

## BELIEF REVISION IN A DYNAMIC EPISTEMIC FRAMEWORK

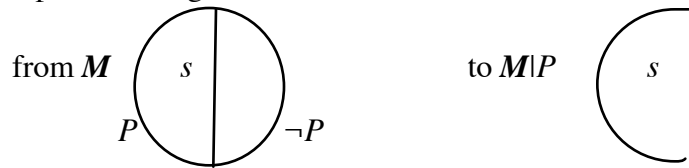
Johan van Benthem, Amsterdam & Stanford, <http://staff.science.uva.nl/~johan/>

Amsterdam–Lausanne–London Workshop LSE, Tutorial on Update Logic, July 2007

**Abstract** Logical Dynamics is about making (inter-)actions first-class citizens in logical systems. Dynamic-epistemic logics treat information update from observed events, changing the current doxastic - epistemic model. Belief revision can be treated in the same format, using update rules for plausibility relations which have been proposed also for preference change. We obtain complete sets of axioms for particular revision mechanisms, as well as a standard modal frame correspondence analysis for revision postulates in general AGM style. Ref. <http://staff.science.uva.nl/~johan/DL-BR-new.pdf>.

### 1 Information update as systematic model change

Card examples: learning  $P$  eliminates the worlds where  $P$  is false. In a picture:



Epistemic content of questions/answers involves iterated and common knowledge.

### 2 Dynamic epistemic logic: compositional analysis of effects

**Static epistemic logic** Language  $p \mid \neg \Box \mid \Box \mid K_i \Box \mid C_G \Box$ , models  $\mathbf{M} = (W, \{\sim_i \mid i \Box G\}, V)$ , with worlds  $W$ , accessibility relations  $\sim_i$ , and valuation  $V$ . Crucial epistemic truth conditions:  $\mathbf{M}, s \models K_i \Box$  iff for all  $t$  with  $s \sim_i t$ :  $\mathbf{M}, t \models \Box$ , and  $\mathbf{M}, s \models C_G \Box$  iff for all  $t$  that are reachable from  $s$  by some finite sequence of  $\sim_i$  steps ( $i \Box G$ ):  $\mathbf{M}, t \models \Box$ .

**Dynamic logic** of public announcement *PAL*: add action expressions:  $!P$  for all formulas  $P$ , and modal operators describing their effects (one simultaneous recursion):

$$\mathbf{M}, s \models [!P] \Box \quad \text{iff} \quad \text{if } \mathbf{M}, s \models P, \text{ then } \mathbf{M}|P, s \models \Box$$

*Theorem 1 (Plaza, Gerbrandy)* *PAL* without  $C_G$  is axiomatized completely by the usual laws of epistemic logic plus the following *reduction axioms*:

$$\begin{aligned} [!P]q & \quad \Box & P \Box q & \quad \text{for atomic facts } q \\ [!P]\neg\Box & \quad \Box & P \Box \neg[!P]\Box & \\ [!P]\Box\Box & \quad \Box & [!P]\Box\Box [!P]\Box & \\ [!P]K_i\Box & \quad \Box & P \Box K_i(P \Box [!P]\Box) & \\ [!P][!Q]\Box & \quad \Box & [!(P \Box [!P]Q)]\Box & \end{aligned}$$

**Methodology** Add dynamic superstructure to static base logic. *Compositional* analysis of all post-conditions. Requires design for *pre-encoding* in static language. Example:  $[!P]C_G\Box$  requires new notion: ‘conditional common knowledge’ with reduction axiom:

$$[!P]C_G(\Box, \Box) \quad \Box \quad C_G(P \Box [!P]\Box, [!P]\Box).$$

By-product: reduction dynamic logic to completeness/validity in static language.

**Program** ‘Dynamification’ of existing logics, making the underlying actions explicit.

### 3 Update by general events with partial observation

*Email*: epistemic-dynamic function of  $cc$ ,  $bcc$ . Computer security. High-light: *Games* designed to manipulate information flow (*Cluedo*). Partial observation of events.

