Towards Epistemic-Doxastic Planning with Observation and Revision in a Lightweight Logic

Andreas Herzig
CNRS-IRIT, Toulouse
(paper with Thorsten Engesser and Elise Perrotin, AAAI 2024)

MAFTEC, Jan. 18, 2024

“Epistemic logic”

- narrow sense: logics of knowledge
  - $\mathbf{K}_i\varphi$ = “agent i knows that φ”
- broad sense: logics of knowledge **or** of belief
  - $\mathbf{B}_i\,\varphi$ = “agent i beliefs that φ”
  - “doxastic logics”
- all are more complex than propositional logic
  - SAT is PSpace-hard
  - model checking unfeasible (Kripke models too big)

“Epistemic doxastic logic”

- logics of knowledge **and** belief
  - $\mathbf{B}_i\,\varphi \land \neg\mathbf{K}_i\varphi$ = “agent *i* beliefs that φ without knowing it”
  - “epidox logics”
  - some conceptual issues: which principles? here:

$\mathbf{K}_i\varphi \to \mathbf{B}_i\,\varphi$	OK
$\mathbf{B}_i\,\varphi \to \mathbf{K}_i\mathbf{B}_i\,\varphi$	OK
$\mathbf{B}_i\,\mathbf{K}_i\varphi \to \mathbf{K}_i\varphi$	OK for observational knowledge
$\neg\mathbf{B}_i\,\varphi \to \mathbf{K}_i\neg\mathbf{B}_i\,\varphi$	OK
$\mathbf{B}_i\,\varphi \to \mathbf{B}_i\,\mathbf{K}_i\varphi$	KO! (inconsistent with $\neg\mathbf{K}_i\varphi \to \mathbf{K}_i\neg\mathbf{K}_i\varphi$)

  - more for the same price: epidox logics are also PSpace complete

Adding dynamcis

- needed: reasoning about evolution of knowledge and belief!
  - reasoning about actions (cf. epistemic SitCalc)
  - planning (cf. multiagent STRIPS)
- logics of knowledge + action
  - dynamic epistemic logics DEL
  - conceptually nice
    - rich account of who observes what ('event models')
  - but computational problems
    - DEL-based planning undecidable
- logics of belief + action
  - computational problems (v.s.)
  - conceptual problems:
    - action may reveal that some belief is false
    - requires revision of beliefs
    - no good solution in DEL

Let's restrict the language

- logics of knowledge + belief + action inherit difficult problems
  - conceptually
  - computationally
- first idea: restrict static epidox language
  - basically: no knowledge/belief about disjunctions
    - $\mathbf{K}_i(p \vee q)$ cannot be expressed
  - *lightweight* epidox logic
  - much better computational properties: SAT in NP!
- second idea: restrict language of actions
  - DEL: not very fruitful
    - except special case of fully public actions (PAL)
  - but works better when combined with lightweight epidox logic
  - here: STRIPS-like 'flip-lists' (instead of add- and delete lists)
- will work nicely for planning tasks involving false belief, revision, deception,. . .

Lightweight logics of knowledge: 'knowing-that' literals

[Demolombe&Pozos Parra; Lakemeyer&Lespérance 2012; Muise et al. 2015; 2021]

$$\lambda ::= p \mid \neg\lambda \mid \mathbf{K}_i\lambda$$

- formula = boolan combination of epistemic literals
  - no conjunction or disjunction in scope of epistemic operators
- complexity: same as propositional logic
  - view epistemic atoms as propositional variables
  - plus theory: $\neg(\mathbf{K}_i\lambda \wedge \mathbf{K}_i\neg\lambda)$, $\mathbf{K}_i\mathbf{K}_i\lambda \leftrightarrow \mathbf{K}_i\lambda$, etc.
- cannot express "I know you know more than me"

$$\neg\mathbf{K}_i p \wedge \neg\mathbf{K}_i\neg p \wedge \mathbf{K}_i(\mathbf{K}_j p \vee \mathbf{K}_j\neg p)$$

  but is fundamental in interaction (precondition of questions)
- sequel: 'knowing-whether' primitive instead [Lomuscio; van der Hoek et al.; Gattinger et al.]

