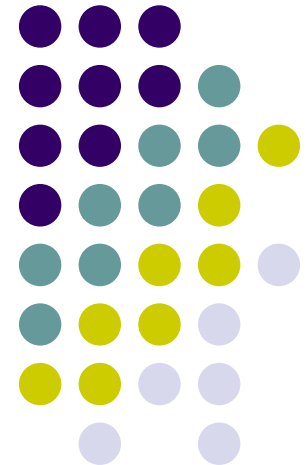
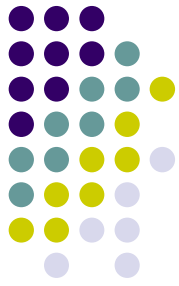


Opponent Modelling

Im Rahmen des Seminars:
**Knowledge Engineering und
Lernen in Spielen**
SS 2006



Thema & Herausforderung

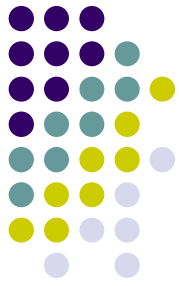


- „Opponent Modelling“ beinhaltet das Lernen wie ein Gegner handelt, um darauf aufbauend eine Gegenstrategie aufzudecken, welche diese Informationen ausnutzt.
- Was ist eine „optimale“ Strategie?
- Optimale Strategie spielt immer gleich gut, egal wie gut oder schlecht der Gegner ist
- Mit Opponent Modelling: Spieler besiegt schwächere Gegner öfter als die optimale Strategie



TECHNISCHE
UNIVERSITÄT
DARMSTADT

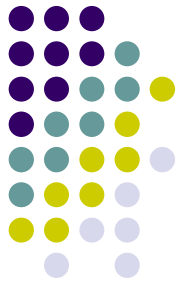
Inhalte zu Opponent Modelling



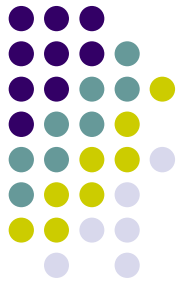
- Einführung Spieltheorie
- Einordnung Spiele, Poker und KI
- Strategiesuchverfahren
- Beispiele anhand Poker
- Fazit & Ausblick



Spieltheorie



- Nash-Gleichgewicht:
„Zustand eines strategischen Gleichgewichts, von dem ausgehend kein einzelner Spieler für sich einen Vorteil erzielen kann, indem er allein seine Strategie verändert.“
- Min. der Verluste statt Max. des Gewinns
- Perfekter Gegner ?



Beispiele Spieltheorie

„Gefangenendilemma“

Gestehen (G) oder

Nicht Gestehen (N)

	G	N
G	-2, -2	-5, -1
N	-1, -5	0, 0

Optimum (N,N) aber

Nash-Gleichgewicht (G,G)

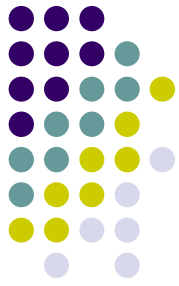
Stein-Papier-Schere

(RoShamBo)

	Stein	Papier	Schere
Stein	0, 0	-1, 1	1, -1
Papier	1, -1	0, 0	-1, 1
Schere	-1, 1	1, -1	0, 0

Nullsummenspiel, optimal ist
gleichverteilte Auswahl (gemischte
Strategie), kein Nash-Gg
vorhanden

Klassifizierung von Spielen

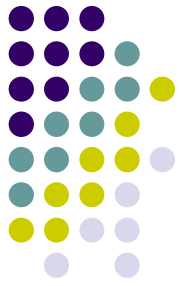


- Spiele bieten guten Rahmen zur Forschung an Künstlicher Intelligenz
„Computer Games are the biggest AI success story to date“ [Schaeffer]
- Klassifizierung nach Zufallseinfluss und Informationsgrad des Spielstandes entscheidend

	Kein Zufall	Zufall
Perfekte Information	Schach	Backgammon
Unvollständige Information	Schiffe versenken	Poker



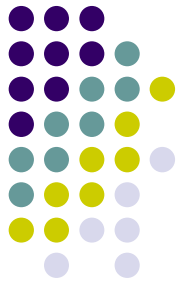
Poker



- Deterministische Welt unrealistisch – Realität beinhaltet häufig Stochastik und unvollständige Information
- Kartenspiel in Stufen - verdeckte und öffentliche Karten – setzen von Geld, um mit Blatt / vorgetäushtem Blatt Profit zu maximieren – bis Showdown jeweils alle Gebote der Mitspieler ausgeglichen
- Zahlreiche Entscheidungen während den Spielrunden beeinflussen Erfolgsrate

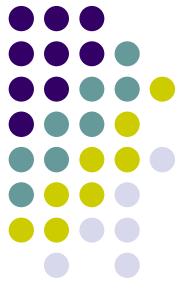


KI Aspekte im Poker



- Entscheidungstheorie (Zufall und Unsicherheit)
- Risikobewertung (stufenweise Karten und Einsätze)
- Gegner Modellierung
(Strategien und Wiederholungen durchschauen)

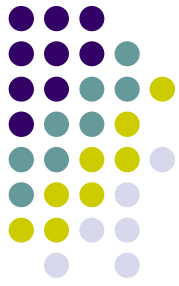
Verbindung zu anderen KI Bereichen



- **Benutzermodellierung** (Verhaltensmuster und Vorlieben erfassen und berücksichtigen)
- **Politiken festlegen und Verhandlungen führen** (Ansätze von Spieltheorie in Gesetz, Politik, Wirtschaft und Militär)
- **Online Auktionen**



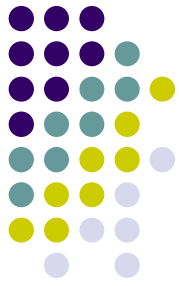
Heuristischer Ansatz



- Regelbasiertes Expertensystem
- Grundlagen für Anfänger (Hand spielbar, passen-mitgehen-erhöhen, eigene Karten verbergen und interpretieren der Gegnerkarten)
- Im Poker zu viele Möglichkeiten durch Unsicherheit, zu viele Entscheidungsszenarien
- Zudem durchschaubar
- Immerhin mittlere Spielstärke bei „full-ring“ (bsp.10 Spieler) Runden



