

# Reinforcement Learning in der PL1-Domäne

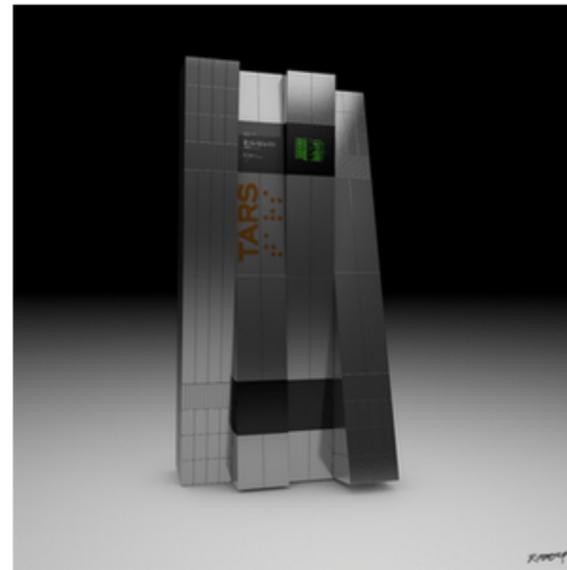
*GSM Vortrag*

# Gliederung

- Motivation
- Auffrischung Reinforcement Learning
- Zustandsraumexplosion
- Lernen mit Features
- Lernen mit PL-1