**Event models**  $A = (E, \{\sim_i \mid i \in G\}, \{PRE_e \mid e \in E\})$ . Scenario: relevant events, relations  $\sim_i$  encode what agents cannot distinguish. I check my card: you cannot tell 'my seeing red' from 'my seeing black'. Events  $e$  have *preconditions*  $PRE_e$  for their execution: my having a red card, not knowing the answer to my question, etc. Update Rule: for any epistemic model  $(M, s)$  and event model  $(A, e)$ , the **product model**  $(M \times A, (s, e))$  has

Domain  $\{(s, e) \mid s \text{ a world in } M, e \text{ an event in } A, (M, s) \models PRE_e\}$ ,

Accessibility:  $(s, e) \sim_i (t, f)$  iff both  $s \sim_i t$  and  $e \sim_i f$ ,

Valuation for atoms  $p$  at  $(s, e)$  is that at  $s$  in  $M$ . (can be generalized to world change)

Product update deals with misleading actions as well as truthful ones, and with *belief* as well as knowledge. Epistemic models can even get *larger* as update proceeds ( $bcc$ )!

**Dynamic-epistemic logic LEA**:  $p \mid \neg \Box \mid \Box \Box \mid K_i \Box \mid C_G \Box \mid [A, a] \Box : (A, e)$  any event model with actual event  $e$ . Semantics:  $M, s \models [A, e] \Box$  iff  $M \times A, (s, e) \models \Box$ .

*Theorem 2 (BMS)* LEA is effectively axiomatizable and decidable.

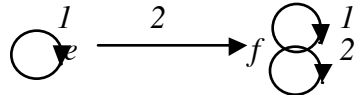
The key reduction axiom is the one extending that for public announcement:

$$[A, e] K_i \Box \quad \Box \quad PRE_e \Box \quad \Box \{ K_i [A, f] \Box \} \mid f \sim_i e \text{ in } A$$

Further issues: extensions to richer languages like *LCC*, or *epistemic  $\Box$ -calculus*.

(a) **Idealized agents**. Product Update: *Perfect Recall*, and *No Miracles*. Diversity?

(b) **Common knowledge** or **common belief** in subgroups: *secrets*. Example (*BEK*):

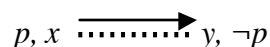


### 4 Belief revision theory, the bare necessities

Update  $T+A$ , revision  $T^*A$ , contraction  $T-A$ . *AGM Postulates*: general constraints. Grove sphere models. Representation theorems. Key topics:  $B \Box T^*A$  close to (but not quite) Ramsey test for  $T \vdash A \Box B$ , issue of iteration, proliferation of revision policies. *Conditional logic*. Restrictions: only factual assertions  $A$ , not epistemic ones – single agent scenarios – recent view: special case of ‘belief merge’ from different sources.

### 5 Belief change under hard facts

**Knowledge and belief** Reinterpret  $K_i \Box$  to weaker  $[ ]_i \Box$  ‘to the best of  $i$ ’s information’. Even better: have both knowledge and belief modalities in base language. Example: a model in whose actual world  $x$ , I believe that  $\neg p$ , though  $p$  is in fact the case:



**Problem with update DEL** does not do true belief revision. ‘Hard announcement’ event  $!p$  turns this into the one-world model  $\{x\}$  where I believe that  $p$ , but even  $B\Box$ !

**Belief and plausibility models** Solution: conditional logic of relative plausibility:

$$M, s \models B_i \Box \quad \text{iff} \quad M, t \models \Box \quad \text{for all worlds } t \text{ minimal in the ordering } \Box xy. \leq_{i, s} xy.$$

Belief change under hard facts:

$$[!P] B_i \Box \quad \Box \quad P \Box B_i ([!P] \Box \setminus P)$$

Conditional belief helps *pre-encode* beliefs we would have if we learnt certain things:

$$M, s \models B_i (\Box \Box) \quad \text{iff} \quad M, t \models \Box \quad \text{for all worlds } t \text{ which are minimal for } \Box xy. \leq_{i, s} xy \text{ in the set } \{u \mid M, u \models \Box\}.$$

Satisfies the standard principles of the minimal conditional logic.