Simulationsbasierter Ansatz



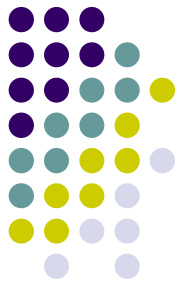
- Unvollständige Information erschwert Entscheidungsfindung
- Raten des Gegnerblattes und Vorhersehen wie Erfolg erzielbar
- Simulation von Handlungsentscheidungen und Berechnung von Erwartungswerten (Bewertungen notwendig)
- Vorteile gg. Heuristiken: automatische Entdeckung dynamischer Setzstrategien bei verschiedenen Spielbedingungen, einheitliches Rahmenwerk, nicht basierend auf Expertenwissen
- **Aber!** Unwissendheit und Nichtberücksichtigung vom Gegner
- Zudem zu viele nicht zu simulierende Nebeneffekte vorhanden

Spieltheoretischer Ansatz



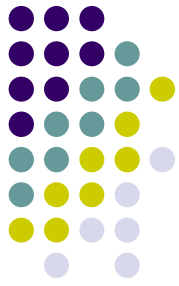
- Beschränkt auf 2 Spieler Variante
- Optimale Strategie führt zu Nash-Gleichgewicht und hat zwei Beweggründe:
 - Führt zu besten Ergebnissen, vorausgesetzt Gegner spielt best-response-Strategie
 - Gegner kann trotz Kenntnis über Strategie keinen Vorteil daraus ziehen
- Komplexitätsproblem führt zu Approximationen, welche jedoch obige Eigenschaften nicht mehr garantieren können

Heuristischer suchbasierter Ansatz



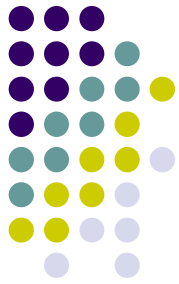
- Vorgehensweise der Rückwärtsinduktion des Spielbaums um Entscheidung zu ermitteln
 - Rekursive Betrachtung aller möglichen Aktionen bis Spielende (Blätter) bildet Spielbaum
 - Ausgang des Spiels (Gewinn, Verlust..) ist bekannt
 - Aktion, die zu diesem Spielstand führt, wird mit dessen Wert bewertet
 - Vermutung welche Aktion Spieler auswählt und Wiederholung für vorherige Entscheidungspunkte

Heuristischer suchbasierter Ansatz (2)



- Minimax-Search
 - Spieler Max wählt jeweils bestverfügbare Aktion
 - Spieler Min ebenso – aufgrund 2 Spieler-Nullsummenspiel gleichzeitig schlechteste Aktion aus Sicht von Max
 - Spieltheoretisch optimale Strategie
 - Eignung bei perfekter Information
 - Vollständige Betrachtung oft nicht möglich
 - Approximation durch Bewertungsfunktionen, Pruning (nicht optimal, aber erfolgreich)
 - Ansätze für Poker: Fokussierung der Entscheidungsmöglichkeiten auf Kartenhand, Zufallberücksichtigung möglich, jedoch Problem der unvollständigen Information

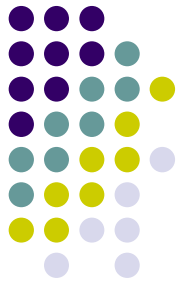
Heuristischer suchbasierter Ansatz (3)



- Expectimax-Search
 - Gegner entscheidet individuell, nicht deterministisch optimal
 - Kein hypothetisch optimaler Gegner mehr sondern spezifischer Gegner
 - Spieler wählt weiterhin bestverfügbare Aktion
 - Für Gegneraktion werden Erwartungswerte basierend auf vorherbestimmten Strategien und deren Wahrscheinlichkeiten berechnet



Heuristischer suchbasierter Ansatz (4)

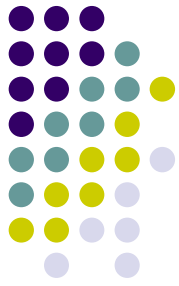


Opponent Modelling (perfekte Information):

- Reibman and Ballard's-Min Search
 - Entsprechend der vorhergesagten Stärke des Gegners ergeben sich Entscheidungswahrscheinlichkeiten mit zugehörigen Erwartungswerten
- Jansen's Probi-max-Search
 - Betrachtet eine bekannte Verliererposition im Schach gegen einen fehleranfälligen Gegner (Heuristiken für Fallen)
- Sen and Aurora's Maximum Expected Utility Player
 - Benutzt Trainingsphase um Gegner zu beobachten und ein opponent model aufzustellen
- Probabilistic opponent-model Search
 - Kenntnis verschiedener Bewertungsfunktionen des Gegners (Gegnertypen) und dementsprechende Schätzungen
 - Experimentelle Analysen im Spiel „Lines of Action“



Beispiel RoShamBo

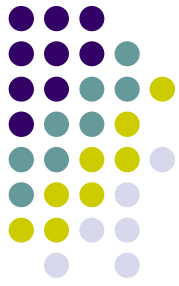


Erfolgsrezepte Opponent Modelling

- Erfolg basiert auf Gegner „durchschauen“
- Gegnermodell erstellen und ausnutzen
- Besonderes Interesse an Schwachstellen (bei unvollständiger Information erst aufdecken!)
- Aktion nur bevorzugt einsetzen, wenn Kenntnis besteht, dass Gegner auch eine Aktion bevorzugt
- Viel gewinnen – Variation beibehalten (defensives Verhalten zum Selbstschutz)

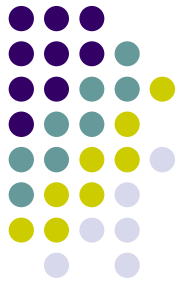


Expectimax Search in Poker



- Miximax und Miximax Anpassung an unvollständige Information
- Entscheidungsknoten, Zufallsknoten und Blattknoten (Spiel vorüber, keine Aktionen mehr)
- Versteckte Informationen führen zu nicht unterscheidbaren Handlungsknoten (Spielzustand unbekannt)- werden als „information set“ zusammengefasst (bei perfekter Information $is =$ ein Knoten)
- Anzahl der Blätter im Spielbaum von 2 Spieler Texas Hold'em zu Beginn ca. 697 Billionen – bei einem Spieler der immer mitgeht ca. 4,55 Billionen

