Intelligent Agents

Doxastic logic and dynamics of beliefs

Özgür L. Özçep

Universität zu Lübeck

Institut für Informationssysteme

Todays lecture based on

- Parts of Lecture notes „EPISTEMIC LOGICS“ by Andreas Herzig, 2017
  https://www.irit.fr/~Andreas.Herzig/Cours/epiLogics.pdf

DOXASTIC LOGIC

Relevance of Knowledge

- When is knowledge the appropriate informational attitude?
- Remember: “knowledge entails truth” principle in epistemic logic: $\vDash_{S5n}\ K_a\phi \rightarrow \phi$
- Relevant for:
  - formal epistemology
    - What is knowledge?
    - Is knowledge possible at all?
    - Are all truths knowable?
  - Distributed processes (Fagin et al 03)

Truth

- Relation to truth less in focus in:
  - philosophy of mind: focus on agent's mental state
  - philosophy of language: effects of speech acts on the participants' mental states: lies, bullshitting
  - implementation of artificial agents
- informational mental attitude not entailing truth: belief
  - "he knows that ϕ, but he is wrong": inconsistent
  - "he believes that ϕ, but he is wrong": consistent
  - however: 'belief aims at truth' (Engel 1998), (Hakli 2006)
- Doxastic logic (Hintikka 2005) (Lenzen 1978, Lenzen 1995)
  - doxa = δοξα = 'belief' (Greek)

Definition (Syntax of Doxastic Logic: KD45n)

- Well-formed formula of doxastic logic are given by BNF:

$$\phi ::= p \mid \bot \mid \neg\phi \mid (\phi \wedge \phi) \mid B_a\phi$$

  where $p \in AP$ and $a \in AGT$.
- Intended reading: $B_a\phi$ ``agent a believes ϕ“
- Dual operator: $\widehat{B}_a$ abbreviates $\neg B_a \neg\phi$

  ``it is possible for a that ϕ“

Example

- $p \wedge B_a \neg p$
- $\mathrm{B}_\mathrm{a} \neg p \wedge B_b B_a p$
- $B_a(B_b p \vee B_b \neg p)$

Doxastic attitudes and situations

- Three possible doxastic attitudes w.r.t. a formula ϕ

$$B_a\phi \quad \hat{B}_a\phi \wedge \hat{B}_a\neg\phi \quad B_a\neg\phi$$

  for ϕ contingent (not tautology and not contradiction) and non-doxastic
- Six possible doxastic situations w.r.t. a formula ϕ

$$\phi \wedge B_a\phi \qquad \phi \wedge \hat{B}_a\phi \wedge \hat{B}_a\neg\phi \qquad \phi \wedge B_a\neg\phi$$

$$\neg\phi \wedge B_a\phi \quad \neg\phi \wedge \hat{B}_a\phi \wedge \hat{B}_a\neg\phi \quad \neg\phi \wedge B_a\neg\phi$$

  for ϕ contingent (not tautology and not contradiction) and non-doxastic

Semantics

Belief explained (as for knowledge) with possible worlds

$B_a\phi$ = „agent a believes that ϕ"

=„ϕ true in every world that is compatible with a's beliefs"

Definition (Models of KD45n)

A $KD45_n$-model is a structure $\mathcal{M} = (W, B, V)$ where

- W nonempty set (of possible worlds)
- $V: AP \rightarrow 2^W$ (valuation)
- $\mathcal{R}: AGT \rightarrow 2^{W \times W}$ such that for every $a \in AGT$:
  - For every w there is some w' such that $(w, w') \in \mathcal{R}_a$ (serial)
  - If $(w, w') \in \mathcal{R}_a$ and $(w', w'') \in \mathcal{R}_a$, then $(w, w'') \in \mathcal{R}_a$ (transitive)
  - If $(w, w') \in \mathcal{R}_a$ and $(w, w'') \in \mathcal{R}_a$, then $(w', w'') \in \mathcal{R}_a$ (Euclidean)

