Comments on van Benthem's "Dynamic Logic for Belief Change"

Gillian Russell

Washington University in St Louis

Formal Epistemology Workshop, 24th May, 2005

The Base Logic

$\langle W, \{\sim_i | i \in G\}, V\rangle$

- a set of *worlds*
- a set of binary relations of epistemic accessibility $\sim_i$ (relative to an agent)
- a propositional valuation function.

$$M, s \models K_i\phi \quad \text{iff} \quad \text{for all } t \text{ with } s \sim_i t, \ \ M, t \models \phi$$

Public Announcement Logic

The language of PAL contains a new expression '!', which when combined with a formula forms a modal operator '$P!$'. ('$[P!]\phi$' is then itself a formula.) We can use the same model theory to give semantics for PAL as we did for the base epistemic logic, and the semantics of [P!] are as follows:

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

Sentences of PAL can be used to make claims about what would be the case, and what would be known, were 'P' to be announced.

Modelling Soft Triggers

Our models are now triples $\langle W, \{\leq_{i,s}\}_{i \in P}, V\rangle$, in which

- W is a set of worlds
- the $\leq_{i,s}$ are ternary comparison relations
- V is a propositional valuation function

$\leq_{i,s}$ *xy* is glossed as 'in world *s*, agent *i* considers *x* at least as plausible as *y*'

The Belief Operator

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

Intuitively, that means that i believes that ϕ at world s iff ϕ is true at all the worlds which are most plausible (with respect to world s)

Responding to Soft Information (1)

Lexicographic Upgrade ($\Uparrow$):

$\Uparrow P$ is an instruction for replacing the current ordering relation $\leq$ between worlds by the following: all P worlds become better than all $\neg P$ worlds, and within those two zones, the ordering remains the same.

Responding to Soft Information (2)

Elite Change ($\uparrow$):

$\uparrow$ *P* replaces the current ordering relation $\leq$ by the following: the best *P*-worlds come out on top, but apart from that, the old ordering remains the same.

A version of an old argument

1. A logic is something that determines an implication relation on sentences.
2. Anything which provides a model for belief revision ought to determine a sensible function from belief states and new information to new belief states.
3. The facts about implication relations between sentences do not determine a sensible function from belief states and new information to new belief states.
4. So no mere logic provides a model for belief revision.

The Argument for Premise 3

The argument for premise 3 is by example:
Suppose A and B together imply C. Does it follow that if you believe A and you learn B that you should believe C?
No, as the following two counterexamples show.
1. Suppose C inconsistent. You shouldn't accept it. What should you do instead? Perhaps give up belief in one of the premises, but which one?
2. Suppose you already believe $\neg C$. Then you might make your beliefs consistent by giving up one of the premises, or by giving up $\neg C$. Or you might suspend belief in all of the propositions and resolve to investigate the matter further on Monday morning.

The Options

The argument implies that van Benthem's model of belief revision either isn't a logic, or doesn't provide a plausible model of belief revision. So one of the following must be the case:

1. van Benthem's model isn't a logic
2. van Benthem's model isn't a plausible model of belief revision
3. or there is something wrong with the argument

Do van Benthem's more sophisticated models imply that if an agent believes p, and she learns $p \to q$, that she should believe q?

For example: Suppose that for some model M, and world s, $M, s \models B_i p$. Is it then the case that $M, s \models [\Uparrow (p \to q)] B_i q$?

Well, to answer the question, we imagine changing the $\leq$ relation on the worlds in M so that all $(p \to q)$-worlds are more plausible than any of the not-$(p \to q)$-worlds, but leave the remaining ordering the same. Then we ask whether $B_i q$ is true at s in the transformed model.

It will be if at each of the most plausible worlds (with respect to s), q is true. And since at each of the most plausible worlds, (with respect to s), $p \to q$ is true, that will be the case if at each of the most plausible worlds p is true.

Now there could be $(p \to q)$-worlds where p is true, and $(p \to q)$-worlds where p is not true.

Claim: since the ordering is otherwise left the same, and since our agent believed p originally, $M, s \models B_i p$, worlds at which p is true will outrank those at which it is not. And in those worlds q is true. So at the most plausible worlds (w.r.t. s)in the new model, p is true, which means $B_i p$ is true at s, which means that $M, s \models [\Uparrow (p \to q)]B_i q$ is true in the original model.