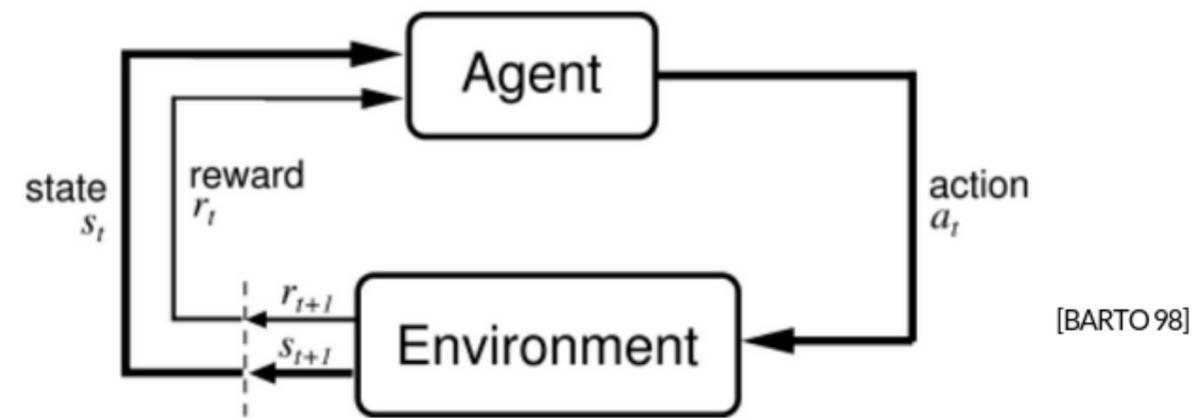
# Motivation

- Faszination Lernen
- Künstliche Intelligenz



<http://clayrodery.tumblr.com/>

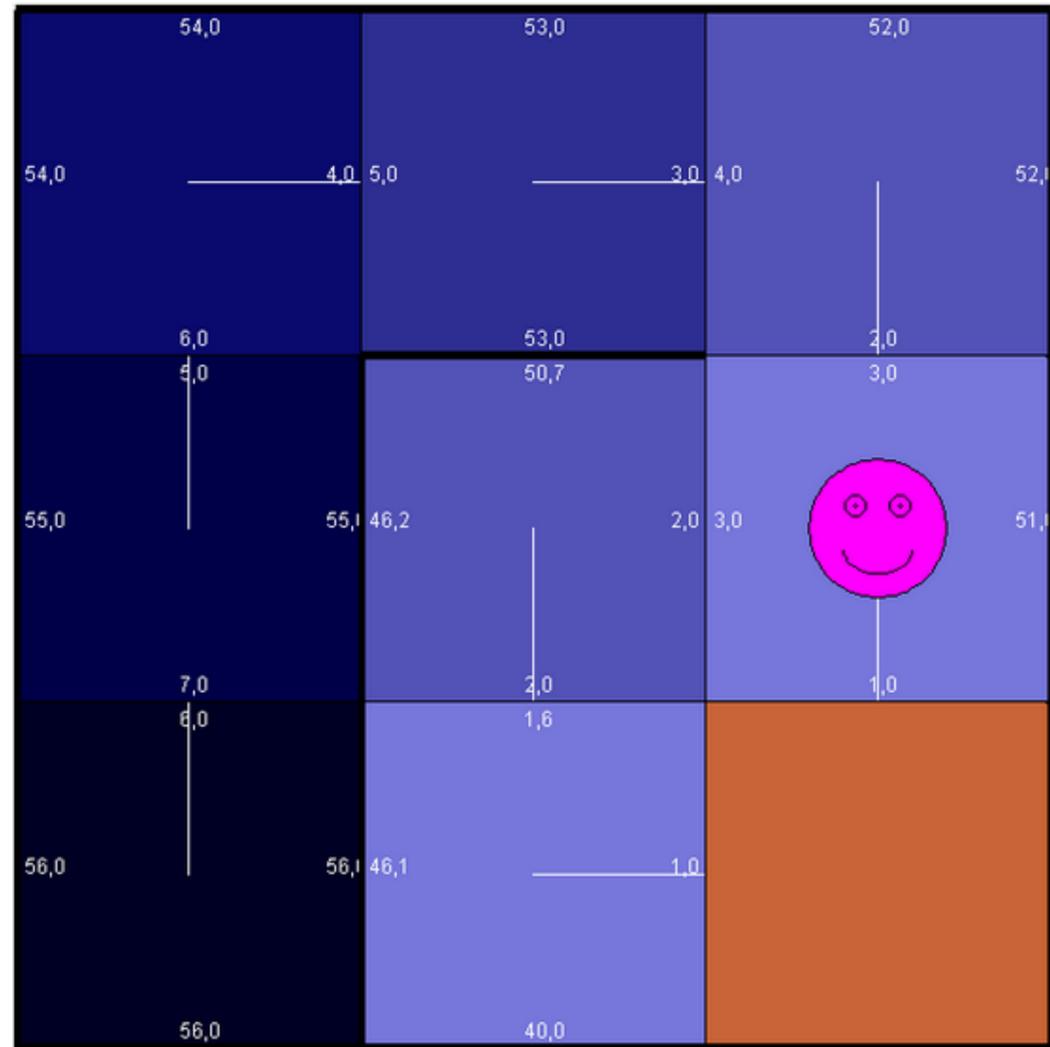
# Reinforcement Learning



# Beispiel

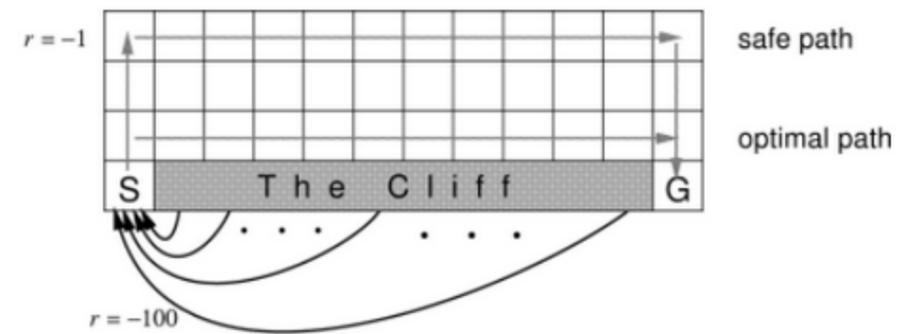
<b>environment</b>	You are in state 65. You have 4 possible actions.
<b>agent</b>	I'll take action 2.
<b>environment</b>	You have received a reward of 7 units. You are now in state 15. You have 2 possible actions.
<b>agent</b>	I'll take action 1.
<b>environment</b>	You have received a reward of $-4$ units. You are now in state 65. You have 4 possible actions.
<b>agent</b>	I'll take action 2.
<b>environment</b>	You have received a reward of 5 units. You are now in state 44. You have 5 possible actions.
...	...

[OTTERLO09]