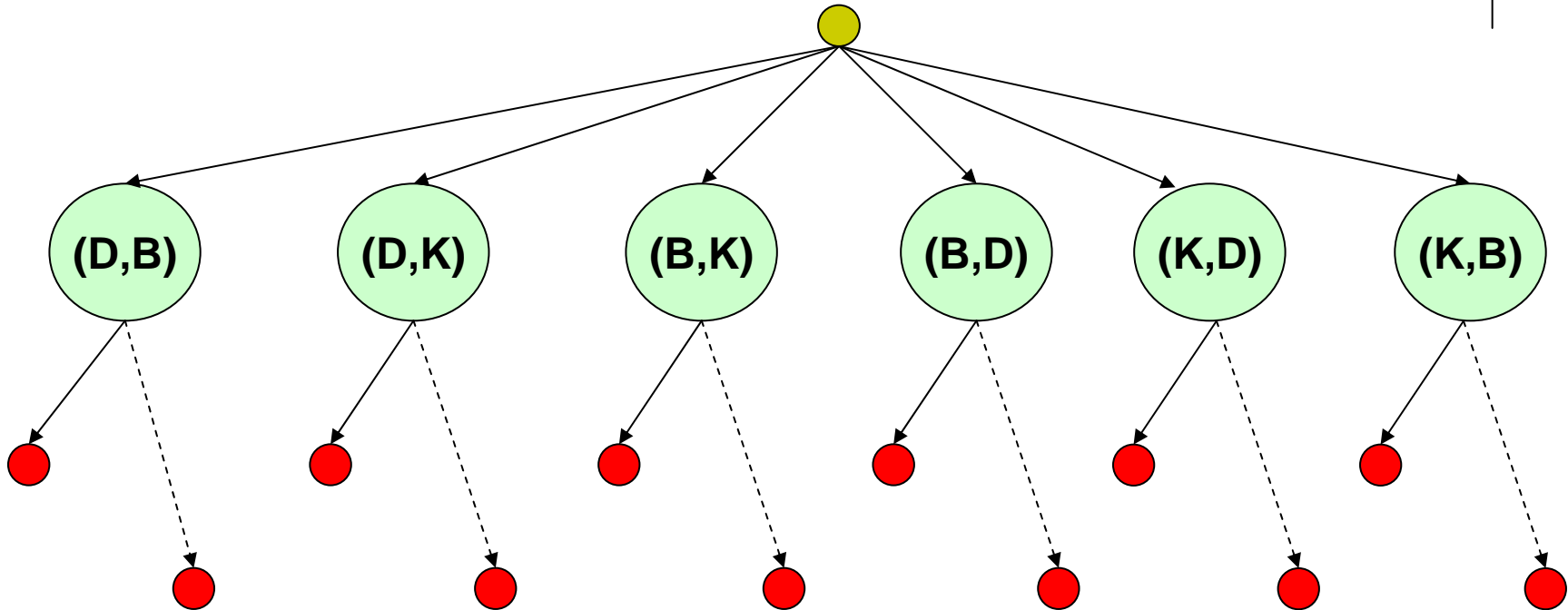
Expectimax Search in Poker (2)



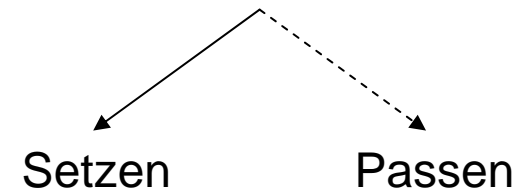
Beispiel:Kuhn Poker

- 2 Spieler mit jeweils 2 GE einzusetzendem Kapital
- 3-er Kartendeck: Bube (B), Dame (D), König (K)
- Spielverlauf
 - Grundeinsatz je 1GE
 - Beide Spieler erhalten 1 Karte
 - In der Setzrunde können beide passen, die verbleibende GE setzen oder aussteigen

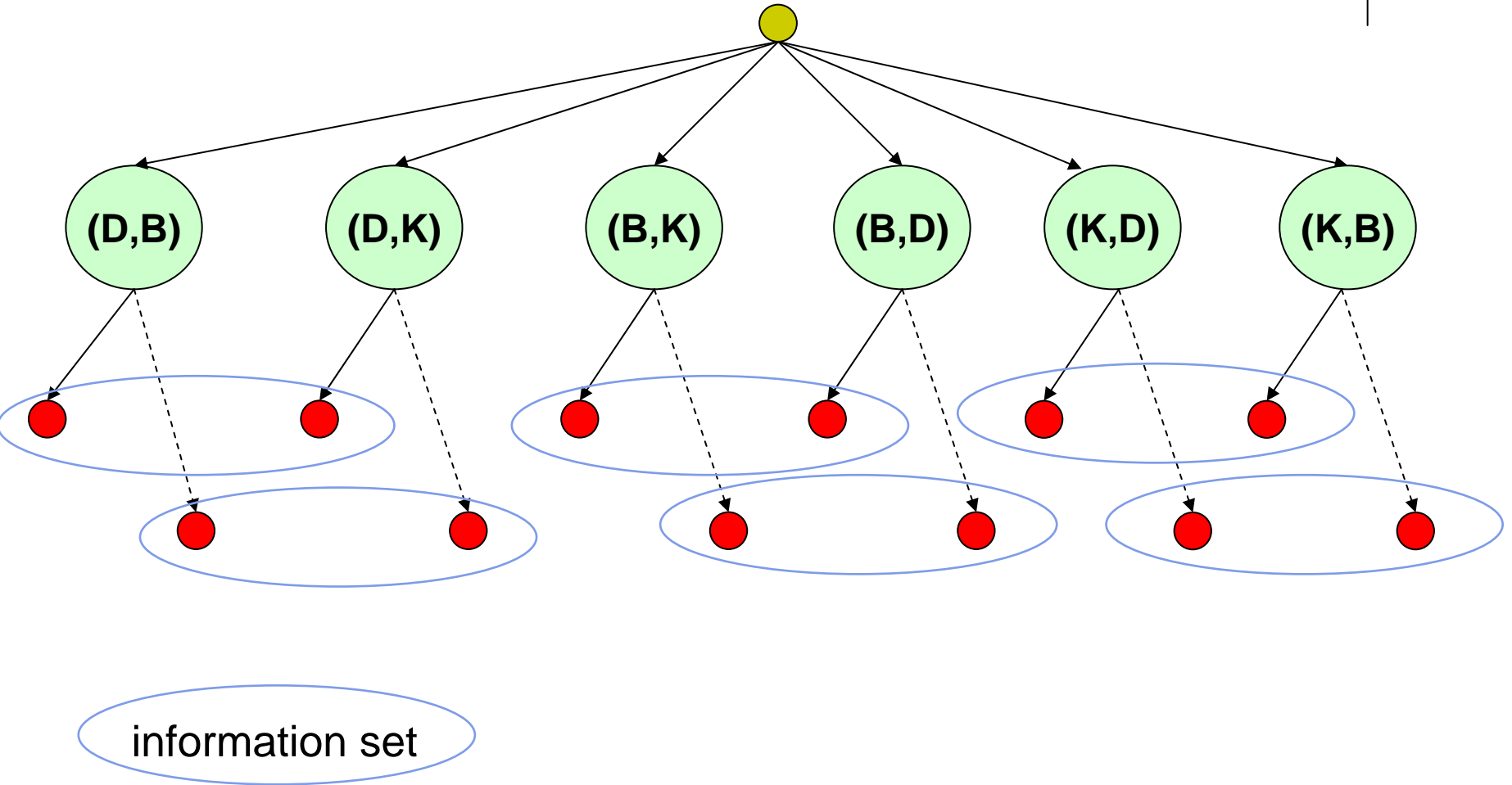
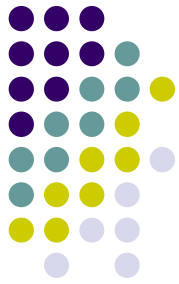
Expectimax Search in Poker (3)



