I thought you didn't know! On belief revision in dynamic doxastic logic

Andreas Herzig, Jérôme Lang, Dominique Longin

I thought you didn't know!
On belief revision in dynamic doxastic logic

Andreas Herzig, Jérôme Lang, Dominique Longin

{herzig,lang,longin}@irit.fr
www.irit.fr/recherches/LILAC, www.irit.fr/recherches/RPDMP

1 Introduction

Suppose agent i believes that agent j does not believe that p. What is i's belief state like after j informs i that p?

Scenarios of this kind can be expressed nicely in dynamic doxastic logics. There, to every action α is associated a dynamic operator $[\alpha]$, to every agent i is associated a belief operator B_i, and to every set of agents $\{i, j, \ldots\}$ is associated a common belief operator $B_{\{i,j,\ldots\}}$. Speech acts can be viewed as particular actions, and we can write j's action of informing i that p as $\mathit{inform}(j, i, p)$.

To simplify our exposition let us make the hypothesis that every agent is sincere: he cannot assert things he doesn't believe himself. This can be expressed by the axiom

$$\neg B_j \varphi \rightarrow [\mathit{inform}(j, i, \varphi)]\bot$$

Then standard KD45 principles entail that $B_i \neg B_j \varphi \rightarrow B_i[\mathit{inform}(j, i, \varphi)]\bot$, i.e. in our scenario i believes j cannot tell him that p.

We moreover make the hypothesis that sincerity is the only precondition of inform-actions. We therefore have the axiom

$$B_j \varphi \rightarrow \langle \mathit{inform}(j, i, \varphi)\rangle \top$$

Such scenarios, where an agent learns that some action α he wrongly believed to be inexecutable has nevertheless occurred, pose problems for the logics of actions that have been proposed in the literature to account for the evolution of knowledge and belief. Some of them [10, 11, 14, 16, 15, 7] deliberately restrict their attention to true belief (alias knowledge) in order to avoid such scenarios. Indeed, for the other approaches in [11, 6, 5, 3, 2, 4] a principle of permutation $B_i[\alpha]\varphi \rightarrow [\alpha]B_i\varphi$ is valid. Semantically, let $B_i(w)$ be the set of worlds agent i considers possible at the actual world w, and suppose $R_\alpha(w)$ is the set of possible outcomes of action α when executed in w. Then in the above approaches the worlds i considers possible after α are obtained by applying α to each of the possible worlds in $B_i(w)$ ("mentally executing α"), and then restricting the resulting set $(B_i \circ R_\alpha)(w)$ in some way.

In the case of actions that do not modify the world but only the agents' belief states ('purely doxastic actions'), the permutation principle roughly speaking expresses that the possible worlds making up i's belief state after α, are a subset of i's possible worlds before α. In terms of AGM belief change operations, every such α is thus an expansion.

Obviously, the permutation principle must be weakened to

$$\neg B_i[\alpha]\bot \wedge B_i[\alpha]\varphi \;\rightarrow\; [\alpha]B_i\varphi$$

Semantically, we should have that $(R_\alpha \circ B_i)(w) \subseteq (B_i \circ R_\alpha)(w)$ only when $(B_i \circ R_\alpha)(w) \neq \emptyset$.

Sure, this prevents i from running into inconsistencies in our scenario. But this does tells us nothing about i's belief state after $\mathit{inform}(j, i, p)$. In this paper we introduce a permutation principle for this case.

2 Action laws and enabling actions

If i believes α is inexecutable, this means that there is no $v \in B_i(w)$ where the preconditions of α's executability laws hold. By an executability law we mean a principle associated to α that has the form $A \rightarrow \langle\alpha\rangle\top$. Let $\mathit{pre}(\alpha)$ be the disjunction of all executability preconditions of α. Then all the executability laws for α can be represented by the single $\mathit{pre}(\alpha) \rightarrow \langle\alpha\rangle\top$.

Now when i learns that α has nevertheless occurred, he cannot just mentally execute α to form his new belief state: he should first change his beliefs about $\mathit{pre}(\alpha)$. We suppose that i does this by adjusting *each* of his possible worlds so as to make $\mathit{pre}(\alpha)$ true, by mentally executing a particular action whose effect is $\mathit{pre}(\alpha)$.

Formally, we associate to every atomic action α an action enable_α, and we say that enable_α makes α executable.