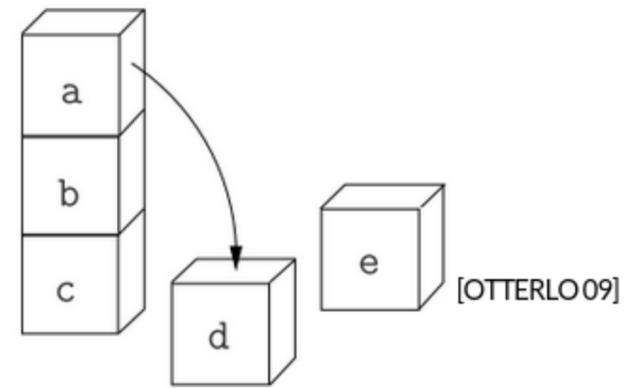
# Zustandsraumexplosion

# Cliff



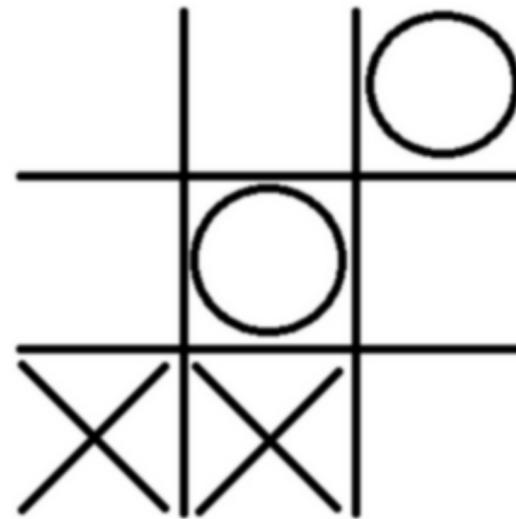
$3 \cdot 12 + 1 = 37 \approx 2^5$  Zustände

# Blocksworld



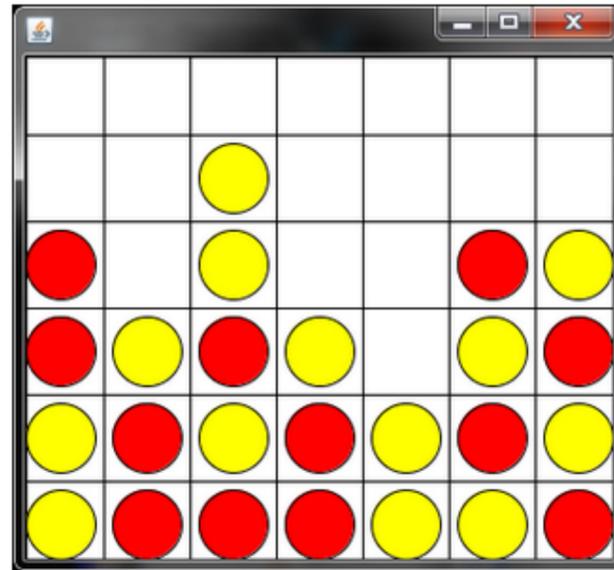
$5! = 120 \approx 2^6$  Zustände

# Tic-Tac-Toe



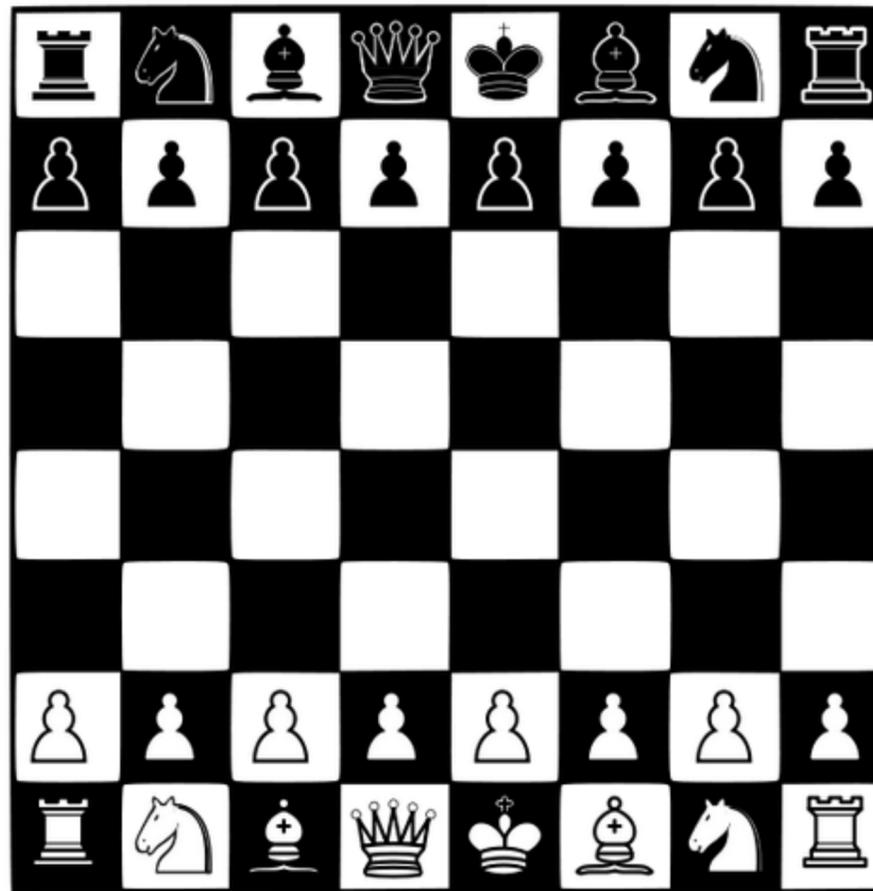
$3^9 = 19683 \approx 2^{14}$  Zustände

# 4-Gewinnt



$3^{6 \cdot 7} \approx 10^{20} \approx 2^{66}$  Zustände

# Schach



[WIKI]

$$\frac{(8 \cdot 8)!}{32! \cdot 8! \cdot 8!} \approx 10^{44} \approx 2^{148} \text{ Zustände}$$