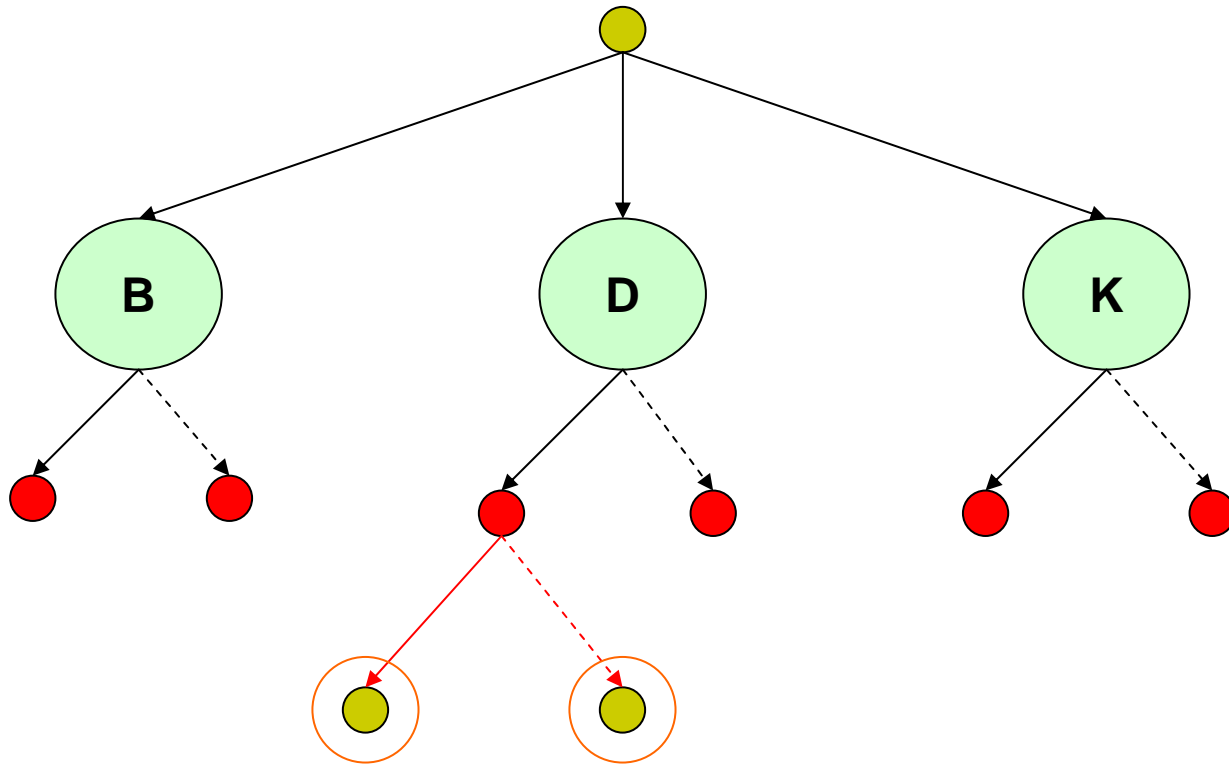
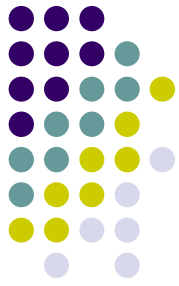
(D,B): Spieler hat die Dame,
Gegner den Buben



Expectimax Search in Poker (4)