Knowledge/belief about a proposition

- ‘know whether’ has no belief-counterpart in natural language (just as the other ‘know wh’ modalities) [Egré, 2008]
- therefore:

$$\mathbf{KA}_i\varphi \quad = \quad \text{“agent } i \text{ has knowledge about } \varphi\text{”}$$
$$\mathbf{BA}_i\varphi \quad = \quad \text{“agent } i \text{ has belief about } \varphi\text{”}$$

'About' modalities: expressivity

1. 'belief about': weaker [Fan et al., 2015]

$$\mathbf{BA}_i\varphi \leftrightarrow \mathbf{B}_i\,\varphi \vee \mathbf{B}_i\,\neg\varphi$$
$$\mathbf{B}_i\,\varphi \leftrightarrow ?$$

2. 'knowledge about': equi-expressive

$$\mathbf{KA}_i\varphi \leftrightarrow \mathbf{K}_i\varphi \vee \mathbf{K}_i\neg\varphi$$
$$\mathbf{K}_i\varphi \leftrightarrow \varphi \wedge \mathbf{KA}_i\varphi$$

but:

- 'knowledge about' can express things more succinctly [van Ditmarsch et al., 2014]
- equivalent presentations may lead to new insights

'Knowledge about' atoms [Herzig et al., 2015, Cooper et al., 2021]

- grammar:

$$\alpha ::= p \mid \mathbf{KA}_i\alpha$$

  where $p \in Prop$

- formula = boolan combination of epistemic atoms
- can express some disjunctions in scope of epistemic operator:

$$\mathbf{K}_i(\mathbf{K}_j p \vee \mathbf{K}_j \neg p)$$

  expressed as

$$\mathbf{K}_i \mathbf{KA}_j p$$
$$= \mathbf{KA}_j p \wedge \mathbf{KA}_i \mathbf{KA}_j p$$

'Knowledge about' atoms: computation

- basically: epistemic atoms can be viewed as propositional logic variables
  - take care of introspection: $\mathbf{KA}_i\mathbf{KA}_i\alpha$ valid
  - simple solution: forbid repetitions
- complexity of reasoning: same as propositional logic
  - satisfiability NP-complete
- can be extended by an operator 'common knowledge about' [Herzig&Perrotin, AiML 2020; forthcoming]

Lightweight logics of knowledge: dynamics

- ‘dual use’ of knowledge about atoms [Cooper et al., AIJ 2020]:
  - $\mathbf{KA}_i\alpha$ = agent *i* sees truth value of α
  - $\mathbf{KA}_i\alpha$ = agent *i* sees truth value changes of α (except if action makes $\mathbf{KA}_i\alpha$ false)
- STRIPS-like actions: preconditions + pos./neg. effects
- complexity of planning: same as propositional logic
  - plan existence PSPACE-complete

Lightweight logics of belief?

- knowledge-about atoms 'work' because there are 4 independent combinations of p and $\mathbf{KA}_i p$:

$p \wedge \mathbf{KA}_i p$	$\neg p \wedge \mathbf{KA}_i p$
$p \wedge \neg\mathbf{KA}_i p$	$\neg p \wedge \neg\mathbf{KA}_i p$

  - in terms of knowledge-that:

$p \wedge \mathbf{K}_i p$	$\neg p \wedge \mathbf{K}_i \neg p$
$p \wedge \neg\mathbf{K}_i p \wedge \neg\mathbf{K}_i \neg p$	$\neg p \wedge \neg\mathbf{K}_i p \wedge \neg\mathbf{K}_i \neg p$

