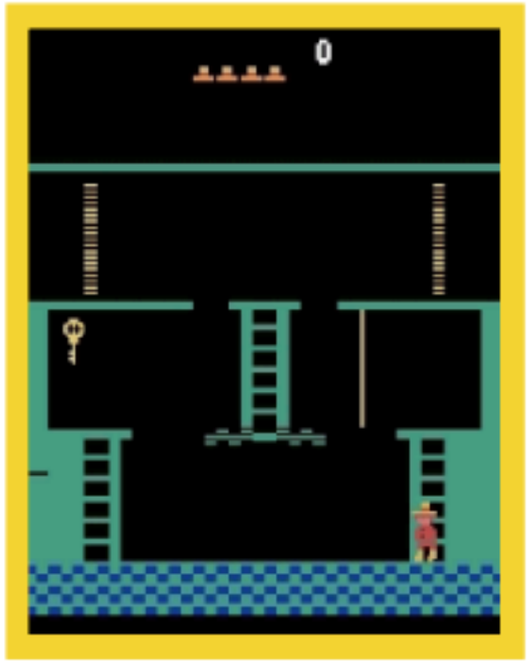
Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation (III)

[Kulkarni, Narasimhan et al, 2016]

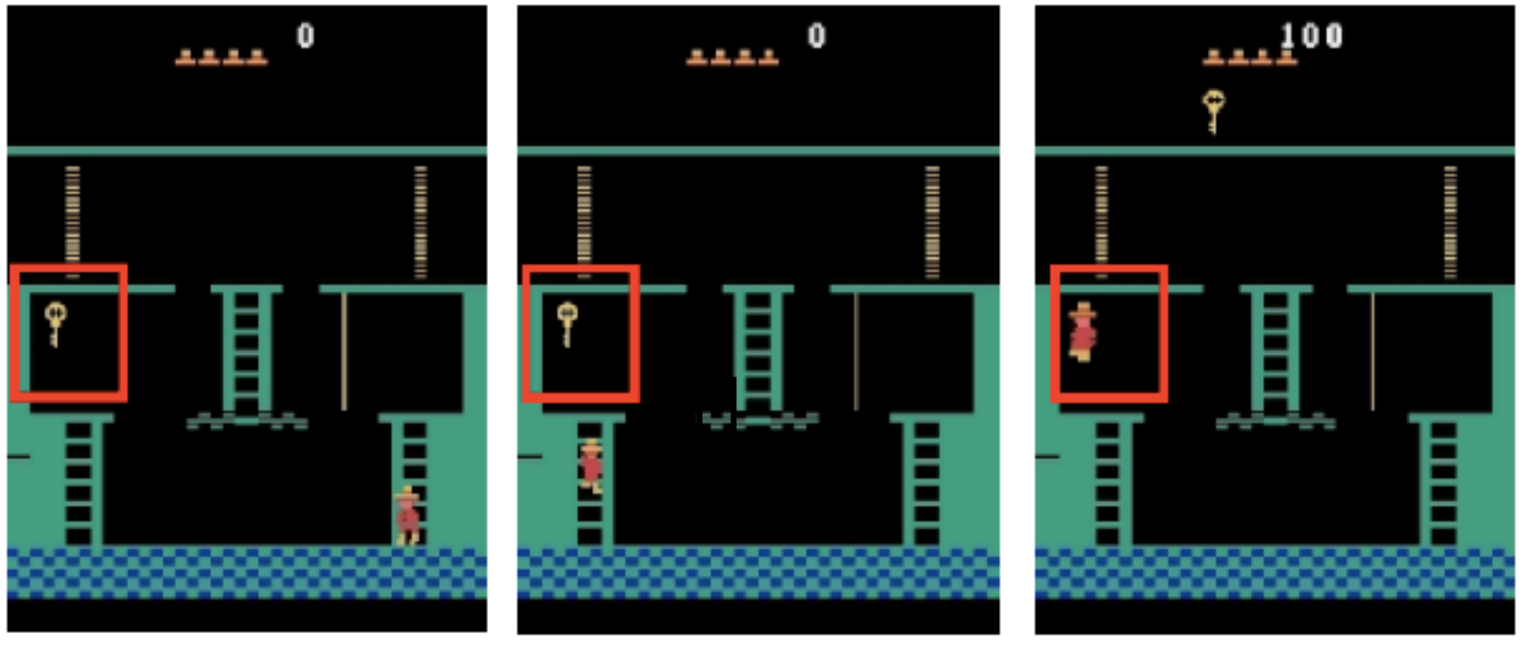


Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation (IV)

