Comment on Sonja Smet's talk

"Iterated Dynamic Belief Revision"

(based on joint work with Alexandru Baltag)

Franz Dietrich

LSE-Groningen Workshop
London
February 2008

Many topics in this talk

Including:

How revise beliefs if I learn a Moore sentence?

Ramsey test: AGM revision fails it (Gaerdenfors' Theorem), dynamic revision satisfies it, AGM revision satisfies a weaker version of the Ramsey test.

(Non-)convergence of beliefs under repeated revision using certain DEL revision policies.

The puzzle about Moore sentences recapitulated

Let L be a suitable[1] logic.

Let $\mathcal{T}$ be the set of deductively closed sets $T \subseteq L$, called 'belief sets' or 'theories'

Let $* : \mathcal{T} \times L \to \mathcal{T}$, $(T, p) \mapsto T * p$ be an AGM belief revision operator

Puzzle: If the agent

- starts from a prior belief set $T \in \mathcal{T}$
- learns a Moore sentence $p \wedge \neg Bp$ ('p and it is not believed that p'), where $p \in L$,

then this Moore sentence

- becomes part of the new belief set $T * (p \wedge \neg Bp)$ (by the 'success' property of AGM operators)
- but becomes false, because Bp becomes true.

[1]We need to express certain sentences, e.g. Moore sentences.

Sonja solves the puzzle...

by clarifying that the beliefs that AGM revision generates are beliefs about the past (i.e. before the revision)
$\rightarrow$ even after learning that $p \wedge \neg Bp$, it is true that p **was** not believed
$\rightarrow$ so the Moore sentence remains true **as a sentence about the past**

In general, AGM revising T by p generates not the belief set $Th(s \otimes p)$ but the belief set

$$T * p = \{q \in L :< before > q \in Th(s \otimes p)\},$$

where s is the ex ante world, $\otimes$ is a 'change of world' operation, and $< before >$ is a past tense operator.

A general remark

Statements like 'I am hungry' or 'I believe that p' lack an explicit time reference.

Two interpretation of the (implicit) time reference:

- a (fixed) time: 'I am hungry at date t', 'I believe that p at date t', ... where t is a fixed time.
  –> such statements have unchanged truth values
- the 'now' time: 'I am now hungry', 'I now believe that p', ...
  –> such statements change their truth value as time progresses, because 'now' changes its meaning.

Under the fixed time interpretation, no paradox

Take the fixed-time interpretation of B (not Sonja's interpretation)

Let's make time explicit: the agent learns $p \wedge \neg B_0 p$, where

$B_t p$ means 'it is believed at time t that p'

$t = 0$: time before learning

$t = 1$: time after learning

No paradox here, because the sentence $p \wedge \neg B_0 p$

- stays true at time 1 ($B_0 p$ stays false, $B_1 p$ stays true)
- can thus be safely taken into the new belief set

Sonja prefers the <before> operator to a fixed-time index...

Under the 'now' interpretation, don't we have a more general problem?

Why look at Moore sentences in particular? Doesn't *every* sentence with an (implicit or explicit) 'now' pose a problem to AGM? Any 'now' sentences can change its truth value as time changes, i.e. as reality (i.e. reality's time centre) changes.
Can AGM revision handle

- only changes coming from new information,
- or also changes from time change?

Example: My current belief set T contains p : 'I am (now) hungry'.
After eating, p becomes false.
How revise T? Revise T into $T * \neg p$?

Back to Moore sentences (with Sonja's implicit 'now' interpretation)

"Solving" the problem by denying that Moore sentences can be learnt

Arguably, just as a rational agent never learns contradictions, he doesn't learn the Moore sentence $p \wedge \neg Bp$
–> because the 2nd conjunct $\neg Bp$ contradicts that $p \wedge \neg Bp$ is being learnt
One might thus restrict the domain of the revision operator, i.e. (re)define $*$ as a mapping $\mathcal{T} \times \{p \in L$: p can be rationally learnt$\} \to \mathcal{T}$
N.B.: also Bayes' updating rule $(\mathrm{Pr}, p) \mapsto \mathrm{Pr}(.|p)$ is defined only for 'certain' sentences, namely those with positive probability
... arguably not a real problem for Bayesian updating, because zero-probability sentences aren't learnt.

Doesn't one conjunct of the Moore sentence suffice?

When learning the Moore sentence $p \wedge \neg Bp$, one of the conjunctions is (In a sense) redundant.

Why?

Case 1 (less plausible): $p \in T$ (i.e. p was already believed before learning $p \wedge \neg Bp$)
Then

- the 1$^{\text{sd}}$ conjunct of the Moore sentence $p \wedge \neg Bp$ is redundant
- the second conjunct $\neg Bp$ is false information... arguably false information won't come

Doesn't one conjunct of the Moore sentence suffice? (ctd.)

Case 2 (more plausible): $p \notin T$ (i.e. p wasn't yet believed when learning $p \wedge \neg Bp$)

Assumption on the prior belief set T:

NI: $p \notin T \Rightarrow \neg Bp \in T$ (the indiv. has negative introspection)

Then $\neg Bp \in T$ (by NI)

So the 2$^{\text{nd}}$ conjunct of the Moore sentence is redundant (is already believed)

Doesn't one conjunct of the Moore sentence suffice? (ctd.)

What could be done given that **learning a sentence p makes the believed sentence $\neg Bp$ false?**

An option (?): Allow lack of introspection, e.g. allow that $p \in T$ but at the same time $Bp \notin T$ or even $\neg Bp \in T$.
–> this allows limited rationality
–> but if the agent happens to be a good self-observer, then after revising to $T * p$ he'll learn that Bp (and revise to $(T * p) * Bp$), then learn BBp (and revise to $((T * p) * Bp) * BBp$), and so on.

Another option (?): Enforce introspection by an extra condition on revision $*$ that requires the revised set $T * p$ to contain Bp (or at least to not contain $\neg Bp$)
–> this deviates from AGM!

Topics:

- revision on Moore sentences
- Ramsey tests
- (non-)convergence of beliefs under different DEL revision rules

There exists no conditional $\rightarrow$ satisfying the **Ramsey Test with AGM revision**: $\phi \rightarrow \psi \in T \Leftrightarrow \psi \in T * \phi$ for all $\phi, \psi \in L$ and all theories $T \subseteq L$.

There exists a conditional $\rightarrow$ satisfying the **Ramsey Test with dynamic revision**: $\phi \rightarrow \psi \in T \Leftrightarrow \psi \in T \otimes \phi$ for all $\phi, \psi \in L$ and all theories $T \subseteq L$.

There exists a conditional $\rightarrow$ satisfying a **Weak Ramsey Test with AGM revision**: $\phi \rightarrow \psi \in Th(s) \Leftrightarrow \psi \in Th(s) \otimes \phi$ for all $\phi, \psi \in L$ and all states of reality s. Here $\phi \rightarrow \psi$ represents conditional belief $B^{\phi}\psi$.