Some derived notions and observations

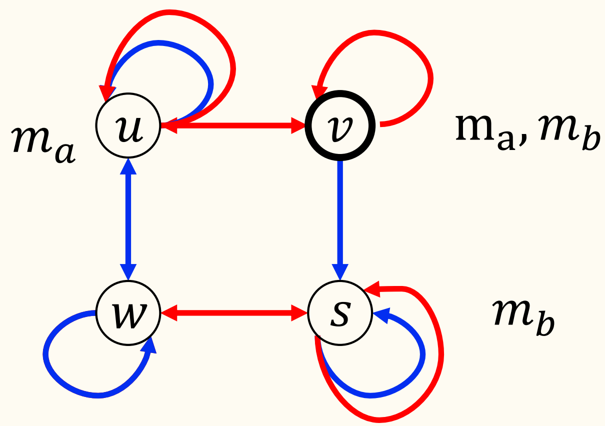
- $\mathcal{R}_a(w) = \{ w' \mid (w, w') \in \mathcal{R}_a \}$
  - $=$ a's alternatives to w
  - $=$ worlds a cannot distinguish from w on basis of beliefs
  - $=$ set of worlds compatible with a's beliefs
  - $=$ belief state of agent a at w
- $\mathcal{R}_a$ serial iff $\mathcal{R}_a(w) \neq \emptyset$
- $\mathcal{R}_a$ transitive and Euclidean iff:
  if $w' \in \mathcal{R}_a(w)$ then $\mathcal{R}_a(w) = \mathcal{R}_a(w')$

Definition (modellig relation in KD45n)

$\mathcal{M}, w \vDash B_a\phi$ iff $\mathcal{M}, w' \vDash \phi$ for every $w' \in \mathcal{R}_a(w)$

Example (Variant of muddy children (a, b) with beliefs)

Child a wrongly believes it is not muddy



$R_a(v) = \{s\}$

$M, v \vDash m_a \wedge B_a \neg m_a$

Axiomatics

Definition (A calculus for multimodal KD45n)

- Axioms for multimodal K
  - Axioms for propositional logic
  - Axiom $\mathrm{B}_a\phi \wedge B_a\psi \rightarrow B_a(\phi \wedge \psi)$ Axiom M(B_a)
  - Rule: $\phi \rightarrow \psi \quad \vdash_{KD45_n} \quad B_a\phi \rightarrow B_a\psi$
- Consistency of Belief: $\neg(B_a\, \phi \wedge B_a\neg\phi)$ Axiom $\mathrm{D}(B_a)$
- Positive Introspection: $B_a\phi \rightarrow B_aB_a\phi$ Axiom $4(B_a)$
- Negative Introspection: $\neg B_a\phi \rightarrow B_a\neg B_a\phi$ Axiom $5(B_a)$

Axiomatics

Theorem (Properties of calculus)

- Sound and complete: $\vdash_{KD45_n}$ iff $\vDash_{KD45_n}$
- Decidable
- Complexity of $KD45_n$-satisfiability
  - NP-complete if $n = 1$
  - PSPACE-complete if $n > 1$
- For $n = 1$ there exists a normal form: modal depth ≤ 1

Discussion: Omniscience problem

- Closure of B_a under inference (see only rule in calculus)
- This is not realistic - in particular for ressource bounded agents.

- (Negative) Introspection also criticised (Lenzen 78)

Discussion: belief and probability

- $KD45_n$'s notion of belief is strong („conviction")
- Weaker version:
  - $\mathrm{B}_a\phi = Prob_a(\phi) > Prob_a(\neg\,\phi)$
  - For classical semantics this amounts to $Prob_a(\phi) > \frac{1}{2}$
- Semantics: $\mathcal{M} = (W, \mathcal{R}, V)$ where

  $$\mathcal{R}: AGT \rightarrow (W \times W)$$

- $\mathcal{M}, w \vDash B_a\phi$ iff among the a-accessible worlds there are more ϕ than $\neg\phi$ worlds
  - $(B_a\phi \wedge B_a\psi) \rightarrow B_a(\phi \wedge \psi)$ not valid!
  - Weakening of Kripke semantics: neighbourhood semantics (Burgess 1969), (Lenzen 1978)