- for belief: 6 possible doxastic situations

$p \wedge \mathbf{B}_i\, p$	$\neg p \wedge \mathbf{B}_i\, \neg p$
$p \wedge \neg\mathbf{B}_i\, p \wedge \neg\mathbf{B}_i\, \neg p$	$\neg p \wedge \neg\mathbf{B}_i\, p \wedge \neg\mathbf{B}_i\, \neg p$
$p \wedge \mathbf{B}_i\, \neg p$	$\neg p \wedge \mathbf{B}_i\, p$

  - requires 3 dimensions $\implies$ cannot be independent

Three dimensions of epidox situations

- 8 possible situations:

$p \wedge \mathbf{K}_i p$	$\neg p \wedge \mathbf{K}_i \neg p$
$p \wedge \mathbf{B}_i\, p \wedge \neg \mathbf{K}_i p$	$\neg p \wedge \mathbf{B}_i \neg p \wedge \neg \mathbf{K}_i \neg p$
$p \wedge \neg \mathbf{B}_i\, p \wedge \neg \mathbf{B}_i \neg p$	$\neg p \wedge \neg \mathbf{B}_i\, p \wedge \neg \mathbf{B}_i \neg p$
$p \wedge \mathbf{B}_i \neg p$	$\neg p \wedge \mathbf{B}_i\, p$

- $8 = 2^3 \implies$ which are the 3 dimensions?

Which epistemic-doxastic situations?

- two new modalities:

$$\begin{aligned}\mathbf{TBA}_i\, p &= (p \wedge \mathbf{B}_i\, p) \vee (\neg p \wedge \mathbf{B}_i\, \neg p)\\ &= \text{“}i \text{ has a } \textbf{true} \text{ belief about } p\text{”}\\ \mathbf{MBA}_i\, p &= (\mathbf{B}_i\, p \wedge \neg \mathbf{K}_i p) \vee (\mathbf{B}_i\, \neg p \wedge \neg \mathbf{K}_i \neg p)\\ &= \text{“}i \text{ has a } \textbf{mere} \text{ belief about } p\text{”}\\ &= \text{“}i \text{ has a falsifiable belief about } p\text{”}\\ &= \text{“}i \text{ has a belief about } p \text{ but does not know whether } p\text{”}\end{aligned}$$

- insensitive to negation:

$$\mathbf{TBA}_i\, \neg p \leftrightarrow \mathbf{TBA}_i\, p$$
$$\mathbf{MBA}_i\, \neg p \leftrightarrow \mathbf{MBA}_i\, p$$

Epistemic-doxastic situations: 3 dimensions

- 2^3 epistemic-doxastic situations:

$p \wedge \mathbf{TBA}_i\, p \wedge \neg\mathbf{MBA}_i\, p$	$\neg p \wedge \mathbf{TBA}_i\, p \wedge \neg\mathbf{MBA}_i\, p$
$p \wedge \mathbf{TBA}_i\, p \wedge \mathbf{MBA}_i\, p$	$\neg p \wedge \mathbf{TBA}_i\, p \wedge \mathbf{MBA}_i\, p$
$p \wedge \neg\mathbf{TBA}_i\, p \wedge \neg\mathbf{MBA}_i\, p$	$\neg p \wedge \neg\mathbf{TBA}_i\, p \wedge \neg\mathbf{MBA}_i\, p$
$p \wedge \neg\mathbf{TBA}_i\, p \wedge \mathbf{MBA}_i\, p$	$\neg p \wedge \neg\mathbf{TBA}_i\, p \wedge \mathbf{MBA}_i\, p$

- needs getting used to, but is natural. . .