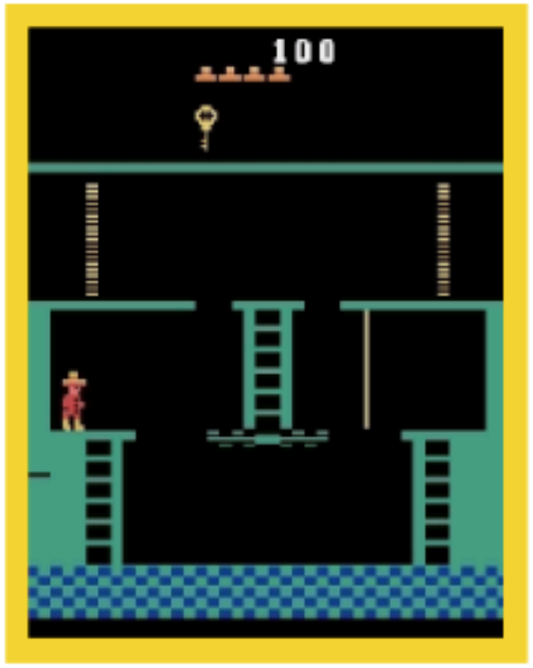
[Kulkarni, Narasimhan et al, 2016]



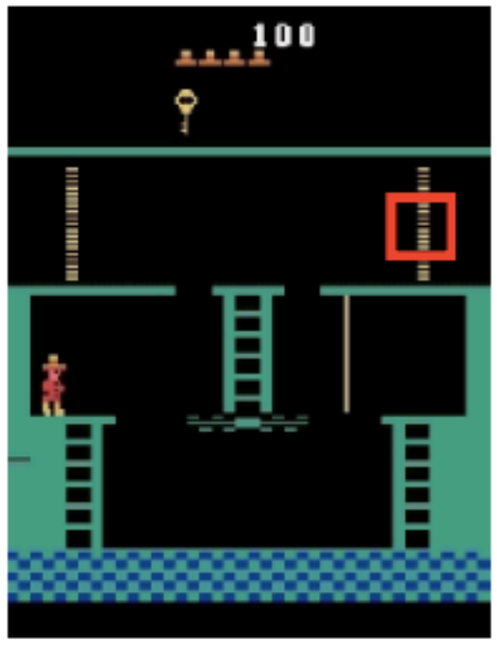
goal reached

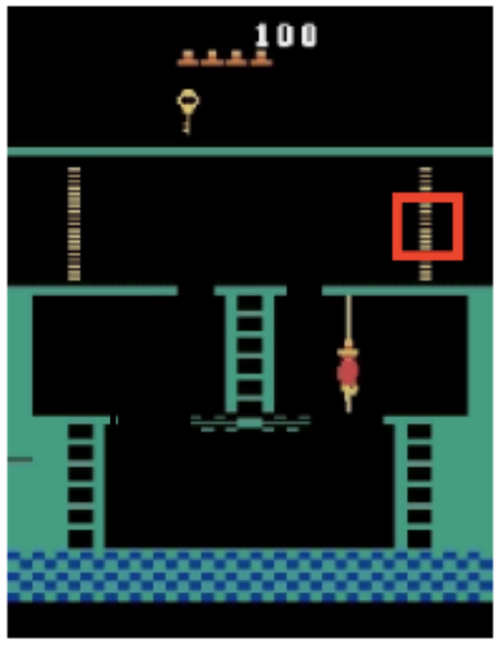


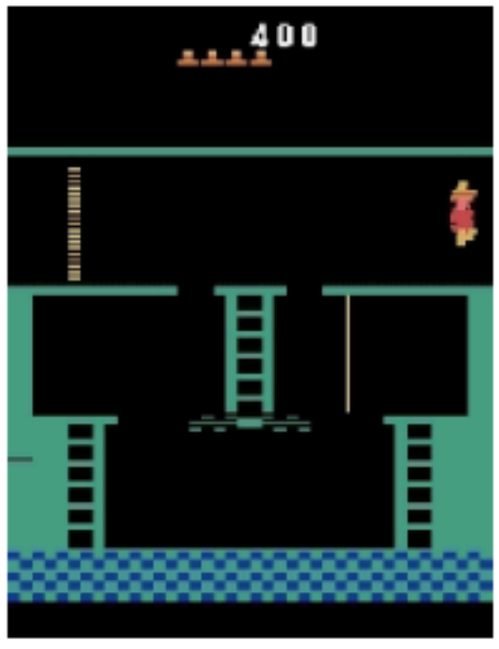
Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation (V)



[Kulkarni, Narasimhan et al, 2016]







Mein weiteres Vorgehen

Grundprojekt	• Finden und beherrschen der Werkzeuge • Lösen eines episodischen Problems am Beispiel eines geeigneten Spiels (Beispielsweise Kniffel)
Hauptprojekt	• Nichtepisodisches, fortlaufendes Problem • Roboter?
Master-Thesis	• Finden einer Methode zur Wiederverwendung bestimmter Strategien • Optimierung der gezeigten Strategien

Konferenzen

- Reinforcement Learning and Decision Making (RLDM2017), Michigan, 12-14 Juni 2017

