# Moderne Methoden der KI

Prof. Dr. sc. Hans-Dieter Burkhard
Vorlesung Sommer-Semester 2008

**Modale Logiken**

## Einführung

Ausgangspunkt: Zustand des Agenten

- Welt-Modell (Umwelt, interne Parameter)
  *Vergangenheit - Gegenwart*
- Ziele, Aufgaben. Pläne
  *Zukunft*

Darstellung als Datenbank von Fakten
(logischer Formalismus)

Annahmen eines Agenten müssen nicht korrekt sein

## Einführung

Agent glaubt, dass Kronos der Vater von Zeus ist

Wie formalisieren?

Als Relation: belief(agent, vater(kronos,zeus))

- nicht mehr klassischer PK1:

Jupiter=zeus

belief(agent, vater(kronos,jupiter))

muss nicht gelten: belief-Logik ist **nicht extensional**

Darstellung durch für möglich gehaltene Welten:

Kartenspiel-Beispiel



## Modale Logik

**Definition: Ausdrücke für modale Aussagen-Logik**

AV := Menge der Aussagenvariablen.

(0) Jedes $p \in AV$ ist ein Ausdruck.

(1) Mit $H_1$, $H_2$ sind auch $\neg H_1$ und $H_1 \vee H_2$ Ausdrücke (und ggf. weitere)

Zusätzlich:

(2) Wenn H ein Ausdruck ist, so auch $\Box H$ ( und ggf. $\Diamond H$ ).

Definieren:

$$\Diamond p := \neg\Box\neg p$$

# Modale Logik

Modale Operatoren für *notwendig* und *möglich*

- □ p oder L p für: Es ist notwendig, dass p gilt.
- ◊ p oder M p für: Es ist möglich, dass p gilt.

Anstelle von *Notwendigkeit* auch

- *Verbindlichkeit, Zulässigkeit* bzw. *Verbot* ⇨ Deontische Logik
- *Wissen = knowledge* ⇨ Epistemische Logik
- *Annahmen, Überzeugung = belief* ⇨ Epistemische Logik
  = subjektiv: *für notwendig halten*
- *„immer"* = zu jeder Zeit *notwendig* ⇨ Temporale Logik
  korrespondiert mit *„manchmal"* = zu einer Zeit *möglich*



# Modale Logik

Möglichkeit bedeutet dabei die Nicht-Notwendigkeit des Gegenteils:

¬ □ ¬ p entspricht ◊ p
¬ know ¬ p entspricht „nicht wissen"
¬ belief ¬ p entspricht „nicht annehmen"
¬ verboten ¬ p könnte entsprechen erlaubt ¬ p
¬ immer ¬ p entspricht manchmal p

In epistemischen Logiken beziehen sich Ausdrücke mit ◊ p auf Aussagen über „nicht wissen" bzw. „nicht annehmen" (!)

# Zeitbezüge

*Am Montag nimmt der Agent an, dass am Dienstag bekannt ist, ob bis Donnerstag alle verfügbaren Flugreisen für den folgenden Montag verkauft sind.*
*Am Montag beschließt der Agent, dass am Dienstag festgelegt werden muss, ob bis spätestens Donnerstag neue Angebote an Flugreisen für den folgenden Montag beschafft werden müssen.*

wann gilt eine Annahme,

wann soll ein Plan ausgeführt werden,

wann wurde eine Annahme getroffen bzw. überprüft,

wann wurde ein Plan erstellt bzw. überprüft,

wann soll eine Annahme überprüft werden,

wann soll ein Plan überprüft werden.

…

**Quelle für Missverständnisse.**



# Modale Logik

- beliebig verschachtelte Modalitäten
  ich weiß, dass ich nicht weiß ... know (¬ know H)
  Fairness: immer gilt, dass irgendwann... immer(irgendwann H)

z.T. Reduktionsaxiome/-regeln, z.B.: $\Box\Box H \leftrightarrow \Box H$ im System S4 (s.u.)

- Verbindungen der Logiken
  er glaubt manchmal, dass er weiß ... irgendwann(beliefs(knows H))
- Gegenseitige Modellierung
  er weiß, dass sie weiß ... $know_a$ ( $know_b$ H)

## Modale Logik

„Wise Men - Puzzle" (J. McCarthy)

Der König will seine drei weisen Männer auf die Probe stellen. Sie müssen sich so in einem Kreis aufstellen, dass jeder die anderen sehen und hören kann. Er erklärt, dass er jedem einen Hut aufsetzen wird, dessen Farbe weiß oder schwarz sein kann, wobei aber wenigstens einer der Hüte weiß ist. Tatsächlich setzt er jedem einen weißen Hut auf. Dann stellt er ihnen der Reihe nach die Frage: „Weißt Du, welche Farbe Dein Hut hat?" Wie antworten die (wahrhaftig) weisen Männer?