We postulate that enable_α can always be executed. Hence

$$\mathit{pre}(\mathit{enable}_\alpha) = \top$$

Semantically, for every action there must exist at least one possible world where it is executable. This forces us to exclude actions that are never executable, such as the action $\mathit{inform}(j, i, \bot)$ of unsincerely asserting an inconsistency.[1]

As usually done in AI, we suppose that to every action α there is associated a set of effect laws describing the change α brings about. Such laws take the form $A \rightarrow [\alpha]C$. For inform-actions we have e.g. the effect law

$$[\mathit{inform}(j, i, \varphi)]B_{\{i,j\}}B_j\varphi$$

[1] This is also the reason why we consider atomic actions only. Indeed, let *toss* be the action of tossing a coin, let its executability precondition be that one must hold a coin ($\mathit{pre}(\mathit{toss}) = \mathit{holdCoin}$), and let one of its effects be that one no longer holds the coin ($\neg\mathit{holdCoin}$). Then it is impossible to execute *toss* twice. This means that $\mathit{toss};\mathit{toss}$ cannot be enabled, illustrating that *enable* cannot be applied to complex actions.

expressing that after j informs i that φ it is a common belief of i and j that j believes φ. We also suppose $\neg pre(\alpha) \rightarrow [\alpha]\bot$ to be among α's effect laws.

What are the effect laws for $enable_\alpha$? As $enable_\alpha$ makes α executable, we must have:

$$[enable_\alpha]pre(\alpha)$$

We have supposed that sincerity is the only precondition of inform-actions. Therefore $pre(inform(j,i,\varphi)) = B_j\varphi$. Hence we have the effect law

$$[enable_{inform(j,i,\varphi)}]B_j\varphi$$

3 Constructing the new belief state

Semantically, if w is the actual world then the worlds i considers possible after a surprising occurrence of α can now be obtained from the worlds in $B_i(w)$ by first applying $enable_\alpha$ and then applying α. That is, if $(R_\alpha \circ B_i)(w) = \emptyset$ then we have

$$(R_\alpha \circ B_i)(w) \ \subseteq \ (B_i \circ R_{enable_\alpha} \circ R_\alpha)(w)$$

Axiomatically, the permutation principle is weakened to

$$B_i[\alpha]\bot \wedge B_i[enable_\alpha][\alpha]\varphi \ \rightarrow \ [\alpha]B_i\varphi$$

Just as it has been done for the strong permutation principle in the AI field of reasoning about actions [11, 14, 7], such a principle can then be combined with existing solutions to the frame problem. (For our permutation principles this has been done in [8].) The result is a constructive characterization of the agents' belief states after an action.

Using our permutation principles and the sincerity axiom we can prove with our action laws for inform- and enable-actions that

$$[inform(j,i,\varphi)]B_{\{i,j\}}B_j\varphi$$

is valid.

4 Discussion and conclusion

4.1 Successor state axioms

Inspired by work of Moore [10], Scherl and Levesque have introduced a situation calculus based framework for resaoning about action and knowledge [11]. They adopt the strong permutation axiom, and even strengthen it to an equivalence that they call a successor state axiom. They show how such a principle allows for regression, which is a powerful reasoning technique in the situation calculus. In [15], Shapiro *et col.* adapt Scherl and Levesque's successor state axiom to belief, integrating a revision-like operation that is based on plausibility orderings. They define $B_i\varphi$ as truth of φ in the most plausible among the possible worlds. If a

doxastic action eliminates these most plausible possible worlds, then previously less plausible worlds become the most plausible ones. The plausibility ordering should be kept fixed.

While being intuitively appealing, their solution has several drawbacks. (1) As the authors note, it is restricted to deterministic actions. (2) "The specification of [the plausibility ordering] over the initial situation is the responsibility of the axiomatizer of the domain."[15] This is particularly demanding because (3) in order to guarantee that after α the set of possible worlds is nonempty, the authors require the set of possible worlds to contain enough worlds initially, restricting thus the agent's 'doxastic freedom'. (4) The approach is unsatisfactory when applied to communication. Consider the following example: agent k is competent at p, and j is not. Agent i is completely ignorant initially, and all possible worlds are equally plausible for i. Then (under adequate hypotheses of cooperation) we can expect that when j asserts p, then i adopts p, i.e. $[\mathit{inform}(j,i,p)]B_i p$. Moreover, as all worlds were equally plausible, p holds in every world possible for i. Therefore when subsequently k asserts $\neg p$, i will unavoidably move to an empty set of possible worlds.