Example: the Sally-Ann Test

false belief task
[Wimmer and Perner, 1983, Baron-Cohen et al., 1985]

1. Sally puts the marble in the basket

$$\mathbf{TBA}_S\, \mathrm{b} \wedge \neg \mathbf{MBA}_S\, \mathrm{b}$$

2. Sally goes out for a walk

$$\mathbf{TBA}_S\, \mathrm{b} \wedge \mathbf{MBA}_S\, \mathrm{b}$$

3. Ann takes the marble out of the basket and puts it into the box

$$\neg \mathbf{TBA}_S\, \mathrm{b} \wedge \mathbf{MBA}_S\, \mathrm{b}$$

Full expressivity

- knowledge:

$$\mathbf{KA}_i\varphi \leftrightarrow \mathbf{TBA}_i\,\varphi \wedge \neg\mathbf{MBA}_i\,\varphi$$
$$\mathbf{K}_i\varphi \leftrightarrow \mathbf{TBA}_i\,\varphi \wedge \neg\mathbf{MBA}_i\,\varphi \wedge \varphi$$

- belief:

$$\mathbf{BA}_i\varphi \leftrightarrow \mathbf{TBA}_i\,\varphi \vee \mathbf{MBA}_i\,\varphi$$
$$\mathbf{B}_i\,\varphi \leftrightarrow (\varphi \wedge \mathbf{TBA}_i\,\varphi) \vee (\neg\varphi \wedge \neg\mathbf{TBA}_i\,\varphi \wedge \mathbf{MBA}_i\,\varphi)$$

... remember: $\mathbf{B}_i\,\varphi$ cannot be expressed with $\mathbf{BA}_i$ alone

An epistemic-doxastic logic

- logic:

KD5($\mathbf{B}$)	the principles of modal logic KD5 for $\mathbf{B}_i$
S4($\mathbf{K}$)	the principles of modal logic S4 for $\mathbf{K}_i$
KiB	$\mathbf{K}_i\varphi \to \mathbf{B}_i\,\varphi$
BiKB	$\mathbf{B}_i\,\varphi \to \mathbf{K}_i\mathbf{B}_i\,\varphi$
BiBK	$\mathbf{B}_i\,\varphi \to \mathbf{B}_i\,\mathbf{K}_i\varphi$

- belief definable from knowledge [Lenzen, 1978, Lenzen, 1995]:

$$\mathbf{B}_i\,\varphi \leftrightarrow \neg\mathbf{K}_i\neg\mathbf{K}_i\varphi$$

  - alternative axiomatisation: S4.2($\mathbf{K}$) plus $\mathbf{B}_i\,\varphi \leftrightarrow \neg\mathbf{K}_i\neg\mathbf{K}_i\varphi$
- complexity of satisfiability: PSPACE-complete [Shapirovsky, 2004, Chalki et al., 2021]

Reducing modalities

- reduction of consecutive modal operators to length 1:

$$\mathbf{TBA}_i\,\mathbf{TBA}_i\,\varphi \leftrightarrow \mathbf{TBA}_i\,\varphi \vee \neg\mathbf{MBA}_i\,\varphi$$
$$\mathbf{MBA}_i\,\mathbf{TBA}_i\,\varphi \leftrightarrow \mathbf{MBA}_i\,\varphi$$
$$\mathbf{TBA}_i\,\mathbf{MBA}_i\,\varphi \leftrightarrow \neg\mathbf{MBA}_i\,\varphi$$
$$\mathbf{MBA}_i\,\mathbf{MBA}_i\,\varphi \leftrightarrow \mathbf{MBA}_i\,\varphi$$

$\Longrightarrow$ suppose formulas are 'repetition-free'

- no $\cdots\mathbf{TBA}_i\,\mathbf{TBA}_i\,\cdots p$
- no $\cdots\mathbf{TBA}_i\,\mathbf{MBA}_i\,\cdots p$
- no $\cdots\mathbf{MBA}_i\,\mathbf{TBA}_i\,\cdots p$
- no $\cdots\mathbf{MBA}_i\,\mathbf{MBA}_i\,\cdots p$

Lightweight epistemic-doxastic fragments: the idea

- epidox atoms:

$$\alpha ::= p \mid \mathbf{TBA}_i\, \alpha \mid \mathbf{MBA}_i\, \alpha$$

- repetition-free

Theorem

If φ is a boolean combination of repetition-free epidox atoms then the following are equivalent:

- *φ is valid in epistemic-doxastic logic;*
- *φ is propositionally valid.*

Corollary

Satisfiability of boolean combinations of epidox atom is in NP.
Plan existence is in PSpace.