**Theorem 3** The logic of conditional belief under public announcements is axiomatized completely by (a) any complete base logic of  $B_i(\Box \Box)$  for favorite model class, (b) *PAL* reduction axioms, plus (c) a reduction axiom for conditional beliefs:

$$[!P] B_i (\Box \Box) \quad \Box \quad P \Box B_i ([!P] \Box \setminus P \Box [!P] \Box)$$

**Discussion:** close to Ramsey test, but difference in ‘dynamics’: explain role of  $[!P]$ . Combines immediately with earlier knowledge axioms. (*Preserve interplay axioms?*)

## 6 Belief change under soft facts

**Soft triggers** Call for belief revision  $*p$  ‘softer’ than a call for world elimination, introducing just a greater ‘preference’ for  $p$ -worlds, without totally abandoning the others. *Example* (Spohn, Veltman). *Default rule*  $A \Box B$  does not say all  $A$ -worlds are  $B$ . It just makes the ‘exceptional’  $A \Box \neg B$ -worlds less plausible. Also: resolving conflict, negotiation, belief merge. Scenarios with ‘soft information’ do not eliminate worlds, they rather *change the plausibility ordering* of the existing worlds. Recurrent instruction:

*Lexicographic upgrade*  $\Box P$  changes the current model  $M$  to  $M\Box P$ :

$P$ -worlds now better than all  $\neg P$ -worlds; within zones, old order remains.

Social revolution: underclass  $P$  now becomes upper class. Other policies (Rott’s ‘27’; or with Macchiavelli’s advice: ‘conservative belief revision’,  $\uparrow P$ , just co-opt *leaders* of the underclass!) Dynamic language in standard *DEL*-style:

$$M, s \models [\Box P] \Box \quad \text{iff} \quad M\Box P, s \models \Box.$$

The static pre-encoding is again done by *conditional beliefs*:

**Theorem 4** The dynamic logic of lexicographic upgrade is axiomatized completely by logic of conditional belief + compositional analysis of effects of revision:

$$[\Box P] q \Box q, \quad [\Box P] \neg \Box \Box \neg [\Box P] \Box, \quad [\Box P] (\Box \Box) \Box [\Box P] \Box \Box [\Box P] \Box \\ [\Box P] B(\Box \Box) \Box (\Box (P \Box [\Box P] \Box) \Box B([\Box P] \Box \setminus P \Box [\Box P] \Box)) \quad B([\Box P] \Box \setminus [\Box P] \Box)$$

Here  $E$  is the existential modality ‘in some world’. Case of just factual assertions:

$$[\Box P] B(\Box \Box) \Box (\Box(P \Box \Box) \Box B(\Box \mid P \Box \Box)) \quad B(\Box \mid \Box)$$

$$[\Box P] B \Box \Box (\Box P \Box B([\Box P] \Box \mid P)) \quad B([\Box P] \Box) \quad (\text{a bit more 'Ramsey-like'})$$

**Conclusion** *Constructive belief revision theory can be studied by standard modal techniques.* We find complete axiomatizations for concrete belief revision policies defined as systematic relation change (e.g., conservative upgrade  $\uparrow P$ ). Many further policies axiomatized at *ILLC* in the past year, also including ‘point assignment’, etc.

## 7 AGM postulates as modal frame correspondence

Abstract constraints on model-changing operations are just modal frame correspondences (as with ‘*K4* – transitivity’). Let  $\clubsuit A$  be any operation from models  $\mathcal{M}$  and sets of worlds  $A$  in it to a new model  $\mathcal{M}\clubsuit A$  with the same worlds but a changed relation  $\leq_s$ .

*Fact* The formula  $[\clubsuit p] B p$  says that the best worlds in  $\mathcal{M}\clubsuit p$  are all in  $p$ .

*Fact* The formula  $B(q \mid p) \Box [\clubsuit p] B q$  expresses ‘rule by the upper classes’.

Now *invert* earlier completeness results to see the content of reduction axioms:

*Theorem 5* Eliminative update fixed by the earlier  $[\clubsuit p] K q \Box (p \Box K [\clubsuit p] q)$ .

*Theorem 6* The formula  $[\clubsuit p] B(q \mid r) \Box (\Box(p \Box r) \Box B(q \mid p \Box r) \quad B(q \mid r))$  holds in a frame iff the operation interpreting  $\clubsuit p$  is lexicographic upgrade.