Discussion: Graded Belief

- Language: $B_a^{\geq d}\phi =$ „a believes ϕ with degree at least d
  (where $d \in [0,1]$)
- Semantics: $\mathcal{M} = (W, \mathcal{R}, V)$ where
  - $\mathcal{R}$: $AGT \times [0,1] \rightarrow (W \times W)$ such that $B_a^{\geq d}\phi \subseteq B_a^{\geq d+d'}\phi$
    Linear chain of accesibility relations
    (-> „system of spheres“)
  - $wB_a^{\geq d}v =$ „for a at w world v has degree of possibility at least d
- Axiomatics:
  - $KD45(B_a^{\geq d})$ for every a and d
  - $B_a^{\geq d}\,\phi \rightarrow B_a^{\geq d'}\phi$ if $d \geq d'$

KNOWLEDGE VS BELIEF

Can knowledge be defined from belief?

- The antique definition according to Platon (Theaetetus)
  - $K_a\phi = B_a\phi \wedge \phi$ ...
    Problem: knowledge by accident
  - $K_a\phi = B_a\phi \wedge \phi \wedge hasJust(a, \phi)$
    „Knowledge is justified true belief“

- Held to be true for more than 2000 years
- And then comes Gettier
  - Fun fact: idea written on napkin
  - leading to a highly influential 2 page paper (in analytical philosophy) (Gettier 1963)

Gettiers two counterexamples

Scenario 1

- Smith and Jones apply for a job
- Smith believes (justifiably):
  (p) Jones will get the Job &
  John has ten coins in his pocket
- Smith believes also in the entailed assertion:
  (r) The one who gets the job has ten coins in his pocket.
- Coincidence : Smith gets the job and Smith has ten coins in his pocket.
- Smith „knew“ (r) only by chance

Scenario 2

- Smith justifiably believes
  (p) Jones owns a Ford
- Smith also believes in entailed assertion
- (r) = (p or q): Jones owns a Ford, or Brown lives in Barcelona (Though Smith has no justification for q)
- Coincidence: Jones does not own Ford, but Brown lives in Barcelon
- Smith „knew“ (r) only by chance

General idea: decouple justification and truth conditions of propositional content of belief

General remarks

- What is a justification at all?
  - „Solutions“ to Gettier's problem deal with this problem
  - A formal treatmant of justification similar to provability logic: (Artemov 2008)

- Gettier's problem formalized
  - Suppose logic of belief and justification such that
    (*) $\phi \rightarrow \psi \vdash \quad hasJust(a, \phi) \rightarrow hasJust(a, \psi)$
  - Suppose: a wrongly but justifiably believes in p
    $$\neg p \wedge B_a p \wedge hasJust(a, p)$$
  - By M(B_a): $B_a(p \vee q\ ) \wedge B_a(p \vee \neg q)$
  - By (*): $hasJust(a, (p \vee q\ )) \wedge \mathrm{hasJust}(a, (p \vee \neg q))$
  - Hence: $\vDash B_a p \wedge hasJust(a, p) \rightarrow (K_a(p \vee q) \vee K_a(p \vee \neg\, q))$

Relation of knowledge and belief not obvious

- Suppose logic of knowledge and belief defined as
  - $KD45(B_a)$
  - $S5(K_a)$
  - $K_a\phi \rightarrow B_a\phi$
  - $B_a\phi \rightarrow B_aK_a\phi$

- Would entail that $B_a\phi \leftrightarrow K_a\phi$
  (intermediate step: $\neg B_a \neg K_a \phi \rightarrow \neg K_a \neg B_a \phi$)
- Culprit: negative introspection for knowledge
  (Lenzen 1978, Lenzen 1995)