4.2 Is this revision?

In our scenario, what i must do is to *revise* his beliefs about the executability of $\mathit{inform}(j,i,p)$. The normative framework for belief revision being the AGM theory [1], Segerberg's DDL [12, 13] integrates AGM revision into a doxastic logic. Linder *et al.* have a similar framework [9].

In the case of an expansion action, the agent's belief state is completely determined in DDL (by an axiom that correspond to the above permutation axiom). In the case of revision actions, apart from the AGM persistence postulate there is no principle relating an agent's belief state before and after α. In a sense, we have integrated into DDL a particular operation of revision, which establishes such a link.

Which of the AGM postulates do we satisfy? With a similar encoding as that of Shapiro *et col.* it can be shown that we satisfy the basic postulates (K*1) – (K*4), and (K*6). (The names of the postulates are as in [15]). If we define update actions as in [15] we satisfy the update postulates (K$\diamond$1), (K$\diamond$2), (K$\diamond$4), and (K$\diamond$5) just as there.

References

[1] Carlos Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *J. of Symbolic Logic*, 50:510–530, 1985.

[2] Alexandru Baltag. A logic of epistemic actions. Technical report, CWI, 2000. available from `http://www.cwi.nl/ abaltag/papers.html`.

[3] Alexandru Baltag, Lawrence S. Moss, and Slawomir Solecki. The logic of public announcements, common knowledge, and private suspicions. In *Proc. TARK'98*, pages 43–56. Morgan Kaufmann, 1998.

[4] Annette Bleeker and Jan van Eijck. The epistemics of encryption. In *Proc. LOFT 4*, 2000.

[5] Jelle Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, University of Amsterdam, 1999.

[6] Jelle Gerbrandy and Willem Groeneveld. Reasoning about information change. *J. of Logic, Language and Information*, 6(2), 1997.

[7] Andreas Herzig, Jérôme Lang, Dominique Longin, and Thomas Polacsek. A logic for planning under partial observability. In *Proc. Nat. (US) Conf. on Artificial Intelligence (AAAI'2000)*, Austin, Texas, August 2000.

[8] Andreas Herzig and Dominique Longin. Sensing and revision in a modal logic of belief and action. In *Proc. ECAI2002*, page 5 pages, 2002.

[9] Bernd van Linder, Wiebe van der Hoek, and John-Jules Ch. Meyer. Actions that make you change your mind. In Armin Laux and Heinrich Wansing, editors, *Knowledge and Belief in Philosophy and Artificial Intelligence*, pages 103–146. Akademie Verlag, 1995.

[10] Robert C. Moore. A formal theory of knowledge and action. In J.R. Hobbs and R.C. Moore, editors, *Formal Theories of the Commonsense World*, pages 319–358. Ablex, Norwood, NJ, 1985.

[11] Richard Scherl and Hector J. Levesque. The frame problem and knowledge producing actions. In *Proc. Nat. Conf. on AI (AAAI'93)*, pages 689–695. AAAI Press, 1993.

[12] Krister Segerberg. Belief revision from the point of view of doxastic logic. *Bulletin of the IGPL*, 3:534–553, 1995.

[13] Krister Segerberg. Two traditions in the logic of belief: bringing them together. Technical Report 9, Uppsala Prints and Preprints in Philosophy, 1996.

[14] S. Shapiro, Yves Lespérance, and Hector J. Levesque. Specifying communicative multi-agent systems. In W. Wobcke, M. Pagnucco, and C. Zhang, editors, *Agents and Multi-Agent Systems - Formalisms, Methodologies, and Applications*, pages 1–14. Springer-Verlag, LNAI 1441, 1998.

[15] S. Shapiro, M. Pagnucco, Y. Lespérance, and H. J. Levesque. Iterated belief change in the situation calculus. In *Proc. KR2000*, pages 527–538, 2000.

[16] Michael Thielscher. Representing the knowledge of a robot. In A. Cohn, F. Giunchiglia, and B. Selman, editors, *Proc. KR'00*, pages 109–120. Morgan Kaufmann, 2000.