Background and motivation

Lightweight logic of knowledge and belief

Lightweight logic of action

Adding actions

- action = precondition + conditional effects
  - precondition = boolean combination of epidox atoms
  - effects = epidox atoms that are flipped

$$\varphi \triangleright \pm\alpha = \text{“if } \varphi \text{ is true then } \alpha \text{ changes its truth value”}$$

- restriction to atoms α of depth ≤ 2
- express STRIPS action with add-list P^+ and delete-list P^-:

$$\{p \triangleright \pm p \ : \ p \in P^-\} \cup \{\neg p \triangleright \pm p \ : \ p \in P^+\}$$

Direct and indirect effects

- direct effects:
  - either on the world (prop.var.s) $\Longrightarrow$ ontic actions
  - or on knowledge/belief $\Longrightarrow$ epistemic actions
    1. observation change/sensing
    2. communication (future work)
- indirect effects:
  - are always epistemic (change knowledge/belief)
  - derived from direct effects
  - depending on agents' observation status

Ontic actions

- direct effects = set of conditional effects

$$\{\varphi_1 \triangleright \pm p_1, \ldots, \varphi_n \triangleright \pm p_n\}$$

  modify the world = the propositional variables p_k
- the main principle deriving indirect effects:

$$(M) \quad \varphi_k \wedge \mathbf{MBA}_i\, p_k \triangleright \pm \mathbf{TBA}_i\, p_k$$

- other principles deriving second-order indirect effects ...

Epistemic actions: starting individual observation

- i starts to observe propositional variable p (without learning about others' belief change):

$$\texttt{startobs}^1(i, p)$$

- direct effects: i has knowledge about p
  1. add $\mathbf{TBA}_i\, p$:
  $$\neg\mathbf{TBA}_i\, p \triangleright \pm\mathbf{TBA}_i\, p$$
  2. delete $\mathbf{MBA}_i\, p$:
  $$\mathbf{MBA}_i\, p \triangleright \pm\mathbf{MBA}_i\, p$$
- indirect effects (obtained via Principle (M)):

$$\{\neg\mathbf{TBA}_i\, p \wedge \mathbf{MBA}_j\, \mathbf{TBA}_i\, p \triangleright \pm\mathbf{TBA}_j\, \mathbf{TBA}_i\, p \;:\; j \neq i\} \cup$$
$$\{\mathbf{MBA}_i\, p \wedge \mathbf{MBA}_j\, \mathbf{MBA}_i\, p \triangleright \pm\mathbf{TBA}_j\, \mathbf{MBA}_i\, p \;:\; j \neq i\}$$

Epistemic actions: starting group observation

- group J starts to observe propositional variable p, learning that the other members of J also do so:

$$\texttt{startobs}^2(J, p)$$

- direct effects:
  - every $i \in J$ has knowledge about p:
    1. add $\mathbf{TBA}_i\, p$, for $i \in J$
    2. delete $\mathbf{MBA}_i\, p$, for $i \in J$
  - every $i \in J$ has knowledge about $\mathbf{TBA}_j\, p$, for $j \in J$:
    1. add $\mathbf{TBA}_j\, \mathbf{TBA}_i\, p$
    2. delete $\mathbf{MBA}_j\, \mathbf{TBA}_i\, p$
  - every $i \in J$ has knowledge about $\mathbf{MBA}_j\, p$, for $j \in J$:
    1. add $\mathbf{TBA}_j\, \mathbf{MBA}_i\, p$
    2. delete $\mathbf{MBA}_j\, \mathbf{MBA}_i\, p$
- indirect effects (obtained via Principle (M)):

$$\ldots$$

Epistemic actions: ceasing to observe a fact

- group J ceases to observe propositional variable p, learning that the other members of J also do so:

$$\texttt{stopobs}(i, p)$$

- direct effect: knowledge about p becomes mere belief

$$\mathbf{TBA}_i\, p \wedge \neg \mathbf{MBA}_i\, p \triangleright \pm \mathbf{MBA}_i\, p$$