Erster weiser Mann: „Ich weiß es nicht.“

Zweiter weiser Mann: „Ich weiß es nicht.“

Dritter weiser Mann: „Mein Hut ist weiß.“



## Semantik

Bel(H) als Gültigkeit von H in allen Welten, die der Agent betrachtet

- „sichtbare“ Welten
- die tatsächliche Welt muss nicht sichtbar sein

Bel(Bel(H)) usw.:

sichtbar aus sichtbarer Welt

Negation: ¬Bel(H)

in einer sichtbaren Welt gilt H nicht

# Kripke-Semantik, „possible-worlds“-Semantik

Rahmen: Mögliche Welten und Sichtbarkeits-Relationen,
formal als Graph ( „frame“, „Rahmen“):

$G = (Z, R)$ mit Knoten-Menge (Welten, Zuständen) $Z$
und Kanten-Menge $R \subseteq Z \times Z$

$R$ ist Sichtbarkeitsrelation (accessibility relation):
„$z_0$ sichtbar von $z$“, falls $R(z, z_0)$ .
Ausgangspunkt: Tatsächliche Welt des „Agenten“

Modell: $[G, \beta]$ mit
Rahmen $G = (Z, R)$ und Belegung $\beta : AV \times Z \to \{ true, false\}$
$\beta(p, z)$ = Wahrheitswert der AV $p$ im Knoten $z$ .
$\beta(p, z) = true$ bedeutet: $p$ ist wahr in der „Welt“ $z$ .



# Semantik

Erfüllbarkeitsrelation $\models$

$< M, z > \models H$ : $< M, z >$ erfüllt den Ausdruck $H$
$H$ ist wahr in der Welt $z$ des Modells $M = [(Z, R), \beta]$

Induktiv definiert durch:

$< M, z > \models p$ gdw. $p \in AV$ mit $\beta(p, z) = true$
$< M, z > \models \neg H$ gdw. nicht $< M, z > \models H$
$< M, z > \models H_1 \vee H_2$ gdw. $< M, z > \models H_1$ oder $< M, z > \models H_2$
$< M, z > \models \Box H$ gdw. $< M, z_0 > \models H$ für alle $z_0$ mit $R(z, z_0)$
$< M, z > \models \Diamond H$ gdw. $< M, z_0 > \models H$ für ein $z_0$ mit $R(z, z_0)$

offenbar gegenseitige Definierbarkeit von $\Box H$ und $\Diamond H$ :

$\Diamond H = \neg \Box \neg H$
$\Box H = \neg \Diamond \neg H$

# Semantik für modale Aussagen-Logik

$H$ ist gültig in $z \in Z$ von $M = [(Z, R), \beta]$ gdw. $< M, z > \models H$

$H$ ist gültig in $M = [(Z, R), \beta]$ gdw. $< M, z > \models H$ für alle $z \in Z$

$H$ ist gültig in einer Klasse $\boldsymbol{M}$ gdw. $H$ gültig für alle $M \in \boldsymbol{M}$

$H$ ist allgemeingültig gdw. $H$ gültig in allen Modellen $M$

(Schreibweise: $\models H$)

Analog:

$H$ ist erfüllbar in $M = [(Z, R), \beta]$ gdw. $< M, z >$ exist. mit $< M, z > \models H$

$H$ ist erfüllbar gdw. Modell $M$ exist. mit $H$ gültig in $M$



# Semantik

**Folgerungen:**

Bei Kripke-Semantik gilt stets:

(1) Wenn $\models H$ , so $\models \Box H$ (Notwendigkeitsregel)

(2) (K) $\Box (H_1 \rightarrow H_2) \rightarrow (\Box H_1 \rightarrow \Box H_2)$ (Distributions-Axiom)

## Weitere mögliche Forderungen

(müssen aber nicht gelten)

Axiom T $\Box H \rightarrow H$ (Notwendigkeits-Axiom)

äquivalent: $H \rightarrow \Diamond H$

Axiom D $\Box H \rightarrow \Diamond H$

äquivalent: $\Box H \rightarrow \neg\Box\neg H$

Axiom 4 $\Box H \rightarrow \Box\Box H$

Axiom 5 $\Diamond H \rightarrow \Box\Diamond H$



## Weitere mögliche Forderungen

Axiom T $\Box H \rightarrow H$

äquivalent: $H \rightarrow \Diamond H$

Axiom D $\Box H \rightarrow \Diamond H$

äquivalent: $\Box H \rightarrow \neg\Box\neg H$

Axiom 4 $\Box H \rightarrow \Box\Box H$

Axiom 5 $\Diamond H \rightarrow \Box\Diamond H$

Für Belief/Knowledge:

Axiom T knows $H \rightarrow H$

$H \rightarrow \neg$knows$\neg H$

(konsistent mit Realität)

Axiom D Bel $H \rightarrow \neg$Bel$\neg H$

(keine Inkonsistenz in Annahmen)

Axiom 4 Bel $H \rightarrow$ Bel Bel $H$

(positive Introspektion)

Axiom 5 $\neg$Bel $H \rightarrow$ Bel $\neg$Bel $H$

(negative Introspektion)

## Modallogische Theorien

**Syntaktisch** bestimmt (Ableitungen)

Th(X) = Ab(X)

Axiomenmenge

Ableitungsregeln

Bei Bezug auf Kripke-Modelle:

„Normale Logiken", enthalten

- Axiom K
- Notwendigkeitsregel

**Semantisch** bestimmte Theorien:

Für gegebene Klasse ***M*** von Modellen:

Th(***M*** ) = { H | H gültig in Klasse ***M***}

Beschreibung von Eigenschaften der Klasse ***M*** ,

z.B. bestimmte Sichtbarkeitsrelationen



## Modallogische Theorien

**Syntaktisch** bestimmte Theorien: Th(X) = Ab(X)

- Axiomensystem X: (als Axiomenschema mit allen Einsetzungen)
  - Tautologien des AK
  - Axiom K
  - weitere Axiome
- Ableitungsoperator Ab durch Ableitungsregeln festgelegt:
  - Abtrennungsregel (modus ponens):
    Wenn $Ab(H_1 \rightarrow H_2)$ und $Ab(H_1)$ , so $Ab(H_2)$
  - Notwendigkeitsregel:
    Wenn Ab(H) , so $Ab( \Box H )$

## Modallogische Theorien

**Syntaktisch** bestimmte Theorien: Th(X) = Ab(X)

- Spezielle Axiome für modale Operatoren
  (als Axiomenschemata mit allen Einsetzungen)

Axiom K $\Box (H_1 \rightarrow H_2) \rightarrow (\Box H_1 \rightarrow \Box H_2)$ (Distributions-Axiom)

Axiom T $\Box H \rightarrow H$ (Notwendigkeits-Axiom)

äquivalent: $H \rightarrow \Diamond H$

Axiom D $\Box H \rightarrow \Diamond H$ (folgt aus T )

Axiom 4 $\Box H \rightarrow \Box \Box H$

Axiom 5 $\Diamond H \rightarrow \Box \Diamond H$

Kombination der Axiome
führt zu speziellen „Systemen“ modaler Logiken, z.B.

System T: Ab(„Tautologien + K + T“)



## Modallogische Theorien

**Semantisch** bestimmte Theorien:

Th(***M*** ) = { H | H gültig in der Klasse ***M*** von Modellen }

Modell-Klassen ***M*** bestimmt durch Eigenschaften
der Sichtbarkeitsrelation R in G = [Z, R] :

- reflexive Sichtbarkeits-Relation $\forall x \in Z$: $R(x,x)$
- serielle Sichtbarkeits-Relation $\forall x \in Z\ \exists y \in Z$: $R(x,y)$
- transitive Sichtbarkeits-Relation $\forall x,y,z \in Z$: $R(x,y) \wedge R(y,z) \rightarrow R(x,z)$
- euklidische Sichtbarkeits-Relation $\forall x,y,z \in Z$: $R(x,y) \wedge R(x,z) \rightarrow R(y,z)$

# Modallogische Systeme

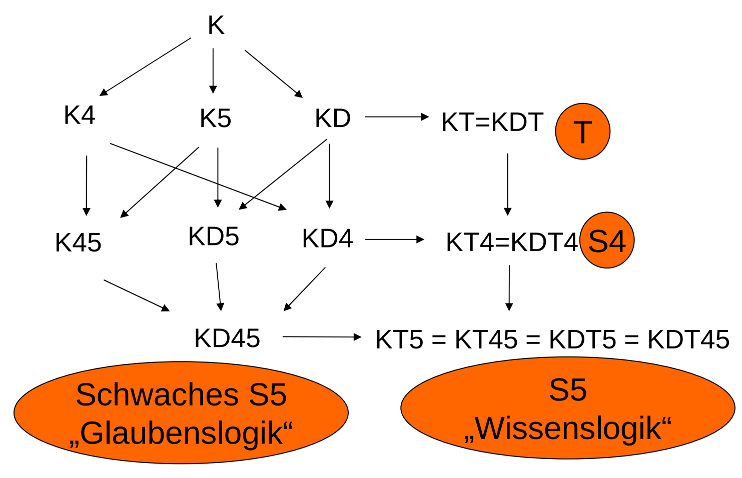
Alle semantisch über Kripke-Modelle definierten Logiken sind normal.
Ferner gelten die Axiome

T  bei reflexiver Sichtbarkeits-Relation R in G = [Z, R]
D  bei serieller Sichtbarkeits-Relation R in G = [Z, R]
4  bei transitiver Sichtbarkeits-Relation R in G = [Z, R]
5  bei euklidischer Sichtbarkeits-Relation R in G = [Z, R]

Umkehrungen gelten in folgender Form:
Sei G=[Z,R] ein Rahmen.
Wenn Axiom XXX bei **jeder** Belegung β (in jedem Modell M=[G,β] ) gilt, dann muss R die korrespondierende Bedingung erfüllen.



# Modallogische Systeme



## Syntaktisch definierte Systeme

Durch Kombination der weiteren Axiome T, D, 4, 5
ergeben sich 11 unterschiedliche normale Systeme:

| | |
|---|---|
| K | |
| K4 | |
| K5 | |
| KD | |
| KT = KDT | System T |
| K45 | |
| KD4 | |
| KD5 | |
| KT4 = KDT4 | System S4 |
| KD45 | schwaches System S5 |
| KT5 = KT45 = KDT5 = KDT45 | System S5 |



## Semantisch definierbare Systeme

mit Bezug auf Rahmen (statt Modellen wie oben) gilt:

System T: durch reflexive Sichtbarkeits-Relation R .