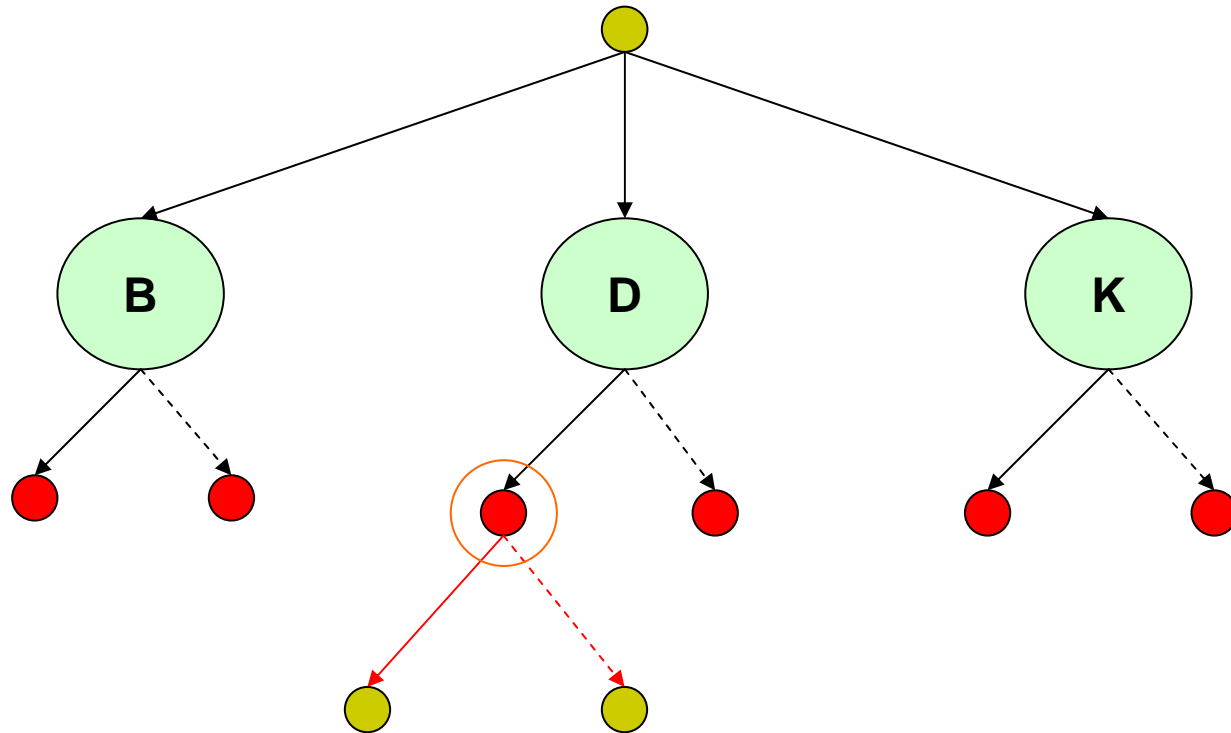
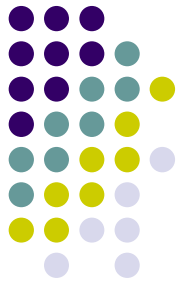


Expectimax Search in Poker (5)



Auswertung der Blattknoten: $E(L) = (P_{\text{win}} \times L_{\text{pot}}) - L_{\text{cost}}$

Expectimax Search in Poker (6)



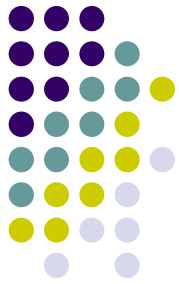
Entscheidungsknoten:

Miximax: $E(V) = \max(E(V_{\text{setzen}}), E(V_{\text{passen}}))$

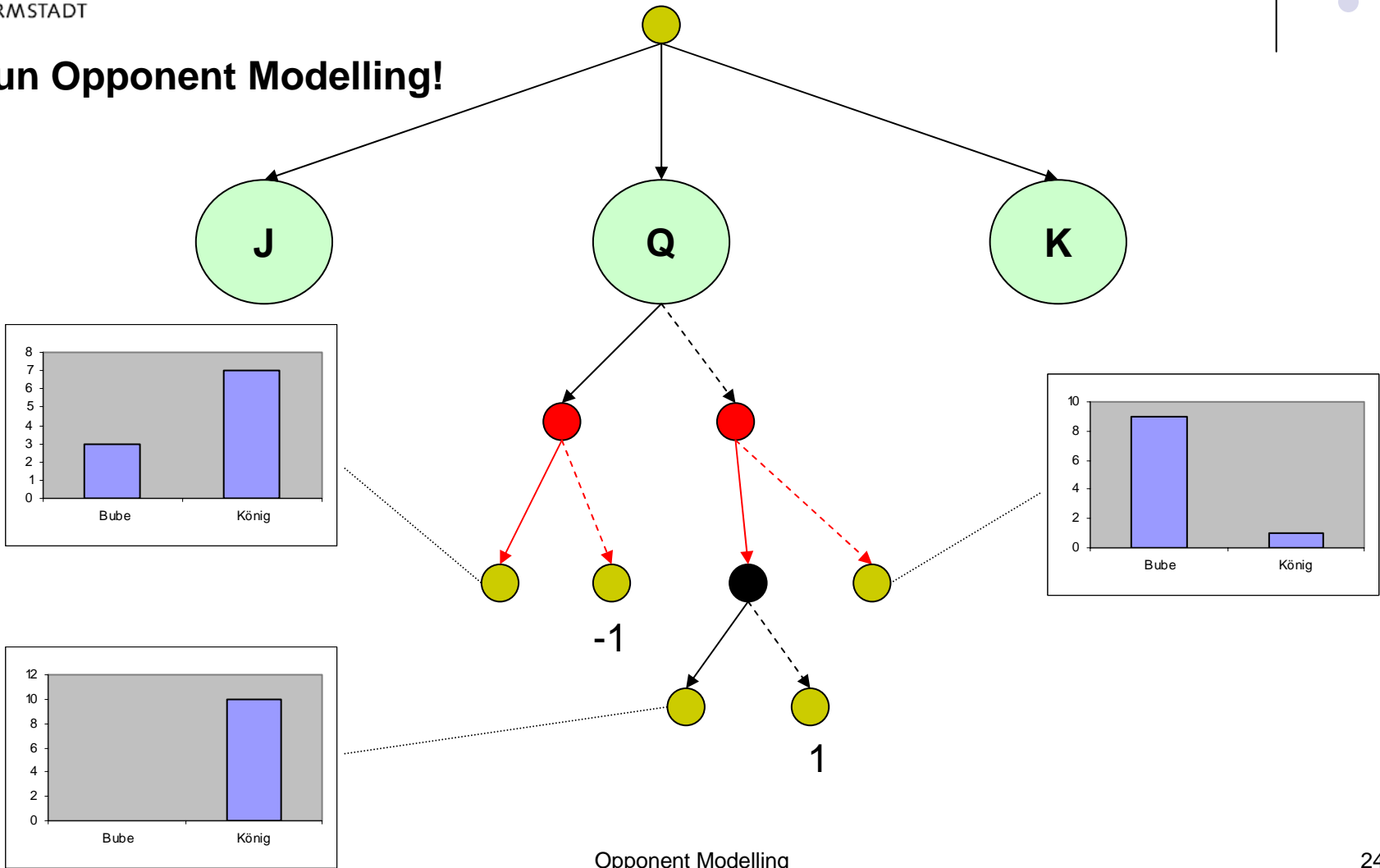
Miximix: $E(V) = P(V_{\text{setzen}}) \times E(V_{\text{setzen}}) + P(V_{\text{passen}}) \times E(V_{\text{passen}})$



