LOGICAL DYNAMICS OF INFORMATION AND EVALUATION

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

SOCREAL Workshop, University of Hokkaido, 27 March 2010

1 Logic and the balance of information and evaluation

Logic has traditionally been concerned with truth, information, knowledge, and belief. Evaluation and other ways of 'colouring' the world as we see it belong to another realm. But one can argue that truly rational agency involves a *harmony* between informational and evaluative attitudes: including the ability to change both as required by circumstances.

2 Logical dynamics of information-driven rational agency

Agency involves a broad range of correct information processing:

Restaurant: how to figure out who has which dish? *Inference, questions*.

Card games: planning moves using theory of mind. *Mutual knowledge*.

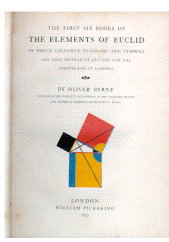
Skype Exam: secret voting on a public channel. *New social procedures*.

To be rational is to reason *intelligently*. Logic of all basic informational processes:

"Zhi: Wen, Shuo, Qin" 知 问 说 亲 (communication, inference, observation)

To be rational is to act *intelligently*. Evaluation, goals, preferences, decisions, actions.

To be rational is to interact *intelligently*. Argumentation, communication, games.



3 The research program

Stage One: charting the relevant abilities of an agent. Knowledge, belief, etc. Different sorts of information: semantic, syntactic. Awareness. Information vs. evaluation.

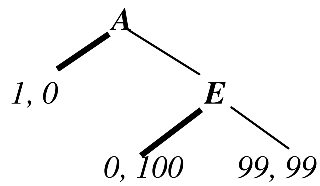
Stage Two: making the local dynamics explicit. Observation, inference, communication. Knowledge and information update, belief revision (learning, self-correction), issue management (questions, agenda), changing preferences and goals.

Stage Three: social features. Groups, collective information, attitudes, action.

Stage Four: mid-term temporal behaviour. Conversation, games and strategies.

Stage Five: long-term temporal behaviour. Dynamic logic meets dynamical systems.

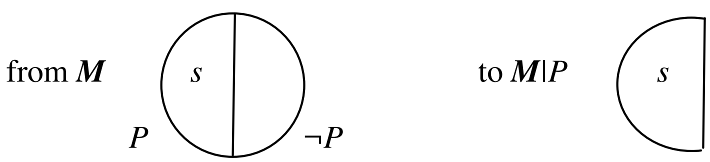
It all comes together naturally in *games*, mathematical, philosophical, computational logic:



4 Pilot system: dynamifying epistemic logic to public announcement logic

Static base logic. Language $p \mid \neg\phi \mid \phi \vee \psi \mid K_i\phi \mid C_G\phi$, models $\boldsymbol{M} = (W, \{\sim_i \mid i \in G\}, V)$, with worlds W, accessibility relations $\sim_i$, and valuation V. Truth conditions ('knowledge as semantic information'): $\boldsymbol{M}, s \models K_i\phi$ iff for all t with $s \sim_i t$: $\boldsymbol{M}, t \models \phi$, and $\boldsymbol{M}, s \models C_G\phi$ iff for all t reachable from s by some finite sequence of $\sim_i$ steps ($i \in G$): $\boldsymbol{M}, t \models \phi$.

Dynamic logics describe key update steps in observation and communication. Pilot system: *PAL*. Hard information update: learning P eliminates worlds with P false:



Language extension: $\boldsymbol{M}, s \models [!P]\,\phi$ iff *if* $\boldsymbol{M}, s \models P$, *then* $\boldsymbol{M}|P, s \models \phi$

Theorem *PAL* axiomatized completely by epistemic logic plus *recursion axioms*:

$[!P]q$	$\leftrightarrow$	$P \rightarrow q$	*for atomic facts q*
$[!P]\neg\phi$	$\leftrightarrow$	$P \rightarrow \neg[!P]\phi$	
$[!P]\phi \wedge \psi$	$\leftrightarrow$	$[!P]\phi \wedge [!P]\psi$	
$\boldsymbol{[!P]K_i\phi}$	$\leftrightarrow$	$\boldsymbol{P \rightarrow K_i(P \rightarrow [!P]\phi)}$	*key recursion axiom*

Aside on 'schematic validities': $[!P][!Q]\phi \leftrightarrow [!(P \wedge [!P]Q)]\phi$

Methodology ***Make actions explicit*** on top of static logic. Compositional analysis effects. *Hunt for right* ***recursion axioms***: also for private information, belief revision, questions. Can describe information flow under many events. Current developments: (a) eventually, no reduction: from single steps to *temporal setting with protocols*, (b) syntactic dynamics.