System S4: durch reflexive und transitive Sichtbarkeits-Relation R .

System S5: durch reflexive, transitive und symmetrische Sichtbarkeits-Relation R .
(R ist Äquivalenz-Relation:
Klassen gegenseitig sichtbarer Welten).

schwaches System S5: durch serielle, transitive und euklidische Sichtbarkeits-Relation R.

## System T

Definiert durch
- semantisch durch reflexive Sichtbarkeitsrelation oder
- syntaktisch als normale Logik mit zusätzlich
  Axiom T $\Box H \rightarrow H$ (Notwendigkeits-Axiom)

Bereits in System T gilt:

$\Box\Diamond H \rightarrow \Diamond H$ (Umkehrung von Axiom 5 : $\Diamond H \rightarrow \Box \Diamond H$ )

$\Box\Box H \rightarrow \Box H$ (Umkehrung von Axiom 4 : $\Box H \rightarrow \Box \Box H$ )

$\Box H \rightarrow \Diamond \Box H$ (Umkehrung folgt aus Axiom 5)

$\Diamond H \rightarrow \Diamond \Diamond H$ (Umkehrung folgt aus Axiom 4)

⇨ Reduktionsmöglichkeiten für verschachtelte Modalitäten in S4, S5.



## Epistemische Logiken

Notwendigkeit interpretieren als
- know p wissen
- bel p annehmen (überzeugt sein, glauben)
- erlaubt sein
- verpflichtet sein (deontische Logik)

Möglichkeit bedeutet dabei die Nicht-Notwendigkeit des Gegenteils.

$\neg \Box \neg p$ entspricht $\Diamond p$

$\neg$ know $\neg$ p entspricht „nicht wissen“

$\neg$ belief $\neg$ p entspricht „nicht annehmen“

$\neg$ erlaubt $\neg$ p entspricht verboten $\neg$ p

Bzw. $\Diamond p$ bedeutet „nicht wissen“, „nicht annehmen“ usw.

## Wissenslogik

Axiom K  $knows\ (H_1 \rightarrow H_2) \rightarrow (knows\ H_1 \rightarrow knows\ H_2)$

Axiom T  $knows\ H \rightarrow H$  (Korrektheits-Axiom)

Axiom D  $knows\ H \rightarrow \neg\ knows\ \neg\ H$  (Konsistenz-Axiom)

Axiom 4  $knows\ H \rightarrow knows\ knows\ H$  (positive Introspektion)

Axiom 5  $\neg\ knows\ H \rightarrow knows\ \neg\ knows\ H$ (negative Introspektion)

Axiome 4 + 5 : „Perfekte Selbstkenntnis“



## Glaubenslogik

Axiome ohne T

Axiom K  $bel\ (H_1 \rightarrow H_2) \rightarrow (bel\ H_1 \rightarrow bel\ H_2)$

Axiom D  $bel\ H \rightarrow \neg\ bel\ \neg\ H$  (Konsistenz-Axiom)

Axiom 4  $bel\ H \rightarrow bel\ bel\ H$  (positive Introspektion)

Axiom 5  $\neg\ bel\ H \rightarrow bel\ \neg\ bel\ H$ (negative Introspektion)

(Möglicher Zusammenhang: $knows\ H \leftrightarrow H \wedge bel\ H$ )

Häufig: System S5 als Wissenslogik
Schwaches System S5 als Überzeugungslogik.

## Omniszenz-Problem

Notwendigkeitsregel:

Wenn H allgemeingültig ist, so ist knows H allgemeingültig .

D.h.: Wissen umfasst mindestens alle Tautologien!

Weiter:

- Sei H Folgerung aus Axiomenmenge ***X*** $= \{H_1,\ldots,H_n\}$

  d.h. $H_1 \wedge\ldots\wedge H_n \rightarrow H$ ist allgemeingültig.
- Notwendigkeitsregel: knows $(H_1 \wedge\ldots\wedge H_n \rightarrow H)$ allgemeingültig.
- Axiome sind bekannt: knows $(H_1 \wedge\ldots\wedge H_n)$ .
- Gemäss Axiom K:

  knows$(H_1 \wedge\ldots\wedge H_n \rightarrow H)$ und knows$(H_1 \wedge\ldots\wedge H_n)$ implizieren knows H.

D.h.: Mit Axiomenmenge ***X*** sind alle Folgerungen bekannt!



## Omniszenz-Problem

In normalen Logiken (possible-worlds-Semantik) gelten stets Axiom K und die Notwendigkeitsregel und damit „Omniszenz“.

Syntaktischer Zugang ohne Axiom K und Notwendigkeitsregel muss folglich andere Semantik haben.

Alternative Ansätze:
Levesque: explizites und implizites (ableitbares) Wissen unterscheiden.
Konolige: Wissensbasis + Inferenzregeln bestimmen das Wissen

In der Praxis häufig Beschränkung der Ausdrucksstärke
(z.B. nur Literale)

# Epistemische Logiken

In Multi-Agenten-Systemen:

Wissen/Glauben jedes einzelnen Agenten a:

Agent a weiß, dass H gilt: $knows_a$ H .

