N. Gratzl D. Klein O. Roy

Introspection, Normality and Agglomeration

MCMP Munich

Universität Bamberg

Universität Bayreuth

Stalnaker

PI	$\vdash B\varphi \to KB\varphi$
NI	$\vdash \neg B\varphi \to K\neg B\varphi$
KB	$\vdash K\varphi \to B\varphi$
D	$\vdash B\varphi \to \langle B\rangle\varphi$
SB	$\vdash B\varphi \to BK\varphi$

If K is an S4 modality then B is KD45.

$\vdash B\varphi \leftrightarrow \langle K\rangle K\varphi$

$\vdash \langle K\rangle K\varphi \to K\langle K\rangle\varphi$

Stalnaker, R.,"On logics of knowledge and belief. *Philosophical studies.* 128 (2006): 169-199.

Baltag, A., Bezhanishvili, N., Özgün, A., Smets, S., "The topology of belief, belief revision and defeasible knowledge". *Logic, Rationality, and Interaction*. Springer (2013): 27–40.

Özgün, A.. *Topological Models for Beliefs and Belief Revision*. Master's Thesis, ILLC, (2013).

Introspection

$$\neg K\varphi \to K\neg K\varphi$$

- Stalnaker rejects negative introspection as a plausible rational requirement for knowledge:
  - If evidence is misleading, I am not irrational if I'm uncertain or mistaken about what I don't know.

$$K\varphi \to KK\varphi$$

- He "provisionally" accepts positive introspection.
- But the principle has been heavily challenged, for instance by Williamson

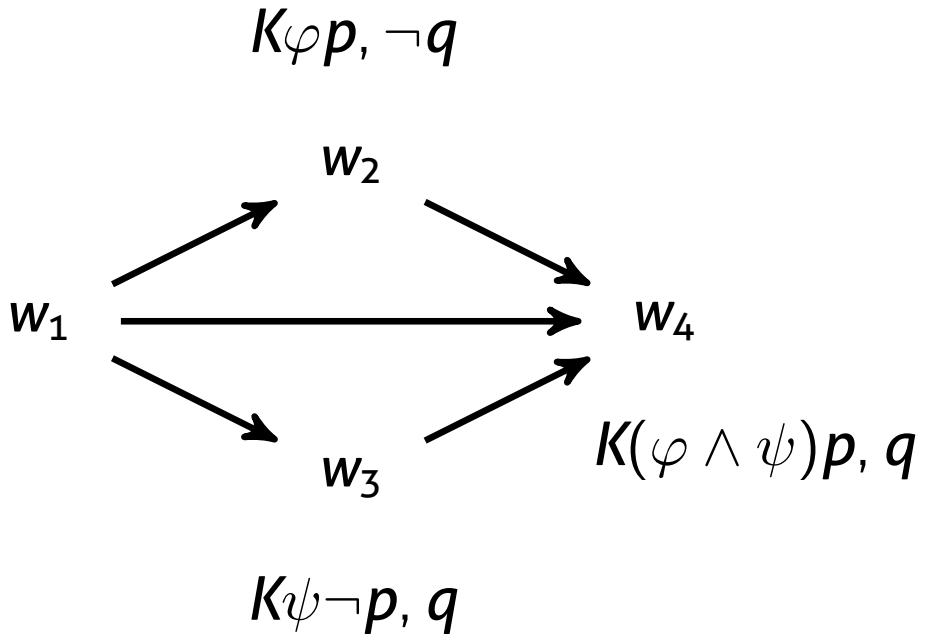
What happens to B if K is a KT.2 modality?

$$B\varphi \leftrightarrow \langle K \rangle K\varphi$$

$$\langle K \rangle K\varphi \rightarrow K\langle K \rangle \varphi$$

Introspection and normality

$$B\varphi \wedge B\psi \leftrightarrow B(\varphi \wedge \psi)$$



Non-normal beliefs

$$\frac{\vdash \varphi}{\vdash B\varphi}\ \text{(NEC)} \qquad \frac{\vdash \varphi \to \psi}{\vdash B\varphi \to B\psi}\ \text{(REG)}$$

$$\vdash B\varphi \to \langle B\rangle\varphi\ \text{(D)}$$

Thm (NEC), (REG) and (D) are sound and complete for B.