Similar methods developed for evaluation dynamics: ***preference, intentions, goals***.

5 Hard information, world elimination and belief change

Belief and plausibility models Conditional logic of relative plausibility:

$\boldsymbol{M}, s \models B_i\phi$ iff $\boldsymbol{M}, t \models \phi$ for all worlds t *minimal* in the ordering $\lambda xy. \leq_{i,s} xy$.

Belief change under hard facts:

$[!P]\, B_i\phi \quad \leftrightarrow \quad P \rightarrow B_i^{\,P}\, [!P]\phi$

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

$\boldsymbol{M}, s \models B_i^{\psi}\, \phi$ iff $\boldsymbol{M}, t \models \phi$ for all worlds t which

are minimal for $\lambda xy. \leq_{i,s} xy$ in the set $\{u \mid \boldsymbol{M}, u \models \psi\}$.

Satisfies the standard principles of the minimal conditional logic.

How beliefs change under hard information (cf. changing obligations):

Theorem The logic of knowledge and conditional belief under public announcements is axiomatized completely by (a) any complete base logic with $B_i^{\psi}\, \phi$ for favorite model class, (b) *PAL* axioms, plus (c) a recursion axiom for conditional beliefs:

$[!P]\, B_i^{\psi}\, \phi \quad \leftrightarrow \quad P \rightarrow B_i^{\,P \wedge [!P]\psi}\, [!P]\phi$

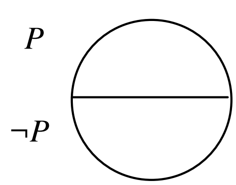
Interesting scenarios Misleading true information. Motivates new notion of 'safe belief' (truth in all more plausible worlds) between knowledge and belief: also in *AI,* game theory, philosophy. So, ***what is natural repertoire of agent attitudes?*** Same issues for evaluation.

6 Belief revision as relation change under soft information

Soft triggers Call for belief change $*p$ 'softer' than world elimination, introducing just greater 'preference' for p-worlds. *Default rule* $A \Rightarrow B$ does not say all A-worlds are B. Just makes exceptional $A \wedge \neg B$-worlds less plausible. 'Soft information' does not eliminate, but *changes the plausibility ordering* of the existing worlds. Typical case:

Lexicographic upgrade $\Uparrow P$ changes the current model $\boldsymbol{M}$ to $\boldsymbol{M}\Uparrow P$:

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 Macchiavelli's conservative advice, $\uparrow P$, just co-opt *leaders* of the underclass!). Logic:

$$\boldsymbol{M}, s \models [\Uparrow P]\phi \qquad \text{iff} \qquad \boldsymbol{M}\Uparrow P, s \models \phi.$$

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

$$[\Uparrow P]\, B^{\psi}\phi \;\leftrightarrow\; (\langle\rangle(P \wedge [\Uparrow P]\psi) \wedge B^{P \wedge [\Uparrow P]\psi}\, [\Uparrow P]\phi) \vee (\neg\langle\rangle(P \wedge [\Uparrow P]\psi) \wedge B^{[\Uparrow P]\psi}\, [\Uparrow P]\phi$$

Here E is the epistemic existential modality. Special cases: factual formulas. We can now axiomatize many policies for relation change, in various formats (cf. Baltag's lecture).

7 Preference and deontic betterness dynamics

Static models: ***betterness order*** on worlds. Deontic setting is social: betterness of the moral authority. Recurrent choice: total orders or *pre-orders*: indifference ≠ incomparability?

Modal language, complete axiomatizations in many flavours (back to Bouilier in AI).

Generic preference $P\varphi\psi$ between propositions by set lifting (Liu 2008, and long literature).

Special features of preference: ceteris paribus reasoning (van Benthem, Girard, Roy 2009).

Dynamic actions: betterness order can change under explicit suggestions, commands, etc. (Yamada, Liu). Example 'suggestion' $\#\boldsymbol{\phi}$: remove all links from $\boldsymbol{\phi}$–worlds to $\neg\boldsymbol{\phi}$–worlds:

For each model $\boldsymbol{M}, w$, the model $\boldsymbol{M}\#\varphi, w$ is $\boldsymbol{M}, w$ with the new relation $\leq' = \leq - \{(x, y) \mid \boldsymbol{M}, x \models \varphi \;\&\; \boldsymbol{M}, y \models \neg\varphi\}$.