Agent a nimmt an, dass H gilt: $bel_a$ H .

Wissen über andere: $knows_a\ knows_b$ H

Gemeinsames Wissen „common knowledge"
(analog: common belief )

(vgl. wise man puzzle)

Absprache-Probleme, Protokolle



# (Statische) Ziele

goal H ebenfalls mittels Kripke-Modell beschreibbar:

$\Box$ p entspricht goal p

Eigenschaften, die unbedingt gelten sollen

$\Box \neg$ p = goal $\neg$ p

Eigenschaften, die keinesfalls gelten sollen

$\Diamond$ p = $\neg \Box \neg$ p = $\neg$ goal $\neg$ p

Eigenschaften, die nicht gefordert werden

Konsistenz bzgl. Annahmen (für den entsprechenden Zeitpunkt) kann in unterschiedlichen Formen und Bedeutungen gefordert werden, z.B.:

- Goal-Modell als Teilgraph des Belief-Modells oder
- goal p$\rightarrow \neg$ bel $\neg$ p (Möglichkeit des Erreichens)
- goal p$\rightarrow$ bel (goal p) (Agent kennt seine Ziele)

# Probleme

**a) Zeitliche Veränderungen** modellieren ( → temporale Logiken)

**b) Seiteneffekte**
Abschluss bzgl. logischer Konsequenzen (vgl. Omniszenz-Problem):

Nach Notwendigkeitsaxiom:
$H_1 \rightarrow H_2$ impliziert goal ($H_1 \rightarrow H_2$)

nach K-Axiom:
goal ($H_1 \rightarrow H_2$) und goal $H_1$ implizieren goal $H_2$
d.h. Konsequenzen sind auch Ziele ...

( → evtl. keine volle logische Inferenz ( → temporale Logiken))



# Zeitliche Abläufe

Aktionen zum Erreichen von Zielen an zeitliche Abläufe gebunden

Vorgaben bzgl. Modellierung zeitlicher Abläufe:
- Verzweigt vs. linear
- Zeitpunkte vs. Intervalle
- Diskret vs. kontinuierlich
- Zukunft (Vergangenheit): beschränkt vs. unbeschränkt
- Zyklisch vs. azyklisch

Modelle analog zu Kripke-Modellen [G,β]
mit Zeitstrukturen G=[Z,R], wobei R zeitliche Nachfolge beschreibt
- R ist transitiv, reflexiv (partielle Ordnung)
- Lineare Zeit: R ist Ordnungsrelation

## Temporale Logik

Operatoren bzgl. linearer Zeit:

$P\,\phi$ Irgendwann in der Vergangenheit galt $\phi$ .

$F\,\phi$ Irgendwann in der Zukunft wird $\phi$ gelten.

$H\,\phi$ Immer in der Vergangenheit galt $\phi$.

$G\,\phi$ Immer in der Zukunft wird $\phi$ gelten.

$P$ und $F$ analog zu $\Diamond$, $H$ und $G$ analog zu $\Box$.

$$P\,\phi = \neg H \neg \phi$$

$$F\,\phi = \neg G \neg \phi$$

(aktueller Zeitpunkt wird immer eingeschlossen)

weitere Operatoren:

$X\,\phi$ ("next") Im nächsten Zeitpunkt gilt $\phi$.

$\phi\, U\, \psi$ ("until") irgendwann gilt $\psi$ , (mindestens) bis dann gilt $\phi$



## Temporale Logik

Definierbarkeit von $F/G$ mittels $X/U$:

$$F\,\phi = \text{true}\ U\ \phi$$

$$G\,\phi = \neg F \neg \phi$$

weitere:

unendlich oft: $F^{\infty}\,\phi =_{Df} G\,F\,\phi$

fast überall: $G^{\infty}\,\phi =_{Df} F\,G\,\phi$

$\phi$ vor $\psi$ : $(\phi\ B\ \psi) =_{Df} \neg\,(\,(\neg\,\phi)\ U\ \psi\,)$

## Handlungen sind zeitliche Abläufe

- Zusammenhang von Absichten/Plänen mit Realität
- Rationaler Agent soll keine „unvernünftigen“ Absichten haben (Übereinstimmung mit Realität, Persistenz, Abbruch)

Beispiele: Roboter Willie, Little Nell´s Problem, Cheating Husbands

Kombinationen in multi-modalen temporalen Systemen:

desire (AF Gewinnen) & intend (EF Loskaufen) & not belief (AF Gewinn)

mögliche Axiome z.B.