- Neural Information Processing Systems (NIPS)2017,Kalifornien, 04 - 09 Dezember 2017

Quellen

- [1] David Silver, Introduction to Reinforcement Learning, http://www0.cs.ucl.ac.uk/staff/d.silver/web/Teaching_files/intro_RL.pdf

- [2] David Silver, AlphaGo, http://www0.cs.ucl.ac.uk/staff/D.Silver/web/Resources_files/AlphaGo_IJCAI.pdf

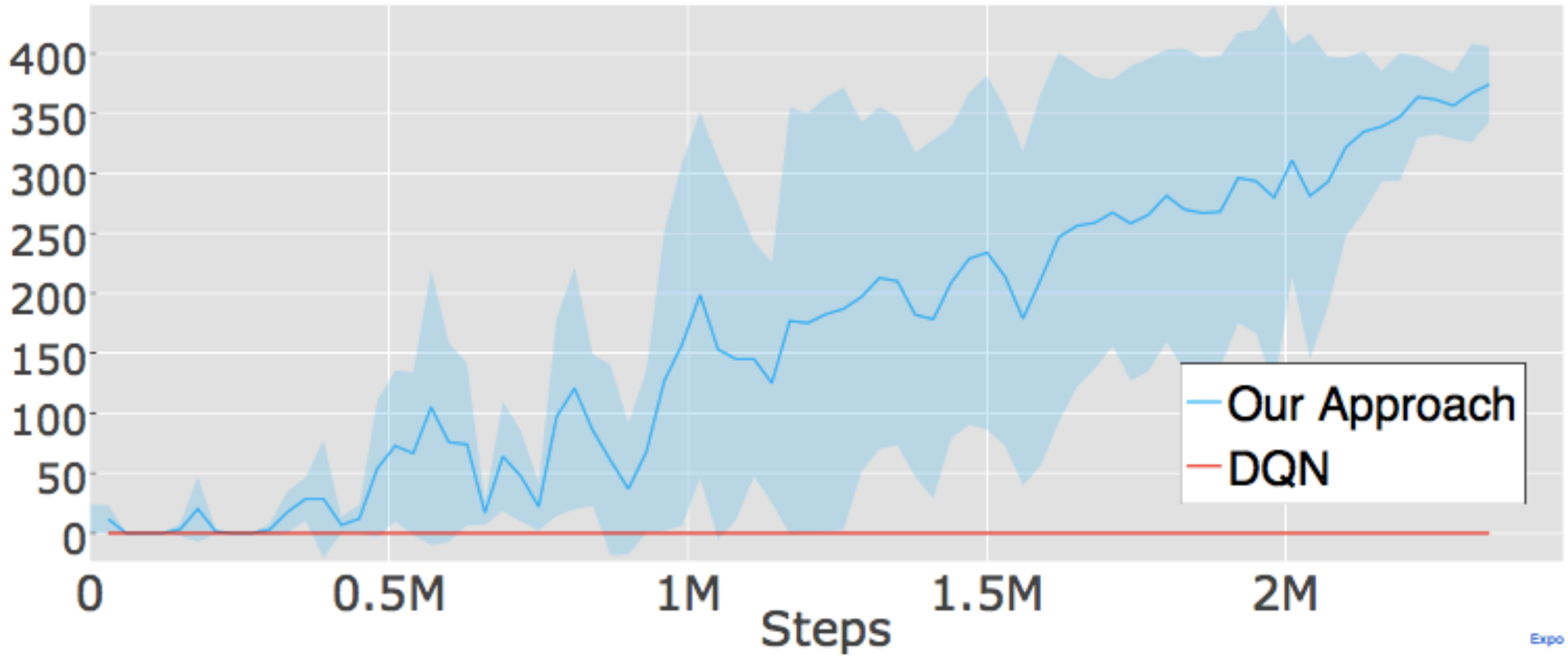
- [Ertel, 2008] Wolfgang Ertel, Grundkurs Künstliche Intelligenz, Eine praxisorientierte Einführung, Friedr. Vieweg & Sohn Verlag, 1. Auflage 2008

- [Silver, Huang et. al. 2016] David Silver, Aja Huang, et al, Mastering the Game of Go with Deep Neural Networks and Tree Search, Google DeepMind, Google, 2016

- [Kulkarni, Narasimhan et al, 2016] Tejas D. Kulkarni, Karthik R. Narasimhan, et al, Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation, MIT, 2016

Danke für die Aufmerksamkeit

Lernen mit Aufwand

	Methode	Trainings-Daten	Ressourcen	Dauer
Policy Network	Überwachtes Lernen	30M Positionen aus Expertenspielen	50 GPUs	4 Wochen
Policy Network	RL	Spiel gegen das eigene Policy-Network	50 GPUs	1 Woche
Value Network	RL	30M Spiele gegen sich selbst.	50 GPUs	1 Woche

Verbesserung durch hierarchical RL