Similar analysis for most exciting AGM Postulates (*two* model changing operations):

- (a)  $[\clubsuit (p \Box q)] B r \Box [!q][\clubsuit p] B r$
- (b)  $[\clubsuit p] E q \Box [!q][\clubsuit p] B r \Box [\clubsuit (p \Box q)] B r$

**Conclusion** *Axiomatic belief revision theory can be studied by standard modal means.*

But of course, we would really want extended AGM postulates for *conditional* belief! Connected to problem of iterated belief revision: what about  $[\Box A] [\Box B] \Box$ ?

## 8 Discussion

**Richer triggers** DEL event models model much richer multi-agent scenarios, where agents have beliefs about events. Graded *strengths* of beliefs (Spohn, Aucher, Liu).

**Real world change** easily added (van Benthem, van Eijck & Kooi). Katsuno-Mendelzon.

**Obstacle to a happy marriage** DEL ‘backward-looking’: computes what we believe after an event takes place, via *preconditions*. AGM: ‘forward-looking’ instructions “come to believe”, “see to it that”. General events via *postconditions* ill-defined. Possible remedy: work over temporal universes which constrain future developments.

## 9 Longer-term processes: conversation, games, learning theory

**Program structures** in conversation: Sequential composition  $;$ , Guarded choice *IF THEN ELSE*, Guarded iteration *WHILE DO*. Even parallel composition  $\parallel$  makes sense.

**Games** Strategic interaction (learning/teaching games: ‘Teaching the Unwilling’). On top of standard game theory, *DEL* gives fine-structure of deliberation and moves.

**Epistemic temporal logics** Branching time: Halpern et al., Parikh et al. For some connections, see van Benthem & Pacuit 2006, ‘*The Tree of Knowledge in Action*’.

**Learning theory** Natural continuation of update and revision logic ‘by other means’.

**Related approaches** Segerberg, Girard, van Ditmarsch, Baltag & Smets, *ESSLLI 2005*.

**References** Stanford seminar page <http://staff.science.uva.nl/~johan/seminar2006.html>.

#### *Addendum 1 Further material on Information Dynamics & Logic and Games*

See the papers on these two topics under ‘Research’ at <http://staff.science.uva.nl/~johan/>. These include a survey of Open Problems in logic & games, and recent work on comparison with other frameworks, such as temporal logic. Major issue raised at the LSE Workshop: how can dynamic logics for *micro-structure* of reasoning and action link up ‘in the limit’ with the elaborate type space ‘macro-models’ presented in Adam Brandenburtger’s tutorial?

#### *Addendum 2 Recent thought: a Social Choice view of Belief Revision*

Idea: **(a)** *Belief revision is a form of belief merge*, between an agent ‘Me’ and a new ‘single issue agent’  $\uparrow A$  who only cares about some proposition  $A$ , and orders worlds in connection with it. Thus, belief revision is like belief merge where we give priority to one of the agents. I even think this is reasonable for human agents, viewed as ‘social collections’ of their sense organs, memory, etc. **(b)** Belief merge is merge of plausibility relations, but we cannot say what it is in general: there is a *missing parameter* or ‘hidden variable’, viz. the *group structure*. This I take to be the true message of the impossibility results in social choice: Voting procedures in flat unordered groups cannot work well in general, and *that explains* why human societies and organizations have developed beneficial hierarchies of information and influence. **(c)** Issue: the most sophisticated preference merge I know (Andreka, Ryan & Schobbens) needs ‘priority graphs’ for a group to find the social ordering. But in order to work, it needs partial rather than total orders, and it has *four* options:  $>$ ,  $<$ , indifference, and *incomparability*. **(d)** Social choice is a two-way process: the social merge, once made, also feeds back, and *changes the original individual preferences*. Maybe we should have a notion of ‘believe as a member of group  $C$ ’ taking this further context into account. Anyway, there is a lot of feed-back dynamics here: which seems highly plausible as an account of actual deliberation (where you would also have to throw in the role of new factual information). **(e)** Multi-agent belief revision and belief merge are more or less the same as social choice, and the proper way of dealing with all this is dynamic logic of *deliberative opinion change*.

[New result, to be checked] We characterize the Priority Update Rule by abstract postulates inspired by Generalized Quantifier Theory and Social Choice/Preference Aggregation.