DYNAMICS OF BELIEF

Getting dynamic with beliefs

- How do a's beliefs evolve when a learns that ϕ is true?
- Extend $KD45_n$ by public announcement operator $[\phi!]$
  - What if agent a wrongly believes that p, but $\neg p$ is announced?
  - This is NOT possible in epistemic logic:
    - $\vdash_{S5_n} K_a p \rightarrow p$ (reflexivity)
    - $\vdash_{S5_n PAL} p \leftrightarrow [\neg p!] \perp$ (reduction axiom)
    - $\vdash_{S5_n PAL} K_a p \rightarrow [\neg p!] \perp$
  - In doxastic logic:
    - $B_a p \wedge \neg p$ is satisfiable
    - $\vdash_{KD45_n PAL} p \leftrightarrow [\neg p!] \perp$ (reduction axiom)
    - $B_a p \wedge \neg[\neg p!] \perp$ should be $KD45_n - PAL$ satisfiable

But in doxastic logic dynamics not trivial

- One can show: inconsistent beliefs possible
  $\vdash_{KD45_n PAL} \ (\neg p \wedge B_a p) \rightarrow < \neg p! > B_a \perp$
- Ways out:
  1. Drop seriality
  2. Modify truth condition for announcements

$$M, w \vDash [\phi!]\psi \text{ iff} \begin{pmatrix} M, w \nvDash \ \phi \quad \text{or} \\ M, w \vDash \hat{B}_a\phi \ \text{ and } \ M^{\phi!}, w \vDash \psi \text{ or} \\ M, w \vDash B_a \neg\phi \text{ and } M, w \vDash \psi \end{pmatrix}$$

   - Reduction axiom
     $[\phi!]B_a\psi \leftrightarrow \neg\phi \vee (\hat{B}_a\phi \wedge B_a[\phi!]\psi) \vee (B_a\neg\phi \wedge B_a\psi)$
   - Believe-contravening input is rejected
  3. Integrate belief revision mechanism

Classical theory of Belief Revision

- We partly follow the presentation of Herzig
- For a more comprehensive treatment see also master course „Information Systems CS4130“ at IFIS

- Landmarking „yellow“ paper of Alchourron, Gärdenfors and Makinson (Alchourron et al 1985)

- Beliefs of an ideal agent = set of Boolean formulas $S \subseteq L$ closed under some consequence operator
  - $S \in BS_L$ is called a belief set

AGM takes an internal perspective

- $\phi \in S$ means: ϕ is believed by the agent
- Internal perspective (S is in agent's head)
- Contrast with external perspective:
  - $\phi =$ ``ϕ'' is objectively true
  - Taken in doxastic logic
- But can „internalize"doxastic logic too (Aucher 2008)
  - Distinguished agent Y (for you)
  - $\phi =$ „Y believes that ϕ"
  - Wanted: $\vdash \phi \leftrightarrow B_Y\phi$
  - Abandond inference rule of necessitation:
    $\models B_Y\phi \rightarrow \phi$ but $\not\models B_a(B_Y\phi \rightarrow \phi)$

Coherentism vs foundationalism

- Two general approaches in epistemology
- Foundationalism:
  - All beliefs rest on some basic beliefs (which do not rest by themselves on others, but are assume to be true
  - Some tribute to foundationalism in post AGM-work: Belief bases are (arbitrary not necessarily closed) usually finite sets of sentences
- Coherentism:
  - Beliefs are justified by their relations (consequence, justification..) to other beliefs in a network
  - Usually there is no notion of truth
  - AGM considers closed sets of beliefs based on a consequence operator (logic not based on a semantics)

Types of Belief Change

- L: Set of well-formed formulas (with at least Boolean operators)
- $Cn: 2^L \rightarrow 2^L$ consequence operator (monotonic, idempotent and conclusive)
- B_L: Sets of belief sets = Cn $-$closed sets in 2^L
  - Single inconsistent belief-set = L
- AGM considers three types of operators op all of signature $op: B_L \times L \rightarrow B_L$
  - Expansion: $X + \psi$
  - Contraction: $X - \psi$
  - Revision: $X * \phi$