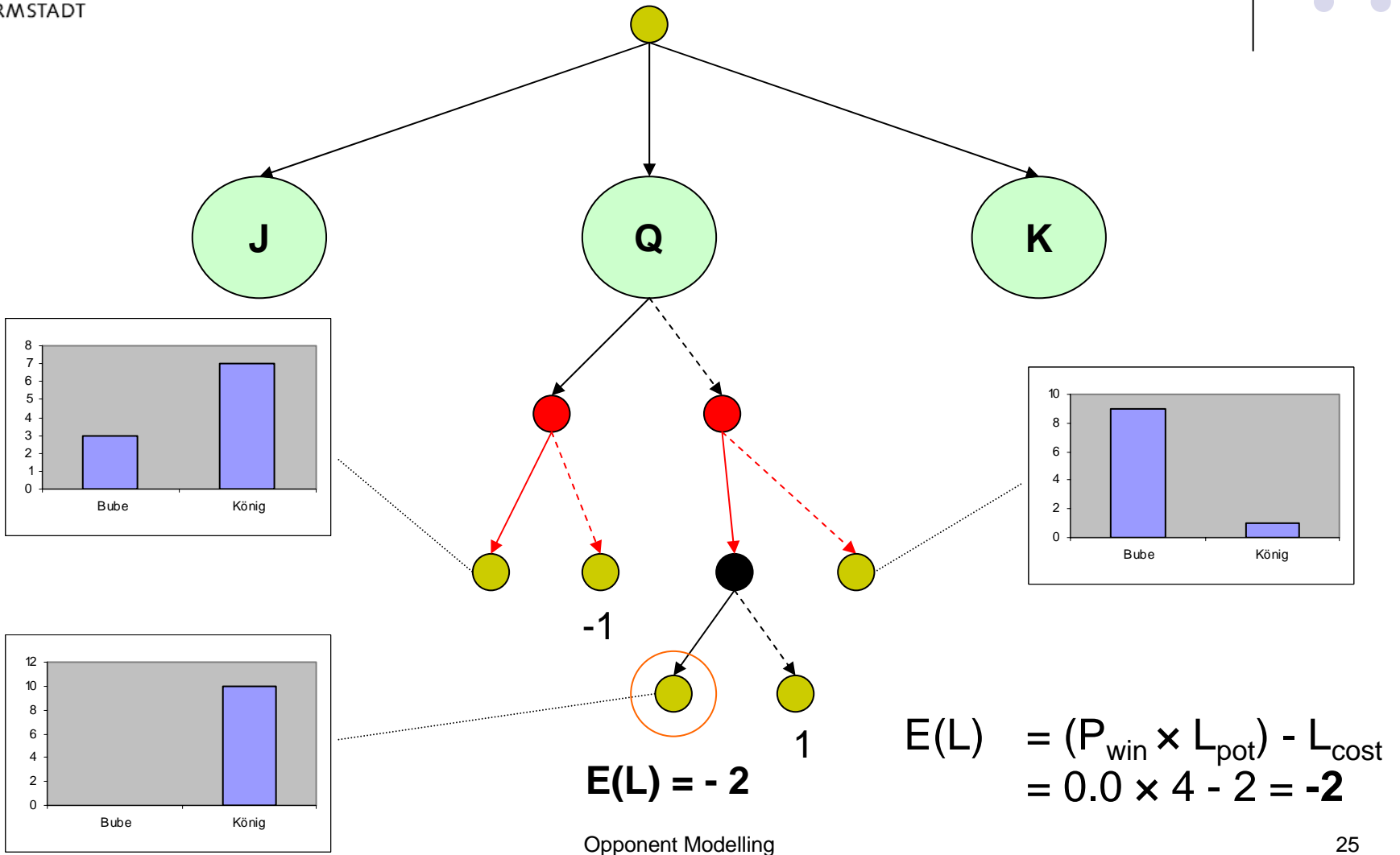
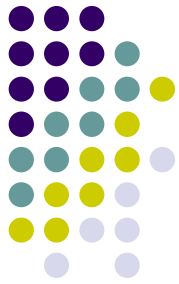
Expectimax Search in Poker (7)