$\text{intend}(\phi) \wedge \text{belief}(\phi \to \psi) \to \text{intend}(\psi)$

$\text{intend}(\phi) \wedge \neg \phi \to X\, \text{intend}(\phi)$

$\text{intend}(\phi) \wedge \text{can}(\phi) \to F\, \phi$

Oftmals problematisch, z.B, bzgl. weiterer Konsequenzen (Seiteneffekte).
Vgl. Frameproblem und Persistenzformeln in „logic of action and change“



## Roboter Willie (in Anlehnung an Cohen-Levesque)

Willie bring Bier - Willie beschäftigt mit anderen Dingen, aber „irgendwann wird er Bier bringen“
*Fehler bzgl. Ziele verfolgen*

Willie bring Bier - Willi wartet vor dem leeren Kühlschrank
*Fehler bzgl. unerreichbare Ziele verfolgen*

Willie bring Bier - Willi holt Bier aber verschüttet es unterwegs absichtlich
*Fehler bzgl. Ziele verfolgen*

Willie bring Bier und Brötchen - Willi findet keine Brötchen und verzichtet auf weitere Arbeit
*Fehler bzgl. Teil-Ziele verfolgen*

Willie bring Bier - Kühlschrank leer, aber Willi bringt Öffner und Glas
*Fehler bzgl. Adaption von Teil-Zielen*

## Cheating Husbands

The queens of the matriarchal city-state of Mamajorca, on the continent of Atlantis, have a long record of opposing and actively fighting the male infidelity problem. Ever since the technologically-primitive days of queen Henrietta I, women in Mamajorca have been required to be in perfect health and pass an extensive logic and puzzle-solving exam before being allowed to take a husband. The queens of Mamajorca, however, were not required to show such competence.

It has always been common knowledge among the women of Mamajorca that their queens are truthful and that the women are obedient to the queens. It was also common knowledge that all women hear every shot fired in Mamajorca. Queen Henrietta I awoke one morning with a firm resolution to do away with the male infidelity problem in Mamajorca. She summoned all of the women heads-of-households to the town square and read them the following statement:

## Cheating Husbands

*There are (one or more) unfaithful husbands in our community. Although none of you knew before this gathering whether your own husband was faithful, each of you knows which of the other husbands are unfaithful. I forbid you to discuss the matter of your husband's fidelity with anyone. However, should you discover that your husband is unfaithful, you must shoot him on the midnight of the day you find out about it.*

What happens?

## Cheating Husbands

*There are (one or more) unfaithful husbands in our community. Although none of you knew before this gathering whether your own husband was faithful, each of you knows which of the other husbands are unfaithful. I forbid you to discuss the matter of your husband's fidelity with anyone. However, should you discover that your husband is unfaithful, you must shoot him on the midnight of the day you find out about it.*

Thirty nine silent nights went by, and on the fortieth night, shots were heard.



## Temporale Logik

Semantik:

- Welten (Mengen gültiger Aussagen) zu unterschiedlichen Zeiten
- Sichtbarkeitsrelation an Zeitmodelle angepasst (z.B. linear)

Umschreibung mit klassischen Mitteln z.B.:

$\exists$ t ( later(t, now) $\wedge$ $\phi$(t) ) für F $\phi$

$\forall$ t ( later(t, now) $\rightarrow$ $\phi$(t) ) für G $\phi$

Probleme (?), z.B.:

a) Unterschied bzgl.: F ($\exists$ x H(x)) und $\exists$x( F H(x))

kein Unterschied bei Umschreibung im PK1

b) F F $\phi$ $\rightarrow$ F $\phi$ ist allgemeingültig, Umschreibung im PK1 nicht.

# Verzweigte Zeit

Operatoren für verzweigte Zeit (bzgl. Zukunft)

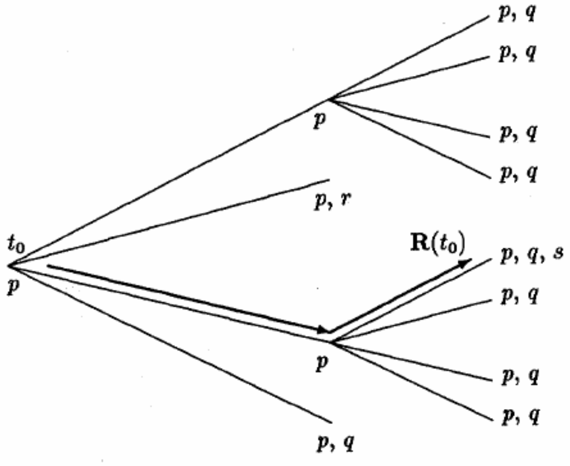
– A ϕ für „always“: In jeder Zukunft gilt ϕ.
– E ϕ für „exists“: In einer Zukunft gilt ϕ.

In Kombination mit Operatoren für lineare Zeit:
„Computational Tree Logic“ (CTL bzw. CTL*)



# Verzweigte Zeit

Beispiel aus M.P. Singh: Multiagent systems



## CTL* (Emerson)

(extended Computational Tree Logic, full branching time logic)

- Operatoren für verzweigte Zeit
  - $A\ \phi$ für „always“: In jeder Zukunft gilt $\phi$.
  - $E\ \phi$ für „exists“: In einer Zukunft gilt $\phi$.
- Operatoren bezüglich eines Weges (lineare Zeit)

  $X\ \phi$ ("next") Im nächsten Zeitpunkt gilt $\phi$.

  $\phi\ U\ \psi$ ("until") Zunächst gilt $\phi$, danach $\psi$ .



## Syntax CTL*

$AV$ = Menge der Aussagenvariablen.

- AK-Ausdrücke wie üblich.
- Zustandsformeln (Betrachtung der Wege von einem Zustand aus):

  (Z1) Jedes $p \in AV$ ist eine Zustandsformel.