# Bewertung

- Keine Abstraktion möglich
- Zustandsraum explodiert
- Lernen dauert dementsprechend lange

# Lernen mit Features

$$s \rightarrow \begin{pmatrix} f_1 \\ f_2 \\ \dots \\ f_n \end{pmatrix}$$

# Beispiel

<i>Ampel grün</i>	<i>true</i>	<i>kuppeln</i> →	<i>true</i>
<i>Kreuzung frei</i>	<i>true</i>		<i>true</i>
<i>Kupplung gedrückt</i>	<i>false</i>		<i>true</i>
<i>Gaspedal gedrückt</i>	<i>false</i>		<i>false</i>
<i>Bremse gedrückt</i>	<i>true</i>		<i>true</i>
<i>Auto fährt</i>	<i>false</i>		<i>false</i>
<i>1.Gang eingeworfen</i>	<i>false</i>		<i>false</i>

# Vergleich zu RL

- Zustandsraum meist auf das Nötigste reduziert
- Rudimentäre Strukturinformationen
- Wissenstransfer für ähnliche Zustände möglich

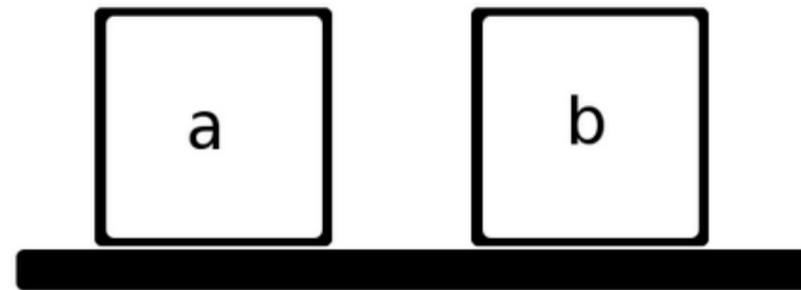
# Lernen mit PL-1

*"It is hard to imagine a truly intelligent agent that does not conceive of the world in terms of objects and their properties and relations to other objects" [KAELBLING01]*

# Konzept

- Zustände: Liste gültiger Atome statt Featurevektor
- Agent lernt mit abstrakten Zuständen
- Aktionen: können an Variablen gebunden sein

# Beispiel



Zustand:

```
on(a,table).  
on(b,table).  
clear(a).  
clear(b).
```

Aktionen:

```
put(a,b).  
put(b,a).
```

Entscheidungsliste:

```
clear(a) ^ clear(b) → put(a,b)  
on(X,a) → put(X,table)  
on(X,b) → put(X,table)  
true → random_action()
```

# Vergleich zu Features

- Volle Strukturinformationen von Zuständen
- Reduzierung auf sehr wenige Zustände
- Variablen abstrahieren Lernerfahrung von konkreten Objekten

# Ziele fürs Grundprojekt

- Implementierung eines ersten PL-1 Agenten für 4-Gewinnt mit vorgegebener Aktionsliste
- Evaluierung sinnvoller Prädikate und Lernmethoden

# Journals / Transactions

- ACM Transactions on Autonomous and Adaptive Systems (TAAS)
- Machine Learning Journal
- Computational Intelligence and AI in Games (IEEE T-CIAIG)

# Quellen

- [BARTO 98] Barto, A. G. (1998). Reinforcement learning: An introduction.
- [OTTERLO 09] van Otterlo, M. (Ed.). (2009). The logic of adaptive behavior: knowledge representation and algorithms for adaptive sequential decision making under uncertainty in first-order and relational domains.
- [WIKI] Wikipedia - Chess:  
*[http://upload.wikimedia.org/wikipedia/commons/2/2b/AAA\\_SVG\\_Chessboard\\_and\\_chess\\_pieces\\_03.svg](http://upload.wikimedia.org/wikipedia/commons/2/2b/AAA_SVG_Chessboard_and_chess_pieces_03.svg)*
- [KAELBLING 01] Kaelbling et al. (2001). Learning in Worlds with Objects, in: The AAAI Spring Symposium.

Vielen Dank für die  
Aufmerksamkeit

Fragen?