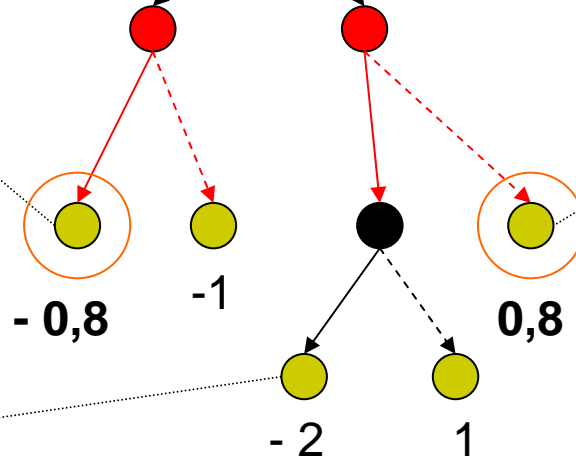
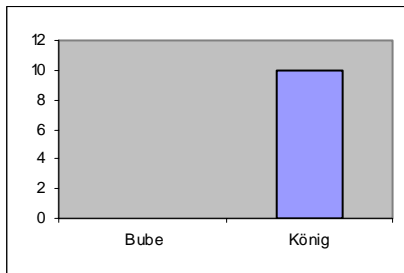
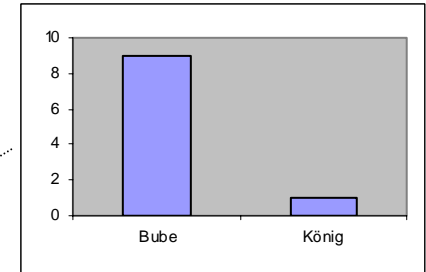
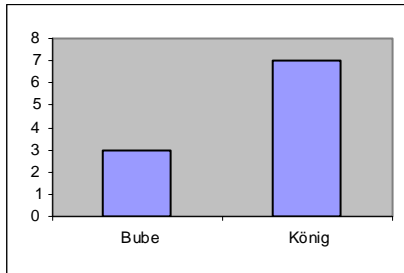
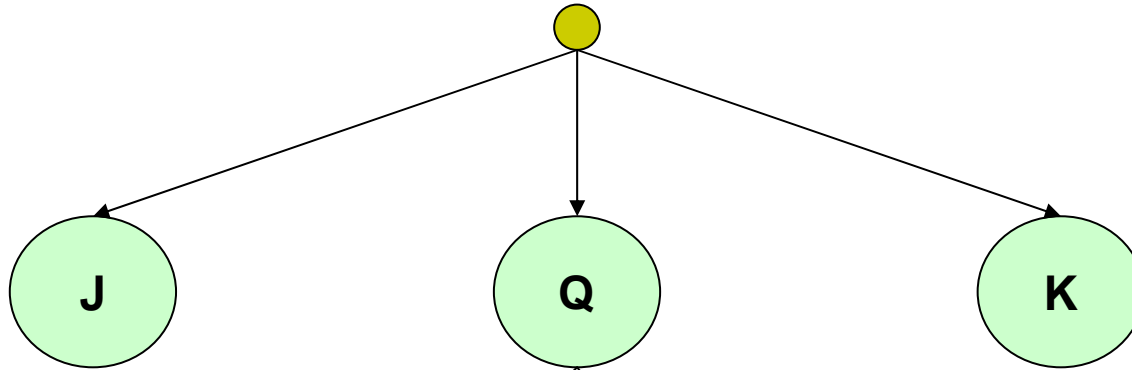
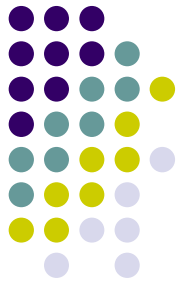
Nun Opponent Modelling!



Expectimax Search in Poker (8)



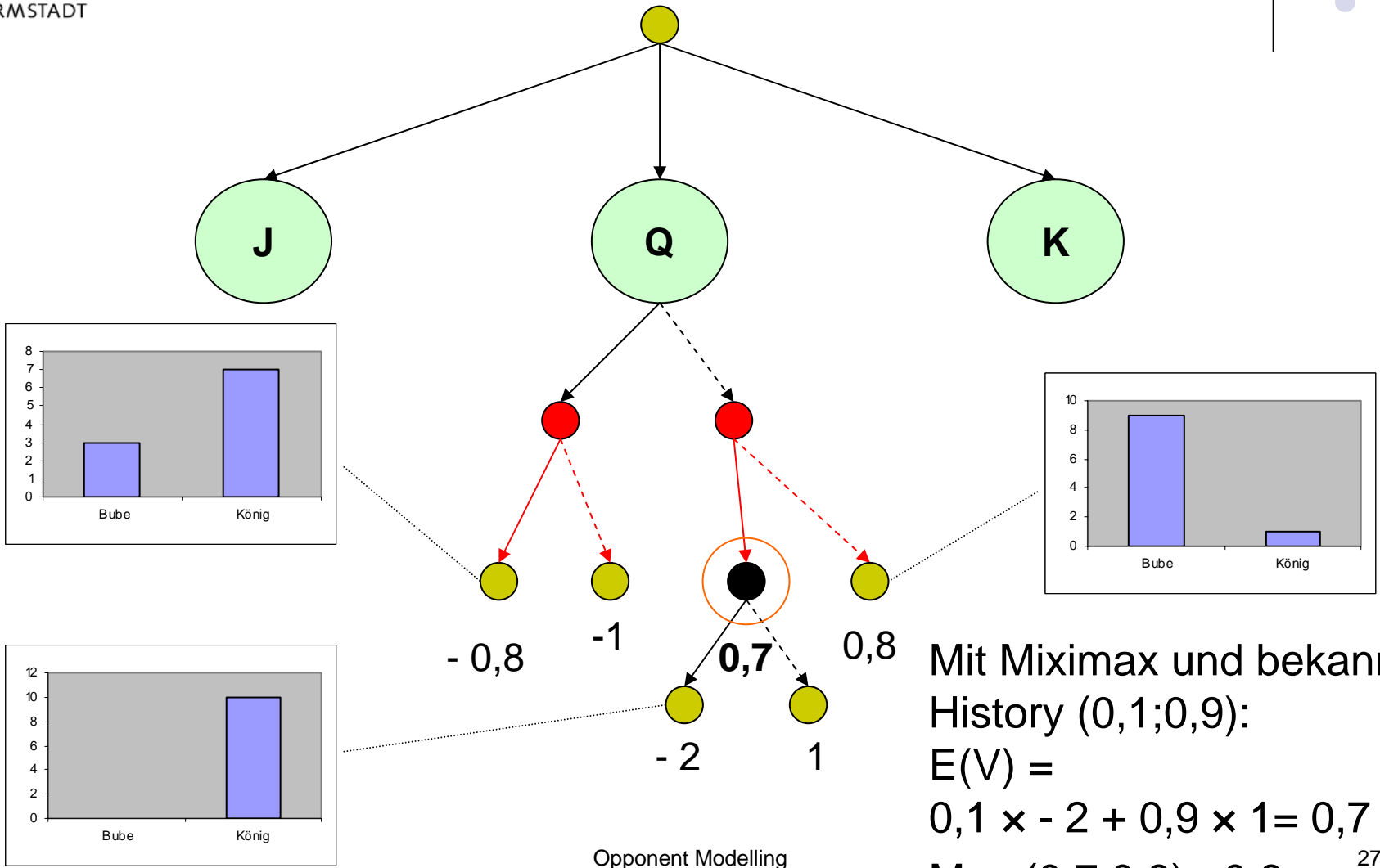
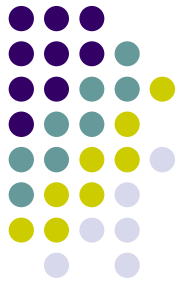
Expectimax Search in Poker (9)



$$E(L) = 0.9 \times 2 - 1 = \mathbf{0,8}$$

$$E(L) = 0.3 \times 4 - 2 = \mathbf{-0,8}$$

Expectimax Search in Poker (10)