  (Z2) Mit $\phi$ , $\psi$ sind auch $(\neg \phi)$ und $(\phi \vee \psi)$ Zustandsformeln.

  (Z3) Wenn $\phi$ Wegformel ist, so sind $E\ \phi$ und $A\ \phi$ Zustandsformeln.
- Wegformeln: (Betrachtung der Zustände auf einem Weg)

  (W1) Jede Zustandsformel ist eine Wegformel.

  (W2) Mit $\phi$ , $\psi$ sind auch $(\neg \phi)$ und $(\phi \vee \psi)$ Wegformeln.

  (W3) Mit $\phi$ , $\psi$ sind auch $X\ \phi$ und $\phi\ U\ \psi$ Wegformeln.

Bemerkung: CTL ergibt sich mit (Z1)-(Z3) sowie

(W) Wenn $\phi,\psi$ Zustandsformeln sind, so sind $X\phi$ und $\phi U\psi$ Wegformeln.

# Semantik für Temporale Logiken

Allgemein Kripke-Semantik: Modell $[G, \beta]$
mit Zeitstruktur $G = (Z, R)$ und Belegung $\beta : AV \times Z \rightarrow \{ true, false\}$

$G = (Z, R)$ mit Knoten-Menge (Welten, Zuständen) $Z$
und Kanten-Menge $R \subseteq Z \times Z$
$R$ ist Sichtbarkeitsrelation (accessibility relation):
hier: $z_0$ zeitlich vor bzw. nach $z$ , falls $R(z, z_0)$

i.a. als (partielle) Ordnung: transitiv, reflexiv, antisymmetrisch

Ausgangspunkt: Aktuelle Welt des „Agenten“

$\beta(p, z)$ = Wahrheitswert der Aussagenvariablen p im Knoten $z$ .
$\beta(p, z) = true$ bedeutet: $p$ ist wahr in der „Welt“ $z$



# Semantik für CTL*

Erfüllbarkeitsrelation $|=$
induktiv definieren gemäß Syntax-Definition Z1-Z3, W1-W3

$< M, z >|= \phi$ : $< M, z >$ erfüllt Zustandsformel $\phi$
$\phi$ ist wahr in der Welt $z$ im Modell $M = [(Z, R), \beta]$

$< M, w >|= \phi$ : $< M, w >$ erfüllt Wegformel $\phi$
$\phi$ ist wahr auf dem Weg $w$ im Modell $M = [(Z, R), \beta]$

dabei ist $w = z_0 z_1 z_2 \ldots$ ein Weg in $G = (Z, R)$

# Semantik für CTL*

(Z1) $< M, z_0 > \models p$ gdw. $p \in AV$ mit $\beta(p, z_0) = true$

(Z2) $< M, z_0 > \models \neg \phi$ gdw. nicht $< M, z_0 > \models \phi$

$< M, z_0 > \models \phi_1 \vee \phi_2$ gdw. $< M, z_0 > \models \phi_1$ oder $< M, z_0 > \models \phi_2$

(Z3 ) $< M, z_0 > \models E\, \phi$ gdw. Ein Weg $w = z_0 z_1 z_2 \ldots$ in $M$ exist.
mit $< M, w > \models \phi$

$< M, z_0 > \models A\, \phi$ gdw. Für alle Wege $w = z_0 z_1 z_2 \ldots$ in $M$
gilt $< M, w > \models \phi$

(W1) $< M, w > \models \phi$ gdw. $w = z_0 z_1 z_2 \ldots$ und $< M, z_0 > \models \phi$

(W2) $< M, w > \models \neg \phi$ gdw. nicht $< M, w > \models \phi$

$< M, w > \models \phi_1 \vee \phi_2$ gdw. $< M, w > \models \phi_1$ oder $< M, w > \models \phi_2$

(W3) $< M, w > \models X\, \phi$ gdw. $w = z_0 z_1 z_2 \ldots$ und $<M, z_1 z_2 \ldots> \models \phi$

$< M, w > \models \phi_1\, U\, \phi_2$ gdw. $w = z_0 z_1 z_2 \ldots$ in $M$ und es gibt $i \geq 0$
mit $<M, z_i z_{i+1} \ldots> \models \phi_2$ und
$<M, z_k z_{k+1} \ldots> \models \phi_1$ f.a. $k$ mit $0 \leq k < i$



# Absichten (Intentions)

Unterschiedliche Begrifflichkeiten (Philosophie):

- Ziele
- Absichten
- Wünsche
- Pläne

Beschreiben: Handlung vs. Zustand

Hier „Absicht“ im Sinne von

- Gewollte Handlungen (self commitment)
  (zum zukünftigen Erreichen eines gewollten Zustandes)
- Plan