  - inertia of beliefs
    - when Sally leaves the room her knowledge about the marble becomes a mere belief
  - more realistic: decaying beliefs
- indirect effects (obtained via Principle (M)):

...

Epistemic actions: ceasing to observe another agent

- group J ceases to observe propositional variable p, learning that the other members of J also do so:

$$\texttt{stopobs}(i, j, p)$$

- direct effects: ...
- indirect effects (obtained via Principle (M)):

$$\dots$$

Epidox planning

- just as in classical planning:
  - initial state = set of epidox atoms
  - goal = boolean combination of epidox atoms
- examples:
  - Sally-Ann test as a planning task (goal: induce Sally's false belief)
  - variants of the grapevine domain
  - tasks involving correction of false beliefs
  - tasks involving deception
  - ...

Theorem

An epidox planning task is solvable iff it is propositionally solvable.

Conclusion: lightweight planning with epidox logic

- lightweight fragment of epistemic-doxastic logic
  - ‘true belief about’ and ‘mere belief about’ modalities
  - repetition-free epistemic-doxastic atoms
  - same complexity as propositional logic

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985).
Does the autistic child have a theory of mind?
Cognition, 21(1):37–46.

Chalki, A., Koutras, C. D., and Zikos, Y. (2021).
A note on the complexity of S4.2.
J. Appl. Non Class. Logics, 31(2):108–129.

Cooper, M., Herzig, A., Maffre, F., Maris, F., Perrotin, E., and Régnier, P. (2021).
A lightweight epistemic logic and its application to planning.
Artificial Intelligence, 298:103437.

Egré, P. (2008).
Question-embedding and factivity.
Grazer Philosophische Studien, 77(1):85–125.

Fan, J., Wang, Y., and van Ditmarsch, H. (2015).
Contingency and knowing whether.
Rew. Symb. Logic, 8(1):75–107.

Herzig, A., Lorini, E., and Maffre, F. (2015).

A poor man's epistemic logic based on propositional assignment and higher-order observation.
In van der Hoek, W., Holliday, W. H., and Wang, W.-f., editors, *Proceedings of the 5th International Conference on Logic, Rationality and Interaction (LORI 2015)*, pages 156–168. Springer Verlag.

Lenzen, W. (1978).
Recent work in epistemic logic.
North Holland Publishing Company, Amsterdam.

Lenzen, W. (1995).
On the semantics and pragmatics of epistemic attitudes.
In Laux, A. and Wansing, H., editors, *Knowledge and belief in philosophy and AI*, pages 181–197. Akademie Verlag, Berlin.

Shapirovsky, I. (2004).
On PSPACE-decidability in transitive modal logic.
In Schmidt, R. A., Pratt-Hartmann, I., Reynolds, M., and Wansing, H., editors, *Advances in Modal Logic 5, papers from the fifth conference on "Advances in Modal logic", held in Manchester, UK, 9-11 September 2004*, pages 269–287. King's College Publications.

van Ditmarsch, H., Fan, J., van der Hoek, W., and Iliev, P. (2014).
Some exponential lower bounds on formula-size in modal logic.

In Goré, R., Kooi, B. P., and Kurucz, A., editors, *Advances in Modal Logic 10, invited and contributed papers from the tenth conference on "Advances in Modal Logic," held in Groningen, The Netherlands, August 5-8, 2014*, pages 139–157. College Publications.

Wimmer, H. and Perner, J. (1983).
Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception.
Cognition, 13(1):103–128.