Types of Belief Change

- $X + \psi$ = expanding X by ψ
  - Result of adding ψ to X without considering inconsistencies
  - Desideratum: $\psi \in X + \psi$ or even $Cn(X \cup \{\psi\}) = X + \psi$
- $X - \psi$ = contracting X by ψ
  - Result of deleting ψ and other sentences such that ψ no longer follows (is contained in the resulting belief set)
  - Desiderata: $\psi \notin X - \psi$; $X - \psi \subseteq X$; ...
- $X * \psi$ = revising X by ψ
  - Result of adding consistently ψ
  - Desiderata: $\psi \in X * \psi$; $X * \psi \neq L$ if $\mathrm{Cn}(\psi) \neq L$

Desiderata captured by AGM postulates

(here for revision)

- (R1) $X * \psi \in BS_L$ (closure)
- (R2) $\psi \in X * \psi$ (success)
- (R3) $X * \psi \subseteq Cn(X \cup \{\psi\})$ (inclusion)
- (R4) If $\neg\psi \notin Cn(X)$ then $Cn(X \cup \{\psi\}) \subseteq X * \psi$ (vacuity)
- (R5) If $Cn(X * \psi) = L$ then $\neg\psi \in Cn(\emptyset)$ (consistency)
- (R6) If $\phi \leftrightarrow \psi \in Cn(\emptyset)$, then $X * \phi = \mathrm{X} * \psi$ (extensionlity)

- (R7) $X * (\phi \wedge \psi) \subseteq Cn((X * \phi) \cup \{\psi\})$ (conjunction 1)
- (R8) If $\neg\beta \notin Cn$, then (conjunction 2)

  $Cn\big((X * \phi) \cup \{\psi\}\big) \subseteq X * (\phi \wedge \psi)$

  (Note: Postulate is not axiom: talks about Cn)

Semantics for AGM

- Postulates generally specify whole classes of operators (exception: expansion)
- How to construct concrete change operators?
- Different design principles
  - Partial meet based on remainder sets (considered here)
  - Orders (epistremic entrenchment)
  - Systems of spheres
  - ...

Remainder Sets: „Maximal Scenarios“

Definition (remainder set)

The remainder set $X \perp \alpha$ of X by α consists of all inclusion-maximal subsets of X not entailing α. The sets in $X \perp \alpha$ are called remainders.

Example

- $\{p, q\} \perp (p \wedge q) = \{\{p\}, \{q\}\}$
- $\{p \vee r, p \vee \neg r, q \wedge s, q \wedge \neg s\} \perp (p \wedge q) =$
  $\{\{p \vee r, p \vee \neg r\}, \{p \vee r, q \wedge s\},$
  $\{p \vee r, q \wedge \neg s\}, \{p \vee \neg r, q \wedge s\}, \{p \vee r, q \wedge \neg s\}$

Selection function

Definition (selection function)

An AGM selection function $\gamma: 2^{B_L} \rightarrow 2^{B_L}$ for X fulfills:

1. If $X \perp \psi \neq \emptyset$, then $\emptyset \neq \gamma(X \perp \alpha) \subseteq X \perp \alpha$
2. $\gamma(\emptyset) = \{X\}$

- As there are many remainders (maximal scenarios) we need to select some of them as possible
- Chooses some remainders (if not empty).

Partial-Meet contraction and revision

Definition

- $X -_\gamma \psi = \bigcap\gamma(X \perp \psi)$ (partial meet contraction)
- $X *_\gamma \psi = Cn((X -_\gamma \neg\psi) \cup \{\psi\})$ (partial meet revision)

Revision operator defined here by so-called Levi-identity from contraction

Theorem (Representation)

An operator $*$ fulfills postulates (R1)-(R6) iff there is a selection function γ sucht that $X * \psi = X *_\gamma \psi$

Notes

- Similar representation result for contraction
- Partial meet revision does not necessarily fulfill (R7) and R(8); need to constrain γ further