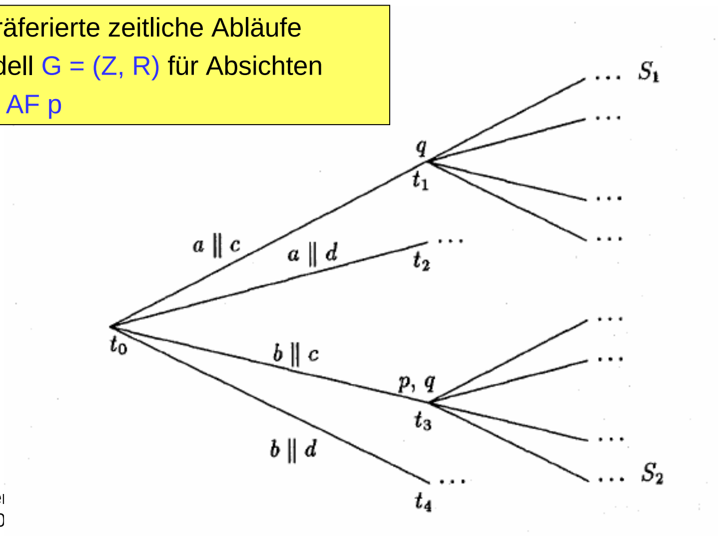
Unterschied:

- Erreichen eines Zustands (achievement)
- Sichern von Bedingungen (maintenance)

# Modell für Absichten

Beispiel aus M.P. Singh: Multiagent Systems

Intentions als präferierte zeitliche Abläufe
z.B. mit Zeitmodell G = (Z, R) für Absichten
intend p := AF p






# Modell für Absichten

Konsistenzforderung bzgl. Annahmen z.B. beschreiben durch :

– Intention-Modell ⊆ Belief-Modell (ebenfalls als Zeitmodell)

$G_{intend} \subseteq G_{bel}$

– intend p ( = $A_{intend} F_{intend}$p ) $\rightarrow E_{bel}\, F_{bel}\;$ p

Weitere Notationen bzgl. Aktionen und Agenten:
does, done, succeeds, ...

(Wissen über) Fähigkeiten:
can, ...

# Absichten (Intentions)

**Rationalitätseigenschaften:**

1. Möglichkeit zum Erfolg (Konsistenz zu Annahmen bzgl. Zukunft, Future directed belief)

   Agent soll nichts unerreichbares beabsichtigen und ggf. aufgeben

   (Problem: wann genau aufgeben?)
2. Konsistenz von parallelen Absichten eines Agenten (folgt aus 1)
3. Konsistenz bzgl. notwendiger Mittel zu einem gewollten Zustand
4. Fähigkeiten zur möglichen Erreichung des gewollten Zustands müssen vorhanden sein

Begrenzte Fähigkeiten, aber optimistisch



# Absichten (Intentions)

Commitment (Verpflichtung) und Persistenz (Dauerhaftigkeit):

(Cohen-Leseque: Intention = choice with commitment)

- Commitment des Agenten auch bei veränderter Umwelt

  Aber: Nicht mehr bei Unerreichbarkeit des gewollten Zustands
- Aufhören, wenn gewollter Zustand erreicht ist.

Absichten sind nicht notwendig erfolgreich

(future belief: gewollter Zustands möglich, aber nicht notwendig)

# Absichten (Intentions)

Strittig:

Abschluss bzgl. logischer Konsequenz bzw. angenommenen Folgen (believed effects)

Sind logische Folgerungen der Absichten auch beabsichtigt?

(Seiteneffekte)

Verantwortung für Seiteneffekte?

Verselbständigung der Seiteneffekte (Schmerz)?



# Wünsche (Desires)

Schwächere Anforderungen bzgl. Rationalität als bei Absichten.

Erlaubt sind:

- Inkonsistenz bei mehrfachen Absichten
- Inkonsistenz bzgl. Annahmen
- Inkonsistenz bzgl. Mitteln

- In einigen Theorien als Vorstufe zu Intention (BDI),
- In anderen nicht vorhanden

# Fazit

Benötigen Modelle
(Entwickler, anderer Entwickler, Nutzer, Agent:
Unterschiedliche Perspektiven!)

Agent benötigt Datenstrukturen für

- Annahmen (belief)
- Ziele, ... (goal, ...)

Die Datenstrukturen müssen

- Weltsituationen beschreiben (zu unterschiedlichen Zeiten)
- zeitliche Aktualisierungen erlauben (eingeschränkte Inferenz, i.a. nicht volle Logik)
- bestimmte Konsistenzbedingungen erfüllen (welche?)



# Fazit

Multi-modale temporale Logiken beschreiben Probleme, z.B.

- verzweigende Zeit
- mögliche Inkonsistenzen
- Konsequenzen (erwünschte, unerwünschte)

Umsetzung in klassischer Softwaretechnik durch

- Datenstrukturen
- Inferenzverfahren (i.a. schwächer als in modaler Logik)

# Fazit

Agentenprogrammierung als Ausführung logischer Spezifikationen:

**Concurrent MetateM** (Michael Fisher)

- Arbeit analog zu Regelsystemen: Update entsprechend Umwelt
- Regeln als temporale Formeln: P -> F
- Interpreter protokolliert bisherigen Verlauf (history):
  - Vorbedingungen bzgl. history testen
  - Nachbedingung als commitment: Agent führt Aktionen aus

- Programm ist Spezifikation des Verhaltens.
- Agent führt Spezifikation aus und erzeugt iterativ ein Modell der Spezifikation