Opponent Modelling

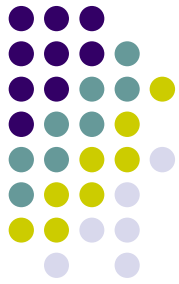
Mit Miximax und bekannter
History (0,1;0,9):

$$E(V) = 0,1 \times -2 + 0,9 \times 1 = 0,7$$

$$\text{Max}(0,7; 0,8) = 0,8$$

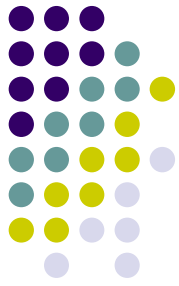


Opponent Modelling im Poker



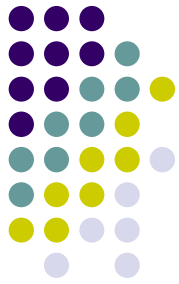
- 2 Klassen der Modellierung
- Strategiemodelle versuchen direkt die Strategie des Gegners zu erlernen – daher nicht anwendbar bei unvollständigen Information
- Beobachtungsmodelle basieren nur auf Beobachtungen aus der Perspektive des Entscheidenden
- Ermöglichen Prognostizieren von Handlungs-, Zufallshäufigkeiten und Aufzudeckenden Karten beim „Showdown“
- Beobachtungsmodelle geeigneter für Poker
 - Beim passen im Gegensatz zu Strategiemodellen kein Informationsverlust
 - Sind einfacher aufzustellen
 - Informationen geeigneter als Input für Expectimax

Opponent Modelling im Poker (2)



- Beobachtungsmodelle für Poker sollen Wahrscheinlichkeitsverteilung über Gegnerhandlungen und Zufall sowie Wahrscheinlichkeiten eines Showdown-Sieges liefern
- Generalisierung des Spielerverlaufs erforderlich
- Hinsichtlich der Gegnerhandlungen werden nur Setzkonstellationen ausgewertet
- Für den Showdown wird eine Gewinnrate aller möglichen Kartenkombinationen des Gegners ermittelt
- Diese wird als Histogramm abgespeichert und mit ähnlich hinterlegten Klassen (10 Vergleichsstufen) abgeglichen

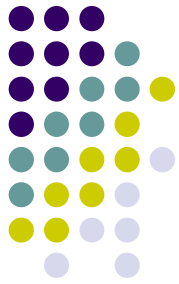
Opponent Modelling im Poker (3)



Test mit Leduc Holde‘m Pokervariante ohne
Generalisierung:

- Gegenspieler der immer mitgeht (CallPlayer)
 - Aufgrund falscher default - Einstellung zunächst schlecht, verbessert aber seine Strategie und ist nach 200.000 Runden auf Best-Response-Level
- Gegenspieler der immer erhöht (RaisePlayer)
 - Ähnlich wie CallPlayer, benötigt aber 1,8 Millionen Runden wegen größerem Entscheidungsraum
- Gegen NashPlayer nach 4 Millionen Runden ausgeglichen

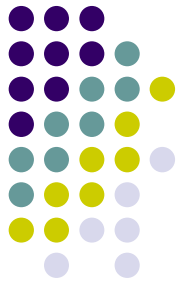
Opponent Modelling im Poker (4)



Test mit Texas Hold'em Variante mit Generalisierung als Approximation:

- Poki basiert auf heuristischem Ansatz
 - bei „full ring“ mittlere Spielstärke
 - schwach bei „heads-up“
 - Chancenlos nachdem (ca. 500 Runden) Schwachstelle aufgedeckt wurde
- PsOpti4 basiert auf spieltheoretischem Ansatz
 - Zunächst (ca. 5.000 Runden) ist PsOpti4 überlegen, ab ca. 150.000 Runden jedoch unterlegen
 - Nachgebessert nachdem Bug entdeckt wurde
- Keine signifikanten Tests gegen Menschen bisher

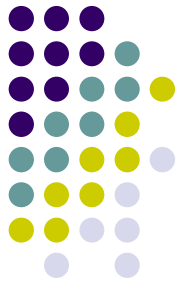
Fazit & Ausblick



- Techniken und Ansätze des Opponent Modelling vorgestellt
- Forschungsbedarf für Erweiterungen und Verbesserungen notwendig (Intuition, Interpretation)
- Beeindruckend gegen Poki, weltklasse Computerpokerspieler noch weit entfernt
 - Schnelleres Opponent Modelling
 - Exploration/ exploitation tradeoff finden
 - Wechselnde Strategien
 - Unabhängig von Spieleranzahl



Quellen

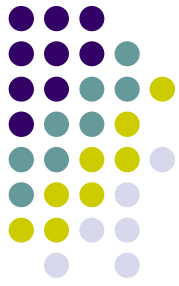


- T. Schauenberg: „**Opponent Modelling and Search in Poker**“. M.Sc.thesis, University of Alberta, 2006
- D. Koller, A. Pfeffer: „**Representations and Solutions for Game-Theoretic Problems**“. In Artificial Intelligence 94(1), 1997
- M. Bowling, M. Veloso: „**Rational and Convergent Learning in Stochastic Games**“. Int'l Joint Conference on Artificial Intelligence, Seattle, 2001.
- D. Billings, M. Bowling et al.: „**Game Tree Search with Adaptation in Stochastic Imperfect Information Games**“. In Computers and Games, 2004



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Vortragsende



- Vielen Dank für die Aufmerksamkeit!
- Fragen ???