AGM: integratio with doxastic logic

- Work of Segerberg (Segerberg 1995, 1996)
  - Modal operators B_a, $[+\psi]$, $[-\psi]$, $[* \psi]$
  - $[* \psi]\phi =$ „ϕ is true after revision by ψ"
- Internal version of doxastic logic (Aucher 2008)
  - Straightforward transfer of AGM representation theorems to multiagent case
- Distinguish several versions of belief (Baltag/Smets 2007, 2009)09
  - Soft beliefs: can be revised
  - Hard beliefs: cannot be revised

GROUP BELIEFS

Group beliefs

- Theory of group beliefs developed in the same way as for group knowledge ...
- $EB_J\phi \coloneqq \bigwedge_{a\in J} B_a\phi$
- $CB_J\phi \coloneqq EB_J\phi \wedge EB_J EB_J\phi \wedge \ \ldots$
- $\mathcal{R}_{CB_J} \coloneqq \left(\bigcup_{a\in J} \mathcal{R}_{B_a}\right)^+$
- Axiomatization of $KD45(B_a)$ with common belief
  - Axiomatics of $KD45(B_a)$
  - Fixed point axiom: $CB_J\phi \leftrightarrow \left(EB_J \wedge EB_J CB_J\phi\right)$
  - Least fixed point inference rule (induction rule)

  $\phi \rightarrow EB_J\,\phi \vdash \ \ EB_J\phi \ \rightarrow CB_J\phi$
- Sound, complete decidable in EXPTIME-c.

Uhhh, a lecture with a hoepfully useful

APPENDIX

References

- R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. Reasoning about Knowledge. MIT Press, 2003.
- P. Engel. Believing, holding true, and accepting. Philosophical Explorations, 1(2):140–151, 1998.
- R. Hakli. Group beliefs and the distinction between belief and acceptance. Cognitive Systems Research, 7:286–297, 2006. J. Hintikka. Knowledge and Belief: An Introduction to the Logic of the Two Notions. Texts in philosophy. King's College London Publications, 2005.
- W. Lenzen. Recent work in epistemic logic. Acta Philosophica Fennica, 30:1–219, 1978.
- W. Lenzen. On the semantics and pragmatics of epistemic attitudes. Knowledge and belief in philosophy and AI, pages 181–197, 1995.
- J. P. Burgess. Probability logic. Journal of Symbolic Logic, 34(2):264–274, 1969.
- E. L. Gettier. Is Justified True Belief Knowledge? Analysis, 23(6):121–123, 06 1963.
- S. Artemov. The logic of justification. The Review of Symbolic Logic, 1(4):477–513, 2008.
- C. E. Alchourron, P. Gärdenfors, and D. Makinson. On the logic of theory change: partial meet contrac- tion and revision functions. Journal of Symbolic Logic, 50:510–530, 1985.
- G. Aucher. Perspectives on belief and change. PhD thesis, Universite de Toulouse, 2008.
- K. Segerberg. Belief revision from the point of view of doxastic logic. Logic Journal of the IGPL, 3(4):535–553, 1995.
- K. Segerberg. Two traditions in the logic of belief: Bringing them together. In H. J. Ohlbach and U. Reyle, editors, Logic, Language and Reasoning: essays in honour of Dov Gabbay, volume 5 of Trends in Logic, pages pages 135–147. 1999.
- A. Baltag and S. Smets. From conditional probability to the logic of doxastic actions. In Proceedings of the 11th Conference on Theoretical Aspects of Rationality and Knowledge, TARK '07, pages 52—61, New York, NY, USA, 2007. Association for Computing Machinery.
- A. Baltag and S. Smets. Probabilistic dynamic belief revision. Synth., 165(2):179–202, 2008.

Color Convention in this course

- Formulae, when occurring inline
- Newly introduced terminology and definitions
- Important results (observations, theorems) as well as emphasizing some aspects
- Examples are given with standard orange with possibly light orange frame
- Comments and notes
- Algorithms