**The bare math** Two ordered sets  $(A, R)$  and  $(B, S)$ , think of the initial plausibility model  $M$  and the event model  $E$ : we seek to order the product  $A \times B$  via some suitable relation  $O(R, S)$ . We order a subset of the full product where a constraint may rule out some pairs ('preconditions', but the precise nature is irrelevant). Let us assume that the relations are connected orders, though this is not essential. This generalizes the usual setting of preference merge, where the individual preferences are over the same domain. This is of more general use anyway, also in social choice: we are really constructing *two things*: both (a) a vector space (the 'social space' of all relevant group outcomes) plus (b) the social order over the latter.

### Choice conditions

(a) *Permutation invariance in some obvious sense.*

Any pair of permutations of  $A$  and of  $B$  leads to an obvious invariance condition on  $O$  on the models  $(A, R)$ ,  $(B, S)$  and their permuted versions, which forces our definitions to be *uniform*. This is the usual structural form of a 'logical definability' condition in many areas.

(b) *Locality:  $O(R, S) ((a, b), (a', b'))$  iff  $O(R \setminus \{a, a'\}, S \setminus \{b, b'\}) ((a, b), (a', b'))$ .*

This is a very strong version of 'Independence from Irrelevant Alternatives'.

Together, (a), (b) force our  $O$  to be definable by just its behaviour in a 3x3-table:

$S$ on $b, b'$	$\square$	$\square$	$\square$
$R$ on $a, a'$	$\square$	-	-
	$\square$	-	-
	$\square$	-	-

That is all, at least, as long as we work on connected orders. Otherwise, we would need four options in the table, including #: genuine incomparability between two options.

Now we rule out all but two combinations by the following appealing conditions:

(c) *Disregarding Abstentions:*

If you vote indifferent ( $\square$ ), then the other(s) determine the outcome.

(d) *Group Alignment:*

'if anyone changes their vote to get closer to the group outcome, the group outcome does not change'.

This fills the diagram as follows:

$S$ on $b, b'$	$\square$	$\square$	$\square$
$R$ on $a, a'$	$\square$	$\square$	$\square$
	$\square$	$\square$	$\square$
	$\square$	$\square$	$\square$

So, we have only two slots remaining. It is also easy to see that, using Group Alignment, that these slots cannot have entries  $\square$  of indifference. So, we must rather have

*Overruling:* both go with one of the two relations, and *always with the same one* (because of the permutation invariance).

***The main Theorem:***

A preference aggregation function is a priority update iff it satisfies Permutation Invariance, Locality, Disregarding Abstentions, and Group Alignment.

Favouring the second relation is the Priority Update Rule, favouring the former a competitor we may either want to rule out with some further principle, or leave as an alternative.

***Possible variations*** Of course, this also shows that the above assumptions are very strong:

*Example 1:* Locality rules out the 'conservative version' of preference update.

Determining whether you are a 'best' *A-item* requires looking at more worlds.

*Example 2:* Group Alignment rules out Democracy.

We cannot put an indifference when the individual agents' preferences *clash*, even though this seems a viable alternative option. I had the latter in the previous version of this note, and it would be good to have a second Theorem allowing this by changing the postulates a bit.

***Further issues:***

\* Extend all this to *partial orders*, where we now need two relations  $\leq$  and  $<$ ; and where the literature on preference aggregation allows for *four* admissible outcomes:  $x < y$ ,  $x > y$ ,  $x \sim y$ , and  $x \# y$  ('incomparable'). In that case, Priority Update should be adapted. One option, in line with the proposal by Andréka, Ryan & Schobbens, is to allow a partial priority order on agents as well, and then say that (this is a two-agent version of something more general):

$(s, e) \leq (t, f)$  iff both agents have  $\leq$  on their component, or if a failure of  $\leq$  occurs for an agent, some higher-placed agent has a  $<$ .

\* *Which rule is more general?* E.g., Priority Update can never make an established strict preference indifferent again, while Democracy can achieve any effect that Priority can achieve by judiciously interpolating some further signals. (In longer belief revision scenarios, I find this observation a somewhat strange feature of Priority Update).

\* Extend to *plausibility merge between more relations* for 'belief merge'?

***Addendum*** At the LSE Workshop, Christian List pointed out connections between the above result with May's Theorem and related literature on social choice. To be continued.