Next, we enrich the formal language by adding action modalities interpreted as follows: [1]

$$M, w \models [\#(\varphi)]\psi \qquad \text{iff} \qquad M\#\varphi, w \models \psi$$

Theorem The dynamic logic of preference change under suggestions is axiomatized completely by the static modal logic of the underlying model class plus axioms as above, plus this key recursion axiom for betterness after suggestions:

$$\boldsymbol{[\#(\varphi)]\langle\leq\rangle\psi \quad \leftrightarrow \quad (\neg\varphi \;\&\; \langle\leq\rangle[\#(\varphi)]\psi) \vee ((\varphi \;\&\; \langle\leq\rangle(\varphi \;\&\; [\#(\varphi)]\psi).}$$

Which triggers for betterness change? Speech acts, juridical acts? Feelings?

[1] Here the syntax is recursive: the formula φ may itself contain dynamic modalities.

8 Priority graphs and two-level dynamics

Idea: derive betterness order from criteria (in many disciplines, even linguistics).

Linear priority sequences ***P***, de Jongh & Liu 2007, ***derived world order***:

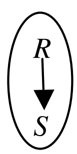
> $x < y$ iff x, y differ in at least one property in ***P***, and
> the first $P \in \mathbf{P}$ where this happens is one with $Py, \neg Px$.

Is lexicographic ordering, if we view each property $P \in \mathbf{P}$ as $x \leq^P y$ iff $(Py \rightarrow Px)$.

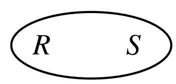
Graphs for ***relation merge*** (Andréka, Ryan & Schobbens 2002):

> Given an ordered *priority graph* $\mathbf{G} = (G, <)$ of indices for relations that may have multiple occurrences in the graph, the *merged group priority relation* is:
> $x \leq_G y$ iff for all indices $i \in \mathbf{G}$, either $x \leq_i y$, or there is some $j > i$ in $\mathbf{G}$ with $x <_i y$

Example



(hierarchy)



(juxtaposition)

Putting R above S, $x \leq y$ iff $x\ R\ y \wedge x\ S\ y$ or there is a difference in the S relation and $x\ R^+\ y$ with R^+ the strict form of R. Putting R alongside S is intersection $x\ R\ y \wedge x\ S\ y$.

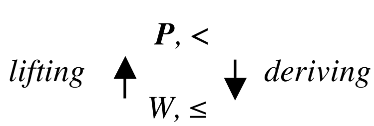
Dynamics at graph level Add propositions in front, behind, in middle (syntactic).

Two natural mathematical operations that change and combine priority graphs:

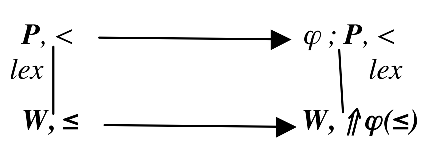
> *sequential composition* $G_1 ; G_2$ (putting G_1 on top of G_2, retaining the same order inside) and *parallel composition* $G_1 \parallel G_2$ (disjoint union of graphs).

Removing items, much harder.

Two-level dynamics *Two-level structures* $(W, \leq, P, <)$ having both worlds with a betterness order $\leq$ and a set of 'important propositions' with a primitive priority order $<$:



Fact The identity $lex(\varphi ; \mathbf{P}) = \Uparrow\varphi\,(lex(\mathbf{P}))$ holds, making this diagram commute:



9 Issues from information dynamics that also play for evaluation

Evaluation Sources for its importance: individual values, decisions, games, 'belonging' and shared agency, society. (Connections to social epistemology and cognitive science.)

Concrete case study: deontics Does this line of thinking apply to deontic logic, norms, laws? Law = priority graph? (See lecture of Fenrong Liu.) Which sort of structure needed to model real deliberation and *judgment*? A few ways of extending the above framework:

Entanglement: preference, belief, obligation, action

Foundations of action: preference ~ beliefs ~ actions. Belief-entangled preference (de Jongh & Liu). Duties and having or gathering information: Pacuit & Parikh.

Information change and *betterness change*, both needed? (Curse: *re-encoding*.)

Games and social interaction

Deontics is social. Legal procedure in court. Laws as social software: Law and Economics.

Fine-grained models: syntactic plus semantic dynamics

Problem of semantic versus syntactic information: informativity of deduction.

Syntactic structure of criteria, laws, proof steps, organization of reasoning.

Moral duty to deliberate well? Duty to think?

10 Conclusion

The dynamic logical tools that we have seem to apply to both information and evaluation. Now the challenge is to look at concrete issues, and see how far they take us.

References

J. van Benthem, 1996, *Exploring Logical Dynamics*, CSLI Publications, Stanford.

2009, 'For Better of for Worse: Dynamic Logics of Preference', in T. Grüne-Yanoff & S-O Hansson, eds., *Preference Change*, Springer, Dordrecht, 57–84.

2010, *Dynamic Logic of Information and Interaction*, Cambridge UP, to appear.