MUD frames

A neighborhood frame $\langle W, n \rangle$ is a MUD frame iff

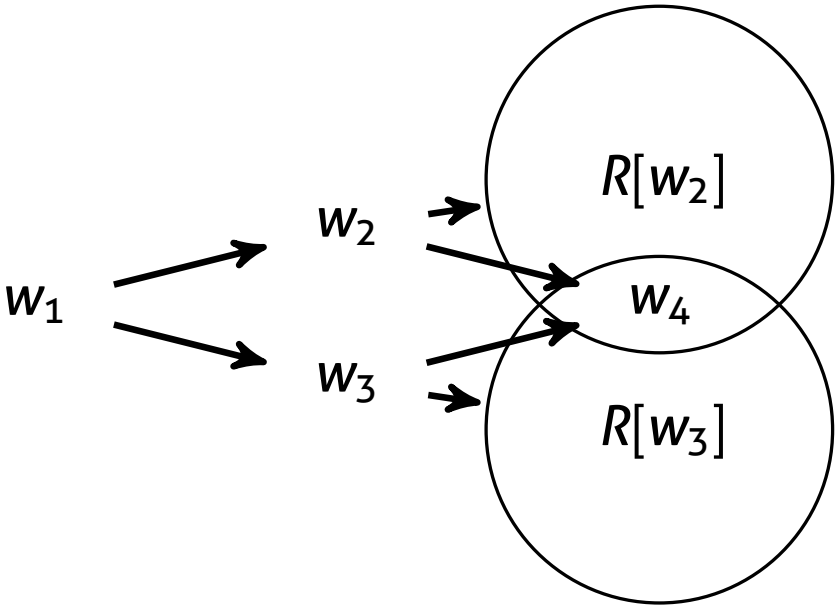
M For all w, if $X \in n(w)$ and $X \subseteq Y$ then $Y \in n(w)$. (Monotonicity)

U For all w, $W \in n(w)$. (contains the Unit)

D If $X \in n(w)$ then for all $Y \in n(w)$, $X \cap Y \neq \emptyset$.

Thm (NEC), (REG) and (D) are sound and complete for the class of MUD frames.

Completeness through representation



$$\mathcal{M}, \mathbf{w} \models \varphi \Leftrightarrow \mathcal{N}_{\mathcal{M}}, \mathbf{w} \models \varphi$$

What kind of belief is this?

- Credence above 1/2?

$$B\varphi \text{ iff } p(\varphi) \geq t > 1/2$$

$$\frac{\vdash \varphi}{\vdash B\varphi}\ (\text{NEC}) \qquad \frac{\vdash \varphi \to \psi}{\vdash B\varphi \to B\psi}\ (\text{REG})$$

$$\vdash B\varphi \to \langle B\rangle\varphi\ (\text{D})$$

Thm For every $\epsilon > 0$ there is a finite MUD frame and a state w such that some $X \in n(x)$ must receive $p(X) < \epsilon$.

Fix a natural number $n \in \mathbb{N}$, and let:

- Possible worlds: All w_{ij} with $i < j \leq n$
- For each $k \leq n$ define the set T_k as

$$T_k = \{w_{ij} | i = k \text{ or } j = k\}$$

- The neighborhood N is defined as $N = \uparrow \{T_1, \dots, T_n\}$.

To illustrate this with an example, for $n = 4$ we have

$$W = \begin{matrix} & w_{12} & w_{13} & w_{14} \\ & & w_{23} & w_{24} \\ & & & w_{34} \end{matrix}$$

$$T_1 = \{w_{12}, w_{13}, w_{14}\} \qquad T_2 = \{w_{12}, w_{23}, w_{24}\}$$
$$T_3 = \{w_{13}, w_{23}, w_{34}\} \qquad T_4 = \{w_{14}, w_{24}, w_{34}\}$$

and finally $N = \{T_1, T_2, T_3, T_4\}$.

What kind of belief is this?

- D in neighborhood models: **pairwise agglomerativity**

$$B\varphi \to \langle B \rangle \varphi$$

  Plausible **consistency requirement** for **resource-bounded** agents.

- Closure under conjunction:

$$B\varphi \wedge B\psi \to B(\varphi \wedge \psi)$$

  - Not a plausible requirement. Cluttering objection applies.

- Not so for pairwise agglomerativity.
- And violation threatens the truth aim.

Note: agents otherwise logically omniscient in this logic!

Strengthening agglomerativity

- It is still plausible that the agent holds beliefs that are pairwise consistent, but for which no triple is consistent.

$$(B\varphi_1 \wedge \ldots \wedge B\varphi_n) \rightarrow \langle B \rangle (\varphi_1 \wedge \ldots \wedge \varphi_n) \quad (n\text{-AGG})$$

- Still plausible for resource bounded agents.
- But nonetheless demanding when it comes to identifying violations and revising.

Unbounded agglomerativity

- What about (n-AGG) for all n? Call that **unbounded agglomerativity**.
- Let $\Box\varphi$ be true at a state w in a neighborhood model if and only if

$$\bigcap_{X \in n(w)} X \subseteq ||\varphi||$$

where $||\varphi||$ is the extension of φ in that model.

Obs $\Box$ is not definable in the present logic of belief.

BAGG logic

A BAGG frame is a MUD frame except that D is replaced by the following requirement.

AGG $\bigcap_{X \in n(w)} X \neq \emptyset$

The new modality $\Box$ (interpreted as in previous slide) is a normal modality.

- Unbounded agglomerativity requirement is D for $\Box$:

$$\Box\varphi \to \Diamond\varphi$$

BAGG Logic

$B\varphi \to \Box\varphi$	Belief Consistency
$\Box\varphi \to \Diamond\varphi$	Unbounded Agglomerativity

Thm These axioms, together with K for $\Box$, Necessitation for B and $\Box$ and REG for B are sound and complete with respect to the class of BAGG frames.

$$(B\varphi_1 \wedge \ldots \wedge B\varphi_n) \rightarrow \langle B \rangle (\varphi_1 \wedge \ldots \wedge \varphi_n) \qquad (n\text{-AGG})$$

Derivable in BAGG logic:

Proof theory of BAGG

$$|\Gamma| \leq 1 \; \frac{\Gamma \Longrightarrow \varphi}{\mathrm{B}\Gamma \Longrightarrow \mathrm{B}\varphi} \; (1\text{-Reg}) \qquad |\Phi| \leq 1 \; \frac{\Gamma, \Delta \Longrightarrow \Phi}{\mathrm{B}\Gamma, \Box\Delta \Longrightarrow \Box\Phi} \; (\mathrm{B}\Box)$$

$$\cfrac{\cfrac{\cfrac{\varphi \Longrightarrow \varphi}{\varphi, \neg\varphi \Longrightarrow}}{\mathrm{B}\varphi, B\neg\varphi \Longrightarrow}}{\mathrm{B}\varphi \Longrightarrow \neg B\neg\varphi}$$

Call LK-BAGG the sequent calculus LK augmented with the rules above.

Thm LK-BAGG admits the following generalization of the CUT rule.

$$(\text{Mix})\ \frac{\Gamma \Longrightarrow \Delta, F^n \qquad F^m, \Phi \Longrightarrow \Psi}{\Gamma, \Phi \Longrightarrow \Delta, \Psi}\ (n, m > 0)$$

Cor. LK-BAGG has the sub-formula property and is consistent.

Thm. Satisfiability for LK-BAGG is decidable.

$\Rightarrow$ Similar result for the logic of B alone.

Probabilistic Interpretation of BAGG?

- Contrary to B alone, BAGG frames are "too easy" to measure:

Obs. Let , w be at countable BAGG frame. Then for every $t \in (0; 1)$ there is a probability function $Prob : W \to [0; 1]$ such that $B \in n(w)$ implies $Prob(B) > t$

- So we do get completeness for models with high enough threshold... but many more as well.
- And, back to Stalnaker, most threshold can't make sense of $B\varphi \to BK\varphi$.

Conclusions

Philosophical points:

- Non-normal beliefs when knowledge is KT.
- Agglomerativity plausible weakening of closure for resource-bounded agents.
- The resulting beliefs are not high credence (nor stable p-beliefs for that matter).

Technical points:

- Sound and complete axiomatizations for logic of Beliefs with bounded and unbounded agglomerativity.
- Well-behaved proof theory.

To do:

- Dynamics.
- Mutli-